2013

An Evaluation of a Nursing Leadership Simulation Experience Using Multitrait Multimethod Matrix

Toby Embry
toby.embry@dominican.edu

Follow this and additional works at: https://repository.usfca.edu/diss

Recommended Citation
https://repository.usfca.edu/diss/65

This Dissertation is brought to you for free and open access by the Theses, Dissertations, Capstones and Projects at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.
AN EVALUATION OF A NURSING LEADERSHIP SIMULATION EXPERIENCE USING MULTITRAIT MULTIMETHOD MATRIX

A Dissertation Presented

to

The Faculty of the School of Education
Learning and Instruction Department

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Toby Embry
San Francisco
May 2013
THE UNIVERSITY OF SAN FRANCISCO

Dissertation Abstract

An Evaluation of a Nursing Leadership Simulation Experience Using Multitrait Multimethod Matrix (MTMM)

The MTMM approach was used to evaluate construct validity of three nursing leadership traits (prioritization, delegation, and patient care management) across four assessment methods (multiple-choice tests—MCT, oral questioning—OQ, high fidelity—HFS, and low fidelity simulation—LFS). Using a correlational descriptive design a 21-item MCT exam, a 21-item oral question instrument, a patient care HFS, and three LFS stations were embedded into a two hour objectively structured clinical examination (OSCE) assessment environment whose aim was to compare traditional assessment methods (MCT and OQ) in nursing education to burgeoning assessment methods (HFS and LFS). Generated scores from 137 senior-level baccalaureate nursing students at a private university located in Northern California were correlated with scores from standardized instruments measuring cognitive abilities (TEAS®) and scores from another outside instrument measuring dimensions of nursing leadership (Kaplan® RN Predictor Exam) to these OSCE scores. Further, a cost comparison and analysis for designing and implementing an OSCE assessment including high- and low-fidelity simulation was compared to the projected costs of similar OSCEs found in the literature.

Results concluded that all four criteria for construct validity were not uniformly met. Yet, the method reliability estimates were high across all four measurement methods. Correlation comparison of items from the TEAS® Exam and items from the Kaplan® RN Predictor Exam indicated that one subtest from each external exam correlated highly with trait items across all methods. Lastly, the estimated budget for the
OSCE assessment study was considerably less than two estimates for similar OSCEs in the literature with the actual OSCE cost ($28,022.04) being significantly less expensive than the estimates found in the literature though higher than the estimated budget.
This dissertation, written under the direction of the candidate’s dissertation committee and approved by the members of the committee has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements for the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Toby Ellis Embry_________________________ June 18, 2013
Candidate Date

Dissertation Committee

Robert Burns ___________________________ June 18, 2013
Chairperson

Xornam Apedoe ___________________________ June 18, 2013

K.T. Waxman ___________________________ June 18, 2013
Dedication

This body of research is dedicated to the fourteen nursing students who endured many long hours as my research assistants during the data collection phase of this study: Amiee Jose, Alexa Holm, Alixandra “Alix” McCuen, Candace Albuquerque, Csarina Reynoso, Daryl Salinda, Elaine Crisostomo, Emily Whitaker, Jamie Evangelista, Mabel Reyes, Molly Howard, Stephanie Hurtado, Stephanie Villanueva, and Venice Francisco. My hope is that you all gained some practical knowledge associated with the inception, planning, and logistics of conducting nursing research. Further, my goal was to perhaps inspire each of you to actively engage in future nursing research at your places of employment. Good luck in the future and I thank you for all your hard work.
Acknowledgements

The decision to further my education was made achievable through support and unwavering positive energy from many Learning and Instruction faculty in the School of Education at the University of San Francisco. Thank you to the faculty who formed my dissertation committee. However, this journey could not be realized without the encouragement and guidance of three colleagues: Barbara Ganley, PhD, RN; Ingrid Sheets, EdD, APRN; and Anita Hunter, PhD, APRN, FAAN.

Barbara, I cannot begin to thank you for your time, encouragement, advice, intellect, and much needed humor. I cherish your friendship; appreciate your emotional support when times were tough, and your ability to make me laugh when I wanted to cry. Ingrid, I must thank you for paving the way down this doctoral path, for lending much need advice, and for having faith when mine was lost. You are an example of grace under fire. Finally, thank you Anita for lifting me up when I needed it and helping me get back on the horse; without you this journey would have ended.

I am especially grateful to my family for their love and unconditional support. Rob, you are my partner in life and my rock. Thank you for understanding when I needed to study during vacations, for your patience when I was stressed, and for the many, many hours of listening to me explain the multitrait multimethod matrix. Rosemary Christine and Dallas, I thank you both for your love and for taking the time to listen. April and Sara, I thank you two for your kisses when I needed them the most. Lastly, I must thank the army without whom this research would not have been possible: Lu Sweeney, Barbara McCamish, Mary Bickler, Miriam Edelman, Alicia Bright, Tommy Anderson, Scott Meehl, Betty Grandis, and Leslie Crane.
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii

Acknowledgements ............................................................................................................. vi

LIST OF TABLES ........................................................................................................................ xi

LIST OF FIGURES ..................................................................................................................... xiii

CHAPTER ONE. STATEMENT OF THE PROBLEM ................................................................. 1

  Purpose of Study .................................................................................................................... 8

  Significance of Study ............................................................................................................. 9

  Theoretical Rationale ......................................................................................................... 11

  Background and Need ....................................................................................................... 16

  Research Questions .......................................................................................................... 22

  Definitions of Terms ......................................................................................................... 22

CHAPTER TWO. REVIEW OF LITERATURE .................................................................. 27

  Nursing Leadership Constructs: Prioritization, Delegation, and Patient Care Management .............................................................................................................................. 27

    Prioritization ......................................................................................................................... 30

    Delegation ............................................................................................................................ 32

    Patient Care Management ................................................................................................... 36

    Summary of Nursing Leadership Constructs Literature .................................................... 44

  Traditional Assessment Techniques .................................................................................... 44

    Multiple-Choice Testing ....................................................................................................... 45

    Oral Questioning .................................................................................................................. 57

    Summary of Traditional Assessment Methods in Nursing Education ............................... 66

  Newer Assessment Methods in Nursing Education: HFS and LFS .................................. 67
Critique of Newer Assessment Methods in Nursing Education ..............................................81
Summary of Newer Assessment Methods in Nursing Education ...........................................86
Summary of the Literature ......................................................................................................87

CHAPTER THREE. RESEARCH METHODS ........................................................................91
Research Design ..................................................................................................................92
Sample ..................................................................................................................................94
Protection of Human Subjects ............................................................................................97
Instrumentation ....................................................................................................................99
Constructs ............................................................................................................................99
Assessment Methods ...........................................................................................................106
The OCSE Assessment Environment .....................................................................................121
Two Additional Standardized Assessment Methods .............................................................127
Procedures ..........................................................................................................................131
Participant Selection ............................................................................................................131
Data Collection .....................................................................................................................132
The OSCE Assessment Environment .....................................................................................135
Budget ..................................................................................................................................139
Training ...............................................................................................................................140
Pilot Study ............................................................................................................................142
Data Analysis .......................................................................................................................145
Multitrait Multimethod Matrix (MTMM) Research Design ................................................146
Campbell and Fiske (1959) Criteria with Associated Literature .........................................147
Data Analyses for Each Research Question ........................................................................149
Summary ................................................................................................................................. 150

CHAPTER FOUR. RESULTS ..................................................................................................... 153

Research Question One ........................................................................................................... 153

Research Question Two Results ............................................................................................. 161

Research Question Three Results ........................................................................................... 164

Summary ................................................................................................................................. 166

CHAPTER FIVE. SUMMARY, LIMITATIONS, DISCUSSION, AND IMPLICATIONS .................. 168

Summary of Study ................................................................................................................... 168

Significance of the Study ........................................................................................................ 169

Theoretical Framework ........................................................................................................... 170

Research Questions and Methodology .................................................................................... 171

Findings .................................................................................................................................... 173

Trends ....................................................................................................................................... 174

Limitations ............................................................................................................................... 174

Discussion ............................................................................................................................... 178

Construct Validity of the Assessment Modalities for the Three Constructs ............... 178

Reliability of the Assessment Measures ................................................................................. 181

Implications ............................................................................................................................. 183

Future Research ....................................................................................................................... 183

Practice ................................................................................................................................... 185

REFERENCES ......................................................................................................................... 187

APPENDICES ......................................................................................................................... 197
Appendix A: Multiple-Choice Test Instrument .................................................................197
Appendix B: Oral Questions ..........................................................................................201
Appendix C: OQ Correct Responses .............................................................................202
Appendix D: Scenario Validation Process Checklist .......................................................203
Appendix E: OSCE Rotation and Stations ......................................................................205
Appendix F: Report Script ............................................................................................207
Appendix G: LFS Rubric – PCM Station 2 (0700 IV push Lasix Administration for
Patient in Room 1A) .....................................................................................................208
Appendix H: LFS Station 3 – Delegation of Nursing Tasks from Recorded Report ......209
Appendix I: LFS Station 4 – Prioritization of Initial Nursing Assessment Immediately
after Receiving Report/Hand-off ...................................................................................210
Appendix J: HFS Rubric – Patient Care Management, Delegation, and Prioritization
Behaviors during Care Delivery ....................................................................................211
Appendix K: Confidential Disclosure Agreement ............................................................212
Appendix L: Expert Plan Review Checklist ....................................................................213
Appendix M: IRBPHS Approval Letter – Dominican University of California ..........215
Appendix N: IRBPHS – University of San Francisco ......................................................216
LIST OF TABLES

Table 1. Years of Hospital Employment of the Participants with Supportive, Directive or Participative Leadership Styles .................................................................35

Table 2. Overview of the OSCE Assessment Environment........................................95

Table 3. Construct Objectives and Indicators – Prioritization........................................101

Table 4. Construct Objectives and Indicators – Delegation ........................................104

Table 5. Construct Objectives and Indicators—Patient Care Management..................106

Table 6. Multiple-Choice Test Blueprint .................................................................109

Table 7. Sample Multiple-Choice Question ...............................................................111

Table 8. Prioritization Tasks, Questions, and Correct Responses Associated with the Recorded Report/Handoff .................................................................114

Table 9. Item Development Blueprint—Oral Questioning Station................................115

Table 10. HFS Rubric Blueprint ...............................................................................117

Table 11. Item Development Blueprint HFS and LFS .................................................120

Table 12. OSCE Assessment Environment Overview ..................................................121

Table 13. Comparison of Actual OSCE Costs with Estimates from the Literature........141

Table 14. Correlations, Reliability, Means (M), and Standard Deviations (SD) for the Pilot Study ..................................................................................................................145

Table 15. MTMM Method Block Average Convergent Validity Coefficients Compared with Average Upper- and Lower Off-Diagonal Block Coefficients .........................157

Table 16. Correlation Comparison of Twelve Measures and Seven Subtest ..................160
Table 17. Correlation Coefficients Between the Twelve OSCE Measures and Seven External Subtests

Table 18. Comparison of Actual OSCE Costs with Estimates from the Literature
LIST OF FIGURES

Figure 1. Structure of a hypothetical Multitrait-Multimethod Matrix with three traits and three methods………………………………………………………………………………13

Figure 2. Multitrait-multimethod Matrix (MTMM) of Three Nursing Leadership Constructs Using Four Methods ..........................................................42

Figure 3. The Room Layout for the OSCE..................................................137

Figure 4. MTMM of Three Nursing Leadership Constructs Using Four Methods......148

Figure 5. MTMM of Correlation Coefficients from the Twelve OSCE Generated Scores………………………………………………………………………………154
CHAPTER ONE

STATEMENT OF THE PROBLEM

Nurse leadership is a desired outcome of most, if not all, undergraduate nursing programs (Thomas, Hodson-Carlton, & Ryan, 2011). Teaching leadership in nursing entails defining what leadership means, who leaders are, what leaders do, and how leadership is different from management (Grossman & Viliga, 2013). One purpose of a baccalaureate nursing curriculum is to address these questions and help students understand the complex, multidimensional concept nurse educators refer to as leadership. While progressing through their nursing program, students are challenged to consider the leadership skills needed in the nursing profession today, think about themselves as leaders, and to reflect on how leadership responsibilities must be integrated into their roles as professional nurses (Grossman & Viliga, 2013; Scully, 2011; Thomas et al., 2011).

Most nursing educational programs are designed to teach three major skill categories associated with leadership: prioritization, delegation, and patient care management. Prioritizing patient care is defined by determining which nursing care activities require immediate attention (Motacki & Burke, 2011). Delegation is defined as the transfer of responsibility, based on priority setting, for the performance of an activity from one individual to another while retaining accountability for the outcome (ANA, 1992). Proper delegation also requires priority setting. Safe patient care management is defined as the nurse’s ability to effectively prioritize and delegate patient care needs (Grossman & Viliga, 2013; Motacki & Burke, 2011). Nursing skills within each of these three categories form a large part of what is meant to determine leadership (Bensfield,
Understandably then, nurse educators place a great deal of importance on evaluating these leadership skills so that feedback can be provided to students to help with the acquisition of these skills (Kaplan & Ura, 2010).

Assessing nursing leadership skills, however, is complex and presents numerous challenges for nurse educators (Bensfield, Olech, & Horsley, 2012; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011). First, leadership is difficult to define and there are almost as many definitions of nursing leadership as there are authors who write about it (Grossman & Viliga, 2013). Second, evaluating nursing leadership is difficult to measure using multiple-choice tests (Kaplan & Ura, 2010), especially if the test writer is unfamiliar with the item writing procedures required to measure higher-order skills (Tarrant & Ware, 2010). Finally, leadership in the clinical setting is difficult to even elicit, making it more difficult for clinical faculty to assess student acquisition of these important skills (Jones, Pegram, & Fordham-Clarke, 2010; Major, 2005; Thomas, Hudson-Carlton, & Ryan, 2011).

In recent years, many nursing programs are meeting these assessment challenges with the use of simulation experiences (Jefferies & Clochesy, 2012; Kaplan & Ura, 2010). Controlled simulation experiences related to problem solving, decision-making, and a variety of opportunities for communication can provide experiences that simulate authentic workplace situations where leadership skills can be elicited (Jones, Pegram, & Fordham-Clarke, 2010; Major, 2005; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011). Simulation offers students a safe learning environment during the teaching/learning process and can also provide a standardized environment for more
formal testing (Jones, Pegram, & Fordham-Clarke, 2010; Major, 2005; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011). If students are provided opportunities to demonstrate leadership skills in simulated, authentic work conditions, then it may be possible to evaluate the quality of student leadership behaviors and offer the feedback necessary for skill acquisition (Kaplan & Ura, 2010; Jefferies & Clochesy, 2012).

Simulation experiences can come in a variety of forms from high-fidelity simulation to low-fidelity simulation (Kaplan & Ura, 2010; Rushforth, 2007). High-fidelity simulation involves the use of written, structured case scenarios constructed around specific student learning outcomes. These scenarios are vetted by faculty savvy in the elements of simulation scenario development that use sophisticated mannequins termed human patient simulators, and/or the scenarios use standardized patients which are actors portraying the “role” of the patient (Jefferies, 2005; Jefferies & Clochesy, 2012; Jefferies & McNelis, 2010). Low-fidelity simulation (LFS) includes simulated environments (typically for knowledge acquisition related to specific skills) that use role-play, simple case studies, and task trainer mannequins to aide in psychomotor skill acquisition (Brewer, 2011; Swenty & Eggleston, 2010).

With the growing interest to use simulation as an assessment method in addition to its use by nurse educators as a pedagogical method, concern has arisen about the reliability and validity of the scores obtained from simulation evaluations. Currently, there is considerable debate within the nursing literature as to the use of simulation to assess student knowledge (Bensfield, Olech, & Horsley, 2012; Byrne & Smyth, 2008; Kardong-Edgren, Hanberg, Keenan, Ackerman, & Chambers, 2010; Mitchell et al., 2009; Rushforth, 2007). Rushforth (2007) has argued that despite considerable research on
simulation in the medical field to support its reliability and validity, there is little research on the measurement properties of simulation as an evaluation metric in nursing education. Without evidence on the reliability and validity of simulation rubric scores, observations of students during such evaluations run the risk of observer bias and measurement error (Byrne & Smyth, 2008; Kardong-Edgren et al., 2010; Major, 2005; Mitchell et al., 2009; Nulty et al., 2011; Radhakrishnan, Roche, & Cunningham, 2007; Rushforth, 2007). Because reliability and validity are two primary criteria used for determining the quality of scores that are generated by any assessment instrument (Conway, Lievens, Scullen, & Lance, 2004; Westen & Rosenthal, 2003), it is important to determine whether assessment of student knowledge by simulation experiences are consistent and accurate when compared to current forms of assessment in nursing education (Rushforth, 2007). This study directly examines some of these measurement concerns.

Measurement specialists in the social sciences almost exclusively use construct validity as a determining factor in assessing the quality of the generated scores from an instrument (Conway, Lievens, Scullen, & Lance, 2004; Westen & Rosenthal, 2003). Construct validity refers to how well a test or instrument actually measures the construct it claims to measure (O’Connor, 2006; Popham, 2000). Generally, there are two procedures for establishing the construct validity of scores obtained from an instrument: experimental procedures and correlational procedures (Popham, 2000). Experimental procedures often take the form of a differential-groups study where scores on the measures of the construct (e.g., intelligence) are compared for two groups: one group that should score high (e.g., university professors) and one group that should score lower
(e.g., manual laborers) (Cizek, 2012; Westen & Rosenthal, 2003). If the group that should score higher actually does score higher, then the result is said to provide evidence of construct validity of the test scores (Popham, 2000). An alternative strategy to determine the construct validity of test scores is through the use of correlational procedures (Cizek, 2012; Westen & Rosenthal, 2003). For example, scores on a new test measuring a construct should correlate with existing tests thought to already measure the construct (Cizek, 2012; Popham, 2000).

An important issue with the use of correlational procedures for providing evidence for construct validity is the fact that every test score is a specific and unique construct-test method pair (Popham, 2000; Westen & Rosenthal, 2003). That is, part of the score is due to the construct being measured and part of the score is due to the test method used to measure the construct. Therefore, test scores can correlate because they measure the same construct, because the same test method is used, or some unknown combination of the two (Cizek, 2012; Conway et al., 2004; Westen & Rosenthal, 2003). Called “shared method variance” in the literature (Brannick, Chan, Conway, Lance, & Spector, 2010; Sharma, Yetton, & Crawford, 2009), it is a problem for construct validity because it is not known how much the magnitude of a correlation coefficient is due to measuring the same construct and how much is due to using the same test method (Brannick et al., 2010; Cizek, 2012; Conway et al., 2004; Westen & Rosenthal, 2003).

Campbell and Fiske (1959) recognized this problem and developed a procedure called the multitrait multimethod matrix (MTMM) as a way to separate construct (trait) variance from method variance. The two key concepts of the MTMM are convergent and discriminant validity. Convergent validity refers to the extent that two test scores
measuring the same construct (trait) correlate; *discriminant validity* refers to the extent to which two test scores measuring different constructs do not correlate (Campbell & Fiske, 1959; Cizek, 2012). The MTMM procedure requires at least two different traits be measured by at least two different methods (Campbell & Fiske, 1959; Cizek, 2012; Sharma et al., 2009). The correlations among the measures are specially arranged in a MTMM matrix where it can be determined if the criteria established for construct validity by Campbell and Fiske are met (Conway et al., 2004; Sharma et al., 2009). Despite being over 50 years old, the MTMM is still used today and considered one of the best procedures for establishing the construct validity of scores (Cizek, 2012; Conway et al., 2004; Sharma et al., 2009).

This study applied the ideas of the MTMM matrix to the measurement of the three key leadership constructs of prioritization, delegation, and patient care management. A four-step procedure developed to test whether the common assessment methods in nursing are in fact measuring prioritization, delegation, and patient care management. First, the three constructs were defined in observable terms. Second, four test methods were designed to measure each of the three constructs based on the definitions. The four test methods were the two commonly used in nursing education—*multiple-choice tests*, and *oral questioning* -- and the two less commonly used methods of *high-fidelity* and *low-fidelity simulation*. Third, scores were obtained from a sample of student nurses on the twelve measures (three traits each measured by four methods), and the correlations among the scores were arranged into a MTMM matrix. Fourth, the MTMM matrix was examined to see to what extent the correlations met the criteria of convergent and discriminant validity established by Campbell and Fiske (1959). Based on the guidelines
identified by Harden (1988), the twelve measures were administered within an assessment environment called an Objectively Structured Clinical Examination (OSCE). In recent years, nurse educators have borrowed and adapted the OSCE assessment environment from medical education (Jones, Pegram, & Fordham-Clarke, 2010; Kardong-Edgren et al., 2010; Major, 2005; Nulty et al., 2011; Radhakrishnan, Roche, & Cunningham, 2007; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011).

Additionally, two external examination scores commonly used in nursing were correlated with the twelve scores to help establish what was being measured by the scores. The first external examination was the Test of Essential Academic Subjects (TEAS®), a standardized test administered to potential nursing candidates prior to admission. The TEAS® has subtest scores in four content areas: reading, mathematics, science, and English language usage. The second external examination used in this study was the Kaplan® RN Predictor Exam. Typically, nursing students are administered a predictive examination that tests knowledge covering several nursing content areas prior to graduation. Three subtests of the Kaplan® RN Predictor Exam, Setting Priorities, Management of Care, and Making Nursing Judgments were judged to measure similar content as the three constructs in this study.

Finally, high-fidelity and low-fidelity simulation are resource intensive. The push in nursing toward using these methods to evaluate student knowledge requires an examination of their associated costs. For years, medical education has integrated the use of high-fidelity simulation (HFS) and low-fidelity simulation (LFS) within the OSCE format. Based on the cost analyses of Palese, Bulfone, Venturato, Urli, Bulfone, Zanini, Fabris, Tomietto, Comisso, Tosolini, Zuliani, and Dante (2012) and Poenaru, Morales,
Richards, and O’Connor (1997) in medical education, a budget for the implementation and scoring of the twelve measures obtained in the OSCE was created for this study. Costs associated with this study were then compared to the budgets from Palese et al. (2012) and Poenaru et al. (1997) to determine if any differences exist.

**Purpose of Study**

The purpose of this study, then, was to implement twelve measures within an OSCE assessment and evaluate the construct validity of the three major nursing constructs of prioritization, delegation, and patient care management. Each construct was measured by the four methods of multiple-choice tests, oral questions, high-fidelity simulation, and low-fidelity simulation. To help in the establishment of the construct validity, subtest scores from two additional external tests were obtained from school records: a test of basic skills and a nursing test measuring some of the same skills as the three constructs considered in this study. Finally, because of the additional costs associated with the use of an OSCE assessment environment, a cost analysis was completed.

The twelve measures were obtained from a convenience sample of 137 senior nursing students enrolled in a nursing leadership course (n=48) and an advanced medical-surgical course (n=89) in the 2012-13 academic year. During a two-hour assessment, the nursing students visited a series of stations where they were asked to respond to a variety assessment methods used to measure cognitive and/or behavioral indicators linked to the three nursing leadership constructs examined in the study. Scores were obtained at each station, and the intercorrelations among the twelve measures were eventually organized into a multitrait-multimethod matrix.
It is important to distinguish between simulation as a method and OSCE as an assessment environment within which simulation (and other forms of measurement) can be embedded. This study used the OSCE assessment environment to deliver the four assessment methods of multiple choice testing, oral questioning, high-fidelity simulation, and low-fidelity simulation, but it was not, per se, a study of the OSCE.

Significance of Study

This study is important for three reasons. First, it fills a void caused by the lack of nursing research on high-fidelity and low-fidelity simulation as assessment methods. Despite the strong push to change the way nursing students are educated and evaluated (Benner et al., 2010), no nursing research exists to examine the construct validity of high- or low-fidelity simulation as an assessment methodology compared to the traditional methods of student assessment (multiple-choice tests and oral questioning) (Luptkar-Flude, Wilson-Keates, & Lapocque, 2012; Rushforth, 2007).

Second, this research study is important because nursing education continues to move toward the inclusion of simulation as a method of evaluating student performance (Major, 2005; Rushforth, 2007). Therefore, it is important to understand the reliability and validity of using simulation in nursing as an assessment methodology (Major, 2005; Morris & Hancock, 2008; Rushforth, 2007; Walsh, Jairath, Paterson, & Grandjean, 2010). Currently, nursing education uses traditional methods to evaluate student performance associated with skills specific to leadership (Gaberson & Oermann, 2010). A transformation of the way nursing students are taught is underway (Benner et al., 2010). However, can the use of high-fidelity and low-fidelity simulation measure nursing student leadership skills associated with the constructs of prioritization, delegation, and
patient management? Current research indicates that new graduate nurses lack competent ability in these specific areas (Benner et al., 2006; Hickey, 2009; IOM, 2003).

A third reason this study is important is that simulation experiences used as an assessment method in an OSCE assessment environment are resource intensive (Poenaru, Morales, Richards, & O’Connor, 1997). Any type of simulation experience used as an assessment method, when compared to multiple-choice, oral questioning and other traditional methods of examination, are more time-consuming and more expensive in terms of human and material cost to develop (Barman, 2005; Palese et al., 2012; Poenaru et al., 1997). In fact, most of the criticism against the use of OSCE assessment environments in medical education has centered on their high cost. Initial reports on developing an OSCE have provided very disparate data, ranging from costs as low as $11 per student to as high as $1200 per student (Poenaru et al., 1997).

Assessing the cost-effectiveness of an OSCE is essential given the economic recession and the internationally-recognized shortage of support for nursing education (Dadgaran, Parvizy, & Peyrovi, 2012). The medical literature on OSCEs considers it to be the most valid and reliable method for assessing safe practice. The issue of the reliability and validity of different assessment techniques in nursing would be less pressing if the commonly used techniques were roughly comparable in ease of use and cost. But this is not the case. The use of OSCE assessment of student knowledge has limited diffusion in United States nursing education specifically related to its cost (Rushford, 2007). Given that nursing students must have their practice skills evaluated, no data exist comparing the cost of the simulation with other assessment methods (Palese et al., 2012). Therefore, it is necessary to provide an estimate of the cost of implementing
a nursing OSCE. The cost of simulation assessment methods could be justified if it can be shown that they measure skills different from that of multiple choice and oral questioning.

**Theoretical Rationale**

The Standards for Educational and Psychological Testing (NCME, 1999) is a joint document produced by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. It represents the endorsed positions of a dozen organizations with regard to the sources and standards of evidence surrounding the validation of test instruments within education and psychology. For nearly 50 years, psychometric traditions that guide the validation of test instruments have been developed, formalized, documented, and disseminated by these organizations (Cizek, 2012). According to Cizek (2012), the unified viewpoints surrounding validity claim that “all evidence that might be brought to bear in support of intended inference is evidence bearing on the construct that the instrument purports to measure” (p. 31).

A psychological construct is a hypothesized attribute, proficiency, ability, or skill used in psychology theories to account for or explain behavior (Moss, 1998). The perspective on construct validity evolved from four seminal papers. These papers are MacCorquodale and Meehl (1948), Cronbach and Meehl (1955), Loevinger (1957), and Campbell and Fiske (1959).

Campbell and Fiske (1959) provided important insights into the nature of validation procedures. Particularly, they propose the multitrait-multimethod matrix (MTMM). The MTMM approach contends that independent measures of traits are
essential information for assessing construct validity, whether the measures are taken at the same or at different times (Campbell & Fiske, 1959; Ferketich, Figueredo, & Knapp, 1991). The basic underlying tenants of the MTMM matrix approach are convergent and discriminant validity. Convergent validity refers to the fact that scores from tests designed to measure the same construct should correlate highly with scores from other tests designed to measure the same construct. Discriminant validity refers to the fact that scores from tests measuring one construct should not correlate with scores from tests measuring other constructs (Ferketich, Figueredo, & Knapp, 1991). Figure 1 shows a hypothetical three trait, three method MTMM, and illustrates the positioning of the reliability diagonal (“R” in the figure), the convergence diagonal (“C” in the figure), same method triangles (“S” in the figure), and different trait-different method triangles (“D” in the figure), all key features of the Campbell and Fiske framework.

The first feature of the matrix in Figure 1 is the main diagonal of the matrix, where the entries are the reliability estimates of each measure, called “R” in Figure 1(Cizek, 2012). The second feature of the matrix is two types of blocks of entries in the matrix, same method and different method. The same method blocks, called “S” in Figure 1 are the correlations of different traits measured by the same method; that is why they are called same method blocks. There are three same method blocks in Figure 1 and they generally show the correlations among measures of the different traits within the same method.
<table>
<thead>
<tr>
<th>Traits</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>S</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>B</td>
<td>D</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>B</td>
<td>D</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

R = reliability of diagonal; C = convergence diagonal; S = same method triangles; D = different trait-different method triangles

*Figure 1. Structure of a Hypothetical MTMM with Three Traits and Three Methods*

The other blocks are the *different method blocks*. The *different method blocks* are formed from the traits measured by different methods. Method blocks contain two types of information. The main diagonal of these blocks is called the *validity diagonal*, and presented by the “C’s” in Figure 1. The validity diagonal shows the correlation among different methods measuring the same traits. The off-diagonals are the *different trait-different method triangles* and are represented by the “D’s” in Figure 1. These entries indicate the correlations of different traits measured by different methods, and generally show the magnitude of correlation associated with each trait measured by different methods.
The MTMM procedure is based on interpreting correlations of the same trait across different methods and correlations of different traits across the same methods (Campbell & Fiske, 1959; Ferketich, Figueredo, & Knapp, 1991). High correlations of the same trait assessed by different methods provide evidence of convergent validity. Conversely, low correlations of different traits assessed by the same methods provide evidence of divergent validity. The MTMM design isolates the correlations between traits and methods and the variances attributable to them. Campbell and Fisk (1959) identified four criteria to assess the extent of convergent and discriminant validity (Ferketich, Figueredo, & Knapp, 1991).

The first criterion is that the correlations in the validity diagonal should be significantly different from zero and be of a sufficient magnitude to encourage further validity exploration. This criterion is usually assessed by simply inspecting the magnitude of the correlations. The second criterion is that the correlations should be higher in the validity diagonals compared to the correlations in the different methods block – the correlation of different measures of the same trait should be higher than different traits using different methods. This criterion is often assessed by averaging the correlations in the validity diagonal and comparing it to the average of the correlations located in the off-diagonals of the different methods block. The third criterion is that the correlations between two measures of the same trait should be higher than correlations between that trait and another trait. This criterion is assessed by inspecting the magnitude of correlations of the specific trait measured by different methods within the same row and column ensuring that the average correlation is higher between the same traits measured by different methods than the average correlation of different traits measured by different
methods. The final criterion is that the correlations between traits (whether in the same method or different method blocks) should indicate patterns of interrelationships between traits and should be similar in the different trait triangles. This criterion is assessed by examining that the magnitude of correlations between different traits measured by different methods are among the lowest correlations located within the entire matrix (Ferketich, Figueredo, & Knapp, 1991).

Correlations on the validity diagonal represent estimates of *convergent validity*: different measures of theoretically similar or overlapping constructs. These types of interrelationships should be strong or highly correlated among methods measuring the same trait (same trait – different method coefficients). The off-diagonal elements of the different method blocks reveal *discriminant validity*: measures of theoretically distinct constructs should not be highly intercorrelated. *Discriminant validity* is evidenced by weaker correlations between different traits measured by different methods (different trait – different method coefficients). Support for discriminant validity is obtained when the observed correlations located in the off-diagonal correlations of the different method blocks are uniformly lower than the validity coefficients.

Evidence for method effects is obtained by assessing the off-diagonal correlations of the same method blocks. The differential magnitude of correlations between different traits measured by the same method (different trait – same method coefficients) relative to the correlations between the same two traits measured by different methods determines the extent of method effects reflected between methods.

The four Campbell and Fiske criteria have been subject to much discussion in the literature. At least three limitations of the MTMM methodology have been noted. First,
there are no hard rules for interpreting the magnitude of the correlations or their patterning, making the analysis of a MTMM somewhat subjective. Second, the reliance on the correlations among observed measures intends for the researcher to draw inferences regarding trait and method factors. Lastly, researchers have used confirmatory factor analysis (CFA) to identify trait and methods factors, but in many cases have met difficulties in obtaining appropriate solutions (Cote, 1995; Marsh & Hocevar, 1983; Marsh & Grayson, 1995). This study used the Campbell and Fiske criteria and the correlation of the twelve measures with two sources of external scores: (a) four basic achievement skills of English language use, mathematics, and science and (b) three measures from the Kaplan® RN Predictor Exam most closely related to the three constructs of prioritization, delegation, and primary care management. The magnitude of these latter correlations will help in the assessment of the three constructs in this study.

**Background and Need**

The OSCE assessment environment has been used in medical education for decades. Because the OSCE evaluates student ability in simulated situations much closer to the desired outcome than what multiple-choice tests might be able to accomplish, the OSCE assessment environment has become a well-established method of assessing medical students (Mitchell, Henderson, Groves, Dalton, & Nulty, 2009; Rushforth, 2007; Ward & Barratt, 2005). The research on OSCE assessment environment use in medical education has been positive and suggests that this evaluation method is reliable and valid (Carraccio & Englander, 2000; Harden, Stevenson, Downie, & Wilson, 1975; Rushforth, 2007). Research by Carraccio and Englander (2000) concluded that with skillful design, reliability and validity can be achieved with a combination of an OSCE
assessment environment using HFS and LFS experiences, standardized board
examinations, and direct observation of medical students in the clinical setting. Similar
medical research on the OSCE assessment environment use with medical students
revealed that reliability and validity of the assessment is based on the number of skill
stations, the timing of student and standardized patient interaction, clear objective
delineation, and faculty training on the use of assessment instrumentation (Harden et al.,

The successful use of the OSCE in medical education is apparently due to the
format originally constructed by Harden et al. (1975). The original OSCE format
comprised a series of 16 to 20 stations with each station lasting approximately 5 minutes
to complete. The stations allowed students to be immersed in a patient-care scenario
crafted to mimic a realistic practice environment with sufficient realism to allow the
learner to believe the fidelity of the situation (Kardong-Edgren et al., 2010). However,
nurse educators in Canada, the United Kingdom, and Australia have modified the Harden
OSCE format to suit the specific needs of nursing (Rushforth, 2007). Moreover, the
change in format by nurse educators has led to a concomitant change in the language
used by these educators to describe the OSCE. The nursing literature often uses the term
OSCE interchangeably with the term simulation to denote student assessment using
simulation techniques, while the medical educational literature use the term OSCE to
define a structured format of student evaluation (Jones, Pegram, & Fordham-Clarke,
2010; Kardong-Edgren et al., 2010; Major, 2005; Nulty et al., 2011; Radhakrishnan,
Roche, & Cunningham, 2007; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011).
The conflated use of these terms by countries that have adopted the use of simulation in
an OSCE assessment environment with their programs of nursing education has likely contributed to the confusion and perhaps to the hesitancy of some to adopt simulation as a means of student assessment (Jones, Pegram, & Fordham-Clarke, 2010; Kardong-Edgren et al., 2010; Major, 2005; Nulty et al., 2011; Radhakrishnan, Roche, & Cunningham, 2007; Scully, 2011; Thomas, Hodson-Carlton, & Ryan, 2011).

Despite the definitional problems, simulation has evolved in nursing into an authentic, innovative pedagogy which has been incorporated into present day hospital staff training and schools of nursing curricula (Kaplan & Ura, 2010; Murphy et al., 2010; Solnick & Weiss, 2007; Swanson et al., 2010; Swenty & Eggleston, 2010; Waxman & Telles, 2010). Nurse educators see simulation as a means of transforming nursing pedagogy and evaluation, and helping to bridge the gap that exists between nursing academia and nursing practice (Benner et al., 2010; IOM, 2003).

However, little literature exists on the use of OSCEs and simulation in the nursing education format. The research that does exist demonstrates that nursing education has varied the OSCE from the original format introduced by Harden et al. (1975). Variations by nursing and other allied health professionals from the medical education format have potentially altered the levels of validity and reliability of this assessment method (Carraccio & Englander, 2000; Chambers, Boulet, & Gary, 2000; Haijazi & Downing, 2008). While broadening the use of OSCE assessment in nursing education in the United States (Kardong-Edgren et al., 2010) may have positive benefits, it is imperative that research be conducted on the OSCE variations used in nursing education, including the various types of simulation. For example, on a continuum from low to high complexity, simulation encompasses role-play, static task-trainers, low-,
medium-, and high-fidelity mannequins, standardized patient actors, and OSCEs using all forms of simulation (Billings & Halstead, 2012). It seems unlikely that all forms of simulation would be equally effective, and not all OSCE formats equally efficient.

This possibility is underscored in a recent literature review by Mitchell, Henderson, Groves, Dalton, and Nulty (2009). They reviewed the literature on the use of simulation in nursing education across the UK, Canada, and Australia. Their review demonstrated simulation use to range from assessment of the purely technical nursing skill to assessment of professional competence, and despite consistent findings of high face validity, the large number of adaptations of simulation assessments using an OSCE environment in nursing education resulted in inconsistencies in the reliability and validity of the assessment methodology.

Research conducted by Walsh, Jairath, Paterson, and Grandjean (2010), Morris and Hancock (2008), and Prion (2008) have concluded that it was essential to provide evidence of construct validity when new assessment methodologies such as simulation are introduced. This was supported by Kardong-Edgren, Hanberg, Keenan, Ackerman, and Chambers (2010), who argued that the push to use simulation as assessment methodologies in nursing education required rigorous research on these new assessment methods.

