Dissertations   Theses and Dissertations

2013-05-01

Psychometric Properties of Standardized Patient and Faculty Rater's Evaluations of Pre-Licensure Nursing Student Competencies

Karen A. Macauley PhD, DNP, FNP-BC, GNP, BC
University of San Diego

UNIVERSITY OF SAN DIEGO

Hahn School of Nursing and Health Science

DOCTOR OF PHILOSOPHY IN NURSING

Psychometric Properties of Standardized Patient and Faculty Rater's

Evaluations of Pre-Licensure Nursing Student Competencies

by

Karen A. Macauley, DNP, FNP-BC, GNP, BC

A dissertation presented to the

FACULTY OF HAHN SCHOOL OF NURSING AND HEALTH SCIENCE

UNIVERSITY OF SAN DIEGO

In Partial Fulfillment of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY IN NURSING

May 2013

Dissertation Committee

Cynthia Connelly, PhD, RN, FAAN, Chairperson

Jane Georges, PhD, RN

Andrea Hazen, PhD

ABSTRACT

Psychometric Properties of Standardized Patient and Faculty Rater's

Evaluations of Pre-Licensure Nursing Student Competencies

The use of simulation as a teaching modality has been rooted in the military,

aviation, space, and engineering for centuries (Bradley, 2006). Clinical simulation allows

for training of healthcare providers that might be too costly, risky, or hazardous in the

clinical setting (Bradley, 2006). A variety of simulation modalities are used including

virtual learning, task trainers, mannequins, and standardized patients (SPs). External

demands for improved accountability of clinical performance is requiring nursing

educators to reevaluate methods of teaching and how we measure nursing competence

(Nehring & Lashley, 2010).

Standardized patients have been used in medical school curricula to teach and

evaluate clinical competence of medical students for decades (Boulet, 2008). Even

though SP programs are used and well-researched in medical schools, the majority of

nursing schools have adopted high-fidelity mannequin simulation programs (Sanford,

2010). Standardized patients contribute to increased realism by exposing students to a

"real" patient with opportunities to practice compassionate and empathetic

communication skills and receive feedback on how to fine-tune their bedside manner.

The capacity to provide compassionate care is the heart and soul of nursing practice as

identified by American Association of Colleges of Nursing and the National League of

Nursing (Rhodes, Morris, & Lazenby, 2011).

SP reliability and validity are well established within medical education, reporting

88-92% agreement on checklists between SPs and faculty. Competency checklists in pre-

licensure registered nursing curricula have not been accompanied with equally rigorous psychometric evaluation thus it is unclear whether SP utility in nursing is equivalent to medical education. This study examines the inter-rater reliability and percent agreement of standardized patients and faculty checklist scores when evaluating pre-licensure nursing students.

Data analysis of SP and faculty scores found signicant agreement (94%-98%) as seen in medical education decades ago. Low internal consistency measures and moderate kappa scores suggest additional research is needed working with multi-site, large sample sizes using the same methodology, cases, and checklists. Nursing programs primarily using mannequins have not been able to realize the potential of using SPs, not only in the evaluation of competence, but also in laying the foundation of practicing and reflecting on humanistic care.

DEDICATION

This dissertation is dedicated to

Wesley Macauley, my loving husband,

and our children, Jennifer and Christopher,

who have been inspirational and

incredibly supportive throughout this journey.

ACKNOWLEDGEMENTS

I owe a debt of gratitude to my committee, Dr. Cynthia Connelly who is an incredible researcher, amazing mentor, and empathic listener and friend, Dr. Jane Georges whose astounding reflective acumen drew deep understanding of our calling focusing on how caring and the spirit and art of our practice contribute to healing, and Dr. Andrea Hazen a statistician with unwavering patience and kindness. They have not only supported me through this dissertation process but also believed in me and encouraged my passion in the pursuit of being a "practice scientist". Thank-you for your collaboration, inspiration, and shared vision over the last two years.

It has taken a team of amazing mentors who have supported me throughout the last several decades throughout my NP, DNP, and now PhD education. I must recognize the innovative thinkers and early developers of our standardized patient program Drs. Louise Rauckhorst and Cheryl Ahern-Lehman. Thank-you Dr. Mary Jo Clark for your steadfast advising throughout my NP and DNP programs and Drs. Susan Instone and Susan Bonnell for your ongoing support of the SP program and research. The staff and faculty of the simulation lab, Dr. Kathleen Sweeney, Dr. Susie Hutchins, Deb, and Paula are an amazing team who work relentlessly to ensure that we are graduating professional, compassionate, competent, and reflective nurses. I am grateful for your dedication, determination, and vision.

Lastly and most importantly, I am thankful for my family. My parents and siblings who have taught me and role-modeled throughout my life that the foundation of our being is grounded in God's presence to do for others with humility and compassion. As taught to me by my mother and sister, both nurses, we are invited into the inner spaces of other people's live. We call that nursing however it was a fundamental philosophy ingrained into my childhood and paid forward to raise our own families. My parent's amazing influence is why our family has a new generation of professions to serve others in law enforcement, physical therapy, lawyers who advocate for social justice, a minister, a university counselor and soon to be paramedic, also those who bring joy to others and share dad's talents of ingenuity through building musical instruments along with sharing his mechanical talents

To my children, Jenny and Chris, who have both endured parents in school for many years, I hope this encouraged you to pursue your own advanced education and dreams early on. I can't be more proud as a mother and be inspired by your inner strength, tenacity, and successes. I love you both!

To my husband who has been by my side for the last 35 years, encouraging and supporting me through this very long journey. Your daily reassurance and incredible gift of love has been a constant source of strength and motivation, I couldn't have accomplished this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter 1

Introduction

The use of simulation as a teaching modality has been rooted in the military, aviation, space, management, and engineering for centuries (Bradley, 2006). Clinical simulation or simulation in practice-based healthcare education, as in other professions, allows for training that might be too costly, risky, or hazardous in the clinical setting (Bradley, 2006). Two seminal Institute of Medicine (IOM) reports *To Err is Human* (Kohn L.T., Corrigan, & Donaldson, 2000) and *Healthcare Professions Education: A Bridge to Quality* (Greiner & Knebel, 2003) which focus on safety and quality of patient care along with healthcare education reform have been the driving force for healthcare educational systems to adopt alternative and innovative ways to teach future generations of healthcare professionals (Harder, 2009).

Background

Historically, nursing education has relied on clinical practicum to provide students with opportunities to practice and refine nursing skills (Henneman & Cunningham, 2005). Recently, because of the IOM's emphasis on safe patient care and healthcare education reform, this tradition of "practicing on patients" in the clinical arena has been replaced by innovative simulated teaching strategies in simulation laboratories

1

(Henneman & Cunningham, 2005). It is the professional responsibility of nursing educators to prepare nursing graduates to transition to a professional practice role, to clearly articulate and evaluate basic nursing competencies, and foster the performance of our practice through nursing education grounded in science. External demands for improved accountability of clinical performance is requiring nursing educators to reevaluate methods of teaching and re-examine how we measure and manage our view of nursing competence (Nehring & Lashley, 2010). Professional regulatory agencies are concentrating on educational systems and methodologies to improve communication, reduce medical errors, and improve competence of all healthcare providers. Nursing education must focus on assuring the competence of our trainees as they enter the professional life of a nurse.

Over the past few years, simulation has been incorporated into nursing education to allow learners to practice in a safe, controlled environment. Simulation, or the replication of clinical scenarios in a laboratory setting, offers teaching-learning opportunities to apply and evaluate theoretical knowledge integrating technical and interpersonal skills (Norman, 2012). Simulation offers an environment for learners to practice and make mistakes without the fear of harming a patient. After simulation scenarios, clinical faculty guide learning opportunities through debrief sessions and allow time for learner self-reflection. Interactive learning is fostered through simulation allowing learning opportunities for refining practice and communication to ensure quality and safe patient care (Ironside, Jeffries, & Martin, 2009). Notably, the National Council of State Boards of Nursing (NCSBN) has explored substituting simulation activities for actual clinical

experience.

Most nursing simulation programs focus on the use of high-fidelity mannequins or human patient simulators (HPS) for nursing students to practice and apply skills and knowledge including communication skills (Nehring & Lashley, 2009). Medley and Horne (2005) identified nursing students have trouble envisioning a mannequin as an actual patient contributing to difficulties in transitioning from the lab to the clinical bedside. Ward, Cody, Schall & Hojat (2012) found a decline in empathy among recent undergraduate nursing students over an academic year. These findings suggest nursing educators must re-examine solely mannequin-based methods of simulation and its impact on future generations of nurses potentially compromising the "art" of nursing. Standardized patients in simulation scenarios contribute to increased realism by exposing students to a "real" patient with opportunities to learn and practice human compassion, empathetic communication skills, and receive feedback from the patient on how to fine-tune their bedside acumen and connection with the patient.

The use of standardized patients (SPs) is slowly transitioning into pre-licensure and advanced practice nursing programs as a formative teaching methodology and for student evaluation (Vessey & Huss, 2002). SPs are individuals coached to consistently portray an actual patient including history, physical findings, and emotions, and feelings. The use of SPs to teach and assess clinical and communication skills has been well documented and researched in the medical literature (Norcini & Boulet, 2003). The reliability and validity of SPs are well established and form the basis for licensure and certification of physicians since 2004 (Boulet, 2008), however their use in nursing is

limited and lacks psychometric rigor (Watson, Stimpson, Topping, & Porock, 2002).

Several articles have been written identifying the strengths and weaknesses on the

utilization of SPs in nursing programs, however to date, there are no studies focused

primarily on comparing faculty and SP inter-rater reliability. The majority of the

research rigor measuring clinical competence has been done outside of nursing. Studies

including instrument development using rigorous psychometric properties are needed

within the nursing profession (Watson et al., 2002). Cronenwett et al. (2007) argue

psychometric rigor is needed in relation to measures of nursing competence and further

longitudinal investigation is necessary with large numbers of nursing students. Studies

are needed to investigate the generalizability of previous findings of acceptable reliability

of SP ratings of performance in medical education to nursing education. As seen in

medical education, large-scale standardized patient nursing research studies can

potentially provide a valid and consistent measure of nursing student competence

relieving faculty of laborious and subjective methods of nursing student evaluation.

Purpose

The purpose of this project is to explore existing nursing literature and the

potential impact of SPs as a simulation methodology in nursing. The importance of this

study is to strengthen and employ rigorous methods of evaluating pre-licensure nursing

student competence. The premise is SPs can consistently and accurately complete

competency checklists similar to nursing faculty.

Specific Aim

Examine the inter-rater reliability and percent agreement of standardized patients

and faculty competency checklist scores when evaluating pre-licensure nursing students

Research Question

Are scores completed by SPs on competency checklists consistent with faculty

scores when evaluating pre-licensure nursing students?

Conceptual Framework

Over several decades, the nursing profession has utilized several theories as

frameworks for nursing programs and experimented in use of various educational

strategies to measure nurse competencies. In recent years, due to external public demand

of clinical competence of all healthcare providers, new frameworks, models, and middle-

range theories are beginning to emerge within the nursing profession. As schools of

nursing begin the task of defining and measuring competence of all nursing students, it is

important that nurse educators employ new frameworks to ensure safe, effective nursing

practice (Nehring & Lashley, 2010).

The application of conceptual frameworks, nursing grand and middle range

theories, and theories from varied disciplines such as educational theories pioneered by

John Dewey's (Dewey, 1916), David Kolb's experiential learning theory (Kolb, 1981)

and Malcolm Knowles adult learning theory (Knowles, 1975) guide nursing educators to

facilitate simulation teaching and learning methodologies that effectively guide students

to provide safe, effective care to patients. Patricia Benner (2001), based on the Dreyfus

Model of Skill Acquisition, identifies five stages of nursing practice ranging from novice

to expert. Benner's definition of competence is characterized by a feeling of mastery and the ability to plan patient care while managing the demands of clinical nursing (Benner, 2001). Her learning theory along with Bandura's self-efficacy middle range theory, constructivist-learning approach have all been used as theoretical frameworks for simulation studies (Kaakinen & Arwood, 2009). More recently, the National League for Nursing (NLN) in partnership with the Laerdal Corporation, a large vendor of low and high fidelity mannequins, developed a simulation model to guide expansion of simulation teaching methodology (Jeffries, 2005). Literature and research applying the NLN simulation model has been primarily grounded in high and low fidelity mannequin-based simulation methods (Jeffries, 2005). Because of the lack of humanistic core values in NLN's simulation model, this study will employ a conceptual framework informed by the works of early experiential educational theorists and Benner's (2001) Novice to Expert theoretical framework focusing on the key concept of measuring competence defined by the Quality and Safety Education for Nursing (QSEN) competencies (see Figure 1.). The empirical referents or measurement tools were developed based on established domains of the QSEN competencies for pre-licensure nursing students identified as knowledge, skills, and attitudes. The QSEN competencies provide a standard based approach and framework for all nursing educators by exploring these three overall encompassing attributes in relation to nursing competence (Cronenwett et al., 2007).

