MEDICAL SIMULATION AS A COMPETENCY-BASED ASSESSMENT WITHIN PHYSICIAN ASSISTANT EDUCATION

by

Michele Toussaint

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Education in Educational Technology

Boise State University

May 2021

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Michele Toussaint

Dissertation Title: Medical Simulation as a Competency-based Assessment within Physician Assistant Education

Date of Final Oral Examination: 31 March 2021

The following individuals read and discussed the dissertation submitted by student Michele Toussaint, and they evaluated their presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Brett Shelton, Ph.D.	Chair, Supervisory Committee
Chareen Snelson, Ed.D.	Member, Supervisory Committee
Jui-Long Hung Ed.D.	Member, Supervisory Committee

The final reading approval of the dissertation was granted by Brett Shelton, Ph.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

DEDICATION

As a first-generation college student, and the first member of my extended family on either side to complete doctoral studies, this all feels a bit surreal. If you had told me twenty years ago when I was just starting my clinical career that I would return to graduate school twice, I never would have believed it. I would like to dedicate this work to my parents and grandparents who imprinted upon me at a young age the value and importance of grit and a strong work ethic. And to Julia and Owen, I can’t wait to see what you will accomplish. When times get tough, remember the words of Michelle Obama, “the only limit to the height of your achievements is the reach of your dreams and your willingness to work for them.”

ACKNOWLEDGMENTS

Many individuals were a great support to me during my doctoral studies and for that I am tremendously grateful.

To my committee, Dr. Shelton, thank you for your patience, and for guidance in helping me shape this research. Dr. Hung, thank you for your insight and sharing your expertise related to methodology and analysis. Thank you Dr. Snelson for making my writing more intentional and precise. These contributions have been critical in shaping the style and content of this manuscript. Finally, thank you to the other members of the BSU Ed Tech faculty for your dedication and support of the students in this program, and helping to shape us to be more innovative and intentional educators.

To my friends and colleagues in the 2017 EdD cohort, and in particular, Andrea, Brandon, Chris, Jenn, and Randall- thank you. Your support through this process has been invaluable and I feel fortunate to have completed the program with you all. I remain grateful for the time spent as collaborators, sounding boards, proof-readers, and for the sharing of ideas, resources and many, many laughs.

To my 'village' of Las Cumbres, I have tremendous gratitude for the use of home offices with freshly baked snacks, hikes, and yoga to clear my head, child entertainment and carpools, meals, companionship, and even the occasional technical disaster assist, or

statistics curbside consult. I feel truly fortunate to have you all in my bubble and you have been instrumental in keeping my family afloat this past four years.

To my MSPA and CVICU colleagues, thank you for your patience and support. Sue, thank you for encouraging me to pursue doctoral studies and for your analytic eye, mentorship, and support of my research interests and professional goals. Nicole, thank you for your diplomacy, guidance, and for always having precise, actionable suggestions to solve the most difficult