Unfortunately, few medical or nursing researchers have employed robust methods for investigating construct validity such as the MTMM introduced by Campbell and Fiske (1959). One of the few studies that did examine construct validity was that of Biag, Violato, and Crutcher (2010). They used the MTMM to assess the construct validity of clinical competence (the extent to which clinicians demonstrate proper judgment and
decision making in a practice setting). Biag et al. (2010) used four methods of assessment: the Physician Achievement Review (PAR), in-training evaluation reports (ITERs), clinical evaluation exercises (miniCEX), and OSCE stations. Three physician traits were evaluated: doctor-patient relationship, in-training evaluation reports, and clinical assessments. Thirty-nine international physicians participated in high-stakes OSCEs.

Evidence for both convergent and divergent validity for clinical competence was determined. The reliabilities of the assessment instruments were in the adequate to good range across all four methods. The clearest evidence for both convergent and divergent validity for clinical competence, followed by doctor-patient relationships, and communications as assessed by the ITER and PAR, but not with the simulation (Biag et al., 2010). But there was evidence of substantial method specificity.

To establish construct validity, it can be helpful to examine the correlations of the measures under investigation with other external measures. In the practice setting, for example, nurses must have the ability to comprehend the written word, to understand English, and to have knowledge about the sciences that ground this profession. How new measures of assessment correlate with such measures of basic skills helps to define what is being measured. Additionally, nursing judgment and the ability to set priorities are the basis for nurses to provide adequate patient care management, as well as to prioritize and to delegate; therefore, any external measures of these nursing skills will help determine the construct validity of the new measures.

Consequently, in addition to examining the MTMM matrix, this study also obtained the basic skills scores of the participants on the Test of Essential Academic
Subjects (TEAS®), a nationally recognized examination used by nursing programs to measure four basic skills content areas deemed essential to the profession: reading, mathematics, science, and English language usage. This study also obtained three subtest scores from the Kaplan® Predictor Exam, a nationally recognized test measuring a variety of nursing concepts thought critical for nursing licensure. The three Kaplan subtest measures were closely related to the three constructs investigated in this study: delegation, prioritization, and patient care management.

This study was primarily about the measurement of these three key nursing leadership skills. However, in part this study was also about the OSCE assessment environment. As mentioned above, it is important to distinguish between simulation as a method and OSCE as a format within which simulation (and other forms of assessment) can be embedded. This study used the OSCE environment to deliver the four assessment methods of multiple choice tests, oral questioning, high fidelity simulation, and low fidelity simulation. But one of the drawbacks to the use of OSCEs is the cost of running such assessments (Major, 2005; Mitchell et al., 2009; Palese et al., 2012; Poenaru et al., 1997).

Barman (2005) and Major (2005) have contended that cost is the single most important reason why OSCE assessments have not been used in nursing programs in the United States (U.S.). Traditional assessment methods like multiple-choice tests are much less expensive to develop and administer. Studies by Barman, (2005), Palese et al. (2012), and Poenaru et al. (1997) have concluded that cost was one of the main criticisms against OSCE assessments in medical education when compared to traditional forms of assessment such as in vivo oral exams. Likewise, the few cost estimates have varied
widely, from $11 per student to $1200 per student. To address this issue, a cost analysis was completed by creating a budget for the OSCE assessment environment used in this study and comparing it to those of Palese et al. (2012) and Poenaru et al. (1997).

**Research Questions**

The Campbell and Fiske (1959) approach (C&F) described four conditions when examining a MTMM matrix. These conditions were used to develop the first research question for this study. The second research question examines the correlation between the scores obtained from this research to scores from the study participant on a national achievement exam and scores from a national predictive exam for nursing licensure. The third research question examines the costs of a nursing OSCE assessment environment compared to similar estimates in the literature. The three research questions are listed below.

**Research Question 1.** To what extent do the twelve OSCE scores generated from using four assessment methods (multiple-choice tests, oral questioning, low-fidelity simulation, and high-fidelity simulation) to measure three constructs (delegation, prioritization, and patient care management) conform to the Campbell and Fiske (1959) criteria for construct validity?

**Research Question 2.** How do the twelve OSCE scores generated from the four methods of evaluating the three constructs correlate with scores of basic skills (TEAS®) and standardize nursing predictive scores of nursing skills (Kaplan® RN Predictive Exam)?

**Research Question 3.** To what extent does the actual cost of designing and implementing an OSCE assessment environment for nursing leadership constructs which
includes high-fidelity and low-fidelity simulation stations compare to the projected costs found in the literature?

Definitions of Terms

Convergent validity: According to Campbell and Fiske (1959), convergent validity is represented when measures of the same trait correlate high.

Delegation: Delegation in nursing is the transferring of the authority to perform selected nursing tasks in selected situations to a competent individual (Motacki & Burke, 2010). The Five “R’s” of delegation exist in nursing practice: right person; right task; right situation; right explanation/communication; and right follow-up.

Discriminant validity: Discriminant validity tests whether concepts or measurements that are supposed to be unrelated are, in fact, unrelated. Campbell and Fiske (1959) introduced the concept of discriminant validity within their discussion on evaluating test validity. They stressed the importance of using both discriminant and convergent validation techniques when assessing new tests. A successful evaluation of discriminant validity shows that a test of a concept is not highly correlated with other tests designed to measure theoretically different concepts.

Method variance: In the social sciences and psychometrics method variance is the spurious variance that is attributable to the measurement method rather than to the constructs the measures represent or equivalently as systematic error variance shared among variables measured with and introduced as a function of the same method and/or source. Studies affected by method variance or method bias suffer from false correlations and run the risk of reporting incorrect research results.

(Kenny, 2012).
Multiple-Choice Questioning: Multiple-choice questioning is a type of choice that has a selected response type of item used to assess knowledge, intellectual skills, or higher order domains of learning. The most common type of multiple-choice item represents the student with the question along with four or five possible answers from which one is to be selected. The initial part of the multiple-choice item will typically be a question or an incomplete statement. This component of the item is known as the stem. The possible answers are referred to as alternatives. In a set of alternatives there are several wrong answers and at least one correct answer. The wrong answers are called distracters used to distract the unknowledgeable or unskilled student from readily guessing the right answer (Popham, 2000).

Objectively structured clinical examination (OSCE): An objectively structured clinical examination is an assessment environment that provides a means of evaluating performance in the simulation laboratory setting rather than in the clinical setting. During an OSCE, students rotate through a series of stations; each station requires students to complete an activity or perform a task which is then evaluated by the examiner. The student performance is rated using a rating scale checklist or rubric. The other stations might require students to be tested on their knowledge and cognitive skills, analyze data, select interventions and treatments, and manage the patient's condition. OSCEs typically are used for summative clinical assessment; however, they also can be used to assess performance and provide feedback to students (Gaberson & Oermann, 2010).

Oral Questioning: Oral questioning is a constant in clinical learning and assessment of nursing students, the act of asking a student questions to evaluate their thinking and knowledge of a specific situation or skill. The primary teaching learning
strategy supporting both critical thinking and reflective practice involves making explicit and verbal that which is thought and unspoken. Wink (1993) contends that critical thinking can be developed through questioning that moves beyond that which stimulates recall of factual information, to ask students why they have reached the conclusions they report or made the decisions upon which they are acting. Further, these levels of intellectual work, involving the construction of new knowledge that incorporates the contextual elements of the clinical experience as well as the assessment of the accuracy of assessments and the effectiveness of interventions, represents the essence of critical thinking (O’Connor, 2006).

*Patient Care Management:* Patient care management in nursing is defined as the act of delegating, prioritizing, and supervising the care of one or more patients (Cherry & Jacob, 2011).

*Prioritization:* Prioritization of nursing care is the act of deciding what care should be done first and what should follow sequentially. There are three levels of prioritization in nursing that establish an ordered method of providing care that is on importance or urgency. These levels are typical taught as: *First Level* – Threats to patient’s immediate survival or safety (ABC’s – Airway, Breathing, and Circulation); *Second Level* – Changes in mental status, acute pain, acute urinary elimination, untreated problems that now require immediate attention (critically high or low lab values), infection risks, safety, or security; and the *Third Level* – Patient needs that do not fit into the other two categories such as monitoring medication side effects, lack of patient knowledge, long-term problems associated with activities of daily living (ADLs) (Cherry & Jacob, 2011).
Simulation: Simulation is “an innovative teaching method that uses technology and informatics, involve faculty guidance and feedback, and has the potential to increase the competency of nursing students and practicing nurses to provide safe patient care” (Gaberson & Oermann, 2010, p. 154).
CHAPTER TWO

REVIEW OF LITERATURE

This literature review is organized into four parts. The first section provides an overview of research associated with the three major nursing leadership constructs of prioritization, delegation, and patient care management. This research demonstrates the unique interaction between this constellation of three leadership skills and the acquisition of clinical competence in nursing. The second section addresses research on the traditional assessment methods used in nursing: multiple-choice testing (MCT) and oral questioning (OQ). The third section provides research on two new assessment methods: high-fidelity simulation (HFS), and low-fidelity simulation (LFS). Lastly, a summary of the literature reviewed concludes the chapter.

Nursing Leadership Constructs: Prioritization, Delegation, and Patient Care Management

“The fundamental tenant of nursing education is to ensure that students are competent” (Walsh et al., 2010, p. 2808). The theoretical context of clinical competence is based on past and present nursing literature. However, there is a lack of concept clarity within the literature regarding the notion of clinical competence. Clinical competence has historically been understood as a behavior or psychological construct (Walsh et al., 2010).

Practice professions such as nurses and physicians have standards of practice that govern their actions. These standards of practice establish the framework from which clinical competence is determined (Biag et al. 2010; Decker, Utterback, Thomas, Mitchell, & Sportsman, 2010; Hinton, Mays, Hagler, Randolph, Brooks, DeFalco,
Kastenbaum, Miller, & Weberg, 2012; NLN, 2010). Biag et al. (2010) defined clinical competence in the physician as a “mix of knowledge, attitudes and skills used to provide patient care and professional services” (p. 19). Mitchell and Sportsman (2010) reported that the National Council of State Boards of Nursing (NCSBN) reaffirmed its definition of competence for registered nurses as “the application of knowledge in the interpersonal, decision-making and psychomotor skills expected for the practice role, within the context of public health” (p. 120).

Benner et al. (2010) proposed educational reform in nursing that included performance assessments during undergraduate programs, licensure, and one year after licensure. The specific goal of these performance assessments is to determine and support decisions regarding a licensed nurse’s (or future licensed nurse’s) competence, continued competence, and remediation needs (IOM, 2010; Hinton et al., 2012; Decker, Utterback, Thomas, Mitchell, & Sportsman, 2010). Clearly, a need existed for the development of rigorous assessment methods of nursing clinical practice performance in educational programs and practice settings to aid in assessing competence (Decker, Utterback, Thomas, Mitchell, & Sportsman, 2010; Hinton et al., 2012).

Prioritization, delegation, and patient care management are three important explanatory constructs within nursing practice that determine clinical competence (Saccomano & Pinto-Zipp, 2011; Walsh et al., 2010). To be successful in their roles, regardless of their experience, registered nurses need to understand how to best prioritize, delegate, and manage the care of multiple patients (Kaplan & Ura, 2009). However, the relationship that exists between these three constructs is not clearly understood.
Prioritization of patient care needs to provide optimal safe nursing care guides the practice of all nurses regardless of their setting. Saccomano and Pinto-Zipp (2011) clearly distinguish the ability of nurses to prioritize care and delegate those needs to appropriate ancillary personnel as a crucial element of safe practice. For the nurse, knowing what care needs to complete first between multiple patients has become an integral characteristic of leadership within the practice setting given the diverse complexity of patient needs today (Kaplan and Ura, 2009).

Delegating a component of nursing care to unlicensed assistive personnel (UAP) is essential in the management of patient care for multiple patients (Saccomano & Pinto-Zipp, 2011). The RN remains accountable for the delegated patient care. However, the UAP to whom the care has been delegated assumes responsibility and answers to the RN. The interplay between leadership skill and confidence associated with delegating is both interesting and difficult to evaluate (Saccomano & Pinto-Zipp, 2011).

The opportunity for students to practice delegation and prioritization in the hospital setting has become increasingly unmanageable due to the lack of staff nurses to engage students in experiential learning opportunities (Walsh et al., 2010). Hospitalized patients have more complex health situations, and the lack of time present few if any occasions for nursing students to develop an understanding beyond the academic application of the terms (Schultz, Shinnick, & Judson, 2012). The advent of high-fidelity simulation provides students with an avenue through which to develop and practice prioritization, delegation and patient care management in a safe environment (Schultz, Shinnick, & Judson, 2012).
Prioritization

Lake, Moss, and Duke (2009) define *nursing prioritization* as “the decision by a nurse as to which nurse-patient interaction to address first among many potentially competing requirements and options” (p. 377). However, *prioritization* is something that nurses as well as other health professionals discuss in the literature as part of health-care delivery. To explore the professions tacit knowledge and understanding of prioritization, Lake, Moss, and Duke (2009) designed a study to more fully define the process of nursing prioritization as it is inferred, described, and discussed in the literature.

Previous research had suggested that a relationship existed between the patient care needs and the nurse’s clinical decision-making. Lake, Moss, and Duke (2009) designed a concept map using iterations of nursing prioritization located in the current literature to explore its origins. Their goal was to conceptually relate the terms and topics associated with prioritization and decision-making (Lake, Moss, & Duke, 2009).

The concept map was built from the main research areas identified in the Encyclopedia of Nursing Research (ENR): potentially relevant terms, and unlikely to be relevant terms. Potentially relevant terms were terms found in the literature that related to prioritization; unlikely to be relevant terms were terms associated with nursing clinical decision making, but not necessarily associated with prioritization. For example, “setting priorities” was a term found often in the literature that was a potentially relevant term and “STAT” was a term used in the literature considered to have implications to clinical decision-making, but not necessarily associated with the care delivery specifically provided by the RN (such as a STAT laboratory specimen draw or x-ray). Exclusion criteria had to be developed by the authors. A search using the Cumulative Index to
Nursing and Allied Health Literature (CINAHL) identified 738 citations. After eliminations and removal of duplicates, 343 citations remained. Books, theses, and papers from 1966 to 2003 (600) were also evaluated resulting in a final dataset of 461 items.

Thematic analysis revealed five main areas: how prioritization is taught; how prioritization is practiced; nursing prioritization in specialized practice and practice settings; the content of clinical decision-making related to nursing prioritization; and discerning nursing prioritization from the language used to describe and discuss clinical decision-making. While prioritization of the patient need for care is not presented as a formal concept in the literature, prioritization of patient care needs in nursing practice is associated with the nurse's understanding of a given patient’s situation and needs relative to other patients under the nurses care. The development of prioritization as an advanced skill has been thought to occur over time and with years of nursing experience. Further development of prioritization skills related to patient care needs has been correlated with nurses going into specialized nursing practice. For example, the oncology nurse specialist has enhanced prioritization skill with respect to the needs of oncology patients than the non-oncology nurse specialist. The implication is that the nurse will reframe prioritization of care needs based on the patient care needs within his or her specialized practice setting.

Three main contextual influences determine the nurse decision-making process regarding prioritization: time as a resource, resource constraints, and multidisciplinary interaction. These influences are associated with specialized practice and practice settings. Some of the language and terms found to describe clinical decision-making in nursing were cues, pattern recognition, heuristics, rules of thumb, and maxims. Research
of Lake, Moss, and Duke (2009) illustrated the need for clarity among nursing scholars surrounding prioritization. The authors discovered that the prioritization of patient care could be referred to as a discussion point within the literature. Further, frequently cited sources were recorded for 350 studies. Most notably, Benner (1984) was cited 182 times. The definition of prioritization according to Benner (1984) was noted as the nurse’s ability to judge the relative importance of different aspects of the situation. Finally, prioritization appeared as an embedded understanding that was discernible throughout the literature; however, it was more readily discussed in an interpretive perspective associated with nursing practice.

**Delegation**

The concept of *delegation* is not new. The act of delegating patient care tasks to others can be dated back to Florence Nightingale (Saccomano & Pinto-Zipp, 2011). Many consider delegation a component within a constellation of nursing leadership constructs (Thomas, Hodson-Carlton, & Ryan, 2011; Patrick, Laschinger, Wong, & Finegan, 2011; Saccomano & Pinto-Zipp, 2011; Kleinman & Saccomano, 2006). Others consider delegation as an entry-level nursing skill (Henry et al., 1994; Billay & Myrick 2008; Lillibridge, 2007; Epstein & Carlin, 2012). However, there is a consensus among the profession of the importance of delegation in providing safe patient care (Henry et al., 1994; Billay & Myrick 2008; Lillibridge, 2007; Epstein & Carlin, 2012; Thomas, Hodson-Carlton, & Ryan, 2011; Patrick, Laschinger, Wong, & Finegan, 2011; Saccomano & Pinto-Zipp, 2011; Kleinman & Saccomano, 2006).

Saccomano and Pinto-Zipp (2011) explored the relationship between RN leadership styles, demographic variables, and confidence in delegation within the
community teaching hospital. The four leadership styles were *directed leadership* (the leader provides guidelines about how to do the task), *supportive leadership* (the leader shows concern and support for the well-being of the subordinate), *participative leadership* (the leader asks for ideas from or participation by subordinates), and *achievement leadership* (the leader sets goals to enhance work performance). The purpose of their cross-sectional survey was to describe the nature of and relationship between the RN leadership style and confidence in delegating patient care tasks to UAPs. Prior to their research, there were no systematic studies in the nursing literature that describe the relationship between RN leadership styles and confidence in delegating patient care tasks.

Prior to the study, the researchers determined that for a medium-effect size of .25, an alpha level of .05, and a power level of .80 the necessary sample size was 158 participants. A convenience sample of 158 RNs employed at one acute care hospital located in the eastern United States provided demographic information and completed two questionnaires: *Path-Goal Leadership Questionnaire (PGLQ)* and the *Confidence and Intent to Delegate Scale (CIDS)*. The PGLQ is a 20-item questionnaire that evaluates the extent to which leadership styles are present using a seven-point Likert-type scale (1 = never, 2 = hardly ever, 3 = seldom, 4 = occasionally, 5 = often, 6 = usually, and 7 = always). The more predominant leadership style was determined by the higher score. The leadership scores could be further refined through examination of the participants’ score as above or below the designated common score. For all leadership styles, a low score was indicated when a participant scored five points below the designated common score and was designated as high if the participant scored five points above the common score. The PGLQ was reported as demonstrating good internal consistency based upon
Cronbach's alpha for the four leadership styles: directive (.83), supportive (.84), participative (.80), and achievement oriented (.87).

The CIDS is a 16-item scale developed by Parsons (1999) measuring three aspects of delegation: confidence with delegation decisions, present use of delegation decisions, and future intent to use delegation decisions in practice. The first seven items required nominal responses and the remaining items ask respondents to evaluate different aspects of their willingness to delegate using a 10-point Likert-type scale. Internal reliability of the CIDS was reported using Cronbach's alpha: confidence with delegation (.94), present use of delegation (.95), and intent to use delegation in practice (.95).

Potential participants answered four questions to determine their eligibility to participate in this study. Participant packets were placed in participants’ nursing unit mailboxes located throughout the hospital. Instructions in the packet included the location of locked collection boxes located on each nursing unit for the participants to deposit their completed survey packets. Data were analyzed using the Statistical Package for the Social Sciences (SPSS). The participants included 14 males (8.9%) and 144 females (91.1%) with an average age of 43.8 years (SD = 10.40). The majority of the nurse participants were Caucasian (68%). Native Americans (10%), African-Americans (10%), Hispanics (10%), and those ethnic groups marked “other” (2%) were also identified as the ethnicity of the participants. The average total years of nursing experience of the participants was 15.6 years (SD = 11.2), with approximately 11.9 (SD = 9.2) years spent at the hospital and an average of 8.1 (SD = 7.7) years spent on a specific unit. Lastly, 66 (41.8%) of the participants reported having earned a baccalaureate degree in nursing, 82 (51.9%) reported holding either a diploma (22 or 13.9%) or an associate degree (64 or
38%), and the final 12 (6.3%) held non-nursing bachelor degrees or master’s degrees but held an RN license.

No significant difference in total confidence in delegating patient care tasks to UAPs was revealed by a one-way ANOVA between subjects when grouped by dominant leadership styles ($F_{2,155} = 1.44, p = .24$). The researcher’s hypothesis that RNs who demonstrate a supportive leadership style would report more confidence in delegating patient care tasks to UAPs than RNs with either directive or participative styles was not supported. Table 1 shows the years of hospital employment of the participating nurses with supportive, directive, or participative leadership styles. The only demographic variable that differed significantly among the supportive, directive, and participative leadership style groups was the number of years that the nurses worked at the hospital ($F_{2,155}= 4.06, p = .02$).

Table 1. Years of Employment for Participants with Supportive, Directive or Participative Leadership Styles

<table>
<thead>
<tr>
<th>Leadership style</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supportive</td>
<td>11.03</td>
<td>8.95</td>
</tr>
<tr>
<td>Directive</td>
<td>14.35</td>
<td>9.47</td>
</tr>
<tr>
<td>Participative</td>
<td>7.23</td>
<td>5.80</td>
</tr>
</tbody>
</table>

*P ≤ 0.05

The supportive leadership style group of nurses was further categorized into low, medium, and high groups. Significant differences in the average total confidence scores emerged within this group of nurses ($F_{2,90} = 4.51, p = .01$). Bonferroni’s corrected post-hoc comparisons indicated that nurses were significantly more confident in delegating patient care tasks to UAPs if they demonstrated high supportive leadership styles than nurses that demonstrated a low supportive leadership style. Researchers used a one-way ANOVA between subjects to compare confidence in delegating patient care tasks in
nurses with less than a baccalaureate degree to those with a bachelor's degree or higher. The outcome of the analysis did not reveal any significant differences in total confidence scores based on educational preparation ($F_{1,156} = .43, p = .51$).

Further, the researchers’ findings did not support their hypothesis that RNs with at least a baccalaureate degree would report more confidence in delegating patient care tasks to UAPs than RNs with educational preparation less than a baccalaureate degree (BSN total confidence score 40.39, SD = 13.65; less than a BSN confidence score 39.24, SD = 7.84). Lastly, a 2 x 2 x 2 ANOVA between subjects evaluating the total confidence scores identified a significant interaction between the nurses educational preparation in years of clinical nursing experience related to their reported confidence ($F_{1,150} = 4.34, p = .04$).

Findings from the Saccomano and Pinto-Zipp (2011) research supported the claim that confidence in delegating patient care tasks is greater for nurses who have at least a bachelor’s degree. Further, their findings indicated that the accumulation of nursing clinical experience increased confidence in delegating patient care tasks for those nurses who possessed less than a baccalaureate degree. This study also demonstrated that a relationship existed between educational preparation, clinical nursing experience, and confidence in delegating.

**Patient Care Management**

The Quality and Safety Education for Nurses (QSEN) project was the result of the Institute of Medicine (2003) report that recognized the importance of transforming healthcare provider education as a means to improve patient outcomes. The competency domains and associated knowledge, skills, and attitudes of QSEN emphasized the need
for students to develop stronger patient management skills. Hassmiller (2010) emphasized quality and safety, evidence-based practice, and leadership along with clinical preparation as basic competencies of nursing to meet the demands of a diverse society. Hendry and Walker (2004) contended that delegation, time management, prioritization, and communication are difficult proficiencies to acquire for nursing students, and that nursing education may not be giving adequate consideration to these areas.

Most baccalaureate nursing programs require some form of a nursing leadership course prior to graduation. The leadership course may establish theoretical foundations for patient management. Yet, traditional delivery of leadership theories provides little opportunity for students to contextualize these concepts. Benner et al. (2010) began a discussion to transform nursing education as a means to bridge the gap between didactic information and clinical experience. The supporters of this conversation urged educators to pursue avenues that provide nursing students with the opportunity to apply leadership principles in managing patient care (Benner et al., 2010).

Sharpnack, Goliat, and Rogers (2011) developed simulated patient care management scenarios using standardized patients to teach senior nursing students leadership competencies. The purpose of their study was to evaluate the ability of baccalaureate nursing students to apply leadership principles and quality and safety competencies to a simulated patient care management scenario that incorporated standardized patients. Additionally, they examined the effects of the simulation experience on standardize computer-based achievement scores. The goal of simulation as a learning strategy was to foster application and synthesis of leadership concepts as well
as quality and safety standards taught in a BSN senior leadership course surrounding patient care management.

A two-group, posttest-only, randomized experimental design was used to evaluate 66 senior-level students on the application of leadership theory learning outcomes with a standardized computer-based assessment. The standardized computer-based assessment was the *Nursing Leadership Content Mastery* assessment developed by Assessment Technologies Institute (ATI). The ATI is a criterion-referenced assessment used to benchmark student performance. Content validity of the ATI was established. A two-tailed *t*-test was used to compare the mean standardized assessment (SA) scores for the group that completed the SA after performing the simulation exercise to those of the group that completed the SA before the simulation.

Development of the scenario occurred through revisions after a pilot study. Eight middle-aged and older adults from diverse ethnic and cultural backgrounds were recruited from the academic faculty, family, and friends to volunteer as standardized patients. Simulation scenarios were constructed through the use of learning modules commercially available from *Elsevier Simulation Learning System*. Minor modifications were made to the learning modules based on demographics of the standardized patients.

Prior to the beginning of the learning experience, student participants were oriented to the nursing unit and provided access to an electronic health record for reference as the scenarios unfolded. The individual patient scenarios in the simulation progressed by the use of laminated cards, which illustrated the progression of the patient's condition. After a hand-off report (report from the nurse being relieved of responsibility
for a patient to the nurse assuming responsibility for that patient) was provided by faculty, students began the scenario.

A planned disruption occurred immediately after students had planned their patient care management priorities for the day (a simulated patient fell out of bed). During this occurrence, faculty guided students in attending to this urgent patient need and reassigning resources soon after the incident. This planned event provided students with the opportunity to complete an incident report and call the house officer to report the patient's condition.

The results indicated that students who completed the simulation prior to taking the SA scored at the 83rd percentile for baccalaureate programs and 73rd percentile nationally on the assessments, while students who completed the SA prior to participating in the simulation scored at the 68th percentile in both cases. Subscale scores for those students participating in the simulation before taking the SA showed evidence of improved capacity to prevent safety errors, collaborate, provide continuity of care, and manage care more efficiently. The mean score for the group that completed the SA after the simulation was 72.3 (SD = 6.09), and the mean SA score taken before the simulation was 67.8 (SD = 4.1). A significant increase was noted for the group that took the SA after the simulation, $t (64) = 3.55, p< 0.01$.

Limitations of this study included the author-developed tool used by the standardized patients and participating interprofessional team members to evaluate the experience required measures of validity and reliability. Further, the small sample size and posttest experimental design failed to address potential group differences that may have influenced the SA outcomes. Lastly, it was found that a more robust study design
and stronger statistical analysis to establish valid and reliable assessment tools as well as analysis of the impact of the simulation on subscale assessment scores would improve and strengthen the research (Sharpnack, Goliat, & Rogers, 2011).

Some researchers perceived patient care management as a component within a constellation of leadership skills similar to that of prioritization and delegation (McCarthy & Murphy, 2008; Saccomano & Pinto-Zipp, 2011; Sharpnack, Goliat, & Rogers, 2011). This point requires the need for a consistent definition of patient care management to allow for comparisons among various studies in the literature. An increased emphasis on broadening leadership skills in baccalaureate education comes from a number of sources (Benner et al., 2010; Patrick et al., 2011). These sources state that new graduate nurses need clinical reasoning skills to manage complex patients. The ability to manage complex patients using a high degree of clinical reasoning is termed clinical competence (IOM, 2003).

Traditionally, the clinical competence of nurses and physicians consisted of obtaining licensure followed by on the job training. In the past, patient care management skills of nursing and medical students were assessed via paper and pencil exams, oral exams, long and short cases, ward observations, supervisor reports, and chart audits (Biag et al., 2010). Harden et al. (1975) quickly changed the landscape of evaluating patient care management competency in medical students by developing the OSCE format of assessment. Biag et al. (2010) took advantage of the Western Alliance for Assessment and International Physicians (WAAIP), which was constructed in an OSCE format to gather data to evaluate clinical competence.
The WAAIP project was created to develop and field test assessment processes for international medical graduates to determine their practice readiness. This Canadian project was implemented over 15 months, started in 2005, and is presented here as a foundational MTMM study design on assessing patient management in health care using MCQ, oral questioning, skill stations, and clinical observations.

Biag et al. (2010) collected data in two parts. The first part was a 150-item multiple-choice exam to test declarative knowledge and a 14-station OSCE for testing clinical as well as communication skills. The second part included supervised clinical practice visits for three months in which participants were assessed using the Physician Achievement Review (PAR), in-training evaluation reports (ITERs), clinical evaluation exercises (miniCEX), and OSCE stations using HFS and LFS.

During the OSCEs, patient management components were assessed by a physician assessor. The communication skills of participants were assessed by both the physician assessor and the standardized patient (SP) using separate instruments. The mini-CEX (used as a measure of doctor-patient relationship) included clinical judgment and medical interviewing skills. In total, 415 clinical evaluation exercises (mini-CEXs) were completed, average 17.3 per participant. From the ITERs, the clinical competence items included assessment of basic science knowledge, clinical knowledge, data gathering skill, physical examination skill, problem formulation and differential diagnosis, diagnostic and therapeutic planning, clinical judgment and decision-making, performance under emergency, procedural skills, and self-assessment ability.
As shown in Figure 2, Biag et al. (2010) constructed the MTMM based on the Campbell and Fiske (1959) technique. The column and row headings represent the method and the trait used for the matrix from that instrument.

The darkest cubes in Figure 2 represent the validity coefficients for the research. The validity coefficients are the correlations of different attributes across different methods and those correlation coefficients ($r$) represent different methods for the same trait. The reliability of scores for each method is represented in Figure 2 as the lightly shaded cubes. The instrument scores are presented here for review: multiple-choice questions (Cronbach's alpha = .86), OSCEs ($K$ = .794 assessors), ITERs (Cronbach's alpha = .859 assessors) and mini-CEXs (Cronbach's alpha = .855 assessors).
range from .34 to .85), miniCEX (Cronbach's alpha range from .54 to .98), and PAR (Cronbach's alpha range from .66 to .96). The reliabilities of the assessment instruments were in the adequate to good range across all four methods. Lastly, the clearest evidence for both convergent and divergent validity was for clinical competence, followed by doctor-patient relationships, and communications as assessed by the ITER and PAR (but not with the OSCE).

The results suggested substantial method specificity. Evidence of divergent validity of traits assessed by the ITERs, mini-CEX, and PAR (4 out of 6; 67%) met the criteria set forth by Campbell and Fiske (1959). Overall, the results presented evidence of discriminant validity. ANOVA was used to further analyze the MTMM to identify method and trait effects. Percentages were calculated on the three constructs from the four methods based on the maximum scores to standardize the results. ANOVA revealed that 23.7% of the variants could be attributed to the method factors in the MTMM. The variance attributed to the physicians was higher (at 35%) than the physician times method variance.

The Biag et al. (2010) study underscored the importance of defining and investigating the validity of assessment methods employing the MTMM approach associated with the evaluation of patient care management traits. Moreover, the MTMM approach could be used to estimate the degree of evidence for validating complex constructs in nursing education such as the leadership skills mentioned. The effectiveness of MCQ, oral examinations, and simulation in evaluating nursing student mastery of leadership skills has significant practice implications and warrant investigation.
Summary of Nursing Leadership Constructs Literature

In summary, literature presented on the three leadership constructs (prioritization, delegation, and patient care management) illuminated the difficulty that exists for nursing faculty to effectively assess student ability to demonstrate behaviors associated with each of the three constructs (Biag et al., 2010; Sharpnack, Goliat, & Rogers, 2011). While researchers have struggled to determine components that delineate behaviors of leadership surrounding patient care delivery (Lake, Moss, & Duke, 2009; Sharpnack, Goliat, & Rogers, 2011), innovative strategies have been developed to elicit opportunities outside the clinical environment to gauge student ability surrounding these three exploratory constructs (Saccomano & Pinto-Zipp, 2011; Schultz, Shinnick, & Judson, 2012; Sharpnack, Goliat, & Rogers, 2011; Walsh et al., 2010).

Research presented by Lake, Moss, and Duke (2009) illustrated three main contextual influences that determined the nurse decision-making process in care delivery. These contextual influences all have specific ties to each of the three constructs. Clear operational definitions of each construct are need to develop avenues with which to teach, methods to assess, and robust metrics with which to research further into these important nursing leadership areas.

Traditional Assessment Techniques

Assessment is an integral part of the learning process in nursing education. The primary techniques used to evaluate nursing students are multiple-choice tests (MCT) and oral questioning (OQ). This section of the literature review is comprised of research that addresses the specific use of MCT as it relates to the assessment of content and OQ as it
related to the assessment of clinical competence. A summary of the literature regarding these two traditional forms of assessment in nursing education concludes the section.

**Multiple-Choice Testing**

Multiple-choice testing is used in nursing education to evaluate student learning primarily in nursing theory courses. Over the last decade, there has been an increase in research into the format, design, and construction of MCTs (Considine, Botti, & Thomas, 2005; Lee, Liu, Linn, 2011; Morrison & Walsh, 2001; Paxton, 2000; Tarrant & Ware, 2008). In recent years, textbook publishers began to provide nursing faculty with a bank of multiple-choice questions to use in constructing tests to evaluate student knowledge (Tarrant & Ware, 2008).

Nursing licensure is also closely connected to MCTs. The format of the national licensing examination – RN (NCLEX-RN) has been modeled after MCTs for decades. Because licensure is so closely connected to MCTs, nursing faculty feel compelled to use this format almost exclusively in evaluating student learning. Further, nursing program success is evaluated by first time student NCLEX pass rates. Consequently, schools of nursing are confronted with the dilemma of structuring curricula specifically for pass-rate success, or structuring curricula to educate nurses who can meet the competency standards as assessed by multiple-choice examination.

The use of MCTs in nursing education is to measure knowledge after an educational intervention (Considine, Botti, & Thomas, 2005). The majority of the literature regarding the format, structure, validity and reliability of MCTs is found in medical education, psychometric testing, and psychology literature. Little research has been conducted regarding the specific use of MCT for assessment in nursing education.
Evidence for construct validity of MCT has not been researched extensively. The focus of research has been on reliability, which is necessary but not a sufficient condition for validity (Berkow, Virkstis, Stewart, & Conway, 2009; Decker, Sportsman, Puetz, & Billings, 2008; Hickey, 2009; McKeon, Norris, Cardell, & Britt, 2009; Williams & Day, 2009).

In most nursing programs, the amount of content requiring assessment can be overwhelming. A considerable portion of nursing faculty time is spent in developing written assessments (Rushton & Eggett, 2003; Masters et al., 2001). Since it has already been established that a substantial portion of those assessments will likely contain MCTs, it is important that educators are basing their practices on the best available research. Additionally, student numbers are generally increasing to meet practice shortages, while the number of available nursing faculty is dwindling (Dadgaran, Parvizy, & Peyrovi, 2012). The result is fewer nursing faculty attempting to construct multiple-choice tests covering large volumes of content (Tarrant & Ware, 2010; Masters et al., 2001).

Test developers have suggested that multiple-choice items can be used to evaluate higher-order thinking ability (Wendt & Kenny, 2009). The proposed method of such assessment is to focus on items that move away from recall or comprehension-level questioning toward constructed response items. Constructed response items require that the answer be written or typed. Items not limited to a single response may move assessment from recall to application-analysis levels of knowledge assessment.

Tarrant and Ware (2010) compared the psychometric properties of three-and four-option multiple-choice questions in nursing examinations. The purpose of their study was to determine if the time required to develop multiple-choice tests could be reduced
without reducing the reliability and validity of the examination. Data for this study were
collected on two cohorts of students in an undergraduate public health nursing course
over two subsequent academic semesters (fall 2006 and spring 2007).

This research compared and examined the psychometric properties of four-option
items with the exact same items rewritten as three-option items. The first examination
consisted of 50 four-option items administered to 36 students. The second examination
consisted of 70 three-option items administered to a subsequent cohort of 106 students.
Using item-analysis data from the four-option examination, authors were able to reduce
the number of options to three by eliminating the least frequently selected distractor.

Tarrant and Ware (2010) examined the item-analysis statistics of item difficulty
and discrimination, the distracter performance statistics, test reliability coefficients, and
mean item scores. The mean item difficulty and the discrimination index of the 50 items
on the two examinations were compared using a paired t-test. For both exams, the authors
evaluated distracter performance. Lastly, the authors evaluated the effect of removing the
least frequently selected option by comparing individual distractor performance on both
examinations using chi-square statistics.

The total number of students tested was 142 over both examinations and
semesters. On the original examinations, the overall mean test scores and the range of test
scores were similar for both the fall 2006 and spring 2007 cohorts (70.3 and 69.7). The
pass rate for the spring 2007 cohort was marginally lower than that of the fall 2006
(94.4% and 97.2%). However, the 41-item three-option items was more reliable ($KR_{20} =
.71$) than the four-option items ($KR_{20} = .65$) (Tarrant & Ware, 2010). Although the data
were not shown in the article, mean item difficulty was reported as the 41 three-option
items as more difficult than the four-option items (.70 +/- .15 versus .73 +/- .14). The difference was not reported as statistically significant ($t = 1.95; p = .06$).