Significance to Nursing

Current clinical performance evaluation methods by faculty in the actual practice environment have been described as subjective, logistically challenging, and time

consuming (Gibbons et al., 2002). Direct observation of students in the clinical setting requires extensive, valuable faculty time that lacks psychometric reliability and validity (Becker, Rose, Berg, Park, & Shatzer, 2006). Factors such as variability of clinical settings, differing levels of patient acuity, decreasing number of clinical sites, and faculty biases limit confidence in clinical performance assessments (Gibbons et al., 2002). Lack of quality assessment instruments and problems with faculty consensus further confound evaluations of students in the real world (Gallagher, Smith, & Ousey, 2012). Nursing educators should reflect and consider the several decades of rigorous SP psychometric research in medical education forming the basis for the Medical Council of Canada (1993) and the United States National Board of Medical Examiners (USMLE) (1995) endorsing the use of SPs to establish a performance-based, clinical-competency examination as part of the USMLE Step II.

With the current nursing shortage, faculty are at a premium especially in light of the added rigor that is becoming increasingly expected of pre-licensure programs. Students have identified standardized patient feedback to be the most valuable evaluation method when comparing faculty, videotape, and SP evaluations (Ebbert & Connors, 2004). This study proposes the inclusion of SP to consistently and accurately record the clinical competencies of pre-licensure students, permitting nursing faculty to attend to other important aspects of student education.

The emphasis of healthcare professional education including nursing education needs to shift from individual knowledge and skill to performance of healthcare teams (Gaba, 2004). The recent IOM reports focus on transforming healthcare educational

systems and methodologies to advance quality of patient care through teamwork, communication, reduction of medical errors, and improve competence of all healthcare providers. Ultimately, the most crucial driver of simulation is the public who are demanding near perfect safety in healthcare settings and within providers of healthcare. Malpractice attorneys have seized upon opportunities of incompetence and negligence of healthcare providers and settings (Gaba, 2004).

Simulation in healthcare has a direct impact on the promotion of best practices to reduce errors and improve the quality of service (Barjis, 2011). Simulation has shown to "successfully enhance and improve clinicians' expertise, accomplish acquaintance with new procedures and processes, and prevent errors that are caused by the lack of training and hands-on experience" (Barjis, 2011, p. 3). Simulations can be used to evaluate competencies of healthcare professionals and teams (Gaba, 2004). Further simulation research will contribute to adult and experiential learning theories and inquiry of human factors in regards to complex system and clinical reasoning.

The increased use of technology adds to the complexity of healthcare systems and the need to prepare future generations of healthcare providers to work within these multifaceted systems. The potential power of informatics to improving healthcare efficiency and quality of care is driving innovative redesigns of healthcare processes and practices. The inclusion of technology into simulation activities prepares healthcare providers for the real world of acute and primary healthcare delivery settings and practices.

Rigorous SP psychometric investigation, as seen decades ago in medicine, are

needed in the nursing domain to provide legitimacy for nursing educators to adopt SP

simulation methodology in nursing programs and eventually serve as consistent

evaluation of competencies for nursing licensure and/or certification

Figure 1. Conceptual Framework
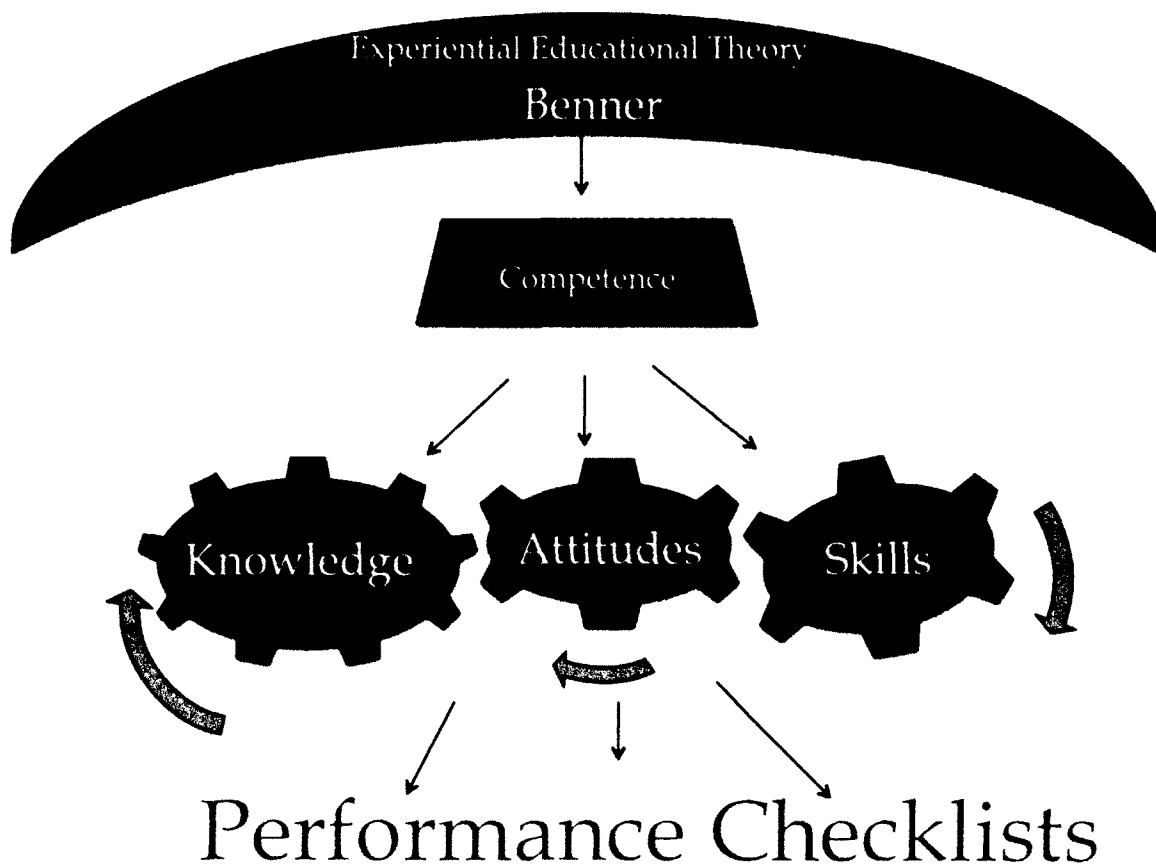
Chapter 2

Review of Literature

This chapter will provide an overview of the history of simulation as a teaching-learning modality, the use of simulation in healthcare and in nursing, use of standardized patients (SPs) in healthcare and in nursing, and review of the nurse competence literature. The conceptual framework for the study will be described. The state of the science in relation to the use of SPs to evaluate competence along with gaps in the literature will also be discussed.

History of Simulation as a Teaching-learning Modality

Simulation: Acting out or mimicking an actual or probable real life condition, event, or situation to find a cause of a past occurrence (such as an accident), or to forecast future effects (outcomes) of assumed circumstances or factors. A simulation may be performed through (1) solving a set of equations (a mathematical model), (2) constructing a physical (scale) model, (3) staged rehearsal, (4) game (such as war games), or a computer graphics model (such as an animated flowchart). Whereas simulations are very useful tools that allow experimentation without exposure to risk, they are gross simplifications of the

11

reality because they include only a few of the real-world factors, and are only as good as their underlying assumptions (Business Dictionary.com, n.d.).

Simulation training has roots in many industries such as commercial aviation, nuclear power, aerospace, engineering, military training, and more recently healthcare (Harder, 2009). The pedagogy of this experiential learning strategy spans over many centuries and has an increased emphasis in education over the last several decades (Wilford & Doyle, 2006). Simulation strategies, used initially in the military to prepare forces for war, included the development of the game of chess and various jousting war games to sharpen battlefield abilities (Bradley, 2006). These initial attempts at using simulation are the foundation for the now complex, digital, and innovative strategies used by the military to prepare soldiers for the battleground (Bradley, 2006). Aviation and the aerospace industries have used flight simulators in preparing astronauts and pilots for space and air travel (Bradley, 2006). Commonalities among all of these industries including healthcare is the use of simulation to replicate serious high-risk conditions without the fear of harming individuals. Simulation provides opportunities for learners to deliberately practice in a safe, non-threatening environment to hone skills and prepare for critical and unusual crisis situations (Rystedt & Lindstrom, 2001).

A groundbreaking study by the Royal Navy in 1975 compared divers who memorized material underwater and on land (Wilford & Doyle, 2006). Findings showed an increased retention of knowledge in the underwater divers than those on land (Wilford & Doyle, 2006). This report supports simulation methodology by showing knowledge and skill can be reproduced and retained if training occurs in realistic environments.

Hence, the more realistic the training environment the more effective the learning will be. "Simulation allows the creation of realistic simulations to allow greater retention of what is learned" (Wilford & Doyle, 2006, p. 605).

Use of Simulation in Healthcare

Healthcare simulation is defined by Gaba as a "…. technique, not a technology, to replace or amplify real experiences with guided experiences, often immersive in nature, that evoke or replicate substantial aspects of the real world in a fully interactive fashion" (Gaba, 2004, p. i2). Because of the successes seen in non-medical industries, the application of simulation shows great promise and acceptance to play a major role in healthcare (Barjis, 2011). With today's extraordinary healthcare expenditures, drastic complex healthcare delivery changes, major technological transitions, and regulatory agencies monitoring quality of care and service, simulation "allows replication of reality allowing exploration of possible changes, experiencing situations that otherwise will not be possible" (Barjis, 2011, p.1). Clinical simulation or simulation in practice-based healthcare education, as in other professions, allows for training that might be too costly, risky, or hazardous in the clinical setting (Bradley, 2006). Simulation, or the replication of clinical scenarios in a laboratory setting, offers teaching-learning opportunities to apply and evaluate theoretical knowledge integrating technical and interpersonal skills (Norman, 2012). A variety of simulation modalities are used in healthcare education including but not limited to role-play, virtual learning, task trainers, mannequins, and standardized patients (SPs).

Use of Simulation in Nursing

Historically, nursing education has relied on clinical practicum to provide students with opportunities to practice and refine nursing skills (Henneman & Cunningham, 2005). Because of the IOM's emphasis on safe patient care and healthcare education reform, this tradition of "practicing on patients" in the clinical arena has been replaced by innovative simulated teaching strategies in simulation laboratories (Henneman & Cunningham, 2005). Over the past few years, simulation has been incorporated into nursing education to allow learners to practice in a safe, controlled environment. Simulation offers an environment for nursing students to practice and make mistakes without the fear of harming a patient. After simulation scenarios, clinical faculty guide learning opportunities through debrief sessions and allow time for student self-reflection. Interactive learning is fostered through simulation allowing learning opportunities for refining practice and communication to ensure quality and safe patient care (Ironside et al., 2009).

It is the responsibility of nursing educators to prepare nursing graduates to transition to a professional practice role, clearly articulate and evaluate basic nursing competencies, and to foster the performance of our practice through nursing education and science (Kohn L.T. et al., 2000). External demands for improved accountability of clinical performance is requiring nursing educators to reevaluate methods of teaching and re-examine how we measure and manage our view of nursing competence (Nehring & Lashley, 2010). Nursing education must focus on assuring the competence of our trainees as they enter the professional life of a nurse (Kohn L.T. et al., 2000).

Nursing educators are constantly challenged to present new knowledge to students through didactic instruction in the classroom ultimately resulting in safe, competent performance of clinical skills in our healthcare communities. Nursing has recently seen a sudden increase in the use of this experiential adult learning and innovative teaching methodology allowing faculty to control the learning environment through deliberate practice, providing feedback, which mimics the reality of a clinical environment. Scenarios are developed for nursing students to focus on patient safety by applying critical thinking and effective communication skills, fostering interprofessional teamwork, preventing medication errors, and developing clinical decision-making and reasoning skills. Breakdown of these high-risk skills have shown to cause preventable errors in many health care systems and institutions (Kohn L.T. et al., 2000).

Traditionally, faculty members visiting community clinical or hospital settings where students are placed for hands-on experience is the current standard for clinical evaluation of nursing student performance (Gibbons et al., 2002). Unfortunately, this is not the most ideal way of evaluating clinical performance. When visiting a clinical site, the faculty member will only see a snapshot view of the nurse/patient interaction with a few select patients. The student may be reluctant to apply newly acquired knowledge, skills, and communication techniques as they may upset or harm the actual patient if not done correctly. Students may be intimidated in these settings, not allowing practice of skills required for active learning. These environments do not allow faculty to evaluate a group of students in a consistent manner due to the highly variable and unpredictable presentation of patients. In these environments, the presentation of patients, types of

diagnoses, and complexity of high-acuity clinical health situations are unpredictable and clinical observation of nursing students can be variable (Gibbons et al., 2002). Mastering clinical skills can be challenging, as it is difficult to find adequate real-life experiences for all students in the clinical sites. The lack of clinical placements along with the limited scope of what student nurses can do in acute care settings are just a few challenges nursing schools face when trying to provide rich learning experiences for students in clinical settings. Simulation labs allow faculty to observe all students with the same simulated case scenario, objectively evaluating the student on set competency criteria (Ironside et al., 2009). The National Council of State Boards of Nursing (NCSBN) has explored substituting simulation activities for actual clinical experience.

Most nursing simulation programs focus on the use of high-fidelity mannequins or human patient simulators (HPS) for nursing students to practice and apply skills and knowledge including communication skills. Medley and Horne (2005) found nursing students have trouble envisioning a mannequin as an actual patient contributing to difficulties in transitioning from the lab to the clinical bedside. Ward et al. (2012) found a decline in empathy among recent undergraduate nursing students over an academic year. These findings argue for nursing educators to re-examine solely mannequin-based methods of simulation and its impact on future generations of nurses potentially compromising the "art" of nursing. While many nursing schools have adopted simulation programs, a research deficiency exists to determine the impact on nursing students (Sanford, 2010).

Use of Standardized Patients in Healthcare Education

Standardized patients are individuals coached to consistently portray an actual

patient including history, physical findings, emotions, and feelings. The use of SPs to

teach and assess clinical and communication skills has been well documented and

researched in the medical literature (Norcini & Boulet, 2003). The reliability and validity

of SPs are well established and form the basis for licensure and certification of physicians

since 2004 (Boulet, 2008). Standardized patients can be used in two principle types of

simulation encounters. Teaching or formative cases are conducted to highlight specific

skills or topics and discuss the case while it is in progress. Formal testing or summative

cases are used to determine and document student competencies. After the SP encounter

(SPE), faculty will debrief the case, discuss appropriate interventions, treatments, and/or

proper skill techniques, and then give feedback to students on their performance. This

methodology allows students to gain confidence in patient care, communication, and

interventions, thus ensuring the students have mastered these skills competently prior to

entering the clinical settings. With multiple exposures to SPs, the faculty is also able to

validate the clinical proficiency and competencies of student groups to clinical agencies

to ensure patient safety of students practicing at all levels.

Medical school education is familiar with standardized patients and first used SPs

in 1963 by Dr. Howard Burrows in a neurology clerkship at the University of Southern

California to validate neurological assessment skills (Wallace, 1997). Barrows also used

SPs to record student performance, as more rigorous clinical skills evaluation tools were

needed at that time. Inspired by Barrows, obstetrician/gynecologist Robert Kretzschmar

developed the first gynecological teaching associates (GTA) at the University of Iowa in

1968 (Wallace, 1997). In 1976, pediatrician Paula Stillman used standardized patients to

teach interviewing skills to medical students using simulated mothers (Rep, 2012). She

also developed the first instrument to assess interviewing skill using a Likert scale. She

taught the simulated mothers to complete the instrument and give performance feedback

to the students. Stillman was also instrumental in teaching SPs to teach and evaluate

medical student doing a comprehensive history and head-to-toe physical exam (Wallace,

1997).

After several decades of piloting SP programs along with SP psychometric

research, the Medical Council of Canada (1993) required SPEs as part of the medical

licensure exam and the United States National Board of Medical Examiners (USMLE)

(1995) endorsed the use of SPEs to establish a performance-based, clinical-competency

examination as part of the USMLE Step II. As of 2005, all U.S. and foreign graduating

medical students are required to take this exam. Since 1963 when first initiated by

Howard Barrows, MD, at the University of Southern California Medical School,

standardized patient programs are now used internationally in most medical schools.

They have been so effective in medical student clinical performance evaluation that

statewide coalitions of medical schools have formed to write cases and formulate

consistent competencies for all medical schools and medical students. In comparison,

nursing schools and nurse practitioner (NP) programs have been slower to adopt SPEs for

student clinical evaluation. It is clear, however, that nursing educators are becoming

increasingly aware SPEs provide controlled situations in which students can be evaluated

at all levels of professional development and enhance objective evaluation of student clinical performance.

Use of Standardized Patients in Nursing Education

The use of SPs is slowly transitioning into pre-licensure and advanced practice nursing programs as a formative teaching methodology and for student evaluation (Vessey & Huss, 2002). The use of SPs and simulation brings life to the classroom and provides nursing students opportunities to practice skills without the possibility of harming an actual patient. These well-controlled laboratory scenarios build student confidence, enabling performance of these practiced skills in the hospital or primary care settings. The encounters are videotaped to allow the students to self-reflect on performance and review with faculty. By utilizing the SPEs, faculty members are able to effectively and consistently evaluate clinical performance of the entire class and revise didactic material as needed. Faculty can immediately debrief, provide coaching, and remediation to avoid potential patient safety issues, poor clinical competence, or improper technique. This patient-centered learning strategy prepares students to become safe, competent, and effective nurses.

A SP program provides a venue for faculty to teach and evaluate students in a consistent, convenient, and comfortable environment, where all students are exposed to the same case presentation. A laboratory setting allowing students to practice and perfect nursing skills provides faculty with a direct comparison of competence among students with an objective evaluation based on the same criteria for all students. The use of SPs also allows for the opportunity for students to receive direct and honest feedback from

their "patient" in real-time. The SP modality allows nursing faculty to develop cases based on real-life clinical experiences that vary in range of difficulty depending on the student's progression in the program. Analysis of clinical course group performance data with SPs provides essential feedback on the effectiveness of didactic content and aids in the development, review, and revision process of nursing curriculum.

Nursing students are videotaped during their interactions with the SPs, allowing faculty to coach, provide feedback, and assess the development of the student's critical thinking and diagnostic reasoning. The use of SPs provides students with the opportunity to apply newly acquired knowledge in a structured laboratory setting within a non-threatening, supportive environment. A standardized patient program also provides a venue to offer clinical experiences, which may not be available or difficult to evaluate in the clinical arena. With current difficulties placing students in the clinical arenas, simulation consistently provides deliberate exposure to common or high-risk situations student may never experience in a clinical site. It also ensures that by graduation all of our nursing students have encountered patients with a certain or range of diagnoses. Standardized patient encounters provide an opportunity for nursing faculty to observe all students delivering patient care to a SP using the same case or patient scenario in a consistent, safe, non-threatening, and convenient environment. This method of instruction and testing allows faculty to develop and introduce a multitude of conditions (cases) frequently seen in clinical settings, and provide time for students to practice with direct feedback or immediate remediation on performance.

Even though SP programs are used and well-researched in medical schools, the

majority of nursing schools have adopted high-fidelity mannequin simulation programs

(Sanford, 2010). Standardized patients contribute to increased realism by exposing

students to a "real" patient with opportunities to learn and practice human compassion

and empathetic communication skills and receive feedback from the patient on how to

fine-tune their bedside manner and connection with the patient. The "art" of nursing is

extremely complex and intricate in regards to providing care to patients. The capacity to

provide compassionate care is the heart and soul of nursing practice as identified by

American Association of Colleges of Nursing and the National League of Nursing

(Rhodes, Morris, & Lazenby, 2011). The ability to develop interpersonal skills through

exploring the ability to view and explore human potential as a healer along with

achievement of empowered caring is a necessary component of nursing education

(Rhodes et al., 2011). Rhodes et al. (2011) describes self-awareness as an important

factor in the development of confidence and competence in nursing students. Human

interaction, a core values in nursing, is experienced primarily with SP simulation

activities as compared to the technical, plastic "dummy" high-fidelity mannequins.

These human interaction opportunities are extremely valuable in teaching students to

develop compassion and empathy to promote healing.

> While working with the standardized patient, the student can experience and
>
> practice clinical medicine without jeopardizing the health or welfare of the real,
>
> sick patients. The value is in the experience of working with a patient. It takes the
>
> process of learning a step beyond the books and away from reliance on paper and

pencil tests. It puts the learning of medicine in the arena of veritable clinical practice-not virtual reality but veritable reality-as close to the truth of an authentic clinical encounter as one can get without actually being there, because there is a living, breathing, responding human being to encounter (Wallace, 1997, p. 6).

Lastly, SP activities can promote the moral knowledge in nursing as defined as the essential morality of knowing the difference between right and wrong in the context of providing care and promoting wellness. This includes ethical behaviors and professional responsibilities of continuous learning, an enduring calling of service to people, respecting life and dignity, and attitudes on relieving suffering and encouraging health (Carper, 1978).

In summary, SP programs have five key advantages for nursing students:

- Development of carefully written cases based on real patients providing opportunities for students to obtain, improve, and integrate teamwork, technology, evidence, knowledge, skills and attitudes into all age and cultural groups in non-threatening healthcare settings.

- Standardized patients are persons trained to portray real-life patient cases, stressing the importance of human interaction, compassion and empathy to prepare students to examine and interact in a professional, ethical, safe, consistent, and convenient manner beneficial to both the student and patient.

- Standardized Patients can evaluate and validate clinical competence and skill performance that ensures that all students meet a minimum level of competence of knowledge, skill, and attitude at each stage of their education.

- Videotaping of the student/SP encounter for evaluation and self-reflection on application of evidence, interpersonal communication, leadership, patient safety, and quality improvement skills.

- Instructional debriefing of the case as a group with direct SP feedback of overall student performance to foster confidence, compassionate bedside manner, appropriate treatment and collaborate ensuring safe patient care.

Nursing Competence Literature

The term nursing competence is commonly used in nursing however, Watson, et al. (2002) found the term to be quite nebulous lacking rigorous psychometric properties and measures. Benner's definition of competence is characterized by a feeling of mastery and the ability to plan patient care while managing the demands of clinical nursing (Benner, 2001). According to Benner (2001), a competent nurse or third stage of nursing, describes a nurse who generally has two to three years of experience. The American Association of Colleges of Nursing (2008) has outlined core competencies for baccalaureate pre-licensure and graduate level nursing. In 2010, the National League for Nursing (NLN) published a report and model to outline outcomes and competencies for graduates including all levels of nursing; vocational, diploma, associates degree, baccalaureate, masters, and practice and research doctorates (National League for Nursing, 2010). Interprofessional competencies have also recently been established to encourage and foster interdisciplinary healthcare education (Interprofessional Education Collaborative Expert Panel, 2011).

More recent work has been conducted to advocate for more holistic nursing approaches incorporating measures and tools to assess knowledge, skills, attitudes, and values of nursing students (Yanhua & Watson, 2011). Quality and Safety Education for Nurses (QSEN) along with the Institute of Medicine (IOM) (Greiner & Knebel, 2003) have established competency standards for pre-licensure and graduate nursing students to clarify the definition of competence in nursing (Cronenwett et al., 2007). Common competence attributes in several fields and various disciplines, including definitions, incorporated a cluster of three critical attributes identified as knowledge, skills and attitudes. The QSEN competencies provide a standard based approach and framework for all nursing educators by exploring these three overall encompassing attributes in relation to nursing competence (Cronenwett et al., 2007). The importance to learn and demonstrate safe performance of skills utilizing patient-centered and evidenced-based care, interprofessional communication, informatics, team work, collaboration, and quality improvement are all addressed within the QSEN, NCSBN, and IOM competencies (Nehring & Lashley, 2010).