The results of this study added to the growing body of research supporting three-option items. The differences in item difficulty and discrimination between four-option items and the same items when written as three-option items were small and statistically non-significant (Tarrant & Ware, 2010). Non-significant results are just as important as significant results (Huck, 2008; Popham, 2000). The finding that three-option items performed equally as well as four-option items can have substantial impact on the practice of item-writing in nursing education in terms of faculty time.

Bailey, Mossey, Moroso, Cloutier, and Love (2011) posited that nursing educators have consistently made the argument that well-constructed multiple-choice tests are central to nursing educational assessments. Research conducted by Leung, Mok, and Wong (2008) evaluated the influences that assessment methods had on the learning approaches of nursing students. The authors contended that students’ experience in assessment can influence how they approach their learning.

Research by Leung, Mok, and Wong (2008) suggests that students adopt a surface approach when assessment predominately focuses on a demand for recall of factual details (rote learning). Students were more likely to employ a deep approach when assessment demanded higher levels of cognitive processing. The deep approach and the surface approach of learning represent the dichotomy in the orientation to learning between an emphasis on understanding the material versus an emphasis on rote recall.

Additionally, Leung, Mok, and Wong (2008) proposed the increasing use of multiple-choice examinations as an assessment method in higher education as having a
strong association to students using the surface learning approach rather than the deep learning approach. The aim of this study was to examine the impact of a high-quality multiple-choice test on the learning approaches of nursing students enrolled in the mental health nursing course. The study used a pretest/post-test design to examine the effect of assessment on the student's approaches to learning. A comparison was then made between the learning approaches they adopted at the beginning of the mental health nursing course and those used by the end of the course. The researchers used focus group interviews that were conducted to solicit students’ feedback on assessments that could facilitate the desired learning outcome in nursing.

To explore this further, the study was designed to investigate student learning approaches toward mental health nursing content at a Hong Kong University. The participants included 136 nursing students from a higher diploma (HD) in nursing program and 142 nursing students from a baccalaureate nursing program (BSN). The nursing assessment components of the course included a group project, problem-based learning, an online quiz, and a final examination. The university’s nursing assessment item test bank provided some of the multiple-choice questions for the online quiz and all of the multiple-choice questions for the final.

The Revised Two-factor Study Process Questionnaire (R-SPQ-2F) was used to assess the learning approaches of the students. It had two main scales, the deep approach (DA) and the surface approach (SA). It also had four subscales, deep motive (DM), deep strategy (DS), surface motive (SM), and surface strategy (SS). The R-SPQ-2F contained 20 items on a five-point Likert-type scale ranging from “always true to me” to “never true to me.”
Confirmatory factor analysis was used to confirm a good fit with the two-factor structure. The calculated Cronbach's alpha for the DA and SA scales were .73 and .64, respectively. Lastly, a five-item questionnaire using a 10-point Likert-type scale (ranging from "strongly disagree" to "strongly agree") was used to evaluate students' perceptions of the assessments they had undergone for the mental health nursing course.

The demographic results included a total of 113 BSN students and 103 HD students completed both the pre-and post-test (R-SPQ-2F). The response rate represented 79.6% (BSN) and 75.7% (HD) of the students enrolled in the course. The BSN students were comprised of 79.6% female (90) and 20.4% male (23). The HD students were comprised of 76.7% female (79) and 23.3% male (24). The ages of both groups ranged from 21 to 24 years old (BSN) and 19 to 27 years old (HD).

A separate two-step cluster analysis was performed for the BSN and the HD students. Three clusters were identified for the BSN students; Cluster “A” (BSN) consisted of 17.7% (n = 20), cluster “B” (BSN) 54.9% (n = 62) and cluster “C” (BSN) 27.4% (n = 31). Three clusters were also identified from the HD students. Cluster “A” (HD) consisted of 29.1% (n = 30), cluster “B” (HD) 24.3% (n = 25), and cluster “C” (HD) 46.6% (n = 48).

The BSN and HD students clustered the same and demonstrated predominately a deep approach to learning. Both group mean scores had no significant difference in the surface learning scores that occurred between the pre-and post-tests in the cluster “A” (BSN), cluster “B” (BSN) and cluster “A” (HD) students.

The relationship between students’ academic achievement and their approach to learning was also examined. The Spearman's rank correlation coefficient indicated that
there was no relationship between BSN students’ achievement mean score and the deep learning score. However, a weak relationship existed between these factors among the HD students (.187). Further, a significant negative correlation existed between the surface learning score (pre-and post-test) and academic achievement for the BSN students (pre-test: \( r = -.208; p = .025 \); post-test: \( r = -.259; p = .003 \)). Lastly, a negative correlation existed between the surface motive score (post-test) and the academic achievement of the HD students (\( r = -.188; p = .032 \)).

A one-way ANOVA was used to compare the academic achievement of students among the three clusters within the BSN and HD students. A statistically significant difference existed in the academic achievements among all three clusters of the BSN students (\( F = 4.8; p = 0.01 \)). Further, the BSN and HD students who predominately took a deep approach to learning (cluster “C”) achieved the highest academic scores. Conversely, students adopting a more surface approach to learning achieved the lower academic course scores.

The results of this research suggest that high-quality multiple-choice tests constructed with case scenarios were regarded to be useful in facilitating the critical thinking of nursing students based upon the higher academic scores of deep approach to learning clusters versus the lower academic scores of the surface approach student clusters. Leung et al. (2008) cautioned that the results did not necessarily indicate the effect of multiple-choice test causing more use of surface learning approach by the students. They proposed that the surface learning approach was a strategy taken by the students to cope with the heavy workload of various nursing courses within the strict time constraints of each program.
The Leung, Mok, and Wong (2008) study is a good example of a program that adopted the uniform approach toward the use of multiple-choice questions for assessment of nursing students. However, it also demonstrates the dilemma faced by nursing faculty regarding construction of high-quality multiple-choice tests to facilitate acquisition of content. Leung, Mok, and Wong (2008) demonstrated that although the faculty objective was to increase the deep learning approach to the content, the students in fact developed a more surface learning approach to survive the course. This could explain why graduate nurses who pass the NCLEX cannot demonstrate nursing leadership skills at the bedside.

**Multiple-choice test critique.** The multiple-choice test (MCT) is a common format used to assess students in nursing and health science disciplines (Tarrant & Ware, 2008; Clifton & Schriner, 2010). The MCT format allows faculty to efficiently assess large numbers of students and to test a wide range of content (Tarrant & Ware, 2008; Downing, 2002; McCoubrie, 2004; Clifton & Schriner, 2010). If properly constructed, multiple-choice questions are able to test higher levels of cognitive reasoning and can accurately discriminate between high- and low-achieving students (Tarrant & Ware, 2008; Downing, 2002; McCoubrie, 2004; Schuwirth & van der Vleuten, 2004; Clifton & Schriner, 2010). The major critique of multiple-choice test use in nursing education is that nursing educators do not have the skill or time to construct high-level items. The results are item writing flaws. Two studies are presented that examine item writing flaws (Tarrant & Ware, 2008; Clifton & Schriner, 2010) in nursing education.

Nursing, psychology, and educational literature were reviewed for the use of MCT using multiple online databases. A consistent theme arose regarding concerns associated with the skill needed to construct a quality multiple-choice examination.
Concerns regarding the potential dangers of test reliability and validity of such exams when deviations from multiple-choice question item writing tenets were also noted (Tarrant et al., 2006; Tarrant & Ware, 2008; Haladyna & Downing, 1989).

Tarrant and Ware (2008) conducted a study to examine the impact of item writing flaws on student achievement within the nursing program at an English-language University in Hong Kong. Ten multiple-choice tests were selected from a larger sample of examinations (n = 121) administered to undergraduate nursing students over a five year period (2001-2005). All test items were reviewed for item writing flaws by a four-person consensus panel.

Tarrant and Ware (2008) discovered item writing violations across all spectrums related to Gronlund and Waugh’s (2009) checklist for evaluating multiple-choice items. For example, eight violations were most commonly found across all tests, accounting for a total of 85% of all violations. These eight common violations included an unfocused stem (n = 70, 17.5%), unnecessary information in this stem (n = 49, 12.2%), a negative stem (n = 52, 13%), no correct or greater than one correct answer (n = 43, 10.7%); implausible distracters (n = 40, 10%), greater detail in correct option (n = 35, 8.7%), logical clues (n = 27, 6.7%); and repeated words (n = 25, 6%). Lastly, violations associated with grammatical clues and lost sequence in options were very rare, accounting for only 2% of all the violations found.

Findings by Tarrant and Ware’s (2008) study illustrated the impact of construct-irrelevant variances on student achievement. Downing (2002) defined construct-
irrelevant variances as the introduction of extraneous variables (e.g. item-writing flaws, test-wiseness) that are irrelevant to the construct being measured and can increase or decrease test scores for some or all students. Construct-irrelevant variants prevent the proper interpretation of test scores therefore reducing the construct validity of the assessment (Tarrant & Ware, 2008; Downing, 2002).

A similar study conducted by Clifton and Schriner (2010) examined data from three adult health courses final examination questions. The purpose was to determine if faculty were paying attention to the quality of their test items. Researchers also sought to determine the cognitive levels of test items and their frequency in the medical surgical nursing courses, compare difficulty levels with cognitive learning levels, and examine discrimination values and their relationship to distractor performance.

Clifton and Schriner (2010) calculated descriptive statistics and the percent of different variables. Comparisons between course levels, cognitive levels, and difficulty levels were examined for expected positive relationships. For example, the researchers looked for a leveling of the items based on the course; the more difficult test items would be in the higher level nursing courses. Discrimination indexes in relationship to distracters and correct options were examined along with the percentage of distracters not chosen by the students.

The results of the 60 multiple-choice questions included 17% knowledge, 37% comprehension, 32% application, and 15% analysis. The goodness of fit analysis determined significant difference between cognitive groups ($p = .038; df = 3, \alpha = .05$), suggesting that there was a significant difference from a normal, or equal distribution of 15 questions or 25% in each cognitive category.
Further, test item mean difficulty levels were calculated for the sample questions and the total number of course questions for each examination. This computation yielded difficulty values of 63% being too easy ($p \geq .80$), 14% being moderately easy ($p = .71$), 21% being desirable ($0.30 \geq p \leq 0.70$), and 2% were too difficult with indications of severe flaws ($p = 0 - .29$) (Clifton & Schriner, 2010).

The discrimination values calculated by Clifton and Schriner (2010) revealed that many test items were in need of closer assessment. Distracters should be plausible, based on common misconceptions of the correct answer, and should distract the informed examinee (Bourke & Ihrke, 2012; Clifton & Shriner, 2010; Fishman & Galguera, 2003; Popham, 2006; Twigg, 2012). Clifton and Schriner’s (2010) discrimination indices affirmed that while nursing faculty were pulled in many directions, proper time allotment for quality item analyzing, reviewing, and rewriting of items needed to remain a high priority if this form of student assessment was to be used.

Multiple-choice test items are by far the most widely used form of student assessment in nursing education (Hickey, 2009; Tarrant & Ware, 2008). Perhaps the popularity of multiple-choice test items use in nursing education is simply due to the close tie that multiple-choice items have with the nursing licensing exam (NCLEX), or due to the ease of use associated with multiple-choice items (Considine et al., 2005; McKeon et al., 2009). Regardless, most literature regarding the use of multiple-choice items to assess student knowledge (in nursing educational literature) was focused on item writing procedure (Tarrant & Ware, 2010) or validity and reliability in medical education, psychometric testing, and psychological literature (Berkow et al., 2009; McKeon et al., 2009; Sportsman et al., 2008; Williams & Day, 2009).
Research presented by Tarrant and Ware (2010) supported that three-option multiple-choice test items performed equally as well as four-option items. While such findings did impact nursing education practice, the substantive result of such research lends to the argument that properly structured multiple-choice test items; no matter three- or four-option items; have validity and reliability, which is central to assessment in nursing education (Bailey et al., 2011). Further, nursing researchers posited that a considerable portion of nursing faculty time is spent in the development of multiple-choice test items (Masters et al., 2001; Rushton & Eggett, 2003) and others caution that over use of multiple-choice questions could produce rote recall versus deep understanding of the presented content (Leung, Mok, & Wong, 2008).

To address the concerns of surface learning and the time constraints of nursing faculty to develop high-quality multiple-choice tests, Leung et al. (2008) designed research to explore the types of learning (surface or deep learning) nursing student employ based on the rigor of multiple-choice test items. While the objective of the Leung et al. (2008) research was to demonstrate an innovative faculty drive to promote deep learning approach to learning nursing content, the results demonstrated that students developed more surface learning approaches as a course surviving mechanism. Research such as those presented on multiple-choice test use in nursing education echo the critique of its use.

Critique of multiple-choice item use in the literature ran the gamut from item writing flaws (Haladyna & Downing, 1989; Tarrant et al., 2006) to faculty use of textbook publisher multiple-choice test item test banks to generate examinations (Clifton & Schriner, 2010; McCoubrie, 2004). Research presented by Tarrant and Ware (2008)
demonstrated the impact of multiple-choice item construct and its impact on student achievement. Research presented by Clifton and Schriner (2010) illustrated the importance of faculty item analysis in determining the quality of their test items. Results presented by Clifton and Schriner (2010) also affirmed that time allotted for multiple-choice test item analysis must be a high priority if this form of student assessment was used with such regular frequency in nursing education.

**Oral Questioning**

A review of the literature on oral examinations used in nursing education was limited. Studies examining the use of oral examinations in nursing education were very rare. Oral examinations date back to antiquity and are widely used by medical schools, residency programs, and credentialing boards throughout the world (Raymond & Luciw-Dubas, 2010). During ancient and medieval times, educators measured and evaluated the work of their students through oral questioning and observation (Ali & Ali, 2010). According to Ali and Ali (2010), the only country in ancient time to use an extensive system of written examinations to evaluate educational achievement was China. As education became more formalized, established European universities during the Renaissance frequently evaluated students through public discussions whose topics were often controversial. In 1836, the University of London established external examinations for degrees. By 1845, written versus oral examination was the controversy in academia.

Current research by O’Connor (2006) contends that oral examinations or questioning remain as an integral component of clinical learning. Clinical instructors ask questions of students to glean insight into the adequacies of their preparation for the clinical assignment, their ability to manage the care demands of the assignment, and their
understanding of the dynamics underlying the patient's situation. Research by Davis and Karunathilake (2005) suggests that oral examinations are a traditional form of assessment in health care related education. Typically in the clinical setting, the oral exam takes the form of an interview or discussion between the examiner and the student away from the patient. The oral examination is proposed to assess knowledge, to probe the depth of knowledge, and to test other qualities such as mental agility of the student (Davis & Karunathilake, 2005). Further, the aspect of assessing higher-order student knowledge was also possible outside the clinical setting with oral (Socratic) questioning. While OQ occurs during each clinical practicum that is connected to a nursing theory course, its most common form is the Socratic Method.

Socratic questions are open-ended, with multiple responses possible. The questions ask students to consider different alternatives and varied points of view and to defend their choices. Usually, no one answer is correct. After exploring these answers with students, the instructor can ask them to make connections to other clinical scenarios and to generalize learning from one patient and clinical situation to others (Billings & Halstead, 2012; Gaberson & Oermann, 2010). Socratic Method of questioning is rigorous, requires active listening, and assists clinical faculty in finding contradictions or inaccurate ideas that need refining and further scrutiny.

The Socratic Method of questioning is rarely discussed in nursing literature. The bulk of literature available on oral examinations centers on medical education and their use of these exams for assessment of clinical competence of students (Davis & Karunathilake, 2005; Lunz & Bashook, 2008; Raymond & Luciw-Dubas, 2010). Gaberson and Oermann (2010) described oral questioning of nursing students in the
practica setting as a practical means of sharing information, developing critical thinking skills, and evaluating student learning. One of the more recent studies using oral examinations in nursing education was conducted by Rushton and Eggett (2003).

Rushton and Eggett (2003) examined the effects of testing style on student learning by comparing five groups of students enrolled in the fifth medical surgical course of a nursing program at a large private university. Group one was comprised of 55 students who were given three written midterm tests and a written final examination. The second group contained 150 students who were given written midterm and final examinations. Group three comprised of 45 students who were given a final oral examination only. Group four comprised of 92 students who were given a written examination and an oral examination. The final group (group 5), comprised of 47 students who were given a written examination with questions resembling NCLEX-type questions and an oral examination.

Rushton and Eggett (2003) examined the effects of testing style on student learning by comparing five groups of students enrolled in the fifth medical surgical course of a nursing program at a large private university. Group one was comprised of 55 students who were given three written midterm tests and a written final examination. The second group contained 150 students who were given written midterm and final examinations. Group three comprised of 45 students who were given a final oral examination only. Group four comprised of 92 students who were given a written examination and an oral examination. The final group (group 5), comprised of 47 students who were given a written examination with questions resembling NCLEX-type questions and an oral examination.
The oral examination was a test consisting of nine case studies containing 132 questions. Questions were designed to test the student’s ability to recall information, synthesize given information into separate concepts, and apply the information and concepts in clinical nursing situations. Students in this group were given six weeks to prepare for the oral examination, which took place in a faculty office and was administered and graded by the same instructor. At the beginning of the semester, students were given the nine case studies and told that the examination question would be taken from the 132 questions posed in the case studies. During the examination, each student chose one question from a hat which contained all 132 questions from the case studies and had 10 minutes to respond.

The oral examination only group scored significantly higher than the other groups on the four-point scale. There was no significant difference in the average grade for the other groups on the four-point scale. The inference from this result is that the students who took only the oral examination performed significantly better than those who took other types of examinations. Further, there was no significant difference between the written examination and oral examination combination and the single midterm with the final combination student groups.

The overall results of the study indicated that more effective study habits were seen in preparing for the oral examination. Improved learning, increased knowledge base, critical thinking, and application of information to clinical situations were noted through significantly higher examination scores with the oral examination in any other form of examination. Conversely, no students said that studying for the oral examination took more time than studying for a written examination.
Two significant disadvantages to the oral examination became obvious. First, faculty stated that it was impossible to be absolutely objective in grading the student during the administration of the oral final. Second, the actual examination administration was very time-consuming for the faculty. Because this was a research project, one faculty member assumed responsibility for administering all of the oral examinations. Even at only 10 minutes for each student, examining 24 students took about four hours. However, this is probably not more time than it takes to grade any test with a subjective or narrative component.

Oral examinations have been used successfully in other professional disciplines and can focus on a variety of different topics. The ability to clearly articulate evidence-based practice skills, such as critique and application, lend themselves well to the oral examination process (Burman et al., 2007). Kelm (2001) contended that oral examinations in medical education provide students the opportunity to show their strengths and provided the examiner's with a clear and complete impression as possible of the candidates knowledge and abilities. However, Lunz and Bashook (2008) conducted a study on the relationship between candidate communication ability and oral certification examination scores of 90 candidates in which they concluded that the candidates communication ability had little relationship to candidate performance on high-stakes, case-based oral examinations.

Burman et al. (2007) examined evidence-based practice with the implementation of an oral examination as an innovative approach of nurse practitioner (NP) education. To use evidence-based practice, NPs must be able to locate research and practice guidelines, understand and evaluate the analysis used in research, and articulate their potential
application of such research. While not a formalized study, the article presented an
innovative approach to integrating oral examinations into a nursing course. Oral
examinations were developed that focused on the students’ ability to use concepts and
skills of evidence-based practice.

Prior to the examination, students were provided a research article to critique. Additionally, students were provided with a patient case study in which they needed to apply the principles of the research article previously distributed. Students could prepare for the oral examination using any resource. During the oral examination, students met with a faculty member to respond to questions requiring critique of the research and thorough explanation of how it would influence their practice.

Implementation of the oral examination process appeared to be simple. Several weeks prior to the examination, students were notified of the general topic. Instructions for the examination and preparation hints were provided at that time. Students were then scheduled for examinations, which occurred on a specific day and time. The oral examinations were scheduled halfway or two thirds of the way through the course to allow students to gain some confidence in the content.

One hour prior to the oral examination, students were provided a case study and a related article to critique. Students were told to focus on how they would or would not use the information from the article given the situation outlined in the case study. Faculty encouraged students to bring their required textbook on evidence-based practice as a reference resource.

The students were graded on their ability to articulate their critique of the article and the application of it to the case study provided. The faculty member grading the oral
exam took extensive notes while the student was talking. However, because this assessment technique was new to the program, the oral examination accounted for only 10% of the final grade for the course.

Most of the NP students received an “A” or a “B” on the oral examination. Although the oral examination represented a small percentage of the overall course grade, it was strongly correlated with the final grade ($r = 0.86$). However, some students who did well on the written course work were challenged by the verbal presentation required in this kind of examination. In addition, other students who did not perform as well on the written work did well presenting orally. Only two students received the grade of either a “C” or a “D.” Both of these students also had trouble on at least one other assignment.

The results of this educational innovation appeared overall positive for the program. The strength of the oral examination is that it not only required students to display expert clinical judgment as well as decision-making, but also required students to be able to articulate their rationale for the decision. Moreover, the oral examination also conveyed to the students that faculty considered evidence-based practice an important clinical skill.

The advantages to oral examinations include the ease with which such an examination can be implemented. Also, faculty had the ability to identify individual and group problems associated with a topic if all of the students in the course are being tested on the same topic. Faculty also was able to evaluate higher-order student knowledge with the oral examination.

Finally, the oral examinations when implemented into a nursing course conveyed an overall sense of rigor. Similar to the addition of standardized patients, OSCEs,
advanced skills training, and clinical faculty mentoring, the addition of oral examinations provided a multi-method approach to student assessment. The multi-method approach of assessment appeared to be a more accurate method of assessing both learners in health professional training programs as well as the effectiveness of the training program (Carraccio & Englander, 2000; Chambers, Boulet, & Gary, 2000; Haijazi & Downing, 2008).

Critique of oral questioning/examinations. The major critique that existed in the literature regarding oral questioning was its poor reliability. Oral examinations for the assessment of clinical competence of medical students in their clinical rotations have long been regarded to have poor reliability (Davis & Karunathilake, 2005; Lunz & Bashook, 2008; Raymond & Luciw-Dubas, 2010). Daelmans, Scherpbier, Van Der Vleuten, and Donker (2001) designed a study for the purposes of investigating the feasibility of increasing the number of oral examinations per student and to estimate the reliability of oral examinations. The researchers became concerned over the expensive alternative to oral examinations which was the objectively structured clinical examinations (OSCEs) using standardized patients or human patient simulators. Their goal was to determine if increasing the number of oral examinations per student would increase the reliability of the examination. The study was conducted using 52 student participants at the Department of Internal Medicine at the Vrije Universiteit in Amsterdam, The Netherlands.

Students participating in this study were required to engage in a patient-based oral examination twice a day for a maximum of five days in a row during their last week of clinical rotation. Participants were also required to have a different examiner for each oral
examination per day. Five examiners were involved in the study. Each examiner was responsible for two oral exams per student, one day during the week of the study.

Statistical analysis associated with the feasibility of raising the number of oral exams and examiners per student was studied by calculating the frequency of oral exams and examiners per student. The estimation of reliability associated with the oral examinations was considered nested within the persons involved since the context varied from oral exam to oral exam. After variance component estimation, generalizability coefficients were estimated as a function of the actual and projected number of the oral exams required to secure reliability of the method.

Daelmans et al. (2001) examined the generalizability coefficients to determine the interval for the number of oral exams that were required for students to have an average grade between the averaged item grades and the global judgment grades at a passing level. The result was coefficients for averaged item grades and global judgment grades at six oral exams were .62 and .72, with an average of both scores being .72. Therefore, it was determined that at six oral exams, students could achieve a higher than average score on both the oral exam in the global judgment score.

Fifty-two students participated in the study; thirty-nine students (75%) engaged in at least six oral exams per day. The results demonstrated that it was possible to moderately increase the number of oral exams and the number of examiners in the day-to-day practice of an outpatient clinic. The goal of six to eight oral exams conducted by four to five examiners during one week was feasible in practice and demonstrated by this study.
The reliability of the oral examination with a small number of cases and examiners remained poor. At least five to eight hours of testing time was required before adequate reliability was achieved (Daelmans et al., 2001). However, this timeframe was comparable to the OSCE, and to other studies that focused on oral examination (Swanson, 1987; Van der Vleuten & Swanson, 1990). Moreover, the OSCE is an examination that cannot be incorporated into the day-to-day routine of clinical practice; whereas the multiple oral examinations can take place in such a setting without modification.

**Summary of Traditional Assessment Methods in Nursing Education**

In summary, research presented on multiple-choice testing and oral questioning methods of student assessment in nursing education illustrated the need for multimethod approaches in evaluating knowledge acquisition (Daelmans et al., 2001; Carraccio & Englander, 2000; Chambers, Boulet, & Gary, 2000; Haijazi & Downing, 2008; Swanson, 1987; Van der Vleuten & Swanson, 1990). While multiple-choice test items are the most popular form of method used to assess nursing students, allied health professions used oral examinations, whether formally or informally, to determine clinical competence and critical thinking abilities (Davis & Karunathilake, 2005; Lunz & Bashook, 2008; Raymond & Luciw-Dubas, 2010). Critiques of both methods of student assessment have been illuminated. The chief concern with multiple-choice test items in the literature surrounds the ability of faculty to adhere to item writing procedures that garner sound instruments with which to gauge student comprehension of presented material (Haladyna & Downing, 1989; Tarrant et al., 2006). Medical education literature posited that the use of oral examinations produced poor reliability (Davis & Karunathilake, 2005; Lunz &
Bashook, 2008; Raymond & Luciw-Dubas, 2010). Yet, the literature presented provided clear methods in which to increase the reliability and validity of multiple-choice test items and oral examinations when using these two methods to assess student ability (Swanson, 1987; Van der Vleuten & Swanson, 1990). Clearly, rigorous methods of evaluation are needed to compare these two forms of student assessment to newer less treaded methods.

**Newer Assessment Methods in Nursing Education: HFS and LFS**

Simulation experiences can come in a variety of forms from high-fidelity simulation to low-fidelity simulation (Kaplan & Ura, 2010; Rushforth, 2007). High-fidelity simulation involves the use of written, structured case scenarios constructed around specific student learning outcomes. These scenarios are vetted by faculty savvy in the elements of simulation scenario development that use sophisticated mannequins termed as human patient simulators, and/or the scenarios use standardized patients which are actors portraying the “role” of the patient (Jefferies, 2005; Jefferies & Clochesy, 2012; Jefferies & Mc Nelis, 2010).

Low-fidelity simulation (LFS) includes simulated environments (typically for knowledge acquisition related to specific skills) that use role-play, simple case studies, and task trainer mannequins to aide in psychomotor skill acquisition (Brewer, 2011; Sweney & Eggleston, 2010). Researchers such as Kaplan and Ura (2010) report that the use of LFS as the primary methodology for teaching patient care skills such as intravenous medication administration has appeared in various forms within the nursing educational literature for decades. In more recent years, the use of LFS has evolved to include its use as a formative assessment technique in courses such as physical
assessment and fundamental nursing skills. Nursing courses using LFS as an assessment methodology employ rubrics developed by faculty as a means to quantify student demonstration of specific critical elements associated with the skill/task being assessed (Kaplan & Ura, 2010; Rushforth, 2007).

In more recent years, the objectively structured clinical examination (OSCE) has become considered a form of high-fidelity simulation when stations embedded in it are comprised of human patient simulators (mannequins), standardized patient actors, and low-fidelity simulation (LFS) stations that are thematically connected to the same simulation scenario (Jefferies & Clochesy, 2012). An OSCE is defined as a standard assessment method used to measure clinical competence (the ability to demonstrate specific care delivery) that focuses on outcomes based upon observable behaviors of learners (Carraccio & Englander, 2000). During a simulation using an OSCE assessment environment, students are placed in clinical scenarios using LFS and HFS stations to assess competency of specific patient care delivery (Chambers, Boulet, & Gray, 2000). Each station or patient scenario is constructed in such a way that specific objectives can be evaluated using methods such as standardization to control many of the biases of conventional observations (Chambers, Boulet, & Gray, 2000). The result is a multi-method approach used primarily to formatively assess medical resident competency and as a more accurate manner of summative assessment for medical training programs.

OSCE using HFS and LFS are widely used in medical education (Alinier, 2003). Research has shown that the use of HFS and LFS in an OSCE format is an effective assessment tool to assess practical skills in a simulated environment (Alinier, 2003; Carraccio & Englander, 2000; Chambers, Boulet, & Gary, 2000; Haijazi & Downing;
Carraccio and Englander (2000) used a MEDLINE search to review the literature relevant to pediatric simulations using HFS and LFS in an OSCE format from 1975 to 2000. The purpose of the article was to evaluate the reliability and validity of all studies regarding the use of HFS and LFS in an OSCE format used in pediatric education of medical residents as a form of formative student assessment. In this article, HFS and LFS in an OSCE format were compared to precertification examinations and to monthly clinical student assessments (Carraccio & Englander, 2000).

Carraccio and Englander found that the greater the number of stations and similarity between tasks at different skill stations an increase was observed in the overall reliability of the HFS and LFS assessments. They also discovered that studies that contained OSCEs with a higher number of stations demonstrated greater validity results. Of the studies examined, a correlation existed between the HFS and LFS stations and precertification examination ranging between .59 and .71, with \( p \leq 0.01 \). Although statistically significant with \( p \leq 0.05 \), the correlation between the HFS and LFS stations and faculty observed monthly clinical assessments was found to be much lower (\( .39 - .57 \)) than that of the HFS and LFS stations and the precertification examination (Carraccio & Englander, 2000). The authors concluded that with skillful design, reliability and validity could be achieved with a combination of HFS and LFS assessments in an OSCE assessment format, standardized board examinations, and direct observation of the student in the clinical setting.

Further research conducted by Chambers, Boulet, and Gray (2000) provided insight into the management of standardized patients during HFS assessment in an OSCE
format. The purpose of the study was to gather information regarding the appropriate length of time that should be allocated for learners to complete specific components of a clinical examination when using standardized patients. Standardized patients (SPs) are actors playing the part of a patient with a standard script. They provide learners with a standardized patient experience from which to be evaluated by faculty observing their interactions and care delivery (Chambers, Boulet, & Gray, 2000). Faculty observers use common marker sheets to assess demonstrated cognitive or behavioral indicators associated with specific learner outcomes embedded within the standardized encounter.

Chambers, Boulet, and Gray (2000) collected data from 1548 learners at the Clinical Skills Assessment Center (CSA) of Educational Commission for Foreign Medical Graduate Candidates (ECFMGC) in Philadelphia. Data were collected on the actual time used by each examinee during their ten SP encounters compared to the amount of information gathered during each encounter. A maximum of 15 minutes was allotted for each encounter. The average time spent with the SP was 13.3 minutes. This average time suggested that a 15-minute time limit was adequate. A positive correlation ($r = .41$) was determined between data gathering and time spent in the encounter, which suggested that participants in this study spent more time in data gathering. No significant difference was determined in time used by the examiner based on gender.

Kirton and Kravitz (2011) examined the value of HFS and LFS in an OSCE format as an effective assessment method for pharmacy students when compared to traditional pharmacy practice examinations. The authors evaluated the use of the HFS and LFS stations of an OSCE on a cohort of 39 pharmacy graduates in the UK. The Royal Pharmaceutical Society of Great Britain (RPSGB) noted that a low correlation
existed between student academic achievement and their performance during the pre-registration year of pharmacy training. To meet the necessary high standards of professional practice, the RPSGP advocated for the inclusion of competency-based learning and assessment in the form of HFS and LFS organized in an OSCE format. Their rationale was that the “measure of competence is contextual and that the assessment of confidence should ideally reflect what the student will habitually do when not being observed” (Kirton & Kravitz, 2011, p. 2).

The purpose of the study was to investigate whether there was a correlation between the students’ success with HFS and LFS stations in an OSCE and their success in other aspects of the course. The format of the HFS and LFS examination was the same at every level throughout the program being studied. However, the complexity and clinical content of the task associated with a given station reflect the student’s level of education at the time of the assessment.

Because OSCE assessments build on academic theory, the initial hypothesis was that performance on the OSCEs assessments would correlate highly with performance on the academic modules associated with pharmacy practice throughout the program. A weaker or no correlation was expected between the OSCE and aspects of the program with no pharmacy-practice content. Cohorts of students were expected to perform better in their final-year OSCE assessments than in first-year examinations. The null hypotheses were that there would be no correlation between OSCE assessments and pharmacy-practice examinations and that there would be no improvement in individual OSCE assessment grades as the student progress through the program. The initial hypothesis proposed that a strong correlation would exist between student grades in the third year
OSCE when compared with their scores in the related pharmacy-practice module for that year.

The chronological order in which examinations took place was used to determine whether the data were dependent or independent. The examination occurring first chronologically was always assumed to be the dependent data, as it would inform faculty of the student success on subsequent examinations. A Pearson correlation coefficient was calculated for the third year OSCE assessment grade (dependent) and *Medicines and Pharmacy Practice Examination* (MPPE3) grade (independent).

Students’ OSCE assessment grades were expected to increase as their experience with clinical situations increased. Therefore, improvements in OSCE assessment grades from the first year to the third year should be observed. When examining the performance of the students in the first year and in the third year OSCE assessments, 100% of the students performed better in the third year. Comparisons of results at the first and fourth (final) year indicated that only 80% attained higher grades in their final-year OSCE assessments compared with their first year results. The correlation data generated a sizeable relationship ($r = .6$) between the two data sets. The mean grades and standard deviations for both assessments (OSCE Assessment 62.9% /- 10.9%; MPPE3 63.7% =/- 9.0%) indicated that there were no great disparities between the dataset.

The findings did not support the hypothesis that students who perform well on examinations in their theoretical aspects of pharmacy practice would also perform well in the clinical aspects. A strong correlation between the two examinations should not be expected. Further, the authors concluded that their results lent credence to the argument that HFS and LFS assessments are an invaluable tool in assessing clinical competence.
They further stated that clinical competence cannot be gauged merely by examining the academic ability of students. Lastly, the authors contended that HFS and LFS assessments are also an important assessment methodology for preparing undergraduates for clinical practice (Kirton & Kravitz, 2011).

To bridge the gap between academic and clinical performance, Cordi et al. (2012) conducted a multi-phased pilot investigation of a single nursing program and followed up with a multisite national study evaluating the reliability of a simulation effectiveness tool (SET). According to Cordi et al. (2012), the development of a multi-item instrument has the intent to measure attributes that represent one or more constructs that are under study. Further, it is vital that the instrument demonstrate psychometric characteristics that contribute to validity and reliability of the tool in order to ensure that the instrument is measuring what it purports to measure and that this happens consistently each time the instrument is administered.

Cordi et al. (2010) the reliability and validity of the SET using a convenience sample of pre-licensure nursing students. The initial pilot study was multi-phased. During the first phase of the study, participants were involved in one or more simulations that were based on learning objectives from their coursework. At the end of each simulation, students were asked to complete a paper and pencil version of a 20-item 5-point Likert-type scale consisting of five responses (5 = strongly agree, 4 = somewhat agree, 3 = undecided, 2 = somewhat disagree, 1 = strongly disagree).

The principal components factor analysis of the 20-items revealed three associated factors: attitudes (factor 1), confidence (factor 2), and learning (factor 3). Based on factor analysis results, the researchers revised the initial simulation instrument
eliminating all items associated with attitude \((n = 7)\). The final concepts of the revised instrument consisted of learning (the student’s self-assessment of the knowledge and skills gained from the simulation experience), and confidence (the student’s self-assessment of their abilities and attributes gained from the simulation experience).

The learning subscale consisted of eight items with the Cronbach’s alpha of .87. The confidence subscale consisted of five items with the Cronbach’s alpha of .84. The final phase of the study was a multi-site investigation of the reliability of the new instrument. A convenience sample of participants from six nursing programs was used. The programs represented the Midwestern, Southern, Southwestern, and Western regions of the country. A total of 23 participants (4%) were post baccalaureate participants, and the remaining 622 participants (96%) were pre-licensure students representing several courses at all course levels in nursing education. There were 599 female participants (93%) and 46 male participants (7%).

Results of the data analysis indicated that the 5-item confidence subscale of the revised 3-point scale instrument had a Cronbach’s \(\alpha = .88\). This finding was similar to the findings in the original pilot study \((\alpha = .84)\). The researchers had similar findings for the 8-item learning subscale. The calculated reliability for the learning subscale was \(\alpha = .87\), similar to the results in the original pilot study (.87).

Based on their findings, it was concluded that the revised instrument met the acceptable criterion for reliability for its total simulation effectiveness score \((\alpha = .80)\) and for the confidence \((\alpha = .88)\) and learning \((\alpha = .87)\) subscale scores. However, the outcome measures reported in the final study used student self-reported data. The participants were both the raters and the ratees of the simulation effectiveness; therefore, data findings had
the potential to be unreliable because they were solely self-reported perceptions (Prion, 2008). This study represents an example of self-reported data used in simulation research to support reliability claims in the nursing literature. There are calls in nursing education for simulation research that moves beyond the focus of student self-assessment and towards studies that measure student competencies as well as clinical reasoning (Kardong-Edgren et al., 2010; Nehring, 2010).

The use of simulation in nursing education and the research regarding its superiority to traditional methods of instruction (lecture) is clearly documented in the literature (Kaplan & Ura, 2010; Paige & Daley, 2009; Reed, Lancaster, & Musser, 2009; Swanson et al., 2010; Solnick & Weiss, 2007; Swenty & Eggleston, 2010; Waxman & Telles, 2009). The use of simulation as a pedagogical modality of nursing instruction is used to engage students in a contextual learning environment. This learning environment combines the use of knowledge, problem solving, psychomotor, and communication skills. Nurse educators have found it necessary to design simulation scenarios using increasingly more complex forms of simulation to standardize the student experience (Kaplan & Ura, 2010; Kardong-Edgren & Michaels, 2007; Paige & Daley, 2009; Reed, Lancaster, & Musser, 2009; Swanson et al., 2010; Solnick & Weiss, 2007; Swenty & Eggleston, 2010; Waxman & Telles, 2009).