Nursing competence is also being investigated extensively with multiple large cohort studies supported by the National Council of State Boards of Nursing (NCSBN). Watson et al. (2002) found nursing lacking rigorous psychometric properties to measure nursing competence and the majority of the research rigor measuring clinical competence has been done outside of nursing. Studies including instrument development using rigorous psychometric properties are needed within the nursing profession (Watson et al., 2002). Cronenwett et al. (2007) argue psychometric rigor is needed in relation to

measures of nursing competence and further longitudinal investigation is necessary with large numbers of nursing students (Cronenwett et al., 2007). The importance of this study is to strengthen and employ rigorous methods of evaluating pre-licensure nursing student competence.

Conceptual Framework

Personal Interest

Teaching efforts and responsibilities as a nurse educator are to ensure graduation of safe, competent student nurses through the use of various innovative methodologies in nursing education and competence evaluation. Research efforts investigating how students learn in classrooms and/or labs and transfer knowledge to apply safe effective care to patients at the bedside are fundamental for the development of competency in nursing aiming to bridge the theory-practice gap (Scully, 2011).

Theoretical/Educational Frameworks

Nurse educators and scientists have adapted and tested many various learning and teaching paradigms along with educational learning theories to simulation studies and research. The application of nursing grand and middle range theories and theories from varied disciplines such as educational theories pioneered by John Dewey's (Dewey, 1916), David Kolb's experiential learning theory (Kolb, 1981), and Malcolm Knowles adult learning theory (Knowles, 1975) guide nurse educators to facilitate teaching and learning methodologies and practices that will effectively guide students to provide safe, effective care to patients. A recent simulation model, developed by the NLN, guides nursing educators to thread experiential learning methods throughout curriculum

development, didactic content, simulation activities, and evaluation of clinical

competence and practice. This model however has not been tested with a standardized

patient simulation modality. Its only application has been with high-fidelity mannequins

in nursing.

Using the Dreyfus Model of Skill Acquisition, Benner (2001) identifies five

stages of nursing practice ranging from novice to expert. This continuum describes five

levels of nursing expertise including: Novice, Advanced beginner, Competent, Proficient,

and Expert. This theory incorporates 7-domains of practice contributing to the

identification of learning outcomes and competency evaluation tools based on the

student's previous knowledge and experience. These five levels of nursing expertise and

seven domains of practice underpin and guide nursing educators with curricular design

including simulation lab and clinical experiences. Benner's more recent work has been

focused on nursing's professional identity of caring. Benner postulates caring is required

for critical thinking to take place (Benner, Sutphen, Leonard, & Day, 2010). She argues

educational strategies fostering the understanding of caring are essential for nursing

programs (Rhodes et al., 2011). Additionally, Benner supports developing clinical

simulation activities intended to incorporate knowledge, skills, and ethical character

(Benner et al., 2010).

John Dewey (1916) first postulated experiential learning theory in the later part of

the Twentieth Century. Dewey founded the field of reflective learning proposing that

learning occurs where experience is fostered. He challenged educators to consider the

environment where teaching and learning activities occur (Dewey, 1916). According to

Kolb (1981), there are four stages reflected in a cycle that influence the experiential

learning process. These four elements are concrete experience, observation and

reflection, the formation of abstract concepts and testing in new situations. All of these

elements are reflected in simulation methods. Both Dewey and Kolb's experiential

models support the ability to fully involve students in simulation activities and labs, to

observe and reflect on performance, create concepts based on their observations, while

concepts transition to theories that influence decisions and the ability to solve problems.

This active-reflective model allows for cognitive growth and learning (Kolb, 1981).

Malcolm Knowles's (1975) adult learning theory is based on environments that support

immediate application of knowledge that draw on student experiences. His theory along

with Dewey and Kolb's support simulation methodologies by fostering structured

learning environments, identifying needs for learning through specific objectives,

designing and managing learning experiences, along with evaluating results. The works

of these early educational theorists inform the foundations of simulation and this study.

The NLN simulation model integrates teacher factors, student factors, educational

practices, design characteristics, and outcomes. According to Jeffries (2005), these

characteristics heighten optimal learning experiences for students and teachers resulting

in the development of necessary skills and knowledge needed for clinical practice

(Jeffries, 2005). The NLN mannequin-based simulation model because of its lack of

focus on caring (a core value of nursing) and application to standardized patient

methodology is not applicable to this study.

Because of the lack of humanistic core values in NLN's simulation model, this study will employ a conceptual framework informed by the works of early experiential educational theorists and Benner's Novice to Expert theoretical framework focusing on the key concept of measuring competence defined by the QSEN competencies. The empirical referents or measurement tools were developed based on established domains of the QSEN competencies for pre-licensure nursing students identified as knowledge, skills and attitudes. The QSEN competencies provide a standard based approach and framework for all nursing educators by exploring these three overall encompassing attributes in relation to nursing competence (Cronenwett et al., 2007).

State of the Science

Use of Standardized Patients to Complete Competency Checklists

Standardized patient methodology in medical education dates back to the early 1960's. As appreciated in nursing today, medical educators and innovators were searching for alternative methodologies to evaluate clinical competence of medical students (Wallace, 1997). Two large instrumental research studies focusing on psychometric properties provided legitimacy to using SP methodology as a basis for clinical performance evaluations (Wallace, 1997). The first study in 1990 by van der Vleuten and Swanson, "Assessment of Clinical Skills with Standardized Patients: State of the Art", provides a comprehensive meta-analysis of psychometric properties of SP performance-based tests (van der Vleuten & Swanson, 1990). This review of the SP literature focused on three main psychometric properties including reproducibility of scores, validity of scores, and the scholastic impact of SP exams (van der Vleuten &

Swanson, 1990). After reviewing 16 studies over a decade focusing on faculty and SP

raters, van der Vleuten and Swanson (1990) found inter-rater reliability acceptable when

checklist training occurred regardless of academic background. They also found test

scores were reproducible if multiple stations were included. The review also suggested

the validation studies were promising, although not highly informative. The authors

concluded further research was recommended to investigate validity incorporating station

organization, scheduling, and SP partiality along with application of generalizability

theory (van der Vleuten & Swanson, 1990).

An ensuing, expanded study in 1994 by Vu and Barrows, "Use of Standardized

Patients in Clinical Assessments: Recent Developments and Measure Findings", also

contemplates the complexity, lack of standardization, and subjectivity of clinical

evaluations of medical students (Vu & Barrows, 1994). Due to the widespread adoption

of SP methodology across medicals schools utilizing the same SP case scenarios, it was

shown that variances in SP based performance scores were not significant across testing

sites (Vu & Barrows, 1994). Well-trained SPs were shown to portray cases and complete

competency checklists consistently and accurately across medical schools and testing

centers (Vu & Barrows, 1994). Decisions to rely on SPs to complete checklist were

motivated by the financial costs of using faculty to evaluate medical students (Vu &

Barrows, 1994).

Multiple early medical education studies also found faculty raters and well-trained

SP raters to have good inter-rater reliability ranging from .68 to .79 (Newble & Swanson,

1988; van der Vleuten, van Luijk & Swanson, 1988), comparable accuracy in completing

checklists (Elliott & Hickman, 1987; van der Vleuten, van Luijk, Ballegooijen, &

Swanson, 1989), and high percentage of agreement between faculty and SP raters ranging

from 80% to 100% (Norman, Mussin, Williams, & Swanson, 1985; Rethans & van

Boven, 1987; Tamblyn, Klass, Schnabl, & Kopelow, 1991b; Vu et al., 1992; Williams et

al., 1987) (Vu & Barrows, 1994).

These efforts over several decades resulted in not only using this methodology to

evaluate medical student performance but became the foundation of establishing a

performance based competence examination for the Medical Council of Canada, USMLE

and the Educational Commission for Foreign Medical Graduates (Becker et al., 2006).

Gaps in the Literature

As seen in medical education decades ago, nursing is now facing the same

monumental issues in evaluating and ensuring clinical competence of nursing students

and practicing nurses at all levels. Nursing is now at the forefront for radical changes in

nursing education due to the complexity of the clinical environment and immense public

and agency scrutiny and demand for safe patient care and protecting patients from error

and system breakdown.

The recent review of the nursing clinical competence literature by Yanhua &

Watson, 2011, states "the definition of competence lacks consensus, remains obscure and

contradictory" (Yanhua & Watson, 2011, p. 832). Efforts to assess clinical competence

are challenging due to lack of nursing consensus on what to assess and subjective

assessment methods lacking reliability and validity (Watson et al., 2002). Studies

including instrument development using rigorous psychometric properties are needed

within the nursing profession (Watson et al., 2002). Cronenwett et al. (2007) argue psychometric rigor is needed in relation to measures of nursing competence and further longitudinal investigation is necessary with large numbers of nursing students (Cronenwett et al., 2007). The majority of the research rigor measuring clinical competence has been done outside of nursing.

Traditional faculty evaluations of clinical performance in the actual practice environment have been described as subjective, logistically challenging, and time consuming (Gibbons et al., 2002). Factors such as variability of clinical settings, differing levels of patient acuity, and faculty biases limit confidence in clinical performance assessments (Gibbons et al., 2002). Lack of quality assessment instruments and problems with faculty consensus further confound evaluations of students in the real world (Gallagher et al., 2012). With the current nursing shortage, faculty are at a premium especially in light of the added rigor that is becoming increasingly expected of pre-licensure programs.

The use of SPs to teach and assess clinical and communication skills has been well documented and researched in the medical literature (Norcini & Boulet, 2003). The reliability and validity of SPs are well established and form the basis for licensure and certification of physicians since 2004 (Boulet, 2008), however their use in nursing is limited and lacks psychometric rigor (Watson et al., 2002). Nursing should consider adopting evidence-based evaluation methods instituted by our medical education colleagues who have contributed decades of psychometric research and adopted rigorous methods of evaluation in education, licensure, and certification.

Chapter 3

Methodology

The purpose of this project was to identify if SPs can consistently and accurately complete competency checklists similar to nursing faculty. This chapter includes a description of the design, sample, sampling, data collection and analytic procedures. The protection of human subjects is also presented.

Specific Aims

Examine the interrater reliability and percent agreement of standardized patients and faculty competency checklist scores when evaluating pre-licensure nursing students.

Design and Methods

A quantitative, descriptive design using dichotomous checklist data was used to identify similarities and/or variances between SP and faculty raters when evaluating a cohort of pre-licensure nursing student competencies. Based on the review of medical education literature, standardized patients can consistently and accurately complete competency checklists however, despite showing great promise, nursing lacks rigorous approaches in the development and testing of evaluation instruments (Miller, Wilbur, Montgomery, & Talashek, 1998).

32

Sample and Sampling

For this study, three consecutive, summative standardized patient cases were

developed and integrated into the curriculum for pre-licensure nursing students over two

semesters. A convenient sample included a cohort of 50 pre-licensure students

commencing a pre-licensure nursing program in August 2012. The sample consisted of

12 males and 38 females. One female student dropped out of the program after the first

case scenario therefore was removed from the study leaving a sample size of 49 (12

males and 37 females). As customary within the SP program, all students and SPs signed

consents to be videotaped and agreed that their videotape could be used for teaching and

research purposes.

Operational Definitions

Standardized patient (SP) rater- Standardized patients are individuals (actors or

laypersons) coached to consistently portray an actual patient including history, physical

findings, emotions, and feelings.

Faculty Rater – Clinical faculty are responsible to teach and evaluate student

competency in the simulation laboratory setting.

Competency checklist – Pre-determined checklists to evaluate competency of

nursing student skills.

Data Collection Instruments/Measures

Data for this study were obtained from three SP cases over a summer and fall

semester in a simulation laboratory. The progression of these three cases represents basic

nursing skills of pre-licensure nursing students in a health assessment and basic nursing

fundamentals course. The cases were designed to assess the clinical skills (history taking and physical exam skills along with basic nursing interventions) of beginning pre-licensure nursing students about to enter acute-care clinical sites. Competency checklists for each case were developed for each scenario by 6 expert clinical faculty members and were based on those skills that pre-licensure students would be expected to identify or perform at their academic level (see Figure 2, 3, and 4). All student participants received didactic lectures and lab practice on each of the skills included in the checklists. According to Norcini and Boulet (2003), SPs can effectively complete dichotomous competency checklists. The skills evaluated in these cases were assessed using dichotomously "yes" or "no" scored checklists.

In order to establish content validity, checklists were developed by a team of 6 "experts" or clinical faculty members to collectively ensure items were covered during the course and included in the overall course objectives. The team of clinical faculty were given the conceptual definitions of each item and asked to retain items based on relevance to practice, the course objectives, material covered in class and/or lab activities, and academic level of student. Wording of each item including descriptors of expected participant behaviors were reviewed for each item for each case. Based on feedback of the team, items were eliminated based on relevance or modified to clarify wording. The clinical faculty who were involved in case and checklist development also served as the faculty raters for each case they were involved with. The team of clinical faculty set the pass score at 80% or above and failures were set at 79% or below in concordance with the class syllabus's minimum grade requirement for all clinical experiences.

Standardized patients received 12 hours of training including the case and checklist. All SPs are trained according to a case template, script, and checklist. SP training focused on patient portrayal of the case including observing and recording skills on a checklist. SPs entered the data and scored checklists immediately following each clinical encounter. Faculty raters scored the checklists on a computer while watching the encounter live, remotely via cameras in the exam rooms.

Students were individually videotaped as they completed the 30 minutes cases. Student competency checklists were completed by one SP and one faculty rater using a checklist of dichotomous (yes or no) responses for history-taking, physical exam, and appropriate nursing interventions. Each student received a total score by one standardized patient and one faculty.

Table 1 outlines the cases in order of succession over the two semesters. The first case consisted of 6 SP raters, 5 faculty, 17 checklist items, and 49 students. The second case involved 5 SP raters, 7 faculty, 19 checklist items, and 49 students. The third case consisted of 3 SP raters, 5 faculty raters, 17 checklist items, and 49 students.

Table 1

Case Details

Case	Date	SP Raters	Faculty Raters	Number of Checklist Items	Students
1	8/13/12	6	5	17	49
2	8/23-24/12	5	7	19	49
3	10/24-26/12	3	5	17	49

Figure 2. Case 1 Checklist

1. Disinfect hands before AND after patient contact	( ) Yes	( ) No
2. Student identified patient by asking name AND checking arm bracelet	( ) Yes	( ) No
3. History - Student asked patient the reason for seeking care	( ) Yes	( ) No
4. Student asked the following questions about primary concern: Character	( ) Yes	( ) No
5. Student asked the following questions about primary concern: Onset	( ) Yes	( ) No
6. Student asked the following questions about primary concern: Location	( ) Yes	( ) No
7. Student asked the following questions about primary concern: Duration	( ) Yes	( ) No
8. Student asked the following questions about primary concern: Severity	( ) Yes	( ) No
9. Student asked the following questions about primary concern: Pattern	( )Yes	( ) No
10. Student asked the following questions about primary concern: Associated Factors	( ) Yes	( ) No
11. Student asked question about past health history	( ) Yes	( ) No
12. Student asked question about family history	( ) Yes	( ) No
13. Student asked question about social history	( ) Yes	( ) No
14. Student assessed vital signs: Temperature	( ) Yes	( ) No
15. Student assessed vital signs: Blood Pressure	( ) Yes	( ) No
16. Student assessed vital signs: Apical OR Radial pulse (minimum 30 seconds)	( ) Yes	( ) No
17. Student assessed vital signs: Respiratory rate (minimum 30 seconds)	( ) Yes	( ) No

Figure 3. Case 2 Checklist

1. Identify patient by asking patient name AND D.O.B. AND checking armband	( ) Yes	( ) No
2. Perform hand hygiene before touching patient	( ) Yes	( ) No
3. Assess level of orientation. Asked at least THREE questions to assess for orientation (person, place, time, or situation)	( ) Yes	( ) No
4. Perform pupil checks bilaterally with penlight for construction	( ) Yes	( ) No
5. Oral exam - inspection of mouth (must be verbalized)	( ) Yes	( ) No
6. Inspect chest (must be verbalized)	( ) Yes	( ) No
7. Auscultate heart at 4 places with diaphragm AND bell under the gown	( ) Yes	( ) No
8. Auscultate the lungs in 2 places on anterior chest under the gown	( ) Yes	( ) No
9. Auscultate the lungs in 5 places on posterior chest under the gown	( ) Yes	( ) No
10. Inspect abdomen (must be verbalized)	( ) Yes	( ) No
11. Auscultate abdomen for bowel sounds in all four quadrants	( ) Yes	( ) No
12. Palpate abdomen for tenderness OR masses	( ) Yes	( ) No
13. Inspect arms AND legs for skin & musculoskelature (must be verbalized)	( ) Yes	( ) No
14. Palpate for capillary refill in bilateral feet	( ) Yes	( ) No
15. Assess grip strength in hands by having patient squeeze examiner's fingers bilaterally	( ) Yes	( ) No
16. Palpate upper and lower extremities bilaterally (must be verbalized)	( ) Yes	( ) No
17. Palpate strength of wrist pulses checked bilaterally	( ) Yes	( ) No
18. Palpate strength of feet pulses checked bilaterally	( ) Yes	( ) No
19. Throughout other parts of exam, inspect condition of skin (inspection of skin must be verbalized)	( ) Yes	( ) No

Figure 4. Case 3 Checklist

1. Introduces self	( ) Yes	( ) No
2. Washes hands when entering room	( ) Yes	( ) No
3. Asks patient to rate pain by using pain scale (1-10)	( ) Yes	( ) No
4. Asks patient to describe right ankle pain (dull, sharp, constant)	( ) Yes	( ) No
5. Assists patient back to bed	( ) Yes	( ) No
6. Elevates right ankle	( ) Yes	( ) No
7. Applies ice to right ankle	( ) Yes	( ) No
8. Checks MAR for pain medication orders	( ) Yes	( ) No
9. Check medication label with MAR from medication drawer. Check medication with MAR second time before coming to bedside. Check medication with MAR at bedside. (Must do all 3 checks to get "YES")	( ) Yes	( ) No
10. Checks arm band prior to giving medication.	( ) Yes	( ) No
11. Gives patient Vicodin after completing the 3 checks above	( ) Yes	( ) No
12. Applies gloves before giving Lovenox (shot)	( ) Yes	( ) No
13. Gives Lovenox after performing three checks	( ) Yes	( ) No
14. Student does not give Lipitor	( ) Yes	( ) No
15. Instructs patient about medication	( ) Yes	( ) No
16. Elevates side rails and places call bell within reach before leaving room (need to do both for a Yes)	( ) Yes	( ) No
17. Clears objects off floor before leaving room	( ) Yes	( ) No

Data Collection Procedures

Case Development. Over the last 4 years, SPs have been used as an adjunct to simulation and skills laboratory educational experiences for pre-licensure nursing students at this academic institution. SPs have been used and trained along with faculty to complete competency checklists. Each SP case was developed from an actual patient case encountered by clinical faculty his/her clinical practice. Anonymity was maintained by blinding the case to patient name, age, date, location of service, and provider. Each case involved a SP with a commonly encountered clinical scenario. All cases were reviewed and endorsed by 6 expert clinical faculty involved in the pre-licensure program and currently in active practice as acute-care nurses. The objectives of each case were congruent with the course syllabus. The complexity of the cases reflected progression of the students along the curriculum continuum utilizing Benner's Novice to Expert framework (Benner, 2001). SP cases were chosen and based on content presented in a didactic lectures including reading materials before each SP case testing event.

Data Management and Analysis

Statistical Analysis. Descriptive statistics will be used to determine the mean participant age and range along numbers and percentage of gender reflected in the sample. To assess for internal consistency, Kuder-Richardson #20 (K-R 20) will be computed for faculty and SP scores on dichotomous (yes/no) case items. Percent of agreement on overall scores for the SP and faculty were obtained by calculating individual item mean percent agreement, adding mean percent for each case, and dividing by the number of case items to determine the mean percent agreement for each case. An

interrater reliability analysis using Cohen's Kappa statistic (Cohen, 1960) will be used to calculate measure agreement between faculty and SP raters and for checklist items beyond that expected by chance alone, with a level of .60 to .75 representing an acceptable level of agreement (Sim & Wright, 2005). The kappa statistic will also be calculated for individual dichotomous (yes/no) case items and an overall mean kappa will be calculated for each case. Data analysis was performed using SPSS 20.0.

Human Subjects Protection

Protection for Human Subjects was obtained through the Institutional Review Board per the protocols of the University of San Diego's institutional review board committee. There were no specific risks or benefits for the student nurse participants in the study. All student nurse performance checklists along with faculty and SP rater data were de-identified to ensure participant confidentiality.

Chapter 4

Findings

The purpose of this project was to examine the interrater reliability and percent agreement of standardized patients and faculty competency checklist scores when evaluating pre-licensure nursing students. The importance of this study was to strengthen and employ rigorous methods of evaluating pre-licensure nursing student competence. This chapter presents the findings and an interpretive discussion of the study results.

Overview

Primary data analysis of three consecutive, quantitative checklists from August, 2012 through October, 2012 was conducted to determine interrater reliability and percent agreement of standardized patients (SP) and faculty competency checklist scores when evaluating a cohort of 49 pre-licensure nursing students. One SP and one faculty rater completed dichotomous (yes/no) competency checklists for each student. In the case of pass/fail disagreement between faculty and SP, a second clinical faculty reviewed the videotape of the participant in question and completed the same checklist.

Descriptive statistics were calculated for age and gender of the participants. Kuder-Richardson formula (K-R20) was calculated for each checklist to determine internal consistency with a range of 0.00 to 1.00. Scores closer to 1.0 represent greater

internal consistency. Percent agreement was calculated by counting the number of

checklist items in which both the SP and faculty agreed and dividing by the total number

of checklist items to determine the percent agreement for each participant for each case.

Once percent agreement was calculated for each participant, an overall mean percent

agreement was calculated for each case. Percent agreement was also calculated for each

checklist item by counting the items in agreement by both the faculty and SP raters

divided by the total number of participants. Cohen's Kappa statistic is an index of

interrater reliability used to measure the level of agreement between raters for

dichotomous or scale ratings correcting for chance agreement (Cohen, 1960). Table 2

outlines the interpretation of the kappa statistic by Landis & Koch (1977). The kappa

statistic was calculated for dichotomous (yes/no) case items and for each case. Data

analysis was performed using SPSS 20.0.

Table 2

Kappa Statistic Interpretation*

Kappa	Interpretation
<0	Poor agreement
0.0-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

*Landis & Koch, 1977

Results

Age and Gender of Student Participants

The sample (n=49) consisted of twelve male (24.48%) and 37 female (75.51%) students who were evaluated by the faculty and SP raters. The mean age for the sample was 28.62 years.

Case Analysis

Faculty and SP raters agreed on a 100% pass rate for case 1 and 2. Checklists completed by the faculty and SP raters identified 4 failures (scores lower than 80%) for case 3. Faculty and SP raters agreed on 2 of the failures and disagreed on 2 failures. A second clinical faculty member reviewed the videotapes of the 2 failures with SP/Faculty disagreement. The second faculty reviewer agreed with the SP rater to pass or fail both of the participants and disagreed with the first faculty rater scores. Table 2 includes mean checklist scores, standard deviation (SD), checklist score ranges, percent score range, and pass/fail numbers.

Internal Consistency

The K-R 20 calculated for internal consistency of the checklists ranged from -0.017 to 0.405. Breakdown of the SP and faculty K-R20 statistic for each case is shown in Table 3.

Table 3

Case Analysis

Case (# checklist items) n=# of case participants	Rater	K-R20	Mean Raw Checklist Score	SD	Checklist Raw Score Range	Percent Score Range	Pass/ Fail
1 (17) n=49	SP	0.193	16.6	0.63	15-17	88-100%	49/0
	Fac	0.405	16.5	0.81	14-17	82-100%	49/0
2 (19) n=49	SP	-0.017	18.6	0.56	17-19	89-100%	49/0
	Fac	0.186	18.3	0.83	16-19	84-100%	49/0
3 (17) n=49	SP	0.302	15.3	1.3	12-17	70-100%	46/3
	Fac	0.254	15.4	1.2	11-17	64-100%	46/3

Percent Agreement

Table 4 reflects the SP/faculty rater percent agreement for all checklist items per case, and mean percentage agreement for checklist items for each case. SP and faculty percent agreement of each checklist items for all cases ranged from 76-100%. Analysis of percent agreement for individual checklist items is shown in Tables 6, 7, and 8.