In the beginning of its emergence into health education, simulation was used at the medical resident and practitioner level as an authentic means of psychomotor skill acquisition through repeated practicing in a setting that was safe from patient and student harm (Kaplan & Ura, 2010; Murphy, Hartigan, Walshe, Flynn, & O’Brien, 2010; Solnick & Weiss, 2007; Swenty & Eggleston, 2010; Waxman & Telles, 2009). Simulation has
since evolved through varying forms into an authentic, innovative pedagogy, which has been incorporated into present day hospital staff training and schools of nursing curricula (Kaplan & Ura, 2010; Murphy et al., 2010; Solnick & Weiss, 2007; Swanson et al., 2010; Swenty & Eggleston, 2010; Waxman & Telles, 2010).

Jefferies (2005) defined the *fidelity* that is extended to simulation as how closely the experience mimics reality. Hence, the categorization of simulation into levels: low-, and high-fidelity. Low fidelity simulation (LFS) examples include the use of role modeling, case studies, task trainers, and mannequins to aide in student acquisition of knowledge surrounding a specific situation (Swenty & Eggleston, 2010). Examples of high-fidelity simulation (HFS) would include written scenarios surrounding a specific learning that combine human standardized patients or manikins set within the context of their truest environment (Kardong-Edgren & Michaels, 2007; Swenty & Eggleston, Waxman & Telles, 2010). Kardong-Edgren and Michaels (2007) use the analogy of a continuum with LFS on one end and HFS on the other end of the spectrum.

A third term related to simulation is *static*. Continuing with the analogy of simulation fidelity as a continuum by Kardong-Edgren and Michaels (2007), static simulation would be located somewhere between low- and high-fidelity on that spectrum. An example of static simulation or models would be that of a practice manikin used to insert a Foley catheter or an angiocatheter with no connected scenario to the skill (Kardong-Edgren, 2007). Given the trueness of higher fidelity of simulation, the expectations that are associated with this type of simulation are often greater and geared more toward the higher order domains of learning such as prioritization, patient care management, and delegation of patient care (Kardong-Edgren & Michaels, 2007).
Kaplan and Ura (2010) posited that while the clinical learning experience is a rich environment for nursing students, it often does not provide students with the opportunity to practice the important exploratory constructs such as delegation of patient care to other members of the healthcare team, prioritization of patient care delivery, and safe patient care management specific to caring for multiple patients. The main factors contributing to this issue are hospital system based: high patient acuity requiring specialized nursing care, and rapid patient turnover based on short lengths of stay in the acute setting (Kaplan & Ura, 2010). The use of multiple patient simulators in a well-constructed environment could provide the avenue for students to practice these important professional skills (Swanson et al., 2010; Kaplan & Ura, 2010). Further such constructed experiences allow faculty to assess higher-order student knowledge associated with the exploratory constructs of prioritization, delegation, and patient care management.

Kaplan and Ura (2010) developed a priority setting and delegating simulation scenario to study the use of simulation to enhance course content on leadership, delegation, management, and prioritization within a nursing program. The simulation exercise used three patient simulators simultaneously for each simulated clinical experience. Each simulator represented a particular patient from the case study with appropriate intervention needs based on findings on the mannequin. Mannequins were set-up to simulate patients receiving intravenous fluids, patients with wounds needing wound care, patients with ecchymosis indicating a potential problem, or patient requiring Foley catheter insertion.

Participants included 97 upper-division baccalaureate nursing students in the final semester of their senior year. Students were assigned into clinical groups (10 to 12
students per group) to participate in the 4-hour simulation-based learning (SBL) experience. Faculty conducted each 4-hour session weekly for one semester until all student groups had rotated through.

The SBL experience was designed to synthesize material from several different courses covering content on leadership, delegation, management, and prioritization. Students were given released clinical time to participate in the simulation experience. Clinical and simulation faculty supervised the training. The SBL experience was exempt from the institutional review board requirements, as the experience served as the novel teaching strategy.

Prior to participating in the SBL experience, students anonymously rated their level of confidence in prioritizing and in delegating care. They also self-rated their confidence in working effectively in a team. Students were also asked to self-rate themselves on the same areas after the SBL experience. An anonymous questionnaire containing two open-ended questions regarding the simulation experience (what they liked most and what they liked least about the simulation experience) was distributed to the students as a method of evaluating the entire as SBL experience. Faculty evaluated student performance based on a priority and decision-making scale, as well as a critical action checklist used in guiding the simulation debriefing process. The actual SBL was not evaluated for a student grade.

On the SBL day, students were provided with a 20 to 30 minute orientation reviewing the format of the day, general instructions for the simulation, and the case study assignment. Two students from the clinical group self-selected the roles as “nurse” or “nurse orientee.” The two participating students listened to an audio tape of change-of-
shift-report on the three simulated patients. The scenario was designed to have the report be incomplete. The incomplete report was to determine whether students would try to collect more information from the mock patient chart or patient assessment. From the report, one of the patients was stable, one had the potential for problems, and one was admitted during the night and considered unstable.

The two students were allowed to discuss the plan of care for the day following the report. The “nurse” was expected to prioritize patient care based on acuity levels of all three patients, competencies of the staff, and stability of the patients. The most urgent needs of the patient were to be identified, and the “nurse” would then delegate and make appropriate adjustments in care. The student portraying the “nurse orientee” would also be expected to prioritize care and confer with the “nurse” when appropriate.

Faculty and simulation staff located themselves in the simulation control room while the students provided care for the three patients. The faculty members would pose as family members or members of the healthcare team from the control booth. The health status of two of the three patients could deteriorate rather rapidly based on the level of assessment data collected and associated interventions of the two students.

Students rotated through the 20 minute simulation until all students in the clinical group had participated. The waiting students discussed case studies on a variety of situations in which delegating and prioritizing care was required. After all students had rotated through the simulation, all students participated in a group debriefing for approximately one hour.

The results indicated that 78% (n = 76) of participants reported more confidence in their ability to work as a team. Sixty-eight percent (n = 66) strongly agreed or agreed
that the SBL increased understanding of prioritizing and delegating patient care. Fifty-five percent of participants \((n = 52)\) reported more confidence in prioritizing and delegating care. While it is important that the participants in the study reported more confidence in their ability to prioritize and delegate care as well as feeling more confident in working as a team, what about the 46% \((n = 45)\) of students who had no opinion, disagreed, or strongly disagreed with feeling more confident in their ability to prioritize and delegate patient care. In addition, the authors did not discuss those 32 (33%) participants that did not feel (no opinion, disagree, or strongly disagree) that the SBL increased their understanding of prioritizing and delegating patient care.

The implications of the Kaplan and Ura (2010) study are that SBL can improve student experience with regard to prioritizing, delegating, and managing the care of multiple patients which is good. Conversely, generalization of the findings would be challenging. Yet, the study adds to the argument that more research regarding this subject is needed.

Although simulation is an effective means to measure psychomotor skill acquisition, little research has examined the reliability and validity of this form of assessment compared to traditional assessment forms (Benner, Sutphen, Leonard, & Day, 2010; Kardong-Edgren et al., 2010; Paige & Daley, 2009; Reed, Lancaster, & Musser, 2009; Swenty & Eggleston, 2010; ). Simulation has become engrained within current models of nursing education. The push to use simulation as a means of content and clinical skill assessment warrants more research (Kardong-Edgren et al., 2010; Paige & Daley, 2009; Reed, Lancaster, & Musser, 2009; Waxman & Telles, 2009).
Critique of Newer Assessment Methods in Nursing Education

Most current research surrounding simulation in nursing education involved student confidence with high- and low-fidelity simulation (Aronson, Glynn, & Squires, 2012; Beyer, 2011; Solnick & Weiss, 2005; Zulkosky, 2012) or psychomotor skill acquisition (Jefferies & Rizzolo, 2006; Linden, 2008; Ravert, 2004; Shepherd et al., 2007; Zulkosky, 2012). Cordi et al. (2012) noted that although a growing body of knowledge surrounding simulation existed in nursing educational literature, a gap existed between published research and simulation practice. Simulation in nursing education is transforming from an effective pedagogical method of instruction to a form of assessment. Multiple studies exist that examined simulation from an affective and psychomotor perspective. Research on the use of simulation as an assessment method of nursing theory content and clinical skill acquisition has begun to appear in the literature. However, more nursing research is needed on the use of simulation as an assessment method. The major critique (aside from cost) regarding simulation as a method of assessment relates to the validity of assessment instruments used during the experience.

Research by Nehring (2010) consisted of a review of the literature where only 26 published studies involving simulation in nursing education were identified. The studies were divided into four categories: simulation as an adjunct to traditional teaching; simulation as a means of assessing competence; simulation used as a substitute for judgment; and simulation as a method of teaching. Nehring (2010) cited the need for studies that supported the efficacy of simulation in nursing education. Further, the current research was inconclusive primarily because the published studies examined had small sample sizes and used instruments that had not been subjected to validity or reliability
testing (Nehring, 2010). Lastly, Nehring (2010) and Sanford (2010) noted that the few published quantitative research studies surrounding simulation in nursing education largely used untested instruments, which resulted in the potential lack of adaptation and progress.

The measurement of clinical skill performance is a continuing challenge for nursing educators (Mitchell, Henderson, Groves, Dalton, & Nulty, 2009). The most common criticism of the traditional clinical examination has been that students simply are evaluated on knowledge and their ability to memorize. Others believed that traditional clinical group examinations ignored other important characteristics of clinical performance including problem solving, critical thinking, and communication skills (Barman, 2005; Poenaru et al., 1997).

Evaluating clinical competence has advanced over the past two decades. Several structured performance tests have enabled limitations of traditional clinical examinations to be overcome. One of the most popular forms of structured performance test in nursing is simulation using an OSCE format (Mitchell et al., 2009). LFS and HFS assessments designed in an OSCE format in nursing education is a greatly modified version of the medical OSCE and typically involves the use of simulation (low fidelity simulation, task trainers, human-patient simulators, standardized patients, or any combination of these methods). The modification of the medical OSCE to suit nursing educational needs has created a strong need for research on the validity and reliability of the nursing simulation assessment in the modified OSCE version (Rushford, 2007; Selim et al., 2012). As a mean to provide insight associated with the validity and reliability of using the OSCE
format of a simulated assessment in nursing education, Selim et al. (2012) conducted a study evaluating its use in an undergraduate psychiatric nursing course.

The OSCE simulation stations constructed by Selim et al. were designed to cover the content and skills of clinical psychiatric nursing from course content and student learning objectives. Thirteen stations were prepared, including eleven working stations and two rest stations. Standardized patients (actors) were trained using role-play to act as patients for two days before the exam. Seventy-six undergraduate nursing students (60 female and 16 male) were randomly chosen to participate in the study during the 2008-2009 academic year. Development of the OSCE simulation for this study took four months.

During the OSCE, stations one, five, and eight, consisted of three interactive simulated patient experiences. Stations two, three, six, seven, and nine, were post stations. The other three stations included a medication classifications and indications station (station four), a medication side effects station (station 10), and a laboratory investigation results station (station 11). The first rest station followed station four, while the second rest station followed station seven.

Students were rated individually by two faculty raters using a checklist during the simulated experience. Each checklist consisted of a series of performance-based observations. Each student’s performance was rated as “done accurately,” “done inaccurately,” and “not done.”

The post stations were writing assignments. Post station number two and number three included writing a patient’s record, writing nursing notes, and writing a nursing care plan based on the interview and assessment of the patient in station number one.
station number six and number seven included writing nursing interventions before receiving electroconvulsive therapy and choosing appropriate activity therapy based on assessment of the patient in station number five. Post station number nine involved writing nursing management goals for hallucinations based on interviewing the simulated patient in station number eight.

Inter-rater reliability was calculated for simulated patient stations number one, five, and eight using non-parametric Spearman's rank order correlation. The results indicated statistically significant positive correlations between the two raters of the stations (station 1, \( r = .67; p = 0.0 \), station 5, \( r = .71; p = 0.0 \), and station 8, \( r = .58; p = 0.0 \)). The reliability of the OSCE simulation assessment, using Cronbach’s \( \alpha \), was higher than .7 for stations two, three, four, five, six, ten, and eleven.

The criterion validity of the OSCE simulation assessment was evaluated using Spearman’s rank order correlation. The correlation between the HFS and LFS simulation stations in the OSCE and the traditional clinical examination, final oral exam, final written exam, and the total grade for the nursing course were statistically significant correlations between the HFS and LFS simulation stations in the OSCE and the clinical evaluation (total score: 60; \( r = .54, p = 0.0 \)), final oral exam (total score: 60; \( r = .34, p = 0.0 \)), final written exam (total score: 80; \( r = .59, p = 0.0 \)), and total grade for the nursing course (total score: 200; \( r = .79, p = 0.0 \)).

The results indicated that all of the HFS and LFS OSCE stations reliably evaluated the psychiatric nursing students. In terms of validity, the HFS and LFS stations scores of the OSCE when compared to student scores on the final written exam (\( r = .59, p = 0.0 \)), scores on the oral exam (\( r = .34, p = 0.0 \), and scores on their clinical evaluation(\( r \))
=.53, p = 0.0). Lastly, elements claimed to increase the validity of the HFS and LFS stations in the OSCE included high objectivity, testing a wide range of skills and competencies, using a wide range of faculty examiners to reduce bias, and consistency. Additionally, face and content validity performed through reviewing the content of stations by experts in the field was reported to add to the overall validity of the OSCE compared to the other evaluation methods used in the study.

The major critique of using an OSCE with embedded HFS and LFS stations as an assessment technique is the expense. The main obstacle to its implementation to a wider spectrum of nursing programs, especially in the United States (U.S.), is its reportedly high cost (Palese et al., 2012). Reznick et al. (1992) established guidelines for LFS and HFS simulation stations in an OSCE assessment in an effort to standardize the estimation of the cost. Those guidelines remained the considered reference in the estimation of HFS and LFS stations used in an OSCE format for assessment (Poenaru et al., 1997; Palese et al., 2012).

Assessing the cost-effectiveness of an OSCE using LFS and HFS simulation stations is essential given the economic recession and the internationally recognized shortage of support for nursing education (Dadgaran, Parvizy, & Peyrovi, 2012). The medical literature on OSCEs using HFS and LFS stations considers it the most valid and reliable method for assessing safe practice. The OSCE format of using HFS and LFS stations has limited diffusion in nursing education specifically related to its cost. Given that nursing students must have their practice skills evaluated, no data existed comparing the cost of the OSCE using HFS and LFS stations to assess knowledge with other assessment methods (Palese et al., 2012). The results of the Palese et al. (2012) study
provided insight into the costs associated with developing HFS and LFS stations in an OSCE format to assess student knowledge in a nursing education. Further research associated with costs of developing such assessments in nursing education is important so that the most effective strategies for reducing those costs could be considered.

**Summary of Newer Assessment Methods in Nursing Education**

In summary, the review of literature clearly identified the need for researchers to begin to closely examine simulation in nursing education, as there is much about the pedagogy that remains unknown (Cordi et al., 2012). Kardong-Endgren et al. (2010) suggested that researchers should cease using indiscriminate assessment instruments and develop assessment tools that measure learning by providing students with feedback on complex learning outcomes in the cognitive, psychomotor, and affective domains. Further, rigorous summative evaluation studies are needed using reliable and valid instruments that distinctly discern learning constructs (Cordi et al., 2012; Jefferies & McNelis, 2010) as well as complex, higher order nursing skills (Kaplan & Ura, 2010).

Given how resource intensive HFS and LFS stations are in OSCE assessments, it is prudent for nurse educators to establish a standardized assessment instrument (rubric) if this form of assessment is to be used in the future.

In addition to the validity, reliability, and objectivity of an assessment technique, practicability and feasibility of the assessment method must be considered as well (Barman, 2005). When selecting an assessment method, educators would need to consider the number of students to be assessed, the number of staff, the availability of equipment, time, and money. An OSCE assessment containing HFS and LFS simulation stations when compared to oral questioning, and other traditional methods of assessment...
are more time-consuming and more expensive in terms of human and material cost. Further, the time required for setting up of a HFS and LFS station in an OSCE assessment is greater than that needed for the traditional assessment methods mentioned (Barman, 2005; Palese et al., 2012; Poenaru et al., 1997).

**Summary of the Literature**

The review of the literature on the nursing leadership constructs of prioritization, delegation, and patient care management demonstrated the difficulty faculty and students have in defining and operationalizing the behaviors that exemplify these important nursing actions. Research presented on the three nursing leadership constructs developed the argument for how challenging such constructs are to evaluate without empirical evidence (Lake et al., 2009; Luctkar-Flude et al., 2012; Thomas et al., 2011). Further, two consistent themes were evident with regard to clinical leadership in nursing: qualitative studies and theoretical papers comprised the bulk of the literature (Patrick et al., 2011). Moreover, nursing leadership tended to be characterized by defining attributes associated to care delivery (Lake et al., 2009).

Literature presented on the four assessment techniques MCT, OQ, HFS, and LFS demonstrated the challenge of developing substantive assessment tools that measure the constructs in nursing education which they are intended to assess. Research presented by Murphy et al., (2010) highlighted the growing body of literature that exists, questioning how prepared new graduate baccalaureate nurses are for competent practice. These researchers posited specific factors within the literature that have been identified globally as problems within nursing education. Factors which include changing patient characteristics within the acute care setting, changing science and information
technology, the increased complexity of health regulations and policies, and changing professional standards (Murphy et al., 2010).

The research presented on multiple-choice testing in nursing education suggested that nurse educators were driven by several factors to use this assessment format (Bailey et al., 2011; Bastable, 2008; Clifton & Schriner, 2010; Rushton & Eggett, 2003; Scheckel, 2012; Tarrant & Ware, 2008). Bailey et al. (2010) argued that in general, multiple-choice testing by nurse educators was accepted across educational systems in all disciplines as a user-friendly strategy to assess knowledge. Further, these types of tests in nursing education have been seen as a reliable, valid, and efficient way to examine learner outcomes; however, researchers from psychology and education have long been concerned with the consequences of multiple-choice testing on learning outcomes (Lampe & Tsaouse, 2010; Lee et al., 2011; Masters et al., 2001).

Additionally, the research presented on multiple-choice testing in nursing education suggested that authors have concerns regarding the skill of nurse educators in the construction of quality multiple-choice questions as well as violations of multiple-choice question guidelines in commonly accessed test banks furnished by textbook publishers (Bailey et al., 2011; Masters et al., 2001). Further, Tarrant and Ware (2008) echoed similar concerns regarding violations of item writing principles and more importantly, the overuse of questions evaluating low cognitive levels of knowledge by nurse educators. Nurse educators have argued that multiple-choice questioning allowed faculty to efficiently assess large numbers of students, tests a wide range of content, and if properly constructed, test higher levels of cognitive reasoning which can accurately discriminate between high-and low-achieving students (Bastable, 2008; Downing, 2002;
McCoubrie, 2004; Schuwirth & van der Vleuter, 2004; Tarrant & Ware, 2008). However, the reality presented by the research is that MCQ, when developed by nursing faculty, rarely measured higher-order student knowledge (Leung et al., 2008). The implications of the research presented were that more research was needed surrounding multiple-choice testing, reliability, and validity within nursing education (Downing, 2002; McCoubrie, 2004; Schuwirth & van der Vleuter, 2004; Tarrant & Ware, 2008).

A review of the literature depicting the use of OQ in nursing education was found to be very rare. Oral examinations for the assessment of clinical competence of medical students in their clinical rotations have long been regarded to have poor reliability (Daelmans et al., 2001; Davis & Karunathilake, 2005; Lunz & Bashook, 2008; Raymond & Luciw-Dubas, 2010). However, the presented research from medical education posited that oral examinations result in more effective study habits, increased motivation to study, improve the students’ learning and knowledge base, critical thinking skills, and ability to apply information in a specific clinical situation (Rushton & Eggett, 2003). The key to discerning reliability of oral examinations is frequency during a clinical day (Daelmans et al., 2001).

The growing momentum behind change in nursing education/competency assessment over the past several years illuminated many areas of further research: 1) implementation of LFS and HFS in an OSCE assessment format as a component of nurses’ and nursing education, 2) the rigorous evaluation of MCT, OQ, HFS, and LFS assessment techniques, 3) estimating the financial resources necessary to implement broad changes in assessing nursing students, and 4) the way nurses and nursing programs are formatively or summatively assessing competency. The overall review of the
literature presented links the incorporation of HFS and LFS assessment methods to assess nursing student knowledge associated with these specific leadership constructs (Kardong-Edgren et al., 2008).

The distinction between nursing leadership and the underlying constructs is a difficult area to substantiate in research, especially regarding prioritization, delegation, and safe patient care management. The Kaplan and Ura (2010) study is one of a scant number of studies within the literature specific to the use of HFS and LFS to gain some insight into the value of using such methodology to assess key elements of nursing leadership. For that purpose, this study is valuable. The study demonstrated that positive results using simulation specific to prioritization, delegation, and safe patient care management can be achievable, but rigorous research in this area is needed as the major criticisms associated with the use of simulation in nursing education as an assessment method are the use of research designed assessment tools and the cost of developing HFS and LFS assessment experiences, especially those constructed in an OSCE format. To that end, this research is significant in that it provided much needed examination under rigorous methods of the use of HFS and LFS as an assessment method compared to the traditional methods of nursing student assessment (MCT and OQ).
CHAPTER THREE
RESEARCH METHODS

This study investigated the construct validity of three major constructs of nurse leadership: prioritization, delegation, and patient-care management. Four nursing assessment methods—multiple-choice tests (MCT), oral questions (OQ), high-fidelity simulation (HFS), and low-fidelity simulation (LFS)—were used to measure each construct and the twelve correlations thus derived were organized into a MTMM matrix as developed by Campbell and Fiske (1959). The three research questions are listed below.

**Research Question 1.** To what extent do the twelve OSCE scores generated from using four assessment methods (multiple-choice tests, oral questioning, low-fidelity simulation, and high-fidelity simulation) to measure three constructs (delegation, prioritization, and patient care management) conform to the Campbell and Fiske (1959) criteria for construct validity?

**Research Question 2.** How do the twelve OSCE scores generated from the four methods of evaluating the three constructs correlate with scores of basic skills (TEAS®) and standardize nursing predictive scores of nursing skills (Kaplan® RN Predictive Exam)?

**Research Question 3.** To what extent does the actual cost of designing and implementing an OSCE assessment environment for nursing leadership constructs which includes high-fidelity and low-fidelity simulation stations compare to the projected costs found in the literature?
This chapter presents the study methodology addressing research design, research sample, protection of human subjects, construct definition, instrumentation, procedures, data analysis, and budget.

**Research Design**

This study used a correlation design consisting of four methods of nursing student assessment (multiple-choice testing, oral questions, high-fidelity simulation, and low-fidelity simulation) measuring the three nursing leadership traits of prioritization, delegation, and patient care management to evaluate the construct validity of these traits using the MTMM matrix approach. The key concepts of the MTMM approach are convergent and discriminant validity and the key evidence is provided by the magnitude of correlations among scores from the measures organized into a MTMM matrix. Evidence for convergent validity is shown in the MTMM matrix when correlations fit the expect patterns that indicate methods used to measure a concept demonstrate strong correlations among one another (Marsh & Hocevar, 1984); evidence for discriminant validity is shown in the MTMM matrix when correlations between two dissimilar measures are low (Kenny, 2012; Marsh & Hocevar, 1984; Muis, Winne, & Jamieson-Noel, 2007). The logic of the MTMM and the procedures for generating the MTMM matrix were an attempt by Campbell and Fiske to develop statistically the concept of a nomological network suggested by Cronbach and Meehl (1955) four years earlier. A nomological network is the theoretical linkages among constructs that, when taken together, provides the theoretical framework for studying a phenomenon.

The four assessment methods were embedded into an objectively structured clinical examination (OSCE) assessment environment. This assessment environment
included seven assessment stations within the 120-minute OSCE environment where each participant was provided information about two patients whose care they were assuming and their knowledge and judgment about how to deal with this information was assessed. Student groups of six or less were staggered every 30-minutes to begin the seven OSCE stations. Participants rotated through the seven stations in the OSCE assessment environment and each participant completed all stations using the same methods in the same order.

The OSCE assessment environment generated twelve scores for each participant (four methods by three constructs). Scores were generated based on participant demonstration of behaviors associated with the three nursing leadership constructs using a 21-item multiple choice test; a 21-item oral questioning instrument; one patient care HFS rubric; and three LFS skill station rubrics, one rubric station for each construct. The station sequence was the following: station one (7-minutes, 2-minute check-in and 5-minute audio patient report), station two (LFS skill station for patient care management – 6-minutes), station three (LFS skill station for delegation – 6-minutes), station four (LFS skill station for prioritization – 6-minutes), station five (HFS patient encounter with a standardized patient – 5-minute orientation and 15-minute patient encounter), station six (OQ station with audio taped and 21-item OQ – 10-minutes for second audio patient report and 10-minutes for 21-item OQ by faculty evaluator), and station seven (MCT station with 21-item MCT instrument administration – 30-minutes).

Table 2 outlines the OSCE stations, objectives, method of measurement, time allotted for each station, and the generated score for each station. Logistics and previous research aided in the decision to embed all evaluation methods into one data collection
date. Further, previous research conducted by Kaplan and Ura (2010) using OSCE assessment guided the design (Kaplan & Ura, 2010; Selim et al., 2012).

**Sample**

A convenience sample of 137 participants was obtained from senior-level, pre-licensure nursing students enrolled at a small private university in the San Francisco bay area. Permission to conduct this OSCE assessment was requested and obtained from the Nursing Department Chair, the Dean of the School of Health, Natural Sciences and Math, and the nursing faculty from the NURS/L4100 (Advanced Medical Surgical Nursing) and NURS/L4200 (Nursing Leadership) courses during the 2012/2013 academic year.
Table 2.
*Overview of the OSCE Assessment Environment*

<table>
<thead>
<tr>
<th>Station</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Participant Check-in Audio Report #1</td>
<td>Safe Med Admin</td>
<td>Delegation of tasks</td>
<td>Prioritizing patient care needs</td>
<td>Provide care to SP</td>
<td>Oral Questioning Audio Report #2</td>
<td>MCT</td>
</tr>
<tr>
<td>Method</td>
<td>n/a</td>
<td>LFS</td>
<td>LFS</td>
<td>LFS</td>
<td>HFS</td>
<td>OQ</td>
<td>MCT</td>
</tr>
<tr>
<td>Trait(s)</td>
<td>n/a</td>
<td>PCM</td>
<td>DEL</td>
<td>PRI</td>
<td>PRI</td>
<td>PRI</td>
<td>PRI</td>
</tr>
<tr>
<td>Time</td>
<td>7 minutes</td>
<td>7 minutes</td>
<td>7 minutes</td>
<td>7 minutes</td>
<td>20 minutes</td>
<td>20 minutes</td>
<td>30 minutes</td>
</tr>
<tr>
<td># of Items</td>
<td>n/a</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Total Score</td>
<td>n/a</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Generated Score</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3 (1 for each trait)</td>
<td>3 (1 for each trait)</td>
<td>3 (1 for each trait)</td>
</tr>
</tbody>
</table>

Note. LFS = low-fidelity simulation, HFS = high-fidelity simulation, PRI – prioritization, DEL – delegation, PCM – patient care management, MCT – Multiple-choice testing, OQ – Oral questioning, SP – Standardized Patient Actor
The researcher explained to student participants during the first class meeting of the semester the study and the process of the OSCE, each instrument and rubric as well as data collection process. All students enrolled in NURS/L4100 and NURS/L4200 for the 2012/2013 academic year were required to complete the OSCE as a component of clinical hours associated with each course.

No student could opt out of participation as the OSCE was completed in lieu of a required clinical day. Data were collected at three times during the 2012-2013 academic year: November 4, 2012, February 4, 2013, and February 9, 2013. The data collection dates were selected in an attempt to level the knowledge associated with the three leadership constructs for all participant groups completing the OSCE. For example, the NURS/L4100 data collection dates occurred three-quarters of the way through the course. Likewise, the Nursing Leadership and Patient Care Management (NURS/L4200) data collection dates occurred during the first three weeks of the semester. The rationale for these data collection dates was simply to capture the NURS/L4200 participants before lectures associated with the three constructs occurred and to capture NURS/L4100 participants at the maximum level of knowledge associated with advanced medical-surgical patients.

The participants in the research were between the ages of 22 – 28 years. Seventy-nine percent of the participants were female and 21% of the participants were male. All participants had exposure to prior HFS and LFS simulation experience via nursing coursework preceding the senior-level. Further, the participants had exposure to multiple-choice testing through prior examinations in preceding nursing coursework (Clifton & Schriner, 2010; Downing, 2002; McCoubrie, 2004; Tarrant & Ware, 2008). Likewise, all
participants had prior oral questioning experience within all of their nursing clinical experience before and including their senior-level clinical practica (Burman, Hart, Brown, & Sherard, 2007).

The study site was chosen because of the researcher’s access to the study population. Senior-level nursing students were chosen because they should possess sufficient working knowledge of both the traits and measurement methods used in the study. Because nursing students need to prioritize patient care, delegate care appropriately, and safely manage patients in complex care environments, students begin to hone these practice skills, which are the basis of clinically-competent care, in practice settings throughout their educational program (Bittner, Gravlin, Hansten, & Kalisch, 2011; Kaplan & Ura, 2010; Selim et al., 2012). Senior-level nursing students within a baccalaureate program had the most experience with both the methods and the constructs tested in this research.

**Protection of Human Subjects**

An application was submitted to the target University Institutional Review Board for the Protection of Human Subjects (IRBPHS) as well as the University of San Francisco IRBPHS, and approval was received from both institutions prior to data collection. Participants were informed that their participation in this research was a component of the assessment process within their respective courses. Therefore, consent was assumed as an element of their enrolled course work. Individual participant scores collected during the research project would remain confidential; all research data were kept in a secure location away from both University settings.
There was no anticipated risk to students participating in this study. All participants at the senior-level at the target University had attended clinical experiences during which time they cared for patients in a hospital setting; they had also participated in skills training using mannequins similar to those used in the LFS stations of the OSCE and had experienced the use of standardized patients (SPs) during HFS experiences during prior coursework which were similar to the HFS station in the OSCE. Because of these experiences, it was anticipated that participants would be comfortable in the simulation lab and with the equipment (LFS task trainers and SPs). However, an orientation to the simulators’ capabilities and to the resources available in the simulation setting were provided at the start of each simulation session. All seven of the stations that comprised the OSCE allotted time for thorough participant orientation prior to data collection.

Faculty within the nursing program at the targeted University used MCT and OQ as assessment methods in their courses. Faculty also used HFS and LFS as teaching methods. Therefore, students were reminded that they had successfully mastered the MCT and OQ forms of assessment and were very familiar with the use of HFS and LFS teaching methods associated with this research project prior to reaching the senior-level. Some participants were concerned that their performance would reflect on their course grade. Students received clinical hours for participating in the research project. No faculty evaluator was associated with assigning final student course grades for the two senior-level classes.
Instrumentation

Four methods of nursing student assessment were embedded in an OSCE environment to evaluate three constructs each, resulting in twelve scores for each participant. Two additional external measures were also obtained from student records.

Because the definition of the three constructs was a critical first step in developing instrumentation for this study, it is presented first in this section. Second, each of the four assessment methods is described. Third, the OSCE assessment environment within which each of the four methods is embedded is described. In the final section, the two external measures, the TEAS® and the Kaplan® RN Predictor Exam, are described.

Constructs

Prior research conducted by Prion (2008), Schultz, Shinnick, and Judson (2012), and Thomas et al. (2011) guided the initial broad definitions of the three nursing leadership constructs for this study. Achievable elements of the conceptual definitions associated with their objectives and cognitive or behavioral indicators were examined closely to ensure that senior-level students had enough prior education to achieve the expected outcomes for each assessment method being studied (Jefferies, 2005; Jones et al., 2010; Kaplan & Ura, 2010; Kardong-Edgren, Adamson, & Fitzgerald, 2010). During the process of defining these behaviors, one screening process used by the researcher was to construct objectives that elicited the cognitive or behavioral indicators associated with each construct as a means to aide in the development of assessment items. This section presents a conceptual definition for each construct, student nurse objective, and cognitive or behavioral indicators found in the nursing literature that were used to develop the items for each instrument.
**Prioritization.** Prioritization is defined in the literature as when a nurse provides safe care using three specific levels to rank patient care needs. The first level of prioritization is associated with threats to a patient’s immediate survival or safety. For example, airway obstructions, breathing problems, and circulation problems presented by a patient during assessment would rank as the most urgent prioritization level. The second level of prioritization is associated with mental status changes, acute pain, acute urinary elimination issues, untreated problems that now require immediate attention (critically high or low laboratory values), infection risks, and patient safety or security risks. The third level of prioritization includes patient needs that do not fit into the other two categories. For example, monitoring medication side effects, patient teaching, or long-term (chronic) problems associated with activities of daily living (ADLs).

The primary objective for the construct prioritization was participant understanding how to rank priorities for individuals and apply this knowledge in the clinical setting (Jones et al., 2010; Kaplan & Ura, 2010). Further, the construct of prioritization is demonstrated when the nurse knows and applies the principles needed to rank priorities for a group of patients.

Several cognitive and behavioral indicators are associated with prioritization. These cognitive and behavior indicators include, collecting appropriate data including vital signs, laboratory values, and physical assessment findings needed to determine the most urgent care needs of a nurse’s patient case load. The nurse must correlate collected data on his/her patients to organize and prioritize nursing tasks or skills to develop a plan of care for each patient. The nurse follows his or her plan of care to manage the urgency of skills and care needs for each patient by completing the highest priority task/skill first. He
or she completes one task before beginning another and then reprioritizes care based upon the remaining tasks or new assessment findings. Table 3 presents the developed objectives for the prioritization construct along with its cognitive or behavioral indicators.

Table 3. Construct Objectives and Indicators - Prioritization

<table>
<thead>
<tr>
<th>Objective(s)</th>
<th>Cognitive/Behavioral Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Know and apply the principles needed to rank priorities for a group of patients.</td>
<td>First Level – Threats to patient’s immediate survival or safety (ABC’s – Airway, Breathing, and Circulation). Second Level – Changes in mental status, acute pain, acute urinary elimination, untreated problems that now require immediate attention (critically high or low lab values), infection risks, safety, or security. Third Level – Patient needs that do not fit into the other two categories such as monitoring medication side effects, lack of patient knowledge, long-term problems associated with activities of daily living (ADLs).</td>
</tr>
</tbody>
</table>

**Delegation.** Delegation is defined by the National Council of State Boards of Nursing (NCSBN, 1997) as designating ancillary personnel the responsibility of carrying out a specific group of nursing tasks in the care of certain patients. Delegation includes the understanding that the authorized person is acting in the place of the registered nursing (RN) and will be carrying out tasks that generally fall under the RN’s scope of practice. However, the person taking on the RN-level task must be qualified to perform the task within the nurse’s state practice act. For example, a RN could delegate walking a patient once around the nursing unit to ancillary personnel. The RN remains accountable for walking the patient around the nursing unit; however, time may not permit him or her to complete the task given the number of assigned patients and the complexity of other nursing tasks.
The objective for the construct of delegation requires that participants know and apply the *Five Rights of Delegation* during the delivery of care to a group of patients. Further, the construct of delegation is demonstrated when the nurse knows and applies the principles needed to delegate tasks to the appropriate team member in the clinical setting.

Typically, there are *Five Rights of Delegation* in nursing that are taught as a standard across most nursing programs. These *Five Rights of Delegation* are used as a mental checklist to assist nurses working in multiple roles to clarify critical elements of the decision-making process associated with patient care delivery. Staff nurses have accountability in assuring that the delegation process is implemented safely and effectively to produce positive health outcomes.

The first of the five rights is the “right task.” The right identified nursing tasks are those that are appropriate for delegation to other licensed or unlicensed assistive personnel (UAP). When a nurse delegates tasks, the outcomes of tasks should be clear and predictable. For example, when unlicensed assistive personnel are assigned to feed a patient who has suffered a stroke and has weakness on one side of their body, the predicted outcome is that the patient will eat and not choke on their food. The task of feeding this patient should not require excessive supervision, complex decision making, or detailed assessment during its performance. If any of these elements are required, the task must be completed by an RN.

The second “right” in delegating is the “right circumstance.” When delegating a task, the nurse must consider the patient setting, available resources, and consideration of other relevant factors such as the staffing mix, community needs, teaching obligations,
and the type of patients being cared for in the setting. For example, different policies for delegation apply regarding what and how an RN must delegate in home care, long-term care, or in community homes for the developmentally disabled or group boarding home for assisted living patients.

The third “right” in delegating is the “right person.” In planning the right person to complete a task, focusing on outcomes is essential. Evidence-based critical thinking indicators which are outcome focused guides appropriate delegation as well as critical thinking in nursing. For example, Patient “X” has an outcome to be clean. The task is a bed bath. The team member delegated to complete the bed bath is the nursing assistant or other UAP. While the RN could complete the bed bath, he/she has more complex tasks that require either patient education, further assessment, or monitoring which take priority; therefore, to meet the outcome for Patient “X,” the bed bath must be delegated.