Table 4

SP/Faculty Rater Percent Agreement for Case Checklists

Case	Percent Agreement range for all checklist items	Mean Percent Agreement
1	76-100%	98%
2	89-100%	97%
3	82-100%	94%

Interrater Reliability

The interrater reliability for the raters was found to be kappa = 0.515 for case 1,

0.369 for case 2, and 0.685 for case 3 as shown in Table 5. The calculated kappa along

with percent agreement per checklist item is listed in Tables 6, 7, and 8. The kappa

statistic could not be calculated on 31 of the 53 checklist items marked as "*" on the

tables as the SP or faculty ratings (or both) were a constant (e.g., all SP ratings for a

given checklist item were recorded as "yes" (therefore, a constant). Across all the cases,

the kappa statistic for individual checklist items ranged from -0.028 to 1.

Table 5

Cohen's Kappa per Case

Case	Cohen's Kappa
1	0.515
2	0.369
3	0.685

Table 6

Case 1 Checklist Item Analyses

Item	Kappa	SP/Faculty Percent Agreement
1. Disinfect hands before AND after patient contact	0.637	94%
2. Student identified patient by asking name AND checking arm bracelet	*	100%
3. History - Student asked patient the reason for seeking care	*	100%
4. Student asked the following questions about primary concern: Character	-0.028	94%
5. Student asked the following questions about primary concern: Onset	*	100%
6. Student asked the following questions about primary concern: Location	0.728	96%
7. Student asked the following questions about primary concern: Duration	*	98%
8. Student asked the following questions about primary concern: Severity	*	100%
9. Student asked the following questions about primary concern: Pattern	0.479	96%
10. Student asked the following questions about primary concern: Associated Factors	0.539	94%
11. Student asked question about past health history	*	100%
12. Student asked question about family history	*	100%
13. Student asked question about social history	*	100%
14. Student assessed vital signs: Temperature	*	98%
15. Student assessed vital signs: Blood Pressure	0.657	98%
16. Student assessed vital signs: Apical OR Radial pulse (minimum 30 seconds)	*	100%
17. Student assessed vital signs: Respiratory rate (minimum 30 seconds)	*	96%

*Kappa statistic could not be calculated as one or both of the ratings (SP, faculty) were a constant.

Table 7

Case 2 Checklist Item Analyses

Item	Kappa	SP/Faculty Percent Agreement
1. Identify patient by asking patient name AND D.O.B. AND checking armband	*	98%
2. Perform hand hygiene before touching patient	*	100%
3. Assess level of orientation. Asked at least THREE questions to assess for orientation (person, place, time, or situation)	0.645	96%
4. Perform pupil checks bilaterally with penlight for construction	1	100%
5. Oral exam - inspection of mouth (must be verbalized)	*	98%
6. Inspect chest (must be verbalized)	0.260	90%
7. Auscultate heart at 4 places with diaphragm AND bell under the gown	1	100%
8. Auscultate the lungs in 2 places on anterior chest under the gown	*	100%
9. Auscultate the lungs in 5 places on posterior chest under the gown	*	98%
10. Inspect abdomen (must be verbalized)	*	94%
11. Auscultate abdomen for bowel sounds in all four quadrants	*	100%
12. Palpate abdomen for tenderness OR masses	*	100%
13. Inspect arms AND legs for skin & musculoskelature (must be verbalized)	*	100%
14. Palpate for capillary refill in bilateral feet	-0.065	88%
15. Assess grip strength in hands by having patient squeeze examiner's fingers bilaterally	0.460	86%
16. Palpate upper and lower extremities bilaterally (must be verbalized)	*	96%
17. Palpate strength of wrist pulses checked bilaterally	*	98%
18. Palpate strength of feet pulses checked bilaterally	*	100%
19. Throughout other parts of exam, inspect condition of skin (inspection of skin must be verbalized)	*	100%

*Kappa statistic could not be calculated as one or both of the ratings (SP, faculty) were a   constant.

Table 8

Case 3 Checklist Item Analyses

Item	Kappa	SP/Faculty Percent Agreement
1. Introduces self	*	98%
2. Washes hands when entering room	0.078	76%
3. Asks patient to rate pain by using pain scale (1-10)	*	100%
4. Asks patient to describe right ankle pain (dull, sharp, constant)	0.747	88%
5. Assists patient back to bed	*	100%
6. Elevates right ankle	*	98%
7. Applies ice to right ankle	0.657	98%
8. Checks MAR for pain medication orders	*	98%
9. Check medication label with MAR from medication drawer. Check medication with MAR second time before coming to bedside. Check medication with MAR at bedside. (Must do all 3 checks to get "YES")	-0.058	88%
10. Checks arm band prior to giving medication.	0.503	82%
11. Gives patient Vicodin after completing the 3 checks above	0.647	96%
12. Applies gloves before giving Lovenox (shot)	1	100%
13. Gives Lovenox after performing three checks	0.539	94%
14. Student does not give Lipitor	*	100%
15. Instructs patient about medication	*	98%
16. Elevates side rails and places call bell within reach before leaving room (need to do both for a Yes)	0.833	96%
17. Clears objects off floor before leaving room	0.898	98%

*Kappa statistic could not be calculated as one or both of the ratings (SP, faculty) were a constant.

Discussion

Case Analysis

Overall participant scores for the three cases were encouraging identifying only 4

(3%) failures in case 3 from 147 participant/SP interactions. Mean raw scores and overall

percent scores pertaining to evaluation of student performance for each case were also

high. Case 1 was very basic, assessing patient safety measures, history taking skills, and

vital signs. Case 2 primarily assessed physical examination techniques. Case 3 was more

complicated, combining multiple tasks including patient safety measures, history-taking

skills for pain assessment, and administration of oral and injectable subcutaneous

medications. All cases were representative of the skills these participants will be required

to perform in acute-care clinical sites during the first semester in the program.

Case 1. Mean raw scores for case 1 were high and consistent for both raters: 16.6

for the SP rater and 16.5 for the faculty rater out of 17 items. The SP and faculty score

percentages ranged between 82-100% with all participants passing with a score $\geq$ 80%.

This was the cohort's first experience with a summative SP event assessing very basic

assessment skills of new pre-licensure nursing students in a beginning health assessment

course. The main objective of this case was for the participants to perform basic safety

and history-taking skills, along with assessing vitals signs. Participants had learned and

practiced all of these skills in laboratory sessions prior to administering case 1. After

reviewing performance from the videos, SP and faculty raters felt the participants needed

additional assistance and review of taking blood pressures. Even though all but one

participant was able to complete the blood pressure, SP and faculty raters felt students

fumbled with the blood equipment and needed a review. An advantage of this scenario, not rated on the checklist, was the ability of the participant to interact with the SP. Several SP and faculty members entered remarks on the checklists commenting on the participant's interaction and connection with the patient. This non-rated attribute is essential to assess the participant's ability to connect with the patient. At this very early stage of new student development, SPs and nursing educators can use these encounters to coach and give feedback to participants on interpersonal and communication skills.

Case 2. Mean raw scores for case 2 were also high and consistent for both raters: 18.6 for the SP rater and 18.3 for the faculty rater out of 19 items. The SP and faculty score percentages were between 84-100% with all participants passing with a score $\geq$ 80%. This was the cohort's second experience with a summative SP event assessing very basic physical examination skills of new pre-licensure nursing students in a beginning health assessment course. The main objective of this case was for the participants to perform an overall physical examination mirroring an initial shift assessment performed on patients during upcoming clinical rotations. These physical examination techniques were taught, demonstrated, and practiced in the simulation lab prior to testing. Participants were asked to perform an overall physical exam including examination of the mental status, eyes, mouth, lungs, heart, abdomen, extremities, and skin. The participants were taught to listen to lung and heart sounds under the gown. Several participants did not receive credit for the lung and heart exam, even though they did the skill, as they listened over the gown. Several of the participants commented they felt uncomfortable asking the SPs to remove their gown potentially exposing sensitive body areas especially

with female SPs. Participants felt as though they were imposing on "the SP's personal space". These types of scenarios provide the students opportunities to talk to patients and communicate appropriately using sensitivity while developing a "safe and trusting" rapport to expose and examine private body areas. This vital communication opportunity, non-scored for this study, is also another benefit for the students to have an actual human in the scenario. If this scenario only used mannequins, this student observation and reflection on practice may have not been realized. When looking at performance of skills in a simulation lab, nursing educators must also evaluate how the task performances might be transferred to the bedside (Scully, 2011). If we had not evaluated this task with an actual human, this "sensitive" issue may have not been addressed leaving the student unprepared on how to develop that "safe and trusting" rapport to examine private body parts of patients in the actual clinical setting.

Faculty raters did comment it was difficult to assess, because of camera angles, if participants palpated pulses, performed capillary refills, and/or did grip strengths. They also found it difficult to assess inspection of a body part if the participant did not verbalize they were doing an inspection.

Case 3. Mean raw scores for case 3 were also promising and consistent for both raters: 15.3 for the SP rater and 15.4 for the faculty rater out of 17 items. The SP and faculty rater score percentages were between 64-100% for all participants with 46 (97%) passing with a score $\geq$ 80% and 4 (3%) failures below the 80%. This was the cohort's third experience with a summative SP event assessing basic fundamental skills of new pre-licensure nursing students in a beginning fundamentals course. The main objective

of this case was to assess fundamental skills the students would soon be performing in the clinical sites. This case was more complicated, combining multiple tasks including patient safety measures, history-taking skills for pain assessment, and administration of oral and injectable subcutaneous medications. Case 3 included all aspects of the prior two cases adding on additional fundamental nursing skills. All participants were taught and practiced all of these skills in laboratory sessions prior to administering case 3. SP and faculty raters had difficulty with the complexity of the checklist items. Three of the items included assessing "three checks" of medication administration. These items, requiring the participants to perform three skills, caused ambiguity and confusion over what the three checks would be. Better wording, along with rater training and agreement on successful performance, would improve these items.

Internal Consistency

The KR 20 calculated for internal consistency of the checklists ranged from -0.017 to .405. These low internal consistency measures can be attributed to the multiple constructs being measured in the checklists instead of all the checklist items measuring one construct. KR20 or internal consistency may have computed higher if checklist items measured only one construct such as a pain assessment. The checklist items for this study measured multiple constructs such as safety, history-taking skills, physical assessment, and appropriate interventions. "Internal consistency reliability approach can be applied to assess the reliability of a summated scale where several items are summed to form a total score. Each item can be considered to measure some aspect of the construct and the items should be consistent in what they indicated about the construct" (Peter & Peter, 2008, p.

991). Other approaches to reliability such as test-retest or alternate-forms reliability may contribute to improved assessment of reliability measurements (Peter & Peter, 2008). Also, "more careful selection of stations [SP encounters] based on their psychometric characteristics can improve reliability" (Newble & Swanson, 1988, p. 325).

Low internal consistency, also seen by Newble and Swanson (1988) in early SP medical research, was due to the amount of time participants spent in multiple station SP encounters. Ideally, for satisfactory internal consistency to occur, test length should last at least 4 hours including 5-7 SP encounters. Newble and Swanson (1988) also reported an "accurate estimation of reliability coefficients requires large sample sizes" (Newble & Swanson, 1988, p. 328). These early psychometric studies included multi-sites, multi-SP encounters, and over 500 students. According to Newble & Swanson (1988), determining clinical competence of students with reliable assessment tools requires multiple methods of evaluation including multi-station SP encounters and a written test over at least a 4 hour time period.

Percent Agreement

The percent agreement between SP and faculty raters on case checklists along with agreement on checklist items in the study is promising showing an overall SP/Faculty mean case percent agreement of 94-98%. Early psychometric measurements in medical education, reported 82% agreement between raters in recording medical student examinee actions (Tamblyn, Klass, Schnabl, & Kopelow, 1991b). De Champlain, Margus, King, and Klass (1997), using the same methodology with medical students as

was employed in this study, found average agreement between SPs and faculty to be 88% to 92%, similar to the findings reported in here.