The fourth “right” in delegating is the “right direction/communication.” In delegating care, the right direction/communication associated with the delegated nursing task is essential. Clearly ask the team member if he/she understands what is being asked of him/her to complete. Concisely provide directions without giving too much unnecessary information. Correctly provide the delegate with directions according to policies, procedures, job descriptions, and state law. Lastly, provide the delegate with all the information needed to complete the task.

Lastly, the fifth “right” in delegation is the “right supervision/evaluation.” Once the RN delegates a task, he/she must evaluate that the task is completed accurately and in a timely manner. Some nursing tasks might require that the RN supervise the correct method of completing a particular task. The RN maintains accountability for all delegated
tasks; therefore, they are responsible for ensuring that each task is evaluated in the appropriate manner and according to policies, procedures, job descriptions, and state law.

Table 4 presents the developed objectives for the delegation construct along with its cognitive or behavioral indicators.

Table 4. 
Construct Objectives and Indicators - Delegation

<table>
<thead>
<tr>
<th>Objective(s)</th>
<th>Cognitive/Behavioral Indicators</th>
</tr>
</thead>
</table>
| Know and apply the Five Rights of Delegation during care delivery for a group of patients. Right Task | • Appropriate delegation activities are identified for specific patients  
• Appropriate care tasks identified for unlicensed assistive personnel (UAP) |
| Right Circumstances | • RN assesses health status of patients  
• RN matches complexity of activities with competency & scope of practice of UAP and LVNs |
| Right Person | • Instruct and/or assess, verify, and identify UAP’s competency on individual basis |
| Right Direction/Communication | • Communicate delegation decision on patient specific and UAP specific basis  
• Detail and method of communication vary  
• Situation specific data to collected and method  
• Specific expected results or potential complications and time lines for communication information |
| Right Supervision/Evaluation | • Supervise performance or assign supervision to another licensed nurse  
• Monitor performance  
• Obtain and provide feedback  
• Intervene if necessary  
• Ensure proper documentation  
• Evaluate the patient  
• Evaluate performance of the activity |
**Patient care management.** The definition of patient care management is the ability to successfully demonstrate an integration of knowledge, skills, experience and attitudes needed to meet the needs of patients and families. The nursing literature associates patient care management typically with knowledge surrounding clinical decision-making, critical thinking and a global grasp of the situation, coupled with nursing skills acquired through a process of integrating formal and informal experiential knowledge and evidence-based practice.

According to Benner (2001) a senior-level nursing student must be able to demonstrate knowledge surrounding clinical decision making at the novice level. The novice needs rules to guide performance and experiences difficulty in applying theoretical concepts to patient care settings. The objective for students at this level is that they are expected to use critical thinking skills that demonstrate a global understanding of the patient situation based upon the information that they gather (assessment) and information that they are provided (nursing report/handoff, patient chart, and interdisciplinary collaboration).

At the student nurse level, the observable characteristics associated with patient care management that are typically exhibited include the collection of basic-level data. Students follow algorithms, decision trees and protocols with all populations of patients and are uncomfortable deviating from these tools. Further, students tend to attempt to match formal knowledge with clinical events to make decisions, tend to question the limits of their ability to make clinical decisions, and delegate the decision-making to other clinicians. Table 5 presents the developed student objectives for the patient care management construct along with its cognitive or behavioral indicators.
Table 5.  
*Construct Objectives and Indicators – Patient Care Management*

<table>
<thead>
<tr>
<th>Objective(s)</th>
<th>Cognitive/Behavioral Indicators</th>
</tr>
</thead>
</table>
| Demonstrates knowledge surrounding clinical decision-making, critical thinking and a global grasp of the situation, coupled with nursing skills acquired through a process of integrating formal and informal experiential knowledge and evidence-based practice. | • Collects basic-level data  
• Follows algorithms, decisions trees and protocols  
• Delegates decision-making to charge nurse or more experienced licensed staff  
• Attempts to match formal education (text-book knowledge) to the situation |

Prion (2008), Schultz, Shinnick, and Judson (2012), as well as Thomas et al. (2011) guided the development of the constructs to be researched. Research by Kaplan and Ura (2009), Schultz, Shinnick, and Judson (2012) guided the operationalized definitions of the three constructs which were then transformed into cognitive or behavioral indicators that were found to be associated with each construct. The cognitive or behavioral indicators associated with each construct were differentiated and vetted by the faculty consensus panel to achieve the minimal number of behaviors for evaluation using a rubric.

**Assessment Methods**

Research by McCaughey and Traynor (2010), Reed et al. (2007), Swenty and Eggleston (2010) as well as Thomas, Hodson-Carlton, and Ryan (2011a) guided the development of the simulation scenario in collaboration with a simulation expert. Participant outcomes were developed to optimize the possibility of evoking the associated behaviors that were linked to each construct in the literature. The highest degree of fidelity was also validated through a process developed by California Institute
for Nursing and Health Care (CINHC). Test blueprints were developed for the MCT and OQ instrument as a means to construct items to measure the specific higher-order domains of Bloom’s Taxonomy. Lastly, the HFS station rubric and the LFS station rubrics were developed specifically to assess observed participant behavior associated with the objectives for the OSCE, the station objective(s), and the research questions.

Information acquired from Baig et al (2010), Decker et al. (2010), Hinton et al. (2012), and the National League of Nurses (2010) was used to develop the test blueprint for the MCT and the OQ Instrument, the HFS rubric, and the three LFS station rubrics. Feedback from a consensus panel of nursing faculty with content and method expertise was used as well to improve the fidelity of the OSCE experience and improve the instruments used in this research.

The first assessment method developed was the multiple-choice test (MCT). Segments of the final MCT were used as quizzes, midterms, and as final exams by the researcher in his NURS/L4200 course over a two year period. This section will describe the instruments used to assess the three constructs in this study, provide the reliability estimates for the multiple-choice test, and inter-rater reliability estimates for the OQ Instrument. The HFS and LFS Rubrics reliability estimates are presented. Further, the OSCE environment is also described. Finally, descriptions of how the two additional test scores (TEAS® and Kaplan® RN Predictor Exam) were obtained for this study.

**Multiple-choice test.** Multiple-choice testing was measured by means of a researcher developed 21-item examination (see Appendix A). A test blueprint was created to guide development of the items using a grid that included learning objective, cognitive process dimension, specific trait (delegation, prioritization, and patient
management), and type of item (Bastable, 2008). Table 6 presents the Multiple-Choice Test Blueprint.

The objectives for the MCT were taken from the course reading text for NURS4200 at the primary investigator’s University. Initially, a 45-item MCT was piloted by the researcher during the spring of 2012. Item analysis was computed from the scores generated following the pilot. The resulting item frequency tables, intercorrelation matrices, and reliability estimates were reviewed to assess overall functioning of the instrument. Using Cronbach’s alpha, the reliability for the 45-item test was .454. A number of potentially ambiguous worded items were discovered. By eliminating these poorly performing items, a maximum alpha of .824 was achieved and a final possible range of scores was 0 to 21. The 21-item multiple-choice test included seven items on prioritization, seven items on delegation, and seven items on patient care management.
Table 6.
Multiple-Choice Test Blueprint

<table>
<thead>
<tr>
<th>Objective (Bloom’s Taxonomy Level)</th>
<th>Prioritization</th>
<th>Delegation</th>
<th>Patient Care Management</th>
<th>Total Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differentiate between leadership and management (applying)</td>
<td>12</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Identify the management structures of patient care (applying)</td>
<td>2</td>
<td>16</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Describe the various modes of patient care delivery systems (evaluating)</td>
<td></td>
<td>14</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Determine the responsibility of the nurse in the various care delivery systems (analyzing)</td>
<td></td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Identify outcome measures of patient care management (applying)</td>
<td>17</td>
<td>9</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Relate a clinical scenario to each of the delivery models (applying)</td>
<td>3</td>
<td>15</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Outline five topic areas that the professional nurse should consider when making delegations decisions (analyzing)</td>
<td>4</td>
<td>20</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>Incorporate principles of delegation and supervision in professional nursing practice to ensure safe and legal patient care (analyzing)</td>
<td>8, 18</td>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Understand how time is managed personally and at the unit level (evaluating)</td>
<td>11</td>
<td>19</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Describe the importance of allowing adequate time for daily planning and priority setting (evaluating)</td>
<td></td>
<td></td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Demonstrate how to build evaluation into planning so that reprioritization can occur when necessary (evaluating)</td>
<td>7</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td><strong>Total Items</strong></td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>21</td>
</tr>
</tbody>
</table>
Items were constructed to test the Bloom’s Taxonomy cognitive process dimensions of applying, analyzing, and evaluating (Bloom & Krathwohl, 1956; Eber & Parker, 2007). These higher order cognitive processes are believed to be connected to the development of nursing leadership which includes prioritization, delegation, and patient management (Bloom & Krathwohl, 1956; Kaplan & Ura, 2010; Savickiene, 2010; Su & Osisek, 2011). For example, the first ten items were associated with the following narrative statement:

You are the leader of the team providing care for six patients. Your team includes yourself (an RN), a Licensed Vocational Nurse (LVN), and a newly hired nursing assistant, who is undergoing orientation to the unit. The patients are as follows:

Mr. Duncan, a 68-year-old with unstable angina who needs teaching for cardiac catheterization scheduled this morning.

Ms. Johnson, a 45-year-old experiencing chest pain scheduled for a graded exercise test later today.

Mr. Richardson, a 75-year-old who had a left-hemisphere stroke four days ago.

Ms. Sampson, an 83-year-old with heart disease, a history of myocardial infarction, and mild dementia.

Ms. Baker, a 93-year-old newly admitted from a long-term care facility, with decreased urine output, altered level of consciousness, and in elevated temperature of 99.5°F (37.5°C).

Mr. Lincoln, a 59-year-old with mild shortness of breath and chronic emphysema.
The first question associated with the narrative is crafted to test the objective requiring the participant to demonstrate how to build evaluation into planning, so that reprioritization can occur when necessary. This objective measures the cognitive process from Bloom’s Taxonomy of evaluating. The question is present in Table 7.

Table 7.
Sample Multiple-Choice Test Question

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Mr. Duncan</td>
</tr>
<tr>
<td>b</td>
<td>Ms. Johnson</td>
</tr>
<tr>
<td>c</td>
<td>Mr. Richardson</td>
</tr>
</tbody>
</table>

The LVN Scope of Practice requires the registered nurse (RN) to retain accountability for all patients assigned to them. When the RN assigns patient care to an LVN on the team, he/she must verify all patient physical assessments made by the LVN and reprioritize care based on his/her findings not those provided by the LVN. Further, it is out of the LVN Scope of Practice to make decisions that impact the plan of care for any patient. Therefore, of the patients described in the narrative associated with this question, only the presented patients with chronic, controlled disease processes can be assigned to the LVN. Those patients from narrative are Mr. Richardson, Ms. Sampson, and Mr. Lincoln.

By retaining the care of those patients who are presenting with complicated acute situations, the participant demonstrates that he/she understands both their scope of practice and that of the LVN. Further, he/she also demonstrates knowledge associated with each patient condition and that further assessment and reprioritization of care for each of the remaining patients is necessary. For example, Ms. Johnson is a 45-year-old
female currently experiencing chest pain who has been scheduled for a graded exercise later that day. The graded exercise is a procedure that determines the potential for blockages of the cardiac vasculature (the vessels that supply nutrient blood to the heart muscle). If a patient admitted for such a procedure is experiencing chest pain, this patient requires immediate assessment and evaluation by the RN. Reprioritize the care planned for this patient during the day is likely; therefore, the LVN cannot be assigned to care for this patient.

**Oral questioning instrument.** A researcher-developed oral questioning instrument was used to measure prioritization, delegation, and patient care management in a simulated clinical setting (Appendix B). A consensus panel of three faculty experts participated in the development of the final instrument (Tarrant & Ware, 2008; Twigg, 2012). Thirty items were presented to the panel for review and revision. The final instrument consisted of ten questions designed to elicit twenty-one items (7 prioritization items, 7 delegation items, and 7 patient care management items). These items were designed to assess participants’ ability to prioritize care for a group of patients, to delegate care appropriately, and to manage the care of four simulated patients.

Participants were provided access to a recorded patient handoff of four patients, mock patient charts on all four patients in the recorded report, and a nursing medication resource book. During the recorded report, participants were allowed to take notes, look up medications, and review the patient charts. Embedded in the recorded report were seven nursing tasks associated with the care for the four simulated patients during the day shift. After listening to the recorded patient handoff participants were asked to prioritize the care for the four patients, delegate skills/care for these patients, and manage the care
of four simulated patients. The seven nursing tasks associated with the prioritization construct, associated questions, and correct responses are presented in Table 8.

Participant responses were audiotaped and scored as correct or incorrect by the researcher. (Appendix C: OQ Expected Responses). A second faculty researcher also scored the audiotaped interview using the same method. Interrater reliability was calculated by adding up the number of cases that received the same rating by both raters and dividing by the total number of cases. Both raters were trained to the point of 100% agreement prior to data collection. The possible range in oral questioning scores was 0 to 7 in each construct subscale (7 items by 3 constructs). Calculated inter-rater reliability for the OQ instrument were $K = .951$ (prioritization), $K = .902$ (delegation), and $K = .892$ (patient care management). Calculated reliability estimates for the items were .77 (prioritization), .76 (delegation), and .85 (patient care management).
Table 8. *Prioritization Tasks, Questions, and Correct Responses Associated with the Recorded Report/Hand-off*

<table>
<thead>
<tr>
<th>Task</th>
<th>Question</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritize initial assessment</td>
<td>1) Which patient from your recorded report would you need to assess first on your nursing rounds?</td>
<td>1) Patient 1A</td>
</tr>
<tr>
<td>Organize &amp; Prioritize Nursing Care</td>
<td>2) Why is this patient’s situation a priority?</td>
<td>2) Verify lab value. Potassium lab value is critically low. Low serum potassium levels could cause arrhythmias.</td>
</tr>
<tr>
<td>Develop a Plan of Care for each patient</td>
<td>7) Speaking aloud, please describe to me the focused assessment priority of patient in room 1A.</td>
<td>7) Auscultate lungs and heart sounds; assessment impairment.</td>
</tr>
<tr>
<td>Review the MAR for all patients</td>
<td>8) Why do you consider this body system as a priority for your focused assessment?</td>
<td>8) The patient was admitted for heart failure exacerbation; CV system is compromised; therefore, critically low potassium could be fatal.</td>
</tr>
<tr>
<td>Verify a.m. lab values for all patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notify MD of 1A’s lab values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain MD orders for potassium replacement &amp; follow-up lab draws</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prion, 2008 proposed that when developing methods of assessing nursing students in simulated environments that exploratory items be developed based on defining characteristics of each construct. Exploratory items were developed for each construct in this study based on defining characteristics of each construct found in nursing literature. An assessment method was then developed for each construct looking at the defining characteristics and exploratory items. The methods used in this study were guided by the assessment methods used in nursing education (Jefferies, 2005; Motacki & Burke, 2010; Prion, 2008). Table 9 presents the blueprint used to develop items for the OQ station.
Table 9.
Item Development Blueprint – Oral Questioning Station

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition</th>
<th>Observable Characteristics</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>prioritization</td>
<td>Understand how to rank priorities for individuals and apply this knowledge in the clinical setting.</td>
<td>Allow time for planning &amp; establishment of priorities. Complete the highest priority task first. Complete one task before beginning another Reprioritize care based on remaining tasks &amp; new assessment information received.</td>
<td>Speaking aloud, please describe to me the steps that you would take immediately after listening to this recorded report if you were in a hospital setting? Which patient from your recorded report would you need to assess first on your nursing rounds? Why is this patient’s situation a priority?</td>
</tr>
<tr>
<td>delegation</td>
<td>Know and apply the Five Rights of Delegation during care delivery for a group of patients.</td>
<td>Right Task Appropriate delegation activities are identified for specific patients Appropriate care tasks identified for unlicensed assistive personnel (UAP)</td>
<td>Is there any nursing task or tasks discovered from your recorded report that you as the RN could delegate to another team member? Which task(s) would you delegate and to whom?</td>
</tr>
<tr>
<td>patient care</td>
<td>Demonstrates knowledge surrounding clinical decision-making, critical thinking and a global grasp of the situation, coupled with nursing skills acquired through a process of integrating formal and informal experiential knowledge and evidence-based practice.</td>
<td>Collects basic-level data Follows algorithms, decisions trees and protocols Delegates decision-making to charge nurse or more experienced licensed staff Attempts to match formal education (textbook knowledge) to the situation</td>
<td>Are there any laboratory values provided to you in your shift hand-off that are concerning to you as a RN? Why do you consider these laboratory values concerning?</td>
</tr>
</tbody>
</table>
**High-fidelity simulation rubric.** The development of the HFS rubric used in this study was guided by medical education and nursing education literature. After the simulation scenario was developed, vetted by a census panel, and validated, a blueprint was created for the rubric. The rubric blueprint development process described by Jones *et al.* (2010) was used to review the simulation objectives; identify the competencies that the participating students were required to achieve; and match the simulation outcomes with the competencies to determine which skills would be examined. From there, a simulation rubric was created to assess participant achievement of the specified outcomes of the experience.

The possible score range for this rubric was 0 to 7 for each of the three constructs. The total possible score for this rubric was 21 points (7-points by 3-constructs). Calculated reliability estimates for the HFS rubric were .80 (prioritization), .81 (delegation), and .82 (patient care management). Table 10 represents the HFS rubric blueprint that was used to create the rubric for this study.

The blueprint used in the development of exploratory items for the HFS and LFS stations were developed differently. Jefferies (2005), Jefferies and McNelis (2010), Kardong-Edgren, Adamson, and Fitzgerald (2010) and McGahie et al. (2006) all indicated that the development of rubrics specifically to assess nursing students in a simulation environment is time consuming and must be linked to the simulation scenario objectives. With that in mind, a scenario for the simulation was created using experts in the pedagogy months before the first data collection date. The scenario required six months of development time to determine just how to achieve measurable outcomes using simulation (HFS or LFS) to assess students using the three constructs being studied.
Finally, outcomes of a simulation scenario are linked to the fidelity of the simulation experience for the participant; therefore, a faculty panel was used to validate the simulation scenario with the Scenario Validation Process Checklist developed by the California Institute for Nursing in Health Care (CINHC).

Table 10.  
*High-Fidelity Simulation (HFS) Rubric Blueprint*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Objective(s)</th>
<th>Desired Student Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritization</td>
<td>Identify the most critical problem of a patient based on patient's clinical information given in report, chief complaints and ABC rounds.</td>
<td>Identifies low potassium results of 3.0 as the first priority and calls physician.</td>
</tr>
<tr>
<td></td>
<td>Do ABCs.</td>
<td>Plans to send patient to surgery, which is on-call, while performing tasks such as IV antibiotics and head-to-toe assessments.</td>
</tr>
<tr>
<td></td>
<td>Utilize patient orders.</td>
<td></td>
</tr>
<tr>
<td>Delegation</td>
<td>Student delegates tasks to team members as appropriate.</td>
<td>When new admission arrives, delegates to LVN to take vitals/make patient comfortable and reports back vital signs.</td>
</tr>
<tr>
<td></td>
<td>Follows the “5 R’s of delegation.”</td>
<td></td>
</tr>
<tr>
<td>Patient Care Management</td>
<td>At 5 min. into the simulation: call from physician with a potassium order-10 mEq IV over three hours. May hang up waiting for the nurse.</td>
<td>Identifies order of actions: takes physician phone call on potassium order including SBAR and VORB; informs family that the nursing assistant will attend to their family member; starts surgical preparation; informs admitting.</td>
</tr>
<tr>
<td></td>
<td>At 7 min. into the simulation: family member complaining that patient is lying in stool and insists they be cleaned immediately.</td>
<td>Calls for delegates the call to respiratory therapy. X line secures Mr. Dunlap in bed; calls RT.</td>
</tr>
<tr>
<td></td>
<td>At 8 min. into the simulation: call from admitting with an admission for the registered nurse.</td>
<td>Calls physician to report condition change of Mr. Dunlap</td>
</tr>
</tbody>
</table>
Appendix D is the Scenario Validation Process Checklist used by the simulation expert in conjunction with recommendations imputed from content experts. Once the simulation scenario was developed and vetted through an expert panel, the objectives created in the simulation scenario were used to develop rubrics specific for the behaviors identified as being associated to the leadership constructs in the nursing literature for the HFS station, and the LFS stations. Table 1 represents a sample of the Item Development Blueprint for the HFS station and the LFS stations.

**Low-fidelity simulation rubrics.** Three researcher developed rubrics (one for each station and construct) that were used during the LFS stations in the OSCE assessment. Each rubric was designed to measure prioritization (station three), or delegation (station four), or manage patient care (station two) based upon a recorded nurse report/handoff. A blueprint of the associated behaviors related to each construct was designed and used to guide the researcher in the development of the rubric used at each specific station. Similar to the development of the HFS, the LFS rubrics used the validated simulation scenario to create exploratory items which were then used to create observable behaviors for assessment. The audio report was used to maintain participant fidelity to the simulation objectives and the OSCE assessment experiences.

Each station had a range of scores from 0 to 7. The highest possible score for each rubric was 7. For example, at station two which was the station that assessed participants on their ability to apply patient care management knowledge to the situation presented from the audio report. Participants received a perfect score of 7 if they were to demonstrate or verbalized that they verified the patient order against the medication administration
record (MAR) for the patient, looked up the medication, assessed the patient
(mannequin), verified the morning laboratory values, notified the physician, received
verbal orders to administer the medication, and administered the medication safely over
2-minutes.
Table 11.  
**Item Development Blueprint – HFS and LFS Stations**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Defining Construct Objective</th>
<th>HFS &amp; LFs Item Development Blueprint</th>
<th>Observable Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>prioritization</td>
<td>Understand how to rank priorities for individuals and apply this knowledge in the clinical setting.</td>
<td>Cognitive Objective</td>
<td>Method: HFS Station</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Displays knowledge of First, Second, and Third Levels of prioritizing care</td>
<td>- Prioritizes patient safety (hand hygiene, verifies right patient, right medication, holds medication based on lab values)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Psychomotor Objective</td>
<td>Method: LFS Station</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Prioritize appropriate nursing interventions</td>
<td>- Check the lab results on the patient in room 1A &amp; inform doctor regarding low lab value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behavioral Objective</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Prioritization at appropriate opportunities</td>
<td></td>
</tr>
<tr>
<td>delegation</td>
<td>Know and apply the Five Rights of Delegation during care delivery for a group of patients.</td>
<td>Cognitive Objective</td>
<td>Method: HFS Station</td>
</tr>
<tr>
<td></td>
<td>Right Task</td>
<td>- Displays knowledge of 5 rights of delegation</td>
<td>- Delegates UAP to address appropriate patient; delegates Charge RN to address OR RN.</td>
</tr>
<tr>
<td></td>
<td>Right Circumstances</td>
<td>Psychomotor Objective</td>
<td>Method: LFS Station</td>
</tr>
<tr>
<td></td>
<td>Right Person</td>
<td>- Delegates tasks using 5 Rights of Delegation</td>
<td>- Verbalizes or demonstrates appropriate application of knowledge in delegating nursing tasks from report.</td>
</tr>
<tr>
<td></td>
<td>Right Direction/Communication</td>
<td>Behavioral Objective</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right Supervision/Evaluation</td>
<td>- Delegates at appropriate opportunity</td>
<td></td>
</tr>
<tr>
<td>patient care</td>
<td>Demonstrates knowledge surrounding clinical decision-making, critical thinking</td>
<td>Cognitive Objective</td>
<td>Method: HFS Station</td>
</tr>
<tr>
<td>management</td>
<td>and a global grasp of the situation, coupled with nursing skills acquired through a process of integrating formal and informal experiential knowledge and evidence-based practice.</td>
<td>- Recognize symptoms of CHF</td>
<td>- Reviews patient MAR, MD order, and labs; Holds medication; notifies MD; Administers med after orders received by MD; Correct Med Administration; hand hygiene; correct documentation and associated physical assessment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Understand therapeutic effects and side effects of medications</td>
<td>Method: LFS Station</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Psychomotor Objective</td>
<td>- Verbalizes or demonstrates Med Admin Safety.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Perform a focused, C/V assessment including VS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behavioral Objective</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Hand Hygiene, safe Med/Admin</td>
<td></td>
</tr>
</tbody>
</table>
Calculated reliability estimates for the LFS prioritization rubric was .85, for the LFS delegation rubric was .78, and for the LFS patient care management rubric was .70. Table 11 is also the blueprint used to create the LFS rubrics used in this study.

**The OCSE Assessment Environment**

Jones, Pegram, and Fordham-Clarke (2010) provided the framework for the step-by-step development of each station of the OSCE used in this study. There were seven stations in this Nursing Leadership OSCE. Table 12 presents an overview of the OSCE Assessment Environment. Station one was the check-in and introduction station. Stations two, three, and four were LFS stations. Station five was a HFS station which used a SP portraying one of the patients from the recorded report. Station six was an oral questioning station related to the handoff report. Station seven was where the participants were administered the MC testing instrument. The entire Nursing Leadership OSCE took approximately 120-minutes for participants to complete. All students received 7.5 hours of clinical time for participating in the OSCE.

**Table 12.**

**OSCE Assessment Environment Overview**

<table>
<thead>
<tr>
<th>Station</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Check-in Audio Report #1</td>
<td>Safe Med Admin</td>
<td>Delegate tasks</td>
<td>Prioritize patient care needs</td>
<td>Provide care to SP</td>
<td>OQ Audio Report #2</td>
<td>MCT</td>
</tr>
<tr>
<td>Method</td>
<td>n/a</td>
<td>LFS</td>
<td>LFS</td>
<td>LFS</td>
<td>HFS</td>
<td>OQ</td>
<td>MCT</td>
</tr>
<tr>
<td>Time Allotment(minutes)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>20</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

*Note. LFS = low-fidelity simulation, HFS = high-fidelity simulation, PRI = prioritization, DEL = delegation, PCM = patient care management, MCT = Multiple-choice testing, OQ = Oral questioning, SP = Standardized Patient Actor*
Participants entered the OSCE as if they were starting a clinical nursing shift. Each participant began the OSCE in the role of a RN on the day shift of a medical/surgical nursing unit in the acute hospital setting accepting responsibility for four patients and awaiting a patient admission to the unit. The participants came dressed in their school nursing uniform with ID, stethoscope, watch, and note taking material. Note taking was allowed and participants had access to patient charts, medication administration records (MARs), and reference books. Each participant group consisted of six students who received an orientation to the laboratory, and the timing of the stations. Group members were assigned randomly and their start times were provided two weeks prior to the OSCE assessment date. The groups of six were staggered by a half-hour to arrive at the Simulation Laboratory on the university campus. After check-in, the group of six was divided into subgroups of two participants with each subgroup proceeding to a six-minute circuit consisting of Stations Two, Three, and Four. Lastly, the OSCE consisted of seven stations where the four measurement methods were embedded. Appendix E represents of the OSCE rotation stations with the times and faculty, research assistants, and standardized patient resources that were used during administration. Each of the seven stations are described in more detail below.

**Station one.** Participant groups entering the OSCE were greeted by a research assistant (RA) at Station One. The RA conducted a two-minute check-in and orientation to the experience. Students signed-in and received a participant number and a packet containing the station rubrics for specific OSCE stations. Participants were asked to place their participant number on their left shoulder for identification and to remove their picture ID or name tag. After the check-in each student was directed to the nursing report
kiosk, handed headphones, and listened to a five-minute recorded patient change of shift report (handoff). The handoff was a simulated nurse report from the night shift. Appendix F is the transcribed recorded shift handoff report that participants received.

**Station two.** Station Two was a LFS station lasting six minutes with one minute for supply gathering and transition time. At this station the participant was presented with patient orders for medication administration. Prior to obtaining the medication from the medication dispensing machine the participant must verify the medication order. Once the medication is verified, the participants were rated on use of five rights of medication administration (right patient, right medication, right route, right time, and right dose). A faculty observer evaluated the participant as the medication was administered to a task trainer set-up for this purpose. The task trainer was a mannequin arm used for intravenous medication administration. The mannequin arm was positioned on a bedside table with a full-sized static mannequin sitting upright in a hospital chair simulating patient Stacey Collins from the recorded report. Appendix G is the LFS Rubric – PCM Station 2 (0700 IV push Lasix Administration for Patient in Room 1A).

The objective of Station Two was to manage safe medication administration related to the 0700 scheduled medication of intravenous push Lasix (furosemide) which is a potassium wasting diuretic to patient Stacey Collins. According to report, Stacey Collins had a morning laboratory potassium value of 3.0 (normal potassium values range from 3.5 to 5.5; 3.0 is a critically low potassium value). A total score of seven was achieved if the participant demonstrated all seven behavioral objects/indicators correctly on the rubric for this station. After six minutes an RA timekeeper called time and the
participants rotated to the next Station (Station Two to Station Three; Station Three to Station Four; Station Four to Station Two).

Station three. Station Three was a LFS station lasting six minutes plus one minute for supply gathering and transition time. At this station the participant was presented with seven nursing tasks from the recorded report. The participant had six minutes to delegate each task using the Five Rights of Delegation to one of the following team members: themselves, UAP, licensed vocational nurse (LVN), or Charge Nurse. The seven tasks that each participant needed to delegate at Station Three are present in Appendix H. The faculty evaluator assessed the participant on delegation skill using the Five Rights of Delegation (right person, right task, right situation, right explanation, right follow-up). Appendix H represents the LFS Rubric – Delegation Station Three.

The participant objective for Station Three was that the seven nursing tasks were delegated to appropriate members of the care team. Participants received a total score of seven if they demonstrated all the behavioral objectives for Station Three. After six-minutes an RA timekeeper called time and the participants rotated to their next Station (Station Two to Station Three; Station Three to Station Four; Station Four to Station Two).

Station four. Station Four was a LFS station lasting six minutes plus one minute to gather supplies and transition time. The participant was provided a written transcript from the recorded patient report that depicted the four patients that as the RN, they were responsible for during this day shift. The participant had six minutes to prioritize the care needs of seven nursing tasks based on the Three Levels of Prioritizing Patient Care. The transcript for the patient report is presented in Appendix F.
Faculty evaluated participant prioritization skills using the LFS Rubric - Prioritization Station Four. The LFS Rubric – Station Four, Prioritization is presented in Appendix I. The objective for Station Four was to prioritize the sequence of nursing rounds immediately after receiving report. The behavioral objectives for Station Four are listed in Appendix I. A total score of seven was possible if the participant correctly delegated the seven tasks to the correct team member. After six minutes an RA timekeeper called time and the participants from Stations Two, Three and Four moved to the Station Five holding area for a two-minute HFS room orientation.

**Station five.** After the two-minute simulation room orientation each participant was escorted to a one of three identical simulation rooms and waited for the signal to begin the simulation. Once the “begin simulation” signal was given participants entered the simulation room and began caring for the patient in Room 1A from their recorded patient report. Each simulation room was identical in set-up and contained one SP portraying the patient Stacy Collins.

The HFS lasted 20-minutes including room orientation and SP feedback. The participant had 15-minutes to assess this patient based on the recorded report. During their patient interaction the participant was interrupted on two occasions. These two interruptions provided participants with delegation opportunities and opportunities for prioritization of care. Further, participants during the HFS were assessed on safe management of patient care surrounding medication administration and appropriate practice related to a patient presented with critically low potassium laboratory values (3.0) and a scheduled medication that was a potassium wasting diuretic.
After the 15-minute interaction between the participant and SP, the SP provided one-minute of feedback to the participant. Each 15-minute interaction was video recorded and two Faculty evaluators reviewed the video recording and assessed the participant using the HFS Rubric (Appendix J). The simulation outcomes and behavioral outcomes for this HFS station are presented in Appendix J. A total score of 21 points were possible for Station Five (0- to 7-points for demonstrating behavioral or cognitive indicators for each of the three constructs). After completing Station Five, the participant was escorted to Station Six.

**Station six.** Station Six was a 20-minute oral questioning station. The participant had 10-minutes to listen to a recorded patient report followed by a 10-minute standardized oral questioning session with a faculty evaluator. The session was audio recorded. At a later date two trained faculty evaluated each participant recording using the Oral Questioning Correct/Incorrect Instrument. The Oral Questioning Correct/Incorrect Instrument is presented in Appendix C. Questions three, four, and five were related to delegation; questions one, two, seven and eight were related to prioritization; and questions six, nine and ten were related to patient care management. A total score of 21 points was possible for this station (0- to 7-points for each of three constructs). After 10-minutes of questioning the RA timekeeper called time for Station Six. Participants were escorted to Station Seven.

**Station seven.** Station Seven was a 30-minute multiple-choice test (MCT) station. During the multiple-choice test component of the OSCE, participants were provided a Scantron® answer sheet, one copy of the 21-item MCT (Appendix A – Multiple-Choice Test), and a number two pencil. Two RAs proctored the administration of the MCT.
Scantron® answer sheets were scored using a University supplied Scantron® machine and appropriate software used for item analysis. The blueprint used to create items for the MCT including the objectives, Bloom’s Taxonomy Level, and the items associated with prioritization, delegation, and patient care management are presented in Table 6. A total of 21 points was possible for this station (0 to 7 points for each construct). After 30-minutes the RA timekeeper called time for Station Seven. Participants were thanked for their participating in the OSCE and were released for the day.

In summary, seven OSCE stations generated twelve scores. The researcher, four adjunct clinical faculty, a simulation expert, and 14 student research assistants administered the OSCE assessment to the participants. All involved in administering the OSCE were asked to sign a confidentiality agreement (Appendix K). Participants were assessed on their prioritization, delegation, and patient care management abilities by the researcher and four adjunct clinical faculty. Eight student research assistants (RA) were responsible for collecting and organizing all of the rubrics and Scantrons®. The 21-Item MCT, the Oral Questioning Instrument, the HFS Rubric, and three LFS Rubrics were used to generate the twelve scores for participants during each 120-minute OSCE cycle. During the OSCE, the simulation expert was available to assist with timekeeping and break relief. The remaining two student research assistants were supply runners.

**Two Additional Standardized Assessment Methods**

Two additional standardized assessment methods were used in this study: the *Test of Essential Academic Skills (TEAS®) Exam* and the *Kaplan® RN Predictor Exam*. This section will provide a description of each exam, the reliability estimates for each exam (according to literature), and the rationale for its use in this study.
Test of essential academic skills (TEAS®). First, the Test of Essential Academic Skills (TEAS®) is a standardized 209-minute, computer adaptive test (CAT), four-option multiple-choice standardized exam containing of 170 questions. The TEAS® measures basic essential skills in the academic content area domains of reading, mathematics, science and English and language usage. The test is intended for use primarily with adult nursing program applicant populations. The objectives assessed on the TEAS® are those measures considered most appropriate and relevant by nurse educators to measure entry-level skills and abilities of nursing program applicants. An additional 20 unscored pre-test items are administered along with the 150 items representing the four content areas (Assessment Technologies Institute ©, LLC, 2009) (ATI). Second, reliability for the TEAS® V which is the latest version was estimated to be .92 (ATI, 2012; Salvucci, 2012). Until recently, all nursing students applying for entrance into the nursing program at the target university were required to take the TEAS® V. While the exam has reported high reliability, the nursing faculty at the targeted university voted to cease using the TEAS® V as a readiness indicator for success in the nursing program for two reasons. The first reason was the cost of using the program series ($770 per student). The second reason was the predictive low predictive reliability for passing NCLEX® RN which was 79% (ATI, 2012; Salvucci, 2012).

Third, the rationale for including the TEAS® scores in this study was twofold: the test is a nationally recognized standardized exam used specifically for nursing students; and the test is a measurement of ability in key content areas. The assumption was that scores from the TEAS® would provide an outside dataset generated from the participants to compare to the twelve generated OSCE scores. Further, these TEAS® scores from
participants when correlated with their twelve OSCE scores provide correlational data as another means to evaluate the methods used in this study.

**Kaplan® RN predictor exam.** First, the *Kaplan® RN Predictor Exam* is a standardized 180-item test that identifies nursing student readiness to take the *NCLEX-RN®* licensing exam. This test provides students with predictive information for their probability of passing the *NCLEX-RN®* licensing exam. Typically, the *Kaplan® RN Predictor Exam* is taken at the completion of a nursing program and before any *NCLEX® RN* review course. According to Kaplan® research, students who score 65% or higher have a 93.3% chance of passing the *NCLEX-RN®* licensing exam the first time they take the exam (Kaplan, 2011).

Second, no data was available on the reliability of this exam on the Kaplan® Nursing Website. However, Kaplan® developed this test following guidelines in the test plan or blueprint for the *NCLEX-RN®* examination published by the National Council of State Boards of Nursing (NCSBN). The *NCLEX-RN®* examination is a secured examination, so significant differences exist between the actual *NCLEX-RN®* examination and the *Kaplan® RN Predictor Exam* (Burckhardt, 2004). The *Kaplan® RN Diagnostic Exam* which is an exam that provides examinees an evaluation of nursing content strengths and weaknesses has reliability estimates of .90 (Burckhardt, 2004; Kaplan, 2011; Lockie, Van Lanen, and Gannon, 2013; Ukpabi, 2008). Further studies conducted by Kaplan® Nursing identified the probability rates of *NCLEX-RN®* pass/fail rates based on scores achieved on the *Kaplan® RN Predictor Exam*.

Third, data collected from nursing school graduates who voluntarily self-reported their first-time *NCLEX-RN®* pass/fail decision from May 2010 through August 2011 was
used by Kaplan® Nursing to generate statistical information regarding the RN Predictor Exam. The sample of 735 NCLEX-RN® examinees presented in the Kaplan® research, also took the Kaplan® RN Predictor Exam between March 2010 and August 2011. All examinees were US trained; except one whose training location was not self-identified. Further, data was collected regarding the type of program in which the examinee was enrolled during the time the Kaplan® RN Predictor Exam was administered.

Fourth, examinees who failed the NCLEX-RN® averaged 49.5% of the Kaplan® RN Predictor Exam items correct, while those who passed averaged 57% of the items correct. This difference in means was statically significant (F = 10.6, p = .001). Logistic regression was used to determine the probability of passing the NCLEX RN®. The Diagnostic Exam score of the examinee was used to predict passing or failing the NCLEX RN®. Overall, 93.1% of the examinees were classified correctly using this model (Kaplan, 2011).

Lastly, the rationale for including the Kaplan® RN Predictor Exam scores in this study is that of the thirty content areas on the exam, three are closely related to the constructs evaluated in this study: Setting Priorities (prioritization), Management of Care (delegation), and Making Nursing Judgments (patient care management). Setting Priorities has a total of twenty-three items on the exam. Management of Care has a total of twenty-five items on the exam; and Making Nursing Judgments has a total of seventy-two items on the exam. Students who participated in this research took the Kaplan® RN Predictor Exam at the end of NURS/L 4200. Participant scores from the three content areas of the Kaplan® RN Predictor Exam when correlated with the twelve OSCE scores
provide another independent form of assessment used to evaluate the measurement methods from this study.

**Procedures**

This section is presented in five subsections. First, participant selection is described. Second, the data collection process is presented. Third, the OSCE assessment environment set-up, location, and resources are described. Fourth, the budget for the OSCE is described. Lastly, the training process to administer the OSCE is described.

**Participant Selection**

The study was conducted at a private, Northern California University. Access to the senior-level nursing students occurred during their NURS/L4100 (Advanced Medical-Surgical Nursing) and NURS/L4200 (Nursing Leadership and Patient Care Management) courses during the 2012/2013 Academic Year. The Advanced Medical-Surgical Nursing Faculty (NURS/L4100) and Nursing Leadership and Patient Care Management Faculty (NURS/L4200) agreed to allow the researcher to contact students at the beginning of each semester to explain this study, data collection processes, and course impact.

Students in both senior-level nursing courses had the data collection process integrated into their nursing course assessment process. Participants were notified of the date and time of their assigned OSCE session at the beginning of the semester and once again two weeks prior to the date. The OSCE rotation for each data collection date is presented in Appendix E. Student groups rotated through the OSCE every thirty-minutes and their participation in the entire OSCE lasted approximately 120-minutes.

Participants in each data collection group were randomly assigned to their starting time using a random number generator. The number generated was matched to the last
digit of each student’s university identification (ID) number. Each time slot contained six slots. Students were assigned times systematically from a student roster for both courses from bottom to top until all students with the last digit of their student ID matching the generated number were exhausted. If the six slots were filled by students from the roster a new number was generated. This sequence continued until all six student slots were filled for each half-hour time group during each data collection date. Students were not allowed to change data collection dates after the dates were announced.

**Data Collection**

Data collected during the OSCE were managed four ways. First, Station Two, Three, and Four required evaluators to mark participant Scantrons® during the OSCE. Packets containing the Scantrons® were presented to each participant upon check-in. The participant and all the packet contents were marked with a number. The contents of the packet were screened at each station by research assistants (RAs) to ensure that no information was misplaced in participant transition from station to station. An RA at each station was assigned to sign the outside of the packet as a visual cue that the participant had completed all stations. The packets were collected at station seven by an RA and stored in a locked file case.

Second, the OQ station data were stored on audio files. The three RAs assigned to this station were responsible for maintaining the integrity of the recording device and its files to ensure that no data was lost. Participants were asked to state their participant number into the recording device prior to the questioning session. The recording devices were stored in a locked file case off-site until reviewed and scored by two evaluators at a later date. During review of the audio files, evaluators marked participant rubrics from
the OQ station which were labeled with participant numbers. Completed OQ mark sheets were reviewed by the researcher for completeness, sorted sequentially by participant number, and stored in a three-hole binder off-site.

Third, the HFS data were stored as video files with audio on equipment stored in the Simulation Laboratory at the University. The Simulation Director and the RAs assigned to that station were responsible for reviewing the video files immediately after each simulation session to ensure that no files were lost. Once all participants completed the OSCE for that date, the Simulation Director transferred the video files to two flash drives that were capable of supporting large files. One flash drive was given to each evaluator to review and mark the HFS rubric at a later date. Completed HFS rubrics were reviewed by the researcher for completeness, sorted sequentially by participant number, and stored in a three-hole binder off-site.

Fourth, the MCT data was collected on a Scantron® Test Sheet (Form number 95679). Participants filled-in the bubble(s) on the Scantron® Test Sheet that corresponded to their selected response(s) on the MCT. The MCTs were numbered. Research Assistants assigned to this station were responsible for ensuring that participants did not collaborate during administration of the MCT. After participants completed the MCT, the RAs would screen the Scantron® Test Sheet for completeness, attach the numbered MCT to the numbered participant Scantron® Test Sheet, and store both documents in a locked file case. The researcher removed the MCT documents from the file case and ran them through a scoring device. Scored Scantron® Test Sheets were sorted sequentially by participant number and OSCE assessment date off-site.
Fifth, participant scores from the TEAS® were collected as a requirement for consideration of admission into the University’s nursing program. The test was proctored by a nursing administrator during specific intervals throughout the academic year. Traditional students (students entering the University immediately from high school) were administered the TEAS® prior to the beginning of their first semester. Non-traditional students were administered the TEAS® during the semester prior to admission into the nursing program. A national normed score was used as a criterion for admission into the nursing program. Scores for the test were stored on a secure website maintained by the Assessment Technologies Institute® (ATI). The researcher received permission to use the TEAS® scores through the Institutional Review Board (IRB) at the University. Retrieved scores were matched to participant OSCE numbers and stored off-site.

Scores generated from the Kaplan® RN Predictor Exam were collected similarly to those scores collected for the TEAS®. While the TEAS® was administered prior to admission into the nursing program, the Kaplan® RN Predictor Exam was administered just prior to graduation from the nursing program. The Kaplan® RN Predictor Exam is a proctored computer adapted test (CAT) that uses a national normed score to predict the graduating nursing student’s success on the national licensing exam for registered nurses. Scores for the Kaplan® RN Predictor Exam were stored on a secure website maintained by Kaplan®. The researcher secured approval to use these scores through the IRB at the University. Retrieved scores were matched to participant OSCE numbers and stored off-site.
The OSCE Assessment Environment

Set-up, location, and resources. The OSCE environment encompassed three areas of the University. The main area used to administer the OSCE was the Simulation Laboratory (SIM Lab) which housed the HFS station and three LFS stations. The SIM Lab at the University consisted of three simulation rooms, each with the capability to video record a scenario, and one very large classroom. Only two rooms were used as the third room had malfunctioning equipment. Further, its close proximity to another simulation room warranted data contamination concerns related to sound and voices traveling from one room to the next. The large classroom was used to house the three LFS stations.

The two simulation rooms used for the HFS were equally sized and contained identical supplies. Each room had a patient chart, a medication administration record, and a drug resource book. Both rooms contained a medication dispensing machine, a hospital bed, bedside table, and telephone. Located above each telephone was a white dry-erase board which contained the names and telephone numbers for the Charge RN, the pharmacist, the unit secretary, and the physician.

Each room was assigned to a standardized patient (SP) who was a trained actor portraying the role of Stacey Collins. The SP was located and remained in the hospital bed during the entire encounter. Each SP was dressed in a hospital gown and had an identification bracelet located on their left wrist as well as a venous access device taped to their right forearm for the administration of intravenous medication. While no actual medications were administered, nor was the venous access device inserted into the SP’s vein, the appearance of each actor was typical of hospitalized patient. Further, each room
appeared to resemble a hospital room located on a medical/surgical nursing unit. After eight simulation encounters, each SP was replaced with another trained actor to maintain standardized portrayal of Stacey Collins.

Both SIM Labs contained two research assistants, the Standardized Patient (SP) Trainer, and either the Simulation Director or the Simulation Expert. These people were located in the control room of each simulation room which was on the other side of a one-way mirror. The purpose of the research assistants was to monitor recording equipment and provide the scripted telephone calls into the room. The Simulation Director and Simulation Expert were on hand to troubleshoot malfunctions with the recording equipment and ensure the simulation fidelity was maintained throughout the scenario. The SP Trainer rotated rooms to ensure that the actors remained in charter, maintained the fidelity of the scenario, and ensured the rotations of SPs after eight encounters as well.

The large classroom was configured to support the three LFS stations, a participant check-in area located in the vestibule, and an area to listen to the audio report. The listening area contained three computers which stored the report audio file. There were also six headphones, sitting for six people, scratch paper for note taking, and alcohol wipes to clean the headphones after each use. Two RAs managed the check-in and listening areas. Figure 4 is a schematic of the Fall 2012 OSCE set-up and configuration which shows the stations, supply areas, and personnel assigned to specific areas.
Figure 3. The Room Layout for the OSCE

Each of the LFS stations contained supplies specific to the construct being tested.

For example, the prioritization station contained a long table, three chairs, seven
laminated 5 x 8 index cards each containing one of seven nursing tasks from the report requiring prioritization, and blue masking taped areas on the table in the configuration of four boxes. Each box represented a level of prioritization and one box labeled “unsure.” The boxes marked off with blue masking tape were further labeled as “A,” “B,” “C,” and “D” which corresponded to the letters on the Scantron® Test Sheet. The delegation station was set-up similarly, but with laminated 5 x 8 index cards representing the seven nursing tasks that required delegating. Description of the patient care station was presented earlier. Lastly, a supply area was set-up between stations which contained extra supplies for medication administration; syringes, examine gloves, alcohol wipes, and saline flush syringes.

The second area used was the Skills Laboratory (Skills Lab). The Skills lab was located in a small build across from the SIM Lab. The OQ station was housed in the Skills lab. This area contained a computer, a large table, seating for eight people, scratch paper for note taking, and a drug reference book. The computer contained an audio file of the recorded nursing report. Three areas containing a small table, two chairs, and an audio recorder were set-up in main room. Each of these “questioning areas” required enough distance between one another to allow privacy and clear recording of the participant and the questioner.

The third area used was a classroom for the MCT station. This area varied between data collection dates. Two research assistants were assigned to this station. One RA was assigned to escort participant groups from the OQ station to the MCT to decrease transition time and minimize discussion of the OSCE. The second RA was assigned to control the distribution of the MCT and manage the completed documents. Resources
needed for this area included seating for six participants, pencils (Number 2 Lead), Scantron® Test Sheets (Form Number 95679), a timer, and a locking case file.

In summary, the OSCE set-up, location, and resources needed required significant planning to manage during each data collection date. Three areas were used on the University to ensure enough space and the logistics required to administer the seven OSCE stations. Further, the resources required to test the four methods being evaluated in an OSCE assessment environment were accessible to the researcher; however, required planning and accrued expense. As a means to manage the planning and expenses needed to administer an OSCE assessment environment, a budget that estimated costs was necessary. The next section describes the budget used to estimate and contain the expenses of such an assessment environment.

Budget

A budget was created for the OSCE based on Poenaru et al. (1997) and Palese et al. (2012). The overall expenditures for the project were based on 206 students which included the pilot study. The projected student number was estimated based upon the students enrolled in the nursing-level courses targeted as the research population (Fall 2012—pilot, senior-two; Fall 2012—senior-one; Spring 2013—senior-one, and senior-two). The budgeted dollar amount for the standardized patients and the simulation consultant were projected to be the largest expense items for this project (SP—$4,000; Simulation Consultant—$4,900). An hourly rate of $60/hour was negotiated to secure the needed support for the entire project. The result of this negotiated hourly rate was the expense was associated costs for faculty evaluators and markers cost which was also a high budget item ($4,000 and $4,000). The budgeted resources for the pilot and research
study are presented in Table 13 as a comparison to budgets presented by Poenaru et al. (1997) and Palese et al. (2012).

**Training**

A large number of personnel were needed to administer the OSCE assessment and evaluate the participants who went through the OSCE during the three data collection dates. Faculty evaluators were trained on four occasions: three prior to the pilot study and once after the pilot study. Faculty training consisted of detailed description of the research process, the methods used to evaluate the three traits of the study, and the logistics of using the developed instruments and rubrics. Each training session lasted four hours. Only trained faculty evaluators were used at each station. Only faculty from the target university who had no current contact with the participants of the study were used as faculty evaluators. Faculty who were trained had adjunct status at the university and taught clinical sections in the nursing program.

There were fourteen research assistants (RAs) used in this study. The RAs were senior-level nursing students enrolled in a directed research course at the University. A directed research course was required for graduation at the target university. Students enrolled in the directed research course were taught by the author of this study. Final selection of the research assistants was based upon grade point average and course grade in a required nursing research course at the University. Students with the highest grade in NURS3103 were selected first.
Table 13.

Comparison of Actual OSCE Costs with Estimates from the Literature

<table>
<thead>
<tr>
<th>Category</th>
<th>Highest Estimate*</th>
<th>Lowest Estimate**</th>
<th>Budgeted Estimate</th>
<th>Actual Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principal Staff – Implementation Phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Leader</td>
<td>10,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standardized Patient Trainer</td>
<td>20,000</td>
<td>20,000</td>
<td>1,500</td>
<td>1,700</td>
</tr>
<tr>
<td>Administrative Support</td>
<td>15,000</td>
<td>15,000</td>
<td>0</td>
<td>6,000</td>
</tr>
<tr>
<td>Simulation Cost</td>
<td>0</td>
<td>0</td>
<td>4,900</td>
<td>3,500</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>45,000</td>
<td>35,000</td>
<td>6,400</td>
<td>11,200</td>
</tr>
<tr>
<td><strong>OSCE Development Phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review of Course Objectives; scenario creation</td>
<td>600</td>
<td>0</td>
<td>4,995</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty honorarium for workshops</td>
<td>12,000</td>
<td>20,000</td>
<td>250</td>
<td>750</td>
</tr>
<tr>
<td>Catering for case development workshops</td>
<td>360</td>
<td>360</td>
<td>0</td>
<td>150.56</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14,200</td>
<td>360</td>
<td>5,245</td>
<td>4,400.56</td>
</tr>
<tr>
<td><strong>OSCE Production Phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient training</td>
<td>2,000</td>
<td>2,000</td>
<td>510</td>
<td>300</td>
</tr>
<tr>
<td>Printing of OSCE material</td>
<td>3,000</td>
<td>500</td>
<td>60</td>
<td>50.35</td>
</tr>
<tr>
<td>Artifacts, props, and supplies</td>
<td>1,000</td>
<td>800</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6,000</td>
<td>3,300</td>
<td>640</td>
<td>450.35</td>
</tr>
<tr>
<td><strong>OSCE Administration Phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized Patients</td>
<td>13,500</td>
<td>13,500</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty Examiners</td>
<td>13,500</td>
<td>0</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty Markers</td>
<td>1,800</td>
<td>0</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Support Staff</td>
<td>2,700</td>
<td>2,700</td>
<td>500</td>
<td>600</td>
</tr>
<tr>
<td>Catering</td>
<td>3,000</td>
<td>3,000</td>
<td>400</td>
<td>752.13</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>34,500</td>
<td>19,200</td>
<td>12,900</td>
<td>11,852.13</td>
</tr>
<tr>
<td><strong>Post-OSCE Analysis and Reporting Phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Entry and Review</td>
<td>1,600</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>1,600</td>
<td>1,600</td>
<td>280</td>
<td>119^</td>
</tr>
<tr>
<td>Report Generation</td>
<td>1,500</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>4,700</td>
<td>59,460</td>
<td>23,780</td>
<td>0</td>
</tr>
<tr>
<td>Total for Entire Implementation</td>
<td>104,400^2</td>
<td>59,460^2</td>
<td>24,060^8</td>
<td>28,022.04</td>
</tr>
<tr>
<td>Total Students</td>
<td>120</td>
<td>120</td>
<td>206</td>
<td>199</td>
</tr>
<tr>
<td><strong>Total per Student</strong></td>
<td>870</td>
<td>496</td>
<td>116.79</td>
<td>140.81</td>
</tr>
</tbody>
</table>


Training of the research assistants began in fall 2012 during the first day of class for NURS4997 which is the directed research course required by all nursing students at the university. Research assistants were informed of the research procedures, and were assigned segments of the study to research and manage. For example, student pairs were
assigned to simulation in nursing education as a research topic. These students would research the topic, review the literature, and manage either the HFS or one of the LFS station(s) during the OSCE assessment.

All student researchers signed a confidentiality agreement and were aware that their course grade was connected to their performance during the training sessions and data collection dates. Eight of the fourteen research assistants were honor students. All students were required to present a poster board and provide an oral presentation of their topic with results. Training of all research assistants occurred during class time each week for one hour. These students were also required to be present during each of the data collections dates to set-up the OSCE stations, collect data at their stations, to manage the data associated with their station, and to aide in clean-up after the OSCEs.

Standardized patient actors (SPs) were recruited and hired to portray Stacey Collins in the HFS station of the OSCE. Training of the SPs occurred on four occasions prior to the pilot in October 2012. All of the actors used as SPs had previous experience in medical education. A script was written by the SPs Expert/Trainer (who was hired to manage the SPs) with supervision by the Simulation Expert, the Simulation Director at the university, and the author of the study. Each SP training session lasted four hours and was similar to a play rehearsal with two OSCE runs in real-time conducted prior to the pilot.

Pilot Study

Following IRBPHS approval, an OSCE assessment was piloted on the Fall 2012 senior-two nursing students at the target University. This was a similar group of students to those students for the intended study sample. The Nursing Department Chair provided
permission to conduct the pilot test on October 25, 2012. Permission of the professor was obtained and the simulation experience was administered. The OSCE piloted had integrated HFS and LFS stations.

Following the pilot, modifications were made to separate HFS and LFS testing into separate stations due to logistics. Further, analysis of item difficulty and discrimination determined that MCT item revision was necessary. The goal for MCT revision was to achieve a level of difficulty that would provide maximum variance in scores ($r = .40$ to $0.60$), and a point biserial correlation coefficient that would provide adequate item discrimination ($r_{pb}$ of $.30$ or above) in accordance to standard objective test principles (Popham, 2000).

Initial reliability for the 30-item MCT piloted was .32. This reliability calculation was below standard objective test principles. Elimination of nine items based on very high or low difficulty or a negative point biserial correlation resulted in the final reliability of .80 with an item count of seven for all three subscales. The Oral Questioning, HFS Rubric, and LFS Rubrics were analyzed using a different method.

Three highly experienced nursing faculty formed a consensus panel that provided assessment for the oral questioning items. The expert responses formed the basis for revising the recorded report (nurse report/handoff) at Station One and Station Six, script revisions, and the HFS and LFS separation from the original OSCE format. The recommendations made by the expert panel were that the use of LFS associated with measuring participant behavior in the three construct areas was actually a fourth method of assessment (Appendix L: Expert Panel Review Checklist). By separating the HFS and LFS as well as revising the recorded RN report/handoff the panel felt that increased
fidelity in the OQ station and the overall student experience would occur. Further, removing the embedded skills from the HFS station improved the logistical flow of the entire OSCE assessment environment.

A summary of the descriptive statistics obtained for each of the methods used to assess the three traits is presented in Table 1. Subset scores from prioritization items of the MCT indicated a moderate level of difficulty ($M = 2.16, SD = 2.05$), with the item means ranging from .16 to .50. Scores on similar subsets from the OQ Instrument indicated delegation items were moderately difficult for the pilot participants ($M = 2.73, SD = 1.59, p = .19$) with strong correlation (.81) and moderate reliability magnitude (.67).

The original simulation rubric used in the pilot indicated prioritization items as moderately difficult for participants ($M = 2.68, SD = 2.07$) with low correlation magnitude (.48) and high reliability (.85). However, the skills which were originally embedded in the simulation experience indicated that the patient care management items were moderately difficult for the pilot participants ($M = 2.34, SD = 1.95$) with moderate correlation magnitude (.51) and high reliability (.82). OQ patient care management items had the lowest reliability (.56) and the highest reliability (.85) associated with prioritization items from the pilot.
Table 14.
Reliability, Means (M), and Standard Deviations (SD) for the Pilot Study

<table>
<thead>
<tr>
<th>Method</th>
<th>N = 62</th>
<th>MCT</th>
<th>OQ</th>
<th>SIM</th>
<th>SKILLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pri</td>
<td>del</td>
<td>pcm</td>
<td>pri</td>
<td>del</td>
</tr>
<tr>
<td>reliability</td>
<td>.67</td>
<td>.63</td>
<td>.67</td>
<td>.59</td>
<td>.62</td>
</tr>
<tr>
<td>M</td>
<td>2.16</td>
<td>2.55</td>
<td>2.45</td>
<td>2.94</td>
<td>2.73</td>
</tr>
<tr>
<td>SD</td>
<td>2.05</td>
<td>2.31</td>
<td>2.05</td>
<td>1.92</td>
<td>1.59</td>
</tr>
</tbody>
</table>

In summary, permission was granted for a pilot study using an OSCE assessment environment with embedded low-fidelity skills into a HFS. The pilot OSCE was administered on October 25, 2012. Originally, a 30-item MCT was used in the pilot. Item analysis determined that the MCT used in the pilot did not meet standard objectives of test principles. Therefore, revisions were made to the MCT that included deletion of poor performing items based on point biserial correlations. The final MCT consisted of 21-items (seven items per construct) with a 0.80 estimated reliability.

An expert panel of nursing faculty reviewed the OQ items, the RN report/handoff, the HFS scenario and rubric. Based upon the expert panel recommendations, revisions were made to each of these OSCE assessment environment components. The result included modifications to the RN report/handoff, separating out the LFS skills that were originally embedded into the HFS, and revision of the HFS rubric and LFS rubrics with the goal of increased fidelity of the overall simulation of a nursing unit environment.

**Data Analysis**

This section presents the data analysis used in this research in four subsections. First, a figure is presented to illustrate the configuration of the MTMM used to arrange the computed correlations and reliability estimations. Second, the criteria developed by
Campbell and Fiske (1959) used to evaluate the MTMM configuration for construct validity are presented along with associated literature. Third, a summary of data analyses for each research question is outlined.

**Multitrait Multimethod Matrix (MTMM) Research Design**

Figure 4 illustrates the MTMM design for this project. The research used a MTMM approach to evaluate three constructs (a, b, and c) using four assessment methods (A, B, C, and D). Criteria established by Campbell and Fiske (1959) for determining construct validity was used to evaluate correlation and reliability computations of the twelve OSCE scores of the participants. Specifics on the analysis used to evaluate the twelve OSCE scores based on this criterion are described in the following section.

Using SPSS 21, Cronbach’s $\alpha$ estimates of the reliability were calculated for each of the methods used to measure the three traits (monotrait-monomethod) in this research to determine the overall reliability statistics of each instrument (Kenny, 2012; Marsh & Bailey, 1991). Calculated estimations of reliability were then analysed and arranged into the MTMM for evaluation of conformity to the criteria established by the Campbell and Fiske Approach (1959) for determining construct validity.

Data collected from each participant group was inputted into a SPSS® 21 Software database for analysis. The specific calculations analyzed for the MTMM were same-trait, different-method correlations (convergent validity, or “validity diagonals” of the MTMM); different-trait, different-method correlations (discriminant validity); and different-trait, same-method correlation (method variance) among all three traits and the four methods used in this research (Kenny, 2012; Marsh & Bailey, 1991).
Campbell and Fiske (1959) Criteria with Associated Literature

Campbell and Fiske identified four criteria to evaluate a MTMM. The first criterion is that the correlations in the validity diagonal(s) (“C” in Figure 4) should be significantly different from zero and be of a sufficient magnitude to encourage further validity exploration. This criterion is usually assessed by simply inspecting the magnitude of the correlations. The second criterion is that the specific correlations should be higher in the “C” diagonals when the same trait, different methods are correlated than when that trait is correlated with a different trait using different methods. This criterion is often assessed by averaging the correlations in the “C” diagonal and comparing the average to the average of the correlations located in the off-diagonals of the matrix. The third criterion is that the correlations between two different methods of measurement of the same trait should be higher than correlations between that trait and another trait. This criterion is assessed by inspecting the magnitude of correlations of the specific trait measured by different methods within the same row and column ensuring that the average correlation is higher between the same traits measured by different methods than the average correlation of different traits measured by different methods. The final criterion is that the correlations between traits (whether in the same method or different method blocks) should indicate patterns of interrelationships between traits and should be similar in the different trait triangles. This criterion is assessed by evaluating the magnitude of correlations between different traits measured by different methods. These correlations usually are among the lowest correlations located within the entire matrix (Ferketich, Figueredo, & Knapp, 1991).
Since the development of the MTMM, researchers have noted a number of problems with the criteria set up by Campbell and Fiske (Campbell & O’Connell, 1967; Marsh & Bailey, 1991; add Widaman 1985). Several alternative procedures for analyzing MTMM data have been proposed. For example, Marsh and Hocevar (1980) compared the four criteria for construct validity developed by Campbell and Fiske (1959) with two
other procedures. The two other procedures were an analysis of variance (ANOVA) model and confirmatory factor analysis. Marsh and Hocevar (1980) determined that the principal advantage of using the ANOVA model for analyzing MTMM data was that it is a convenient summary and test of convergent, divergent and method/halo effects. However, limitations of the ANOVA approach are numerous, and so the ANOVA approach should only be used to supplement other procedures (Kenny, 2012).

Research by Kenny (2012), Widaman (1985), and Marsh and Hocevar (1980), among others, posit that confirmatory factor analysis (CFA) provides a direct test of the statistical significance and importance of various trait and method factors. Unfortunately, the CFA models have been difficult to estimate. In many cases, models fail to converge, contain unreasonable or inconsistent estimates, and/or fail to fit the data. Consequently, this study will use the original Campbell and Fiske criteria as well as the correlations with the two external tests.

**Data Analyses for Each Research Question**

The first research question was answered using calculated reliability and correlation computations arranged into a MTMM matrix. Convergent and discriminant validity were determined based upon the Campbell and Fiske criteria. The second research question was answered by correlating the twelve MTMM scores with the four subtests of the *TEAS®* and the three subtests of the *Kaplan®*.

The third research question was answered by comparing OSCE budget estimates from two primary literature sources (Palese *et al.*, 2012; Poenaru *et al.*, 1997) with the actual costs of administering the OSCE for this research. Comparative analysis of the cost
savings or expenditures is presented for review in the results section of Chapter Four. Discussion of these findings is presented in Chapter Five.

**Summary**

The research methods described for this study included the use of an OSCE assessment environment which included MCT, OQ, HFS, and LFS stations. The OSCE assessment environment was considered the most logistically feasible mechanism to study the four nursing student assessment methods (MCT, OQ, HFS, and LFS) used to measure the three leadership constructs (prioritization, delegation, and patient care management). Construct blueprints based on Jones et al. (2010) were used to develop operationalized definitions of the three constructs. The three constructs were defined and cognitive and/or behavioral indicators associated with each construct were developed in order to measure and study this phenomenon.

These cognitive and/or behavioral indicators were then embedded into the objectives associated with the seven stations of the OSCE assessment environment. The entire OSCE lasted 120-minutes for each participant. This time included time for transition between stations and buildings on the campus. A convenience sample of students enrolled at a private, four-year university in two senior-level nursing courses (NURS4100 and NURS4200) were participants in this study. The participants rotated through seven stations which simulated aspects of the medical/surgical day shift in an acute hospital setting. Each participant received report (handoff) and was immersed into the role of registered nurse during the entire OSCE assessment.

First, twelve scores were generated by each participant during the OSCE assessment (four methods and three constructs). These twelve OSCE scores were
generated from a 21-item MCT, a 21-item OQ Instrument, one HFS rubric, and three LFS rubrics. The researcher estimated reliabilities of internal consistency for the MC testing items and LFS rubrics. Inter-rater reliability was calculated for the OQ Instrument and the HFS Rubric item scores based upon the two faculty evaluators who listened or viewed the participant experience, then scored participants based on four training sessions prior to the OSCE assessment. The MTMM approach was constructed to evaluate construct validity using estimations for reliability and correlation coefficients associated with the four methods used to generate the twelve OSCE scores. The calculated reliability estimations and correlations coefficients were evaluated using criteria developed by Campbell and Fiske (1959) to determine construct validity.

Second, the twelve OSCE scores were correlated with TEAS® content area scores and scores collected from three content areas of the Kaplan® RN Predictor Exam. The assumption was that the twelve OSCE scores would correlate highly with content areas from the TEAS® that were closely related to items that used any of the essential areas tested by the exam. High correlations observed from correlating Kaplan® RN Predictor Exam scores and the twelve OSCE scores would indicate that items from these two methods of measurement were evidenced as being similar to what each were measuring. Conversely, low correlations would evidence that these two methods were dissimilar in what each was measuring.

Lastly, a budget was developed to evaluate the feasibility of OSCE assessment use in nursing education based on parameters denoted by Poenaru et al. (1997) and Palese et al. (2012). Expenditures for the pilot (Fall 2012) and study (Spring 2013) were presented. Comparative information from the Poenaru et al. (1997) estimates, the Palese
et al. (2012) estimates, and the budget projected for this study were examined to determine the differences between the cost estimations to administer an OSCE assessment. The results for all findings are presented in Chapter Four.
CHAPTER FOUR

RESULTS

This chapter reports the results of the study in four sections. First, the correlation coefficients among the twelve OSCE scores are arranged in a matrix and standard evaluation criteria for interpreting the MTMM are presented as results for research question one. Second, the correlation coefficients between the twelve scores and scores from the TEAS® Exam and from three content areas of the Kaplan ® RN Predictor Exam are presented as results for research question two. Third, the costs associated with administering the OSCE assessment to collect scores generated for this study are presented and compared to the cost estimates for running similar OSCEs presented in the literature by Palese et al. (2012) and Poenaru et al. (1997) as the results for research question three. Lastly, the chapter concludes with a summary of the results.

Research Question One

To what extent do the twelve OSCE scores generated from using four assessment methods (multiple-choice tests, oral questioning, low-fidelity simulation, and high-fidelity simulation) to measure three constructs (delegation, prioritization, and patient care management) conform to the Campbell and Fiske (1959) criteria for construct validity? To address this question, the reliability estimates and the intercorrelation coefficients from the twelve OSCE scores were arranged into the MTMM matrix presented in Figure 5, and considered in light of the four MTMM criteria.
### Figure 5. MTMM for the Twelve OSCE Scores

<table>
<thead>
<tr>
<th>Method 1 MCT</th>
<th>Method 2 OQ</th>
<th>Method 3 HFS</th>
<th>Method 4 LFS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trait</strong></td>
<td><strong>Trait</strong></td>
<td><strong>Trait</strong></td>
<td><strong>Trait</strong></td>
</tr>
<tr>
<td>PRI 1</td>
<td>PRI 2</td>
<td>PRI 3</td>
<td>PRI 4</td>
</tr>
<tr>
<td>Del 1</td>
<td>Del 2</td>
<td>Del 3</td>
<td>Del 4</td>
</tr>
<tr>
<td>PCM 1</td>
<td>PCM 2</td>
<td>PCM 3</td>
<td>PCM 4</td>
</tr>
</tbody>
</table>

**Legend**

- PCM = Patient Care Management
- MCT = Multiple-Choice Testing
- LFS = Low Fidelity Simulation
- Del = Delegation
- OQ = Oral Questioning
- PRI = Prioritization
- HFS = High Fidelity Simulation using Standardized Patients
**Criteria one.** The first criterion is that the correlations in the validity diagonal(s) should be significantly different from zero and be of a sufficient magnitude to encourage further validity exploration. Failure of this criterion means that the measures lack convergent validity. This criterion is usually assessed by simply inspecting the magnitude of the correlations and making a subjective judgment as to whether the criterion was met. Correlations greater than .20 and .26 are statistically significant at the .05 and .01 levels of significance for two-tailed tests with a sample size of 100; for a sample size of 150, the same correlations are .16 and .21 (Popham, 2000).

There are six heteromethod blocks with validity diagonals in the MTMM, each with three convergent validity coefficients. All 18 validity coefficients are statistically significant at the .01 level of significance when considering a sample size of 150, and 17 of 18 when considering a sample size of 100. Most are moderate in magnitude; the only exception is the validity coefficient for HFS PCM and LFS PCM, which was only .23. Overall, this criterion is met.

It is important to note that the reliabilities of the twelve OSCE scores are relatively consistent across the twelve measures, ranging from a low of .70 to a high of .85. This is important because two measures can only correlate to the extent that they are reliable measures. That is, unreliability attenuates correlation coefficients. Thus, if the reliabilities are relatively consistent, whatever attenuation occurs is done fairly evenly and differences in correlations cannot be attributed to uneven reliability estimates.

**Criteria two.** The second criterion is that validity coefficients should be higher than the correlations in the same row and column in the heteromethod block. That is, the correlation between two measures of the same trait should be higher than the correlation
of that trait with other traits using the same methods. Failure of this criterion implies lack of discriminant validity. This criterion is sometimes assessed by averaging the correlations in the convergent validity diagonal and comparing the average to the average of the correlations located in the off-diagonals of the heteromethod block. Table 14 presents the average correlation coefficients for the method blocks compared to the average off-diagonal correlation coefficients.

Table 14. *MTMM Method Block Average Convergent Validity Coefficients Compared with Average Off-Diagonal Block Coefficients*

<table>
<thead>
<tr>
<th>Method Block</th>
<th>Average CV Coefficient</th>
<th>Average Off-diagonal Block Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>OQ x MCT</td>
<td>.49</td>
<td>.39</td>
</tr>
<tr>
<td>HFS x MCT</td>
<td>.55</td>
<td>.53</td>
</tr>
<tr>
<td>LFS x MCT</td>
<td>.59</td>
<td>.49</td>
</tr>
<tr>
<td>HFS x OQ</td>
<td>.49</td>
<td>.47</td>
</tr>
<tr>
<td>LFS x OQ</td>
<td>.43</td>
<td>.42</td>
</tr>
<tr>
<td>HFS x LFS</td>
<td>.56</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note. CV = convergent validity, OQ = oral questioning, MCT = multiple-choice test, HFS = high fidelity simulation, LFS = low fidelity simulation.*

The second C&F criterion was met by all six method blocks in MTMM. The differences between comparing the average validity diagonal coefficients and the average off-diagonal coefficients in the six blocks, while meeting the criterion, are not sizeable differences. The differences for the six blocks in Table 14 are .10, .02, .10, .02, .01, and .03. This suggests that while this criterion for construct validity was strictly met, the off-diagonal correlations are almost as high as the validity diagonal and that the evidence for construct validity is relatively weak.
A more specific procedure for examining C&F criterion two is to compare each validity diagonal correlation to the four other correlations in the same row and column.

Table 15 does this for each of the three validity coefficients in each of the six
heteromethod blocks, where the validity coefficient is shown first and then the four
correlations in the same row and column. In this table, the correlation between two
different methods of measuring the same trait (the validity coefficient) is compared to the
correlations of that same trait to different traits using the same methods. One would
expect that two different methods of measuring the same trait would correlate higher than
the correlations of that trait with different traits.

In Table 15, the row and column correlations that exceed the validity correlation
are given in bold face for each of the six heteromethod blocks. The blocks are identified
by the two methods used to measure the traits, starting with methods 1 and 2 and ending
with methods 3 and 4. As can be seen there are quite a few off-diagonal correlations that
are greater than the validity diagonal. Of the 72 off-diagonal correlations, 27 are higher
than expected. Across heteromethod blocks, prioritization had the fewest (6), followed by
delegation (9), and then patient care management (12). The method pair with the fewest
off-diagonals that were too high was MCT x LFS (3 of 12 were too high), followed by
three method pairs – MCT x OQ, OQ x HFS, and OQ x LFS (each with 4 of 12 too high),
and finally the method pair with the most was HFS x LFS with 6 of the 12 too high. To
help summarize, Table 15 also gives the method pair in parenthesis for the number of too
high correlations by trait. There does not appear to be much consistency, as some method
pairs do well with some traits, while the same method pairs less well. For example, OQ
shows up six times as one of the methods when there are none or only one correlation too
high, but twice as a method when three or four correlations are too high. Overall, there is
not a great deal of support for discriminant validity among the constructs.
**Criteria three.** The third criterion is that the validity coefficients should be higher than correlations between that trait and other traits that happen to share the same method. This is a second criterion focusing on discriminant validity. Failure of this criterion suggests that the traits may be highly correlated or that they share considerable method variance, or that both factors are contributing to the lack of discriminant validity. This criterion is assessed by comparing the magnitude of the validity diagonal correlation of a specific trait (in each of the six heteromethod blocks) to the correlations of that same trait with other traits within the monomethod triangle. That is, two measures of the same construct should correlate higher than the trait correlates with other traits that share the same method.

Table 16 provides the relevant comparisons, sharing the same structure as the previous table. In this table, the validity coefficients from each of the six heteromethod blocks are compared to the correlations in the monomethod triangle of the same trait with other traits that happen to share the same method. The basic idea is that two measures of the same trait should correlate higher than the correlations of each with other traits that share the same method. As in the previous table, the correlations that are too high are shown in bold.

As before, there are many correlations that are too high. Of the 72 correlations, 31 are too high. Across the six heteromethod blocks, delegation had the fewest correlations that were too high (8), followed by prioritization (11), and then patient care management (12). The method pair with the fewest monomethod triangle correlations that are too high is MCT x LFS (3 that are too high), followed by MCT x HFS (4 that are too high), MCT
x OQ and HFS x LFS (5 that are too high), OQ x HFS (6 that are too high), and OQ x LFS (8 that are too high).

To summarize, Table 16 presents the same kind of information as Table 15. Table 16 presents the method pair for the number of too high correlations by trait. Again, there are inconsistent results across method pairs. It does appear that the method pair of MCT with any other method works well. Once again, there is not much evidence for discriminant validity. Most correlations are uniformly moderately large and may share considerable method variance.

### Table 16.

**Heteromethod Block Validity Coefficient Comparisons with Correlations in the Monomethod Triangles**

<table>
<thead>
<tr>
<th>Method Block</th>
<th>Trait</th>
<th>Validity Coefficient</th>
<th>Monomethod Triangle Coefficient 1</th>
<th>Monomethod Triangle Coefficient 2</th>
<th>Monomethod Triangle Coefficient 3</th>
<th>Monomethod Triangle Coefficient 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCT x PRI</td>
<td>Del</td>
<td>.50</td>
<td>.54</td>
<td>.48</td>
<td>.30</td>
<td>.57</td>
</tr>
<tr>
<td>OQ PCM PRI</td>
<td>.58</td>
<td>.54</td>
<td>.51</td>
<td>.30</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>MCT x Del</td>
<td>.46</td>
<td>.51</td>
<td>.54</td>
<td>.27</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>HFS PCM PRI</td>
<td>.60</td>
<td>.48</td>
<td>.51</td>
<td>.55</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>MCT x Del</td>
<td>.63</td>
<td>.51</td>
<td>.54</td>
<td>.61</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>LFS PCM PRI</td>
<td>.65</td>
<td>.48</td>
<td>.51</td>
<td>.55</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>OQ x Del</td>
<td>.53</td>
<td>.39</td>
<td>.30</td>
<td>.62</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>HFS PCM PRI</td>
<td>.41</td>
<td>.57</td>
<td>.39</td>
<td>.55</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>OQ x Del</td>
<td>.51</td>
<td>.76</td>
<td>.39</td>
<td>.39</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td>LFS PCM PRI</td>
<td>.23</td>
<td>.57</td>
<td>.39</td>
<td>.39</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>OQ x Del</td>
<td>.57</td>
<td>.62</td>
<td>.27</td>
<td>.61</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>HFS x Del</td>
<td>.57</td>
<td>.62</td>
<td>.27</td>
<td>.61</td>
<td>.39</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** MCT = Multiple-Choice Testing, OQ = Oral Questioning, HFS = High-Fidelity Simulation, LFS = Low-Fidelity Simulation. PRI = prioritization, Del = delegation, PCM = patient care management. Numbers denote method: 1 = MCT, 2 = OQ, 3 = HFS, 4 = LFS.

**Criteria four.** The final criterion is that the correlations between traits (whether in the mono method triangle or in the two off-diagonal triangles of the heteromethod
block should indicate similar patterns of correlation independent of method. Failure to meet this criterion suggests that correlations are affected by method bias. This criterion is difficult to assess, however, as it is unclear how to quantify patterns. Generally, one can look for the largest and smallest correlations within the monomethod triangles and the two heteromethod triangles on either side of the validity diagonal in the heteromethod blocks. Traits should show similar patterns of correlation across these triangles, meaning that the same traits should correlate highest or lowest across the triangles.

Take the four monomethod triangles first. It is apparent that the patterns are different across the four triangles, with prioritization and delegation sometimes having the highest and sometimes the lowest correlations. The other two pairings show similar lack of patterning. More consistency is seen in the lower heteromethod triangles, where the delegation-patient care management correlations show the highest correlation in four of the six heteromethod blocks and prioritization-patient care management correlations show the lowest correlations in four of the six blocks. However, the prioritization-delegation correlations sometimes show the largest, sometimes the middle, and sometimes the lowest correlations. Similar inconsistency is shown in the upper heteromethod triangles. Overall, the patterns are not very consistent, suggesting that this criterion is not met and that considerable method bias exists in the OSCE measurements.

**Research Question Two Results**

How do the twelve OSCE scores generated from the four methods of evaluating the three constructs correlate with scores of basic skills (TEAS®) and standardize nursing predictive scores of nursing skills (Kaplan® RN Predictive Exam)? The twelve scores were correlated with the scores from the TEAS® Reading, Math, Science, and English
subtests) and scores from three subtests of the Kaplan® RN Predictor Exam content areas: Setting Priorities, Nursing Judgment, and Management of Care. Table 17 presents these correlations.

The Kaplan® Predictor Exam is administered during the last semester of a nursing program just prior to graduation. Only forty-eight nursing students from this research sample were eligible to take the exam prior to data analysis. The remaining students from the research sample are eligible to take the exam in December 2013.

Table 17. Correlation Coefficients Between the Twelve OSCE Measures and Seven External Subtests

<table>
<thead>
<tr>
<th>Measures</th>
<th>TEAS® (N = 127)</th>
<th>Kaplan® (N =48)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reading Comp.</td>
<td>Math</td>
</tr>
<tr>
<td>MCT</td>
<td>PRI .42** .03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>DEL .50** .11</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>PCM .75** .17</td>
<td>.17</td>
</tr>
<tr>
<td>OQ</td>
<td>PRI .32** .08</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>DEL .41** -.01</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td>PCM .38** .15</td>
<td>.15</td>
</tr>
<tr>
<td>HFS</td>
<td>PRI .37** .11</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>DEL .51** .03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>PCM .57** .13</td>
<td>.13</td>
</tr>
<tr>
<td>LFS</td>
<td>PRI .44** .04</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>DEL .43** .03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>PCM .55** .15</td>
<td>.15</td>
</tr>
</tbody>
</table>

Note. MCT = multiple-choice test, OQ = oral questioning, HFS = high fidelity simulation, LFS = low fidelity simulation, PRI = prioritization, DEL = delegation, PCM = patient care management. **significant at p = 0.01.

None of the TEAS® subtest scores correlated with the twelve OSCE scores except Reading. Apparently reading ability is required to complete the twelve OSCE measures
correctly. For the Kaplan®, Setting Priorities correlated well with all twelve OSCE measures. It was thought that the Setting Priorities measure most closely resembled the measures of Prioritization and as such, it should correlate the highest with these OSCE Prioritization measures. This it does for only the MCT; for the other three methods, Setting Priorities actually correlates slightly less. The Kaplan® Nursing Judgment was thought to be closely aligned with Delegation, but none of the twelve correlations were statistically significant. Clearly, whatever is measured by Nursing Judgment is not part of the OSCE scores. Finally, Management of Care was thought to align with the OSCE Patient Care Management. Management of Care does show its highest correlations with Patient Care Management using the MCT and OQ, but not for HFS or LFS. In fact, Management of Care shows statistically significant correlations with the LFS and HFS measures of Delegation while the LFS and HFS correlations with Patient Care Management are not statistically significant.

These correlations with the two sets of external measures suggest that the OSCE measures all require reading ability, but not math, science, or English skills as measured by the TEAS®. The Kaplan® Setting Priorities subtest measures ability to synthesize patient information and determine the order in which care tasks need completion, and it is clear that all twelve of the OSCE measures require this set of skills. The Kaplan® Nursing Judgment measures specific responses for patient situations and presentation during assessment, but this set of skills is not required to do well on the OSCE measures. Finally, there is some evidence that the OSCE MCT and OQ measures of Patient Care Management measure skills similar to that of the Kaplan® Management of Care. But this is only true for the MCT and OQ measures. The LFS and HFS measures of Delegation
appear to measure more of what is measured by the Kaplan® Management of Care than the LFS and HFS measures of Patient Care Management.

**Research Question Three Results**

To what extent does the actual cost of designing and implementing an OSCE assessment environment for nursing leadership constructs which includes high-fidelity and low-fidelity simulation stations compare to the projected costs found in the literature? A comparison of budget estimates from the literature and estimates for the OSCE used to collect data for this study is presented in Table 18. The initial budget for the OSCE was $24,060 and was based upon 106 pilot participants plus 100 study participants. The initial cost per student was $116.79 per student. The actual cost to run the OSCE was $28,022.04 or $140.81 per student (based on $= 199). Total cost difference between the high-end estimation from Poenaru et al. (1997) was $76,377.96. The comparative low-end estimation by Palese et al. (2012) was $59,460, which remained significantly more expensive than the original estimated cost ($24,060) as well as the actual OSCE cost ($28,022.04) for this study.
Table 18.
Comparison of Actual OSCE Costs with Estimates from the Literature

<table>
<thead>
<tr>
<th>Category</th>
<th>Highest Estimate*</th>
<th>Lowest Estimate**</th>
<th>Budgeted Estimate</th>
<th>Actual Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Staff – Implementation Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Leader</td>
<td>10,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standardized Patient Trainer</td>
<td>20,000</td>
<td>20,000</td>
<td>1,500</td>
<td>1,700</td>
</tr>
<tr>
<td>Administrative Support</td>
<td>15,000</td>
<td>15,000</td>
<td>0</td>
<td>6,000</td>
</tr>
<tr>
<td>Simulation Cost</td>
<td>0</td>
<td>0</td>
<td>4,900</td>
<td>3,500</td>
</tr>
<tr>
<td>Total</td>
<td>45,000</td>
<td>35,000</td>
<td>6,400</td>
<td>11,200</td>
</tr>
<tr>
<td>OSCE Development Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review of Course Objectives; scenario creation</td>
<td>600</td>
<td>0</td>
<td>4,995</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty honorarium for workshops</td>
<td>12,000</td>
<td>0</td>
<td>250</td>
<td>750</td>
</tr>
<tr>
<td>Catering for case development workshops</td>
<td>360</td>
<td>360</td>
<td>0</td>
<td>150.56</td>
</tr>
<tr>
<td>Total</td>
<td>14,200</td>
<td>360</td>
<td>5,245</td>
<td>4,400.56</td>
</tr>
<tr>
<td>OSCE Production Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient training</td>
<td>2,000</td>
<td>2,000</td>
<td>510</td>
<td>300</td>
</tr>
<tr>
<td>Printing of OSCE material</td>
<td>3,000</td>
<td>500</td>
<td>60</td>
<td>50.35</td>
</tr>
<tr>
<td>Artifacts, props, and supplies</td>
<td>1,000</td>
<td>800</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>6,000</td>
<td>3,300</td>
<td>640</td>
<td>450.35</td>
</tr>
<tr>
<td>OSCE Administration Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized Patients</td>
<td>13,500</td>
<td>13,500</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty Examiners</td>
<td>13,500</td>
<td>0</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Faculty Markers</td>
<td>1,800</td>
<td>0</td>
<td>4,000</td>
<td>3,500</td>
</tr>
<tr>
<td>Support Staff</td>
<td>2,700</td>
<td>2,700</td>
<td>500</td>
<td>600</td>
</tr>
<tr>
<td>Catering</td>
<td>3,000</td>
<td>3,000</td>
<td>400</td>
<td>752.13</td>
</tr>
<tr>
<td>Total</td>
<td>34,500</td>
<td>19,200</td>
<td>12,900</td>
<td>11,852.13</td>
</tr>
<tr>
<td>Post-OSCE Analysis and Reporting Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Entry and Review</td>
<td>1,600</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>1,600</td>
<td>1,600</td>
<td>280</td>
<td>119(^{#})</td>
</tr>
<tr>
<td>Report Generation</td>
<td>1,500</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4,700</td>
<td>59,460</td>
<td>23,780</td>
<td>0</td>
</tr>
<tr>
<td>Total for Entire Implementation</td>
<td>104,400(^{2})</td>
<td>59,460(^{2})</td>
<td>24,060(^{*})</td>
<td>28,022.04</td>
</tr>
<tr>
<td>Total Students</td>
<td>120</td>
<td>120</td>
<td>206</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>64(pilot) + 142</td>
<td>62 (pilot) + 137</td>
<td>(study) = 206</td>
<td>(study) = 199</td>
</tr>
<tr>
<td>Total per Student</td>
<td>870</td>
<td>496</td>
<td>116.79</td>
<td>140.81</td>
</tr>
</tbody>
</table>

The most expensive budgeted items occurred during the Implementation Phase and the Administration Phase in the Palese et al. (2012) and Poenaru et al. (1997) estimates. Each of the largest budget items from the estimates were associated with standardize patients (SP). The SP Trainer was allocated $20,000 by Poenaru et al. (1997) and by Palese et al. (2012). This study budgeted $1,500 and spent $1,700, which was $18,300 less than the high- and low-end estimates. The largest expense in this study accrued as a result of the unforeseen administrative support ($6,000) that was needed to run the OSCE and manage the data collected. Table 18 presents the actual cost of administering the OSCE assessment compared to the cost estimates reported in the literature.

**Summary**

The analysis of the data answered the research questions. A summary of the results is present in this section.

In summary, four findings were evidenced. First, the findings from data analysis indicated that the twelve scores generated from the OSCE provided evidence to support internal consistency through good reliability estimates. This was evidenced by the reliability diagonals of the constructed matrix presented in figure 5.

Second, *convergent validity* was demonstrated through findings associated with analyzing the data generated by the twelve OSCE scores using C&F Criteria. However, findings regarding *discriminant validity* indicated that the four methods used to generate the twelve OSCE scores in this study did not conform uniformly to the C&F Criteria.

Third, significant positive correlations evidenced between content items scores on the *TEAS® Exam* and the twelve OSCE scores. However, several negative relationships
were observed between *Science* content item scores of the *TEAS® Exam* and all trait scores measured by all methods. Further, *Nursing Judgment* content item scores from the *Kaplan® RN Predictor Exam* also evidenced negative relationships when correlated with several trait item scores measured by three of the four methods used in the study indicating that deeper analysis of these findings is warranted.

Lastly, administering the OSCE assessment over the three data collection dates was $3,962.40 over the budgeted amount. However, the final OSCE cost remained significantly less expensive than the high- and low-end estimates for similar OSCE assessments found in the literature. These findings indicated that the costs associated with OSCE assessment are more expensive than traditional methods of assessment in nursing education; however, if the use of HFS was found to be the best method to assess nursing student achievement surrounding leadership constructs as it appeared based upon the data collected from this study, such assessment methodology costs should be factored into the educational budget of nursing programs. Further discussion of these finding are presented in Chapter Five.
CHAPTER FIVE

SUMMARY, LIMITATIONS, DISCUSSION, AND IMPLICATIONS

This chapter represents the findings, limitations, discussion, and implications of the study. The chapter is divided into six subsections. First, a summary of the study is presented. Second, findings from the study are summarized. Third, trends that were identified from the study are presented. Fourth, limitations of the study are identified. Fifth, a discussion of study’s limitations is provided. Lastly, the implications of the study’s findings and the contribution these findings make to the state of the science on this subject matter are presented.

Summary of Study

The purpose of this study was to determine the construct validity of and identify which of four different evaluation methods of multiple-choice testing (MCT), oral questioning (OQ), high fidelity simulation (HFS), and low fidelity simulation (LFS) could assess the three nursing leadership traits of prioritization, delegation, and patient care management. The evaluation methods were embedded in an OSCE assessment environment.

Construct validity criteria established by Campbell and Fiske (1959) were applied to twelve OSCE scores generated during the four-pronged assessment of the three traits. The three traits were measured by the commonly used methods of nursing student evaluation (MCT, and OQ) and by two burgeoning forms of evaluation: HFS and LFS. Lastly, the twelve OSCE scores were correlated with scores from two other commonly used achievement examinations (TEAS® Exam and Kaplan® RN Predictor Exam) to determine if there was a correlation between the reliability and validity of the traits.
measured in this study by four different methodologies to the reliability and validity of these nationally standardized exams.

**Significance of the Study**

The significance of this study was noteworthy for three reasons. First, Benner et al. (2010) postulated that there was a lack of rigorous nursing research related to forms of assessing nursing knowledge. Research conducted by Luctkar-Flude, Wilson-Keates, and Larocque (2012) echoed statements made by Benner et al. (2010) and called for ways to change and improve the education and evaluation of nursing students. Therefore, the aim of this study was to fill the void in nursing literature related to the lack of nursing research to evaluate the methods used to assess nursing knowledge so as to propose different and improved ways to educate and evaluate nursing students.

The second reason for this study was to answer the call by nurse educators to rigorously evaluate new methods of student assessment. Research conducted by Walsh, Jairath, Paterson, and Grandjean (2010), Morris and Hancock (2008), and Prion, (2008) concluded that it was essential to provide evidence of construct validity when new assessment methodologies such as HFS and LFS are introduced. This was supported by Kardong-Edgren, Hanberg, Keenan, Ackerman, and Chambers (2010), who argued that the push to use HFS and LFS as assessment methodologies in nursing education required rigorous research on these new assessment methods to determine if they were any better than the traditional methodologies of MCT and OQ. The lack of adequate assessment of the cognitive and clinical skills of nursing students was highlighted in the literature discussing the lack of competencies in leadership skills such as prioritization, delegation, and patient care management being exhibited by new nursing graduates. Campbell and
Fiske’s (1959) MTMM approach to assess construct validity of three nursing leadership traits (prioritization, delegation, and patient care management) measured by four measurement methods (MCT, OQ, HFS, and LFS) was used in this study.

The third reason for this study was to identify the cost of conducting simulation evaluation and to determine if such costs were warranted, given the reliability and validity of the methodology. Barman (2005) and Major (2005) contended that cost was the single most important reason why OSCE assessments were not used in United States (U.S.) nursing educational programs. Further, traditional assessment methods used in the U.S. are less expensive when considering associated labor (human) and material costs. Studies such as those conducted by Barman, (2005), Palese et al. (2012), and Poenaru et al. (1997) also concluded that cost was one of the main criticisms against OSCE assessments in medical education when compared to traditional forms of assessment such as in vivo oral exams. Likewise, there was no agreement on the typical cost associated with OSCE assessments in the literature, which estimates per student costs to be between $496 and $870. This study aimed to determine the cost associated with an OSCE assessment in nursing education compared to costs identified in the literature, as well as to examine whether there is a benefit to assume this financial burden, given that the benefits of assessment by either HFS or LFS methods are equally reliable and valid methods of measuring constructs.

**Theoretical Framework**

The foundation to determine construct validity of multiple assessment measures measuring different traits was grounded in Campbell and Fiske’s (1959) MTMM. This study aimed to provide evidence as to the construct validity of the four methods of
evaluating nursing students. While MTMM is not a unique approach to establishing construct validity, it is a new approach in evaluating nursing assessment modalities and has great value in assessing student assessment methods. In this correlational study, the MTMM approach was used to determine the degree to which the four assessment modalities could adequately evaluate the complex constructs of prioritization, delegation, and patient care management.

The four conditions of the Campbell and Fiske (1959) criteria provide sufficient evidence for reliability and validity that an instrument or a method is or is not measuring what it intends. The four conditions are: 1) the correlations across the validity diagonal should be different than zero and of sufficient magnitude to warrant further investigation; 2) the specific construct correlations should be higher along the validity diagonal when the same trait-different methods are correlated and when the trait under investigation is correlated with a different trait using a different method; 3) even when methods of measurement are identical, the correlations between the different methods of the same trait should be higher than the correlations between the trait under investigation and another trait; and 4) the correlations between traits (whether in the same method or different method blocks) should indicate patterns of interrelationships between traits and should be similar in the different trait triangles.

**Research Questions and Methodology**

The research questions that guided the study were:

**Research Question One.** To what extent do the twelve OSCE scores generated from using four assessment methods (multiple-choice tests, oral questioning, low-fidelity simulation, and high-fidelity simulation) to measure three constructs (delegation,
prioritization, and patient care management) conform to the Campbell and Fiske (1959) criteria for construct validity?

**Research Question Two.** How do the twelve OSCE scores generated from the four methods of evaluating the three constructs correlate with scores of basic skills (TEAS®) and standardize nursing predictive scores of nursing skills (Kaplan® RN Predictive Exam)?

**Research Question Three.** To what extent does the actual cost of designing and implementing an OSCE assessment environment for nursing leadership constructs which includes high-fidelity and low-fidelity simulation stations compare to the projected costs found in the literature?

This was a correlational descriptive design study that used the MTMM approach to analyze the three constructs (prioritization, delegation, and patient care management) across the four methods of assessment (multiple-choice tests, oral questioning, HFS, and LFS). The conceptual foundation of the MTMM approach was developed by Campbell and Fiske (1959). The study aimed to compare traditional assessment methods of MCT and OQ in nursing education to burgeoning assessment methods of HFS and LFS simulation. Further evaluation of reliability of these four modalities was to compare the results of standardized instruments that measured cognitive abilities needed for leadership and the other that measured dimensions of leadership to the twelve generated OSCE scores.

For this study, MCT, OQ, HFS, and LFS assessment methods were used to evaluate participant knowledge associated with three key nursing leadership constructs. The participants in this study generated twelve OSCE scores from stations designed to
measure knowledge and skill in prioritization, delegation, and patient care management through one of four methods of assessment. These methods were a 21-item MCT exam, a 21-item oral question instrument, a patient care HFS, and three LFS stations. Selim et al. (2012) and Kaplan and Ura (2010) provided guidance for embedding the four measurement methods into one 120-minute OSCE assessment in which each student participated.

**Findings**

There were five major findings in this study. Each of these findings is outlined below.

1. Reliabilities of the twelve OSCE scores are relatively consistent across the twelve measures. These measures ranged from a low of .70 to a high of .85. This is important because two measures can only correlate to the extent they are reliable measures. Therefore, if the reliabilities are relatively consistent, attenuation occurs fairly evenly and differences in correlations cannot be attributed to uneven reliability estimates.

2. Construct validity was strictly met per Campbell & Fisk criteria. However, the off diagonal correlations were nearly as high as the validity diagonal correlations. This suggests evidence for relatively weak construct validity.

3. There is not a great deal of support for discriminate validity among the constructs as there are many off validity diagonal correlations that are greater than the validity diagonal correlations within the MTMM. Method pair correlations by trait demonstrated inconsistent results across method pairs. It does appear that the method pair of MCT with any other method worked well.
However, there is not evidence for discriminant validity as most correlations are uniformly moderately large, and share considerable method variance.

4. Correlations with the two sets of external measures suggest that the OSCE measures all require reading ability, but not math, science or English skills as measured by the TEAS. There is some evidence that the OSCE measured skills similar to the Kaplan RN Predictor Exam subsets.

5. The actual cost to run the OSCE was significantly lower than estimates from the literature. The average cost was $140.81 per student (based on n=199).

**Trends**

There were three trends found in this study that should be considered by nurse educators. First, the use of HFS and LFS as valid assessment tools is now supported by research. Since these are expensive modalities, such evaluation may need to be saved as a summative assessment process with an integration of the other modalities used during the formative assessment phases. The second trend relates to assessing nursing students’ college achievements and their cognitive abilities (grades and GPA) in reading comprehension, writing and communicating in English. A solid foundation in the math and sciences is important to determining whether such students have the ability to assume the scope of practice responsibilities of a registered nurse. Lastly, assessing the students’ achievements on their *Kaplan® RN Predictor* exams can be key indicators of their understanding of the skills needed to operationalize nursing leadership.

**Limitations**

Despite the research findings, there are limitations to the study that need to be considered when determining whether these findings should be used as a foundation to
make changes in nursing education and assessment of students. There were six
limitations identified in the methodological specification, analysis, and interpretation of
this study. The first limitation was an overarching limitation of this study, or any study
using MTMM matrix methods. Despite Campbell and Fiske’s (1959) standardized
methodology, there is no standard for “good” results. Similarly, another overarching
limitation is the lack of precision of MTMM matrix methods. There are no statistical tests
to determine whether or not a given outcome meets or does not meet the four criteria set
out by Campbell and Fiske (1959). Instead, general guidelines are set forth and judgment
is used to determine whether the results meet the criteria or not. As a specific example,
there are no proportions of trait and method variance to determine observed results

The second limitation was the logistical constraint of time as a factor. High-
fidelity simulation experts strongly recommend that students have a debriefing session to
discuss how they did and what areas could be improved. Debriefing following the HFS
was preferable and that practice remains the accepted standard for instructional pedagogy
(Bastable, 2008; Jefferies, 2005). The time constraints in this study were impacted by the
need for students to complete the other method stations in a reasonable amount of time.
For this reason, there was no post-hoc participant debriefing by faculty evaluators. The
inclusion of faculty evaluators conducting debriefing sessions would have increased the
depth of the study as well as the cost; however, the time constraints of the study did not
warrant such a debriefing. The participants in this study were evaluated without feedback.

The third limitation of the study was the sampling size and methodology. The
convenience sampling method was used. The participants selected for inclusion of the
study attended a private, four-year university with a nursing curriculum. While this nursing curriculum was similar to that of other programs at other private four-year universities, the diversity of the student body population may have affected the validation procedure of the evaluation methods. Aside from demographic and cultural differences across universities, the possibility that students in associate degree programs or students attending other four-year universities with a different curriculum might respond differently to the OSCE conducted in this study exists. Ideally, the preferred sampling method would have been random sampling drawn from a representative population of nursing students across a variety of universities offering nursing education programs (Popham, 2000). Since this method was not feasible, nor available for this study, the convenience sampling method was used in its place. The sample size was perhaps too small. A total $n = 137$ were evaluated. A power analysis was not appropriate to use in determining a sample size that would support study outcomes as statistically significant; however, if one were to use the $4 \times 4$ chi-square table proposed by Krejcie and Morgan (1970), it may be concluded that a sample size of 137 in a nursing program with 450 students may be adequate to assume statistical significance. Unfortunately, the sampling process and the singular institution from which the sample was recruited did not make generalizability of the findings possible. Despite this limitation, the outcome of the study has merit for consideration in changing some of the evaluation processes in nursing education.

The fourth limitation to the study was that possible data contamination might have existed, since the RAs associated with this research were also student members of the cohort of senior students who participated in this study. While the RAs were not research
participants in the OSCE, there was a risk of data contamination if any RA gave
dvanced notification about the study to the student participants. To mitigate this risk, the
RAs were asked to sign a confidentiality agreement. Further, the researcher for this study
was a faculty member, which could have influenced how students responded to any of the
OSCE stations. To mitigate for this potential, the researcher did not participate in the
OSCE activities.

The fifth limitation to this study was the instrumentation of the study as a possible
limiting factor. While a consensus panel of experts reviewed the rubrics, and while edits
were made by the researcher based on the feedback from the panel, there was no formal
validity and reliability testing conducted on the rubrics following the pilot test completed
several weeks before the actual study. The MCT contained items that were used in
previous exams conducted by the researcher. An item analysis was conducted and a
consensus of experts from within the nursing department was recruited to determine the
face validity of each item related to the constructs with which they were associated. The
editing and elimination of items was conducted based on panel feedback. Although each
of the measures was designed to evaluate higher-order domains of nursing leadership
knowledge, this was a factor that influenced internal consistency reliability (Popham,
2000). Since the reliability of the instruments for this study was not known \textit{a priori}, the
findings of the study may be limited.

Lastly, consideration may need to be given to instrumentation as a limiter.
Simulation case studies, though valid procedures established by the Bay Area Simulation
Consortium (BASC) Scenario Development Committee protocol, were not validated \textit{a
priori}, even though the case studies were reviewed and approved by the researcher and a
panel of nursing faculty colleagues as appropriate to assess the leadership constructs in the study. This limitation was not a crucial one, since from an *ex-post perspective*, there was no threat to the validity of the process.

**Discussion**

This section provides a discussion of the findings with respect to the construct validity and reliability of the four measurements.

**Construct Validity of the Assessment Modalities for the Three Constructs**

The three constructs being assessed by the four assessment modalities were the leadership skills of prioritization, delegation, and patient care management. In terms of prioritization, the studies that were previously conducted were assessing achievement of this dimension in theory courses or in practice settings. Patrick et al. (2011) conducted a thorough review of the literature regarding prioritization and the authors’ findings demonstrated a need to show construct validity for assessing prioritization in nursing leadership education. While results of the research study by this author did not demonstrate the convergent validity expected for prioritization, the range of reliability magnitudes for the measurement methods were of sufficient magnitude to warrant further validity investigation; therefore, this investigation did address the literature gap about the need for quantitative research in measuring leadership construct.

The literature around the construct of delegation concentrated on the definition of delegation and the relationship that effective delegation has with such factors as earning a baccalaureate and with the ability to ensure task completion and safety oversight while adhering to the registered nurse scope of practice (Billay & Myrick 2008; Lillibridge, 2007; Patrick, Laschinger, Wong, & Finegan, 2011; Saccomano & Pinto-Zipp, 2011;).
Research on the relationship between leadership styles, demographic variables, and confidence in delegation was conducted by Saccomano and Pinto-Zipp (2011). The authors found that a relationship existed between leadership styles and confidence in delegation. In addition, the authors found that there existed a relationship between the number of years worked in a hospital and confidence in delegation (Saccomano & Pinto-Zipp, 2011).

The findings in the research for delegation provided a basis for the importance of including delegation skills in an effective nursing program. This study, however, examined the construct validity with respect to skill measurement across four evaluation techniques. The findings of the study indicated that there was not a high degree of construct validity. Only the first data collection group had low values within this construct. This finding indicated that, overall, there appeared to be a sufficiently weak degree of validity of the construct across the evaluation methodology.

For patient care, the literature regarding the importance of this construct relating to nursing leadership is vast. Despite the large number of studies related to patient care, Biag et al. (2010) underscored the importance of defining and investigating the validity of assessment methods for measuring patient care in undergraduate students. Biag et al. (2010) also indicated that there is a gap in the literature related to the use of MTMM matrix methods associated with the evaluation of patient care management traits. The authors found that the approach used in their study could estimate the degree of evidence for validating complex constructs in nursing education such as patient care, along with prioritization and delegation. Overall, Biag et al. (2010) noted that the effectiveness of multiple-choice tests and oral examinations in evaluating clinical competence had
significant practice implications and should be investigated. This study provided that investigation using the MTMM matrix method and found that based upon C&F Criteria, correlation coefficients located in the different trait — different methods triangles did not conform to criteria establishing discriminant validity. Patterns across the matrix for these off-diagonal correlation coefficients were not consistently lower than those correlation coefficients located in the validity diagonals. Kenny and Kashy (1992) suggested that these off-diagonal correlation coefficients were among the lowest within the matrix as an indicator that similar items are evidencing discrimination across methods. This was not the case for the MTMM of this study. Some off-diagonal coefficients evidenced stronger correlations than coefficients located within the same trait - different method triangles.

Only 38% of the off-diagonal correlations were lower than the coefficients located in the validity diagonal. High off-diagonal coefficients typically indicate poor reliability or are the result of sample size. Reliability across all measurement methods for the three traits being investigated in this study was good, averaging > .80 for each method.

Finally, there has been a general need for quantitative research regarding assessing the effectiveness of HFS and LFS as evaluation methods in nursing education (Schultz, Shinnick, & Judson, 2012). These authors discussed innovative simulation exercises; however, they did not provide quantitative findings of construct validity across evaluation metrics and selected samples. As a result, the research study completed by this author found that using HFS and LFS as evaluation methods could measure nursing leadership traits with reliability (HFS methods averaging .81; and LFS methods averaging .77) and evidenced higher discrimination patterns among the four assessment
methods used in this study; 85% of the off-diagonal correlations lower than the validity coefficients were associated with HFS or LFS measurement methods. HFS and LFS items were just as effective at measuring the three leadership traits as the traditional measure methods in this study. These findings added to the literature on the construct validity of prioritization, delegation, and patient care management, using simulation modalities as measurement methods.

**Reliability of the Assessment Measures**

This section compares the reliability findings about the four assessment modalities (MCT, OQ, HFS and LFS) found in the literature with the findings from this study. Clifton and Schriner (2010) provided evidence for the validity of MCT usage alone; however, the authors did not discuss the importance of finding multiple means of addressing the same construct in an evaluative framework using other studies about MCT reliability. All reliability scores were .70 (the standard) or above (Campbell & Fiske, 1959; Marsh & Bailey, 1991; Popham, 2000). The overall range of reliability in these studies was .70 to .85, with no measurements figure below .70.

The reliability of the MCT in this author’s study was .78 and supporting the validity of the MCT measurement; however, its value was more evident when assessing patient care management items. It was not the most reliable measure across all three leadership constructs. The next measurement modality was the oral examination (questioning). Daelmans et al. (2001) examined the generalizability coefficients to determine the interval for the number of oral exams that were required for students to be successful on global judgment and pass the oral exam. Their findings supported the benefit of oral examination. Research studies assessing the benefit of using oral
examinations in nursing education were rare. While oral examinations are well established dating back to Socrates, the use of Socratic Questioning, a component of an oral exam, is a method of evaluating nursing student knowledge in the clinical setting. The results of this author’s study found that the oral exam had a high reliability, especially when measuring the constructs of delegation (.82) and patient care management (.85). When comparing the reliability of OQ to the other assessment modalities used in this study, it ranked as one of the most effective assessment measures (tied with HFS .81).

The final two measures investigated in this study were HFS and LFS. In the beginning of its use in health education, simulation was used at the medical resident and practitioner levels to engender psychomotor skill acquisition through repeated practicing in a setting that was safe from patient and student harm. Simulation (HFS and LFS) has since evolved through varying forms into an innovative pedagogy, which has been incorporated into present day hospital staff training and schools of nursing curricula (Murphy et al., 2010; Swenty & Eggleston, 2010; Waxman & Telles, 2010). The goal of an effective simulation is to match reality to the largest degree possible (Jefferies, 2005). This fidelity to the experience results in simulations being ranked by how high of a representation they provide with respect to the real experience. The LFS metric, when associated with skill acquisition, is a standard assessment method used to measure clinical competence that focuses on outcomes based upon observable behaviors of learners (Carraccio & Englander, 2000).

In this study, HFS and LFS evaluation methods were used and the results indicated a similar sufficiently high degree of reliability across methods (HFS - .80
(prioritization), .81 (delegation), and .82 (patient care management); LFS - .85 (prioritization), .78 (delegation), .70 (patient care management). As a result, the reliability of the HFS and LFS evaluation methods was found to be sufficient, based on the results of the study where the MTMM matrix method of construct validity was used. When compared to the reliability coefficients of MQT and OQ across all three leadership constructs, it was found as the most reliable measure method across all dimensions (tied with OQ .81).

Given these reliability findings and considering the cost of the various assessment modalities, nurse educators can consider where to use these modalities within their programs to conduct formative and summative evaluations. Based on what the nursing program learning outcomes are can lend itself to determining the best modality for assessing all skills and knowledge required of nurse graduates.

Implications

Future Research

It is important that nurse educators continue to conduct formal research into the area of assessing program assessment modalities and disseminate those findings in the nursing literature. This study was limited only by cost and time. The sample size of about 50 subjects per group was sufficient for the statistical analysis, but generalization could not be made; therefore, to achieve statistical significance and generalization of the findings replication of this study with participants across a variety of settings offering nursing education would be important.

A longitudinal study using these modalities as a summative process that assesses the graduating senior’s achievement of leadership traits would be important. Such a study
could also be conducted across one or more other nursing education programs. Further post-hoc analysis (univariate confirmatory factor analysis) could be conducted to determine the number of factors that load from the dataset. Another study could be conducted to revise the instrument further or delete unacceptable items that cross-loaded (< .25 difference). This instrument alteration could also include a revision and deletion of poorly performing rubric items within the three subscales of each method rubric based on confirmatory factor analysis results.

Before data was collected for this study, the outcome was expected to be consistent with the existing literature where the different trait-same method triangles would be lower than the validity diagonal coefficients, and there would be clear patterns of differentiation among the off diagonal coefficients within the MTMM. However, once the coefficients were arranged in the MTMM and evaluated using Campbell and Fisk criteria congruent validity was weakly demonstrated via criteria one and two. Because there are no established interpretive guidelines for the MTMM, interpretation is more subjective on the part of the researcher. Convergent validity was met, though by minimal statistical means. Discriminant validity was not demonstrated, which was a surprising finding, and indicative of some strong method variance that wasn’t an expected result, and there was no plan to mitigate the method variance. Further research will need to account for method variance among these four specific methods of student assessment.

When the coefficients from the twelve OSCE scores were compared against the external exams, the TEAS comparison showed a high degree of statistical significance across all four methods used in the OSCE with the reading subtest. This was likely due to the amount of reading required at each of the OSCE stations, no matter the method of
evaluation. Because the reading section of the TEAS deals with comprehension, this is intuitively consistent. The other three subsets of the TEAS were not required for participants to perform specific tasks within the OSCE stations. The surprising finding was the Kaplan comparison. Setting priorities was statistically significant across all four methods used in the OSCE. The range was from .34 in the LFS method to as high as .93 in the MCT method. Clearly what the Kaplan measures in the subtest of setting priorities was present throughout all four methods of the OSCE and with the traits that were being evaluated. Conversely, what was originally thought to be measured by the Kaplan subtests of nursing judgment and management of care were not components of the OSCE base on the correlation data. Future research would require the use of other subtests within the Kaplan for correlational analysis of these three traits.

**Practice**

This study provided a number of practical implications. First, no significant differences were observed among the different evaluation methods. All four assessment modalities are reliable; however, there may be more validity in using simulation to assess the application of skills over using MCT or OQ. In clinical practice, successful application of knowledge is critical. Controlling for the cost of conducting simulation assessment is important while still being able to conduct an appropriate evaluator process and this research study has demonstrated such cost containment.

In summary, the evidence of this research supports the reliability of using all four of the metrics used in the evaluation of the nursing leadership constructs presented. Though not all four of the conditions for construct validity, as defined by Campbell and Fiske’s (1959) MTMM methodology, were achieved, this research adds to the body of
knowledge regarding the use of the MMTM approach to evaluation methods of nursing student assessment. The study validates through research the value of the various assessment methods and thus is usable for nursing faculty in their pursuit of making evidence-based changes in education and practice.
REFERENCES


Definition of terms

Delegation: the transferring of the authority to perform selected nursing tasks in the selected situation to a competent individual (Motaki & Burke, 2010).

Prioritization: the act of deciding what care should be done first and what should follow sequentially. Establishing an ordered list or arrange items based on importance or urgency. The method used to determine what actions need to be accomplished ahead of others; Also known as priority setting (Cherry & Jacob, 2011).

Patient Care Management: the act of supervising the care of one or more patients to ensure patient safety (Cherry & Jacob, 2011).

Directions: (questions 1 – 10)
Read the narrative below and then identify the choice that best completes the statement or answers the questions.
You are the leader of the team providing care for six patients. Your team include yourself (an RN), an LPN, and a newly hired nursing assistant, who is undergoing orientation to the unit. The patients are as follows:

• Mr. Duncan, a 68-year-old with unstable angina who needs teaching for cardiac catheterization scheduled this morning.
• Ms. Johnson, a 45-year-old experiencing chest pain scheduled for a graded exercise test later today.
• Mr. Richardson, a 75-year-old who had a left-hemisphere stroke four days ago.
• Ms. Sampson, an 83-year-old with heart disease, a history of myocardial infarction, and mild dementia.
• Ms. Baker, a 93-year-old newly admitted from a long-term care facility, with decreased urine output, altered level of consciousness, and in elevated temperature of 99.5°F (37.5°C).
• Mr. Lincoln, a 59-year-old with mild shortness of breath and chronic emphysema.

1. Four of the six patients in your team should be assigned to the LVN to perform nursing care tasks, under your supervision. Which of the six patients should you assign the LVN? (Select all that apply).
   a. Mr. Duncan
d. Ms. Sampson
   b. Ms. Johnson
e. Mr. Lincoln
c. Mr. Richardsonf. Ms. Baker

2. Which key point would you be sure to include when teaching the patient about the post-procedure care after cardiac catheterization?
   a. “There are no restrictions after the procedure.”
   b. “We will get you out of bed within two hours after the procedure.”
   c. “You will have to stay flat in bed for 4 - 6 hours after the procedure.”
   d. “Family visitors will be restricted until the next day.”

3. Which patient should you assess first?
   a. Mr. Duncan
c. Ms. Baker
   b. Ms. Johnson
d. Mr. Lincoln

4. Which of the tasks listed below should you delegate to the nursing assistant?
   a. Asking Mrs. Sampson memory-testing questions.
   b. Telling Ms. Johnson about treadmill exercise testing.
c. Performing pulse oximetry for Mr. Lincoln.
5. The nursing assistant asks you why it is important to notify someone whenever a patient with heart disease complains of chest pain. What is your best response?
   a. “It is important to keep track of the chest pain episodes so that we can notify the physician.”
   b. “The patient may need morphine to alleviate the chest pain.”
   c. “Chest pain may indicate coronary artery blockage and heart muscle damage.”
   d. “Our unit policy includes specific steps to take in the treatment of patients with chest pain.”

6. The physician’s orders for Mr. Richardson (the patient with the stroke) include assisting the patient with meals. Which of your team members should be assigned to this task?
   a. physical therapist
   b. nursing assistant
   c. LVN
   d. occupational therapist

7. The nursing assistant tells you that Mr. Lincoln, the patient with chronic emphysema, says he is feeling short of breath after walking to the bathroom. What action should you take first?
   a. notify the physician
   b. increase oxygen flow to 4 L per minute via nasal cannula
   c. assess oxygen saturation by pulse oximetry
   d. remind the patient to cough and deep breathe

8. The oral temperature of Ms. Baker, the newly admitted patient from the long-term care facility with decreased urine output and altered level of consciousness, is now 102.6°F (39.2°C). What is your best action?
   a. Notify the physician.
   b. Administer acetaminophen (Tylenol) two tablets orally.
   c. Ask the LVN to give an acetaminophen (Tylenol) suppository.
   d. Remove the extra blankets from the patient’s bed.

9. Which factor most likely precipitated Ms. Baker’s elevated temperature?
   a. bladder infection
   b. increased metabolic rate
   c. kidney failure
   d. nosocomial pneumonia

10. The LVN reports that Ms. Sampson will not leave the chest leads for her cardiac monitor in place and asks if the patient can be restrained. What is your best response?
    a. “Yes, this patient had a heart attack and we must keep her on the cardiac monitor.”
    b. “Yes, but be sure to use soft restraints so that the patient’s circulation is not compromised.”
    c. “No, we must have a physician’s order before we can apply restraints in any situation.”
    d. “No, but try covering the lead wires with a sheet so that the patient doesn’t see them.”

Directions: (questions 11 – 17)
Read the narrative below and then identify the choice that best completes the statement or answers the questions.

Ms. Jefferson is a 63-year-old woman who is admitted directly to the medical unit after visiting her physician because of shortness of breath and increased swelling in her ankles and calves. She is being admitted to rule out a diagnosis of chronic kidney disease (CKD). Ms. Jefferson states that her symptoms have become worse over the past two to three months and that she uses the bathroom less often and urinates in smaller amounts. Her medical history includes hypertension (30 years), coronary artery disease (18 years), and type II diabetes (14 years).

Ms. Jefferson's vital signs on admission were the following:
- Blood pressure: 162/96 mmHg
- Heart rate: 88 bpm
- Oxygen saturation: 91% on room air
- Temperature: 97.8°F (36.6°C)

Admission laboratory tests for which patient samples are to be collected on the unit include serum electrolyte levels, renal function tests, complete blood count, and urinalysis. A 24-hour urine collection for determination of creatinine clearance has also been ordered.
11. During your admission assessment, Ms. Jefferson has all of these findings. For which findings should you notify the physician immediately?

a. bilateral pitting ankle and calf edema rated 2+
b. crackles in both lower and middle lobes of the lungs
c. dry and peeling skin on both feet
d. faint but palpable pedal and post tibial pulses

12. Which task associated with the 24-hour urine collection is appropriate to delegate to the nursing assistant?

a. Instructing Ms. Jefferson to collect all urine with each voiding
b. Teaching Ms. Jefferson the purpose of collecting urine for 24 hours
c. Ensuring that all urine obtained for the test is kept on ice
d. Assessing Ms. Jefferson’s urine for color, odor, and sediment

13. You review Mrs. Jefferson’s laboratory results and note the following values:

- Serum potassium level of 7.1 mmol/L
- Serum creatinine level of 15 mg/dL
- Blood urea nitrogen level of 180 mg/dL
- Serum calcium level of 7.8 mg/dL

Which medication should you be prepared to administer to lower the patient’s potassium level?

a. furosemide (Lasix) 40 mg intravenous push
b. Epoetin alfa (Epogen), 300 units per kilogram subcutaneously
c. calcium one tablets by mouth
d. sodium polystyrene sulfonate (Kayexalate), 15 g by mouth

14. You are the team leader, supervising an LVN. Which nursing care action for Mrs. Jefferson should you delegate to the LVN?

a. insert a catheter intermittently and note the amount of residual urine
b. plan restricted fluid amounts to be given with meals
c. assess breath sounds and note increased presence of crackles
d. discuss renal replacement therapies with the patient

15. As team leader, you observe the nursing assistant performing all of these actions for Mrs. Jefferson. Which of the following action performed by the nursing assistant must you intervene?

a. assisting Mrs. Jefferson to replace the oxygen nasal cannula
b. measuring Mrs. Jefferson’s vital signs after the patient drinks fluids
c. ambulating with Mrs. Jefferson to the bathroom and back
d. washing Mrs. Jefferson back, legs, and teach with warm water

16. You are supervising a new nurse on orientation to the unit who is providing care for Ms. Jefferson after her return from surgery to create a left forearm access for dialysis. Which action by the nurse requires that you intervene?

a. The nurse monitors the patient’s operative site dressing for evidence of bleeding.
b. The nurse obtained a blood pressure reading by placing the cuff on the right arm.
c. The nurse draws blood for laboratory studies from the temporary dialysis line.
d. The nurse administers oxycodone (Roxicodone) by mouth for moderate postoperative pain.

17. Six months later, Ms. Jefferson is readmitted to the unit. She has just returned from hemodialysis (HD). Which nursing care action should you delegate to the nursing assistant?

a. measure vital signs and post dialysis weight
b. assess the HD access site for bruist and thrill
c. check the access site dressing for bleeding
d. instruct the patient to request assistance getting out of bed

Directions: (questions 18 – 19)
Read the narrative below and then identify the choice that best completes the statement or answers the questions.
You are admitting Ms. Cooper, a 91-year-old patient, to the coronary care unit (CCU). Ms. Cooper, who has a history of mitral valve regurgitation and left ventricular enlargement, came to the emergency department (ED) with symptoms of increasing shortness of breath over the last week. The ED registered nurse tells you that the patient received furosemide (Lasix) 100 mg intravenously (IV) and that her dyspnea has improved. She is receiving oxygen via a nasal cannula at 3 L per minute. According to the ED nurse, Ms. Cooper now has crackles in both lungs bases, and her cardiac monitor shows a sinus rhythm, at a rate of 94 to 96, with occasional premature ventricular contractions (PVCs).

18. You reviewed the orders written by the ED physician. Which order is most important to clarify at this time?
   a. Infuse D5W at 10cc/hr
   b. Administer oxygen per CCU policy
   c. Give MS 2 - 4 mg IV PRN dyspnea or pain.
   d. Start nitro drip per protocol PRN chest pain

19. The physician rounds later in the shift and writes the following orders. Which of the physician orders listed is best to delegate to an experienced LVN who is assisting you?
   a. Give enalapril 2.5 mg by mouth.
   b. Administer furosemide (Lasix) 100 mg IV push
   c. obtain blood potassium level
   d. insert a number 16 French Foley catheter

Directions: (questions 20 – 21)
Identify the choice that best completes the statement or answers the question

20. Which patient should you, as the charge nurse, assigned to the care of and LVN, under the supervision of the RN team leader?
   a. a 51-year-old who has just undergone bilateral adrenalectomy
   b. an 83-year-old with type II diabetes and COPD
   c. the 38-year-old with myocardial infarction preparing for discharge
   d. a 72-year-old with mental status changes admitted from a long-term care facility

21. You are the charge nurse and are making assignments for the shift. A nurse with 10 years of experience in the Neurological ICU is floated to your unit. What is the most appropriate assignment for this nurse on your unit (a pediatric oncology unit)?
   a. An 18-month-old infant postsurgery with a neuroblastoma
   b. A 3-year-old boy with leukemia who has contracted bacterial meningitis
   c. A 6-year-old boy with osteogenic sarcoma
   d. A 14-year-old girl with a pheochromocytoma and adrenalectomy
Appendix B: Oral Questions

The following ten questions should be asked to the participant AFTER affording the participant 10 minutes to listen to recorded shift hand-off of four patients:

1. Which patient from your recorded report would you need to assess first on your nursing rounds?
2. Why is this patient’s situation a priority?
3. Is there any nursing task or tasks discovered from your recorded report that you as the RN could delegate to another team member?
4. Which task(s) would you delegate and to whom?
5. If you delegated this task, what then would be your responsibilities as the RN?
6. Speaking aloud, please describe to me the steps that you would take immediately after listening to this recorded report if you were in a hospital setting?
7. Speaking aloud, please describe to me the focused assessment priority of the patient in room 1A.
8. Why do you consider this body system as a priority for your focused assessment?
9. Are there any laboratory values provided to you in your shift hand-off that are concerning to you as a RN?
10. Why do you consider these laboratory values concerning?
## Appendix C: OQ Correct Responses

### Participant ID#:
Rater:

**Rater Directions:** Indicate participant verbal response to the ten oral questions associated with patient care management (PCM), delegation (DEL), and prioritization (PRI) by placing an “x” in the box directly under the behaviors.

<table>
<thead>
<tr>
<th>PCM Verbal Responses</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews patient MAR, MD order, and labs (organizes patient care needs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holds 0700 medication for patient 1A until MD notified of labs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connects pathophysiology of 1A patient to lab values.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbalizes need to receive MD orders for patient in 1A based on lab values.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduces self to all assigned patients during initial safety rounds.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auscultate lung &amp; heart sounds, obtain vital signs, and assess patient 1A before calling MD.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(6) Speaking aloud, please describe to me the steps that you would take immediately after listening to this recorded report if you were in a hospital setting? (9) Are there any laboratory values provided to you in your shift hand-off that are concerning to you as a RN? (10) Why do you consider these laboratory values concerning?

<table>
<thead>
<tr>
<th>DEL Verbal Responses</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delegates task to Charge RN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retains MD Call</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delegates task to LVN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delegates task to UAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right communication &amp; follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(3) Is there any nursing task or tasks discovered from your recorded report that you as the RN could delegate to another team member? (4) Which task(s) would you delegate and to whom? (5) If you delegated this task, what then would be your responsibilities as the RN?

<table>
<thead>
<tr>
<th>PRI Verbal Responses</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priorizes patient 1A as the priority assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate patient data collected (vital signs, labs, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most urgent care needs identified (Lasix held until MD notified)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization and prioritization of tasks / skills based on patient response or assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develops a plan of care and follows the plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiates call to MD regarding Lasix and potassium value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensures that MD orders are received and delivered to appropriate person to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Which patient from your recorded report would you need to assess first on your nursing rounds? (2) Why is this patient’s situation a priority? (7) Speaking aloud, please describe to me the focused assessment priority of patient in room 1A. (8) Why do you consider this body system as a priority for your focused assessment?

Embry/Ganley 2012
Appendix D: Scenario Validation Process Checklist

SCENARIO VALIDATION CHECKLIST

SCENARIO TITLE:  _STACY COLLINS IN-pATIENT NURSING LEADERSHIP RESEARCH VERSION__  VALIDATED BY: _N. SWEENY, MSN, RN__ DATE: _10/05/2012__

SECTION I: SCENARIO OVERVIEW

<table>
<thead>
<tr>
<th>x Yes</th>
<th>No</th>
<th>Scenario Title, Authors &amp; credentials</th>
<th>x Yes</th>
<th>Original date Validation level/Revision date</th>
<th>x Yes</th>
<th>Estimated time Target group Brief summary of case</th>
</tr>
</thead>
</table>

SECTION II: CURRICULAR INTEGRATION

<table>
<thead>
<tr>
<th>x Yes</th>
<th>No</th>
<th>Learning Objectives</th>
<th>x Yes</th>
<th>Secondary Objectives</th>
<th>x Yes</th>
<th>Critical Elements</th>
</tr>
</thead>
</table>

Evidence Base

<table>
<thead>
<tr>
<th>x Yes</th>
<th>No</th>
<th>Citation in APA format</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>x Clinical Expert Review</th>
<th>x  R. Zucker, MD  eSignature</th>
</tr>
</thead>
<tbody>
<tr>
<td>x Pharmacology Review</td>
<td>x  C. Carlucci, MD, RN  eSignature</td>
</tr>
<tr>
<td>x Simulation Expert Review</td>
<td>x  N. Sweeney, MSN  eSignature</td>
</tr>
</tbody>
</table>

Pre-Scenario Learner Activities

X Pre-scenario assignments appropriate for learners at program level
X Learner resources clearly identified prior to simulation
Psychomotor and cognitive competencies identified

Research on simulation as an evaluation method. Debriefing omitted due to logistics. SP debriefing in place after participation in simulation.

X Scenario Data

Debriefing method specified (SP only)

<table>
<thead>
<tr>
<th>x Yes</th>
<th>No</th>
<th>Debriefing guidelines and groundrules available</th>
<th>x Yes</th>
<th>Suggested debriefing questions (SP training x 3; no faculty debriefing)</th>
</tr>
</thead>
</table>

Core concepts included

X Scenario plausible (realistic)

X Patient Safety
X Leadership/Delegation
X Communication
X Priority Setting
X Cultural Diversity

X Critical Thinking/Decision Making skills required appropriate to level

<table>
<thead>
<tr>
<th>x Yes</th>
<th>No</th>
<th>Suggested debriefing questions</th>
</tr>
</thead>
</table>

X Origin of case scenario
X Case data appropriate
X Medications appropriate (revised per MD consultation)

203
<table>
<thead>
<tr>
<th>X Scenarios Script</th>
<th>X Lab results appropriate</th>
<th>X Dosages accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Summary - Sufficient data to present overview to those running scenario</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contextual Details - Scenario cues/triggers based on desired outcomes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Scenario Cast - Specifications</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Type of Simulator specified – Standardized Patient Actors with simulated clinical (acute care) environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Profile x Yes □ No Learner Roles specified *evaluation x Yes □ No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sufficient information for realistic scenario</td>
<td>X Yes □ No</td>
<td>Comments: SP training x 3; SP expert on-site during data collection (B. Grandis)</td>
</tr>
<tr>
<td>Simulation Team Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical appearance (see scenario)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant missing data: (Specify) X K+ out of range (critical value) with 0700 Lasix dose due.</td>
<td>X VS Monitor display □ No</td>
<td>IV lines SL □ Yes</td>
</tr>
<tr>
<td>Non-invasive monitors B/P monitor; thermometer; stethoscope</td>
<td>Other simulator monitors SPs to provide “cards” with written values/auscultation for assessment findings to standardize participant findings.</td>
<td></td>
</tr>
<tr>
<td>Setting specified X Essential props with triggers identified</td>
<td>X Confederate placement with triggers identified</td>
<td>X Essential equipment specified to be available</td>
</tr>
<tr>
<td>Environmental, Equipment, Essential Props Complete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running the Scenario X Documentation forms specified with data (PT chart; MD orders; MAR)</td>
<td>X Medications available with correct/incorrect options if specified</td>
<td>X IV delivery systems available with correct/incorrect options</td>
</tr>
<tr>
<td>Case Flow/Triggers/Scenario Development States</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sufficient detail in each case flow section</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X Resources available to run the scenario SIM Expert on-site; SIM Director; 14 RAs; 6 SIM Techs</td>
<td>□ Teaching points identified for debriefing N/A – evaluation research using simulation as a method of evaluation</td>
<td>X Learner Actions clearly identified</td>
</tr>
<tr>
<td>Recommendations: (reviewed with PI of research; SIM expert; SIM Director; SP Trainer/Expert; RAs; Faculty Evaluators) X Programming complexity at sim staff level</td>
<td>X Sufficient detail for non-authors to run scenario</td>
<td></td>
</tr>
<tr>
<td>Validation Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations: (Received &amp; reviewed with PI; SIM Expert; SIM Director; MD consultants; and Faculty Clinical Experts)</td>
<td>READY FOR PILOT TESTING – 10/25/2012</td>
<td>X Yes □ No</td>
</tr>
</tbody>
</table>
### Appendix E: OSCE Rotation and Stations

| Round 1 | Start 0800 | sim/SP 8:45-9:05 | End 0945 | Break 1040 | # of students 6 |
| Round 2 | Start 0830 | sim/SP 9:15-9:35 | 1020 | Break 1115 | 6 students |
| Round 3 | Start 0900 | sim/SP 9:45-10:05 | 1050 | Break 1145 | 6 students |
| Round 4 | Start 0930 | sim/SP 10:15-10:35 | 1120 | Break 1215 | 6 students |
| Round 5 | Start 1000 | sim/SP 10:45-11:05 | 1150 | Break 1245 | 6 students |
| Round 6 | Start 1030 | sim/SP 11:15-11:35 | 1220 | Break 1315 | 6 students |

**BREAKS**
- **TA's**: break in shifts
- **HA**: 11:15 - 11:40
- **SP**: 11:40 - 12:30
- **Skills**: 12:15

| Round 7 | Start 1200 | sim/SP 12:45-13:35 | 1350 | Break 1440 | 6 students |
| Round 8 | Start 1230 | sim/SP 13:15-13:35 | 1420 | Break 1510 | 6 students |
| Round 9 | Start 1300 | sim/SP 13:45-14:05 | 1450 | Break 1540 | 6 students |
### Round 10

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1330</td>
<td></td>
<td>14:15 - 14:35</td>
<td></td>
</tr>
<tr>
<td>1520</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1610</td>
<td></td>
<td></td>
<td>6 students</td>
</tr>
</tbody>
</table>

### Round 11

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1400</td>
<td></td>
<td>14:45 - 15:05</td>
<td></td>
</tr>
<tr>
<td>1550</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1640</td>
<td></td>
<td></td>
<td>2 students</td>
</tr>
</tbody>
</table>

3SPs 0800 – 1030
3SP 1200 - 1550
Appendix F: Report Script

(Adapted from Schultz, Shinnick, & Judson, 2012)

Report Script:
“This is ______________ reporting on patients 1A, 1B, 2A, 3A, and 4A. Room 1A (state age) (state gender) was admitted for HF yesterday, feeling much better after diuresis. Slept well, morning labs sent. IV Saline Lock, flushed 1 hour ago. Lots of medications; 0700 Lasix due now. Daily I&Os. Fluid restriction of 1200 mL/day. Called Dr. about the 3.0 potassium, but he did not call back yet. Nothing further to report.”
“Room 1B is a (state age) (state gender) admitted for asthma exacerbation yesterday. At 6 AM during last check, lung sounds were clear. He slept well without complaints. Last breathing treatment at 4 AM. No IVs, diet ad lib. Nothing further to report.”
“Room 2A is a (state age) (state gender) Admitted three days ago for bowel obstruction; early dementia. Daughter at bedside 24/7; helps some. Patient confused during the night and did not sleep much. Has NG tube to low intermittent suction, IV D5 W 1/2 normal saline@100 mL/hour. NPO; Bedrest, scheduled for exploratory laparotomy 0900. Nothing further to report.”
“3 A is a (state age) (state gender) admitted for pneumonia yesterday. Currently on IV antibiotics and D5 W 1/2 @ 75 mL/hour. Bedrest, MRSA positive in sputum. Nothing further to report.”
“4 A the patient was just transferred at change of shift to the ICU with the rapid response team, room not cleaned.”
Appendix G: LFS Rubric – PCM Station 2 (0700 IV push Lasix Administration for Patient in Room 1A)

Participant ID#: Rater:

**Rater Directions**: indicate participant demonstration of the seven (7) required behaviors associated with patient care management by placing an “x” in the box directly under the behaviors.

<table>
<thead>
<tr>
<th>Demonstrated behaviors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviews patient MAR, MD order, and labs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holds medication until MD notified of labs. Proceeds with MD approval.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulls medication using patient rights.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Looks up medication to determine administration time (how fast to push the medication).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduces self to patient and explains procedure.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand hygiene before performing associated assessment (auscultate lung &amp; heart sounds).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pushes medication over 2 minutes, flushes access device before and after administration. Documents administration and patient response.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Place an “x” in the box directly under the numbered box of demonstrated behaviors

**Examiner Global Rating**

- Good □
- Pass □
- Borderline Pass □
- Borderline Fail □
- Fail □

Characteristics of a borderline student include limited attention to detail; disorganized approach; lack of awareness of findings; poor articulation when required to give information; limited engagement; reduced professional presentation; shows concern for the patient but main emphasis is on performing the skill(s).

Embry/Ganley 2012
Appendix H: LFS Station 3 – Delegation of Nursing Tasks from Recorded Report

**Rater Instructions to Participant:** Instruct participant that he/she is the RN assigned to the patient load from the recorded hand-off that they just listened to/received. They are working the day shift (0700 to 1500) on a busy medical/surgical nursing unit at Dominican Medical Center. Seven (7) nursing tasks were identified from the recorded hand-off (report). There are seven (7) laminated cards; each of the cards represents one task that could be delegated to members of the nursing care team. There are four members of the nursing care team on this medical/surgical unit: RN (you), LVN, UAP, and the Charge RN. Each member of the nursing care team has been designated a labeled area on the table before you. Each member of the nursing care team has a specific scope of practice which limits which tasks that they can legally perform. With this in mind, delegate a task or tasks to any member of the nursing care team by placing the laminated card containing the task in the area on the table indicating which team member you would assign to carry out the task. Tasks that you will retain to carry out should be placed in the area labeled for the RN.

<table>
<thead>
<tr>
<th>Nursing Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
</tr>
<tr>
<td>Check the lab results on the patient in room 1A &amp; inform doctor regarding low lab value</td>
</tr>
</tbody>
</table>

**Rater Directions:** Once time has been called for the station, have participant turn away from the station. Remove the scantron from the participant packet. Each laminated card has a number from “1” to “7” which corresponds to the item number on the scantron. Each nursing team member has been designated a letter “A” to “D” which corresponds to the response to each item on the scantron. The participant has indicated which tasks would be delegated to which team member by placing the laminated card corresponding to the task into the area associated with the team member. With assistance from the RA, complete the scantron by filling in the letter indicated by the participant of the team member who would be assigned the task. Check scantron for accurate transcription and return it to the participant packet.

Embry 2012
Appendix I: LFS Station 4 – Prioritization of Initial Nursing Assessment Immediately after Receiving Report/Hand-off

**Rater Instructions to Participant**: Instruct participant that he/she is the RN assigned to the patient load from the recorded hand-off that they just listened to/received. They are working the day shift (0700 to 1500) on a busy medical/surgical nursing unit at Dominican Medical Center. Seven (7) RN assessments were identified from the recorded hand-off (report). Based on the notes that he/she took during handoff, seven (7) RN assessments need your attention. Each of the cards represents one RN assessment that requires your attention. If you must complete all the assessments, prioritize each assessment from highest (“A” area labeled on the table in front of you) to lowest (“D” area labeled on the table in front of you).

A = *Emergent* (assessment needs to be completed immediately after receiving handoff)  
B = *Urgent* (assessment should be completed within 30-minutes of receiving handoff)  
C = *Important* (assessment should be completed within 60-minutes of receiving handoff)  
D = *Routine* (assessment should be completed within 90-minutes of receiving handoff)

<table>
<thead>
<tr>
<th>Nursing Assessments</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess the patient in room 1A</td>
<td>Assess patient in room 1B</td>
<td>Assess patient in room 2A</td>
<td>Assess patient in room 3A</td>
<td>Assess status of room 4A</td>
<td>Assess MAR of all assigned patients</td>
<td>Assess morning lab values</td>
<td></td>
</tr>
</tbody>
</table>

**Rater Directions**: Once time has been called for the station, have participant turn away from the station. Remove the scantron from the participant packet. Each laminated card has a number from “1” to “7” which corresponds to the item number on the scantron. Each prioritization level has been designated a letter “A” to “D” which corresponds to the response to each item on the scantron. The participant has indicated which assessments have higher or lower priority based on where he/she placed the laminated card containing the RN assessment. With assistance from the RA, complete the scantron by filling in the letter indicated by the participant of the priority of each RN assessment. Check scantron for accurate transcription and return it to the participant packet.

Embry 2012
Appendix J: HFS Rubric – Patient Care Management, Delegation, and Prioritization

Behaviors during Care Delivery

Participant ID#: Rater:

**Rater Directions**: Indicate participant demonstration of the seven (7) required behaviors associated with patient care management (PCM), delegation (DEL), and prioritization (PRI) by placing an “x” in the box directly under the behaviors.

<table>
<thead>
<tr>
<th>PCM Demonstrated behaviors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reviews patient MAR, MD order, and labs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Holds medication until MD notified of labs. Proceeds with MD approval.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pulls medication using patient rights.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Looks up medication to determine administration time (how fast to push the medication).</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Introduces self to patient and explains procedure.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hand hygiene before performing associated assessment (auscultate lung &amp; heart sounds).</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pushes medication over 2 minutes, flushes access device before and after administration. Documents administration and patient response.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Place an “x” in the box directly under the numbered behavior.

<table>
<thead>
<tr>
<th>DEL Demonstrated behaviors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delegate call from Admission to Charge RN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Retains MD Call</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Delegates task to LVN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Delegates task to UAP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Right person</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Right task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Right communication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Place an “x” in the box directly under the numbered behavior.

<table>
<thead>
<tr>
<th>PRI Demonstrated behaviors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prioritizes patient safety (hand hygiene, verifies right patient)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appropriate patient data collected (vital signs, labs, etc.)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Most urgent care needs identified (Lasix held until MD notified)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Organization and prioritization of tasks / skills based on patient response or assessment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Develops a plan of care and follows the plan while communicatin g with patient</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Initiates call to MD regarding Lasix and potassium value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ensures that MD orders are received and delivered to appropriate person to expedite implementatio n</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Place an “x” in the box directly under the numbered behavior.

Embry 2012
CONFIDENTIAL DISCLOSURE AGREEMENT

regarding the protection of Confidential Ideas, Responses, or Information to be disclosed by the Primary Investigator or any participant in the below mentioned research to the Student Research Assistant.

Research Title: A MULTITRAIT MULTIMETHOD STUDY OF LEADERSHIP SKILLS IN TWO SENIOR-LEVEL BACCALUREATE NURSING COURSES

PRIMARY INVESTIGATOR _______________________________________________

STUDENT RESEARCH ASSISTANT________________________________________

EFFECTIVE DATE of Agreement _________________________

In return for possible benefits to be received in the future, both parties (above) agree to the following Terms of Disclosure:

1. Primary Investigator agrees to disclose Confidential PARTICIPANT RESPONSES which are believed by the RESEARCH PARTICIPANT to have been held in confidence by the Primary Investigator.

2. Student Research Assistant agrees not to disclose any of these Confidential PARTICIPANT RESPONSES to any others in any way without the prior consent of the Primary Investigator.

3. Student Research Assistant agrees not to use any of these Confidential PARTICIPANT RESPONSES in any way without the prior consent of the Primary Investigator.

4. On request of the Primary Investigator, the Student Research Assistant shall immediately return all documents and other items associated with this Disclosure and shall not retain any unauthorized copies or likenesses.

It is understood that this Agreement does not cover ideas which were associated with the segment of research that the Student Research Assistant must complete in fulfillment of his/her Directed Research Project or which were already known to the Student Research Assistant prior to their disclosure by the Primary Investigator.

As with all agreements, this Agreement can be terminated or amended at any time by mutual agreement of both parties.

SIGNING FOR THE PRIMARY INVESTIGATOR ______________________________

Date ______________________

SIGNING FOR THE STUDENT RESEARCH ASSISTANT ______________________

Date ______________________
## Appendix L: Expert Plan Review Checklist

**EXPERT PANEL REVIEW / VALIDATION CHECKLIST**

<table>
<thead>
<tr>
<th>STACY COLLINS IN-PATIENT NURSING LEADERSHIP RESEARCH ORAL QUESTIONING, RN HANDOFF/REPORT, MULTIPLE-CHOICE ITEM</th>
<th>DATE: 10012012</th>
</tr>
</thead>
</table>

### SECTION I: INSTRUMENT OVERVIEW

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Instrument Title, Authors &amp; credentials</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>1) MCQ Instrument 30-items</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Original date 10012012</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>Validation level/Revision date 10312012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Estimated time: OQ = 20 minutes; MCQ = 30 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>Target group: senior-level BSN students (med/surg III – leadership)</td>
</tr>
</tbody>
</table>

---

### TEST PLAN CONSTRUCTION/COMPONENTS

#### Learning Objectives

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Primary Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>Evaluate participant application of three nursing leadership skills (PCM, DEL, PRI) using MCQ and OQ items.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Secondary Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>Reliability and Validity of Instruments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Critical Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>Level of student; prior knowledge; clinical practice with constructs.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Evidence Base (see literature review by author; both instruments developed for doctoral research)</th>
</tr>
</thead>
</table>

**Primary Text/Chapter Source (Primary Learning Objectives):**


<table>
<thead>
<tr>
<th>x Yes</th>
<th>Citation in APA format</th>
</tr>
</thead>
</table>


---

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Assesment/Statistical Expert Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>B. Ganley, PhD, RN eSignature</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Pharmacology Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>E. Klich-Heartt, DNP, RN, CNL eSignature</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x Yes</th>
<th>Leadership Expert Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ No</td>
<td>J. Wilson, PhD, RN eSignature</td>
</tr>
</tbody>
</table>

Comments: eSignature for use in electronic version of document for publication. Original signed documents on file with Research PI.

---

<table>
<thead>
<tr>
<th>X Pre-test assignments appropriate for learners at program level</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Learner resources clearly identified prior to administration of instruments</th>
</tr>
</thead>
</table>

**Bloom’s Taxonomy Levels identified: Applying, Analyzing, Evaluating**

---

### Item Construction/Components

<table>
<thead>
<tr>
<th>X Narrative plausible (realistic)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Critical Thinking/Decision Making skills required appropriate to level</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Patient Safety</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Leadership/Delegation</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Communication</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Priority Setting</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Cultural Diversity</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Patient Teaching</th>
</tr>
</thead>
</table>

**DIRECTIONS - Sufficient instruction/information for all items to be completed during the timeframe with target**

<table>
<thead>
<tr>
<th>X Origin of case</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Case data appropriate</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>X Medications appropriate (revised per DNP consultation)</th>
</tr>
</thead>
</table>

---

213
### Patient Profile

<table>
<thead>
<tr>
<th>Items used previous? Date: Spring 2011 MCQ items; N/A OQ items</th>
<th>Lab results appropriate</th>
<th>Dosages accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>x Yes □ No Narrative(s) sync with item taxonomy level</td>
<td>x Yes □ No</td>
<td></td>
</tr>
<tr>
<td>NCLEX format (MCQ items) Socratic in nature (OQ items)</td>
<td>x Yes □ No</td>
<td></td>
</tr>
</tbody>
</table>

### Physical Administration Environment

**SIGNIFICANT SPACE FOR ADMINISTRATION OF INSTRUMENT:**

(Specify)

OQ – Skills Lab; three participants per round; three RN evaluators present; three RAs present for each round

MCQ – Reserved Classroom; max number allowed per round = 6 participants; two RAs present with instrument, scantron, pencil.

**ESSENTIAL EQUIPMENT/RESOURCES COMPLETE** □ YES □ NO

**Specify:** Audio recording device; recorded 5-minute report; seating for three participants; resource texts (OQ)

### Item Analysis

<table>
<thead>
<tr>
<th>Total Possible Points: 30</th>
<th>Participants in Group: 64</th>
<th>Lowest Score: 12</th>
<th>Measure of Skewness:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median:</td>
<td>Mode:</td>
<td>Highest Score:</td>
<td>Degree of Kurtosis:</td>
</tr>
<tr>
<td>Mean: 23.1</td>
<td>Variance: 2.68</td>
<td>STANDARD DEVIATION: 1.64</td>
<td>KR20 (reliability coefficient): .45</td>
</tr>
</tbody>
</table>

**Number of Items:** 30

**Point Biserial**

- # of items with point biserial .30 or above (very good item): Item #’s: 2, 3, 6, 10, 14, 18
- # of items with point biserial .20 to .29 (reasonably good item): Item #’s: 19, 22, 24, 27
- # of items with point biserial .09 to .19 (items that need improvement): Item #’s: 5, 11, 21, 30
- # of items with point biserial below .09 = poor item: Item #’s: 9

**Comment:** REDUCE THE NUMBER OF ITEMS GIVEN THE TIME ALLOTTED (MCQ); CONSIDER REVISING RECORDED REPORT FOR OQ TO IMPROVE FIDELITY OF HANDOFF.

**Recommendations:** (MCQ) OMIT ITEM 9; REVISE ITEMS: 5, 11, 21, & 30

<table>
<thead>
<tr>
<th>Validation Level</th>
<th>Revision of Items Required</th>
<th>Reevaluate Instrument After Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Omission of Specific Items Required</th>
<th>No Omission/Revision/Reevaluation Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td>□</td>
</tr>
</tbody>
</table>


Appendix M: IRBPHS Approval Letter – Dominican University of California

October 9, 2012

Toby Embry
48 Lochness Lane
San Rafael, CA, 94901

Dear Toby:

I have reviewed your proposal (entitled, A Multitrait Multimethod Study of Nursing Leadership Skills in Two Senior-Level Baccalaureate Nursing Courses) submitted to the Dominican University Institutional Review Board for the Protection of Human Subjects (IRBPHS Application, #10037). I am approving it as having met the requirements for expedited review.

In your final report or paper please indicate that your project was approved by the IRBPHS and indicate the identification number.

I wish you well in your very interesting research effort.

Sincerely,

Martha Nelson, Ph.D.

Chair, IRBPHS
Institutional Review Board for the Protection of Human Subjects
Office of the Associate Vice President for Academic Affairs  50 Acacia Avenue, San Rafael, California
95901-2298  415-257-1310
www.dominican.edu
December 17, 2012

Dear Toby Embry:

The Institutional Review Board for the Protection of Human Subjects (IRBPHS) at the University of San Francisco (USF) has reviewed your request for human subjects approval regarding your study. Your study has been deemed to be exempt from IRB review based on the following conditions:

Unless otherwise required by department or agency heads, research activities in which the only involvement of human subjects will be in one or more of the following categories are exempt from this policy:

1) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects, and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation.

This application does not require IRB review.

On behalf of the IRBPHS committee, I wish you much success in your research.

Sincerely,

Terence Patterson, EdD, ABPP
Chair, Institutional Review Board for the Protection of Human Subjects

--------------------------------------------------
IRBPHS ˆ University of San Francisco
Counseling Psychology Department
Education Building ˆ Room 017
2130 Fulton Street
San Francisco, CA 94117-1080
(415) 422-6091 (Message)
(415) 422-5528 (Fax)
irbphs@usfca.edu

http://www.usfca.edu/soe/students/irbphs/