Interrater reliability

Cohen's Kappa statistic is a more robust measure of interrater reliability taking into account agreement occurring by chance and represents a measure of "true" agreement (Cohen, 1960). A value of +1 implies perfect agreement between two raters, while -1 implies perfect disagreement (Sim & Wright, 2005). If the kappa assumes the value of 0, then this implies that there is no relationship between the ratings of the two raters and any agreement or disagreement is due to chance alone (Sim & Wright, 2005). According to Landis & Koch (1977), the kappa statistic for the three cases in this study is interpreted as 0.515 for case 1 (moderate agreement), 0.369, (fair agreement) for case 2, and 0.685 (substantial agreement) for case 3. A kappa statistic was also calculated for each checklist item. Over all three cases, the kappa statistic could not be calculated for 31 of the total 53 checklist items as one or both of the ratings were a constant. Many of these items, where the kappa could not be calculated, showed high percent agreement between the two raters.

Case 1 shows a range of kappa from -0.028 to 0.728 for individual checklist items. Only one item showed very low (negative) agreement when participants asked their patient to described pain character. The low agreement may have resulted from the interpretation of character of pain by faculty and SP raters. Interesting enough, there was 94% agreement between faculty and SP on this item. Pain character can be subjective in interpretation by either rater as it is describing the character of the pain including

throbbing, stabbing, dull, and/or aching. If those specific words were not used during the scenario the raters might have not agreed on the interpretation as pain character. Improved rater training for interpretation may improve the kappa on this item if used in future studies when assessing pain. The remainder of the other checklist kappas for case 1 is interpreted from Landis & Koch (1977), as fair to substantial agreement.

Case 2 checklist item kappas range from -0.065 to 1. Only one item showed very low (negative) agreement when participants were assessed on palpating capillary refill of both feet. The SP/faculty percent agreement was also low at 88%. This skill was difficult for faculty raters to see due to camera angles. It was also difficult for item interpretation as some faculty raters felt the participants should be testing capillary refill of all toes on both feet however the SP raters were taught that if the participants performed it on one toe on both feet the participant received a "yes". Some participants checked capillary refill on the fingers causing increased confusion among the raters. The remainder of the other checklist kappas for case 2 is interpreted from Landis & Koch (1977), as fair to perfect agreement.

Case 3 checklist item kappas are overall more promising despite the complexity of the case. The overall case kappa is also stronger than the other two cases showing improvements with checklist development over time. Improvements in SP and faculty training and better interpretation of checklist items may account for the higher kappa. The 3 SPs in case 3 are very experienced and have worked in the SP program for over 4 years. Future studies accounting for years of experience as a SP or faculty rater may provide insight into the interpretation of checklist items potentially improving kappa

scores. The raters also agreed on variances in performance, strengthening internal consistency measurements. The lowest (negative) kappa was seen on a medication administration item where the participant was required to perform 3 checks with the medication administration record prior to giving medication. As described in the case 3 analyses, these items requiring multiple steps tend to cause confusion for both the faculty and SP raters. It was difficult for the faculty and SP raters to see the participants doing these "checks" as the medication drawer where the participants were required to go was not in camera or SP view. These items, with multiple steps, cause ambiguity and will need revision when used in SP events in the future. Only two checklist items were assessed on all three cases. The first checklist item on all three cases assessed for hand washing but interpretation of the item varied over the three cases. Case 3 required hand washing when entering the room, case 2 before touching the patient, and case 1 before and after touching the patient. The second checklist item assessed on all three cases was checking the patient's armband. This item also varied on each case. To receive a "yes" answer to this item in case 1, the participant was also required to verbally ask for the patient's name. Case 2 required asking the name, verifying the birthday, and the armband. Case 3 required the participant to only check the armband before giving medications but was confusing as two medications were administered. If participants only checked the armband before giving one of the medications and not both, raters can be confused on how to rate this item resulting in poor agreement. Faculty must agree on the timing of hand washing and checking identification for each case and remain consistent so participants are not confused to what is appropriate and be realistic in terms of actual

practice. In case 3, participants entered the scenario finding their patient in pain and in an unsafe disposition. Was it realistic to think the student would wash their hands before tending to the patient safety issue first? Most participants in case 3 tended to the safety issue first then washed hands causing confusion among the SP and faculty raters. These are case design issues requiring faculty agreement on appropriate care. Faculty must agree and determine the appropriate care first before SP training. The remainder of the other checklist kappas for case 3 is interpreted from Landis & Koch (1977), as moderate to perfect agreement.

Summary of Results

Much has been learned about the psychometric measures needed to ensure reliable, valid methods of evaluating nurse competence. This study is the first attempt to look at SP and faculty raters in evaluating pre-licensure nursing students. Percent agreement statistics met or exceeded the percent agreement rates seen in medical education decades ago. Low internal consistency measures and moderate kappa scores suggest additional research is needed working with large sample sizes, at multiple sites, using the same methodology, cases, and checklists. The benefits seen from using SPs in this study was the opportunity for the participants to engage in human interaction, practice effective communication, and develop a rapport with a patient.

In summary, nursing educators must rethink traditional methods of competence evaluation and learn from our colleagues in medicine who have applied many years of rigorous research in the quest for graduating students competent to practice in the real world. Most nursing programs primarily using mannequins have not been able to realize

the potential of using SPs, not only in the evaluation of competence, but also in laying the

foundation of practicing and reflecting on humanistic care. Nursing lacks this type of

rigorous research most likely due to lack of federal funding, low percentage of nurses

who obtain doctorates, and/or low interest or support in nurses studying nursing

education.

Chapter 5

Discussion of Findings

The goal of this study was to examine the interrater reliability and percent agreement of standardized patients and faculty competency checklist scores when evaluating pre-licensure nursing students. The importance of this study was to strengthen and employ rigorous methods of evaluating pre-licensure nursing student competence. Nursing has not adopted consistent widespread measures of competence in pre-licensure nursing programs other than a licensing exam. Given the external demands for improved accountability of clinical competence from professional agencies, nursing educators must reevaluate methods of teaching and re-examine how we measure and evaluate nursing competence (Nehring & Lashley, 2010). Nursing education must focus on assuring the competence of our graduates. Studies in nursing, as seen decades ago in medicine, are needed to provide legitimacy for nursing educators to adopt SP simulation methodology in nursing programs and eventually serve as consistent evaluation of competencies for licensure and/or certification.

This study adds a glimpse of the necessary large-scale rigorous research required to evaluate performance and competence of pre-licensure nursing students. The notion

59

standardized patients can accurately complete checklists, as seen in medical education, requires rigorous psychometric analysis of developing competency checklists and training raters of expected participant behaviors and skills. Meticulous training of SPs is required for the accurate and reproducible presentation of the case along with extensive training of checklists. Detailed training for faculty is also required to develop robust clinical cases and checklists. Case design, SP portrayal, and faculty training can influence the reliability, and subsequently the validity of competency-based assessments. Multi-site, large sample size standardized patient nursing research studies can potentially provide a legitimate and consistent measure of nursing student competence relieving faculty of laborious and subjective methods of nursing student evaluation. As Newble (2004) discusses, it is important to determine whether the final score on checklists reflect the ability of students to practice competently.

Research Strengths

The project is original to nursing education, providing insight into development of innovative and objective methods for evaluating student nurse competence without the fear of harming an actual patient in clinical practice sites. Based on gaps in the nursing literature, this is the first attempt in nursing to determine inter-rater reliability and percent agreement of standardized patients and faculty competency checklist scores when evaluating pre-licensure nursing students. Kappa scores and percent agreement seen in this study were promising despite a small sample size and the use of only three cases.

An abundance of literature exists on the value of SPs to evaluate competencies in medical students and physicians. Studies in nursing only focus on the strengths and

weaknesses on the use of SPs and OSCEs, but lack the rigor of comparing nursing faculty and SP interrater reliability using valid and reliable tools. This study adds new nursing knowledge in our current pursuit to determine basic fundamental competencies of all pre-licensure nursing students. This study also questions the expansive growth of mannequin-based methods of simulation and its impact on future generations of nurses potentially compromising the "art" of nursing. The study adds to nursing knowledge by concentrating on the use of standardized patients in simulation scenarios and contributing increased realism by exposing students to a "real" patient providing opportunities to learn and practice human compassion, empathetic communication skills, and receive feedback from the patient on how to fine-tune their bedside acumen and connection with the patient.

Research Limitations

The findings of this study should be cautiously interpreted in light of several limitations. First, the analyses were conducted on a small set of cases (three) as well as a relatively small sample of students (49 students for each case). This study occurred in one school of nursing, limiting the ability to generalize results. Also only three cases were used in this project where multiple Objective Structured Clinical Examination (OSCE) stations in multiple medical schools were the foundation of SP psychometric research in medical education. Epstein (2007) recommends that a minimum of ten standardized patient exams over a three to four hour period are necessary to achieve a reliable measure of competence. Secondly, the SPs relied on memory recall to complete the checklists while the faculty scored the checklists in real time. For high stakes summative

examinations, using SPs as separate observers who record actions in real time might provide optimal accuracy. Faculty raters might have had difficulty ascertaining detailed physical exam maneuvers due to camera angle limitations.

Much was learned to improve tool validation and training of raters. Development of standardized training videos of SP encounters to reflect poor, average, and good performance would be helpful to assist with interrater agreement of variances in performance. Raters should practice rating performance on each of these training videos independently then come together to determine relevance, description, and wording of items to agree on expected participant behaviors. Other factors might account for incongruities such as length of the checklist and subjectivity of the faculty when evaluating correct technique of a skill such as taking a blood pressure.

Implications for Nursing

Nursing education is now at the forefront for radical changes to prepare our students for complex clinical environments, immense agency scrutiny, and public demand for safe patient care. It is the goal of all nurse educators to use effective methods to assist nursing students to become competent clinicians (Nehring & Lashley, 2009). Educational and nursing theories guide educators in the use of engaging and experiential methods of learning to the classroom, labs, and clinical practice sites to ensure deliberate practice and exposure to the complexities of our patients within complicated healthcare systems. With a looming nursing shortage, nursing programs are accepting additional students while facing decreasing numbers of clinical sites and faculty to teach in all programs across the spectrum of nursing education (Nehring & Lashley, 2009). Because

of these factors, nursing programs have had to adopt new innovative methods of allowing students to practice skills, apply critical thinking and effective communication strategies, and reflect on role transition. All simulation methods (including SPs methods) provide opportunities for student nurses to acquire these skills in a safe environment without the fear of doing harm. This study highlights the need for all nursing students (like medical students) to experience the concept of human connectedness through active involvement and practice with "live" patients or SPs. As seen in the early 1960s in medical education, nursing educators are currently struggling to define or measure competence of nursing students at all levels. Research is needed to support nursing educators in exploring reliable, innovative ways for objective clinical evaluations and to give students opportunities to work with "live patients" in a safe environment. Technology based education (use of manikins) has exploded in most nursing simulation labs over the last 5 years. Will future generations of nurses lack the "art" of human connectedness and interaction associated with healing and caring? In nursing, caring is competence and how can we measure this with plastic mannequins?

Future Nursing Research

It is hoped the findings of this study will foster future research in nursing education regarding accuracy rates of SPs and faculty in relation to competency evaluations. More importantly, what specific aspects of nursing education should be evaluated consistently of all pre-licensure nursing students in a simulation lab and do encounters with SPs impact patient outcomes in clinical practice? Currently, the National League of Nursing is involved in a multi-site simulation study primarily sponsored by

mannequin vendors (Jeffries & Rizzolo, 2006). Nursing educators must interpret vendor-sponsored research with caution and consider potential implications of bias as with drug-sponsored research. To improve reliability and validity of competency tools, further nursing research is needed to look at large cohorts of nursing students at multiple schools of nursing, assessing the same skills at the same academic level. Future research applying other approaches to reliability such as test-retest or alternate-forms reliability may contribute to higher internal consistency measurements. More importantly, future nursing research on assessing bedside manner, caring, and human connection would be needed to provide legitimate reasons for nursing educators to move away from predominately non-human simulation methods.

Conclusions

Nursing competence is the driving force and fundamental of all schools of nursing and clinical practice. This study supports the inclusion of SP to consistently and accurately evaluate the clinical competencies of pre-licensure nursing students. Nursing educators must rethink traditional educational and evaluation methods and consider other schools of thought to ensure the public with competent, safe nurses and maintain a trusting appreciation for our profession. Currently, with all the changes in healthcare delivery, nursing is in a pivotal position to advocate for safe, effective, and competent patient care in all healthcare settings. As schools of nursing begin the task of defining and measuring competence of all nursing students, it will be important that nurse scientists test new frameworks and methods to ensure safe, effective nursing practice (Nehring & Lashley, 2010). Nursing is just now embarking on the task of coming to an overall

consensus on what needs to be taught, when, how to teach it and how to measure it (Cronenwett et al., 2007). The medical profession has spent the last two decades instituting standards and milestones of competency for all medical schools and defining and measuring competencies from the first year of medical school through residency including licensure and certification (Cronenwett et al., 2007).

Most importantly, nursing must not deviate from the foundation of our practice. The "art" of nursing is very complex and intricate in regards to providing care to patients. The capacity to provide compassionate care is the heart and soul of nursing practice as identified by American Association of Colleges of Nursing and the National League of Nursing (Rhodes, Morris, & Lazenby, 2011). The ability to develop interpersonal skills through exploring human potential as a healer with the achievement of empowered caring is essential for all nursing students (Rhodes et al., 2011). Rhodes et al. (2011) describes self-awareness as an important factor in the development of confidence and competence in nursing students. The ability to practice these unique skills in nursing simulation labs only reinforces the basic need and importance of human connection to generations of future nursing students.

References

American Association of Colleges of Nursing. (2008). *The essentials of baccalaureate education for professional nursing practice.* Washington, DC: Author

Barjis, J. (2011). Healthcare Simulation and its Potential Areas and Future Trends. *Magazine of the Society for Modeling & Simulation International, 2*(1).

Becker, K. L., Rose, L. E., Berg, J. B., Park, H., & Shatzer, J. H. (2006). The teaching effectiveness of standardized patients. *Journal of Nursing Education, 45*(4), 103-111.

Benner, P. (2001). *From novice to expert: Excellence and power in clinical nursing practice.* Upper Saddle River, NJ: Prentice Hall Health.

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

Boulet, J. R. (2008). Summative assessment in medicine: the promise of simulation for high-stakes evaluation. *Academic Emergency Medicine, 15*(11), 1017-1024. doi: 10.1111/j.1553-2712.2008.00228.x

Bradley, P. (2006). The history of simulation in medical education and possible future directions. [Historical Article]. *Medical Education, 40*(3), 254-262. doi: 10.1111/j.1365-2929.2006.02394.x

Carper, B. (1978). Fundamental patterns of knowing in nursing. *Advances in Nursing Science, 1*(1), 13-23.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 41*(1), 37-46. doi: org/10.1177%2F001316446002000104

Cronenwett, L., Sherwood, G., Barnsteiner, J., Disch, J., Johnson, J., Mitchell, P., . . . Warren, J. (2007). Quality and Safety Education for Nurses. *Nursing Outlook, 55*(3), 122-131. doi: 10.1016/j.outlook.2007.02.006

De Champlain, A. E., Margcus. M. J., King, A., & Klass, D. (1997). Using standardized patients for teaching and assessment. *Academic Medicine, 72*(10), S58-S87.

Dewey, J. (1916). *Democracy and Education.* New York: Macmillan.

Ebbert, D. W., & Connors, H. (2004). Standardized patient experiences: evaluation of clinical performance and nurse practitioner student satisfaction. *Nursing Education Perspectives, 25*(1), 12-15.

Elliott, D. L., & Hickman, D. H. (1987). Evaluation of physical examination skills: Reliability of faculty observer and patient instructors. *Journal of the American Medical Association, 258*, 3405-3408.

Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine, 356*(4), 387-396. doi: 10.1056/NEJMra054784

Gaba, D. M. (2004). The future vision of simulation in healthcare. *Quality and Safety in Health Care, 13*(1), 2-10. doi: 10.1136/qshc.2004.009878

Gallagher, P., Smith, T., & Ousey, K. (2012). Problems with competence assessment as it applies to student nurses. *Nurse Education in Practice* 1-3.

Gibbons, S. W., Adamo, G., Padden, D., Ricciardi, R., Graziano, M., Levine, E., &

Hawkins, R. (2002). Clinical evaluation in advanced practice nursing education:

Using standardized patients in health assessment. *Journal of Nursing Education,*

41(5), 215-221.

Greiner, A. C., & Knebel, E. (2003). Health Professions Education: A Bridge to Quality.

In E. C. o. t. H. P. E. Summit (Ed.). Washington D.C.: Institute of Medicine.

Harder, B. N. (2009). Evolution of Simulation Use in Health Care Education. *Clinical*

Simulation in Nursing, 5(5), e169-e172. doi: 10.1016/j.ecns.2009.04.092

Henneman, E., & Cunningham, H. (2005). Using clinical simulations to teach patient

safety in an acute/critical care nursing course. *Nurse Educator, 30*(4), 172-177.

Interprofessional Education Collaborative Expert Panel. (2011). *Core competencies for*

interprofessional collaborative practice: Report of an expert panel. Washington,

D.C.: Interprofessional Education Collaborative.

Ironside, P. M., Jeffries, P. R., & Martin, A. (2009). Fostering patient safety

competencies using multiple-patient simulation experiences. *Nursing Outlook,*

57(6), 332-337. doi: 10.1016/j.outlook.2009.07.010

Jeffries, P. R. (2005). A framework for designing, implementing, and evaluating

simulations used as teaching strategies in nursing. *Nursing Education*

Perspectives, 26(2), 96-103.

Jeffries, P. R., & Rizzolo, M. A. (2006). Designing and Implementing Models

for the Innovative Use of Simulation to Teach Nursing Care of Ill Adults and

Children: A National, Multi-Site, Multi-Method Study: National League of

Nursing.Kaakinen, J., & Arwood, E. (2009). Systematic review of nursing

simulation literature for use of learning theory. *International Journal of*

Educationals Scholarship, 6(1) doi: 10.2202/1548-923X.1688

Kaakinen, J., & Arwoond E. (2009). Systematic review of nursing simulation literature

for use of learning theory. *International Journal of Educationals Scholarship,*

6(1), doi:10.2202/1548-923X.1688

Knowles, M. (1975). *Self-Directed Learning.* New York: Association Press.

Kohn L.T., Corrigan, J. M., & Donaldson, M. S., eds for the Committee on Quality

Health Care in America (2000). *To Err Is Human: Building a Safer Health*

System. Washington, DC: Institute of Medicine.

Kolb, D. A. (1981). Learning Styles and disciplinary differences. In A. Chickering (Ed.),

The Modern American College (pp. 232-255). San Francisco, CA: Jossey Bass.

Landis J. R., & Koch. G. G. (1977). The measurement of observer agreement for

categorical data. *Biometrics,* 33(1), 159-174. doi: org/10.2307%2F2529310

Medley, C. F., & Horne, C. (2005). Using simulation technology for undergraduate

nursing education. *Journal of Nursing Education, 44*(1), 31-34.

Miller, A. M., Wilbur, J., Montgomery, A. C., & Talashek, M. L. (1998). Standardizing

faculty evaluation of nurse practitioner students by using simulated patients.

Clinical Excellence for Nurse Practitionters, 2(2), 102-109.

National League for Nursing. (2010). *Outcomes and competencies for graduates of practical/vocational, dipoloma, associate degree, baccalaureate, master's, practice doctorate, and research doctorate in nursing.* New York: National League for Nursing.

Nehring, W. M., & Lashley, F. R. (2009). Nursing Simulation: A review of the past 40 years. *Simulation & Gaming, 40*(4), 528-552.

Nehring, W. M., & Lashley, F. R. (2010). *High-fidelity patient simulation in nursing education.* Sudbury, Massachusetts: Jones and Bartlett Publishers.

Newble, D. I. (2004). Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education, 38*, 199-203.

Newble, D. I., & Swanson, D. B. (1988). Psychometric characteristics of the objective structured clinical examination. *Medical Education 22*, 325-334.

Norcini, J., & Boulet, J. (2003). Methodological issues in the use of standardized patients for assessment. *Teaching and Learning in Medicine: An International Journal, 15*(4), 293-297. doi: 10.1207/S15328015TLM1504_1

Norman, G. R., Muzzin, L. J., Williams, R. G., & Swanson, D. B. (1985). Simulation in health sciences, *Journal of Instructional Development, 8*, 11-17.

Norman, J. (2012). Systematic review of the literature on simulation in nursing education. *Association of Black Nursing Faculty Journal, 23*(2), 24-28.

Peter, R., & Peter, V. (2008). *An assessment of the internal consistency of measures of constructs used to revise the innovation decision framework.* Paper presented at the Academy of World Business, Marketing, & Management Development.

Rethans, J., & van Boven, C. (1987). Simulated patients in general practice: A different look at the consultation. *British Medical Journal, 294,* 809-812.

Rep, M. A. (2012). Standardized patients in education. *Radiologic Technology, 83*(5), 503-506.

Rhodes, M. K., Morris, A. H., & Lazenby, R. B. (2011). Nursing at its best: Competent and caring. *The Online Journal of Issues in Nursing, 16*(2). doi: 10.3912/OJIN.Vol16No02PPT01

Rystedt, H., & Lindstrom, B. (2001). Introducing simulation technologies in nurse education: a nursing practice perspective. *Nurse Education in Practice, 1*(3), 134-141. doi: 10.1054/nepr.2001.0022

Sanford, P. G. (2010). Simulation in Nursing Education: A Review of the Research. *The Qualitative Report, 15*(4), 1006-1011.

Scully, N. J. (2011). The theory-practice gap and skill acquistion: An issue for nursing education. *Collegian: Journal of the Royal College of Nursing Australia, 18*(2), 93-98. doi: org/10.1016/j.colegn.2010.04.002

Sim, J., & Wright, C. (2005). The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257-268.

Simulation. (n.d.). In *BusinessDictionary.com*. Retrieved from http://www.businessdictionary.com/definition/simulation.html

Tamblyn, R., Klass, D. J., Schnabl, G. R., & Kopelow, M. L. (1991b). Sources of unreliability and bias in standardized patient ratings. *Teaching and Learning in Medicine: An International Journal, 3,* 74-85.

Vessey, J. A., & Huss, K. (2002). Using standardized patients in advanced practice nursing education. *Journal of Professional Nursing, 18*(1), 29-35. doi: 10.1053/Jpnu.2002.30898

van der Vleuten, C. M., & Swanson, D. B. (1990). Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine: An International Journal, 2*(2), 58-76. doi: org/10.1080/10401339009539432

van der Vleuten, C. M., van Luijk, S., Ballegooijen, A., & Swanson, D. B. (1989). Training and experience of medical examiners. *Medical Education, 23,* 290-296.

van der Vleuten, C. M., van Luijk, S., & Swanson, D. B. (1988). *Reliabilty (generalizability) of the Maastrict Skills Test.* Proceedings of the 27th Annual Research in Medical Education Conference.

Vu, N. V., & Barrows, H. S. (1994). Use of Standardized Patients in Clinical Assessments: Recent Developments and Measurement Findings. *Educational Researcher, 23*(3), 23-30. doi: 10.3102/0013189x023003023

Vu, N.V., Marcy, M. L., Colliver, J.A., Verhulst, S. J., Travis, T. A., & Barrows, H. S. (1992). Checklist characteristics and length of testing: Their effects on standardized patients' simulations. *Journal of Medical Education, 26,* 390-395.

Wallace, P. (1997). Following the threads of an innovation: the history of standardized patients in medical education. *Caduceus, 13*(2), 5-28.

Ward, J., Cody, J., Schall, M., & Hojat, M. (2012). The empathy enigma: An empirical study of decline in empathy among undergraduate nursing students. *Journal of Professional Nursing, 28,* 34-40. doi: 10.1016/j.profnurs.2011.10.007

Watson, R., Stimpson, A., Topping, A., & Porock, D. (2002). Clinical competence

assessment in nursing: a systematic review of the literature. *Journal of Advanced*

Nursing, 39(5), 421-431. doi: 10.1046/j.1365-2648.2002.02307.x

Wilford, A., & Doyle, T. (2006). Integrating simulation training into the nursing

curriculum. *British Journal of Nursing, 15*(11), 604-607.

Williams, R., Barrows, H. S., Vu, N., Verhulst, S., Colliver, J., Marcy, M., & Steward, D.

E. (1987). Direct standardized patient assessement of clinical competence.

Medical Education 21, 482-439.

Yanhua, C., & Watson, R. (2011). A review of clinical competence assessment in

nursing. *Nurse Education Today, 31*(8), 832-836. doi: 10.1016/j.nedt.2011.05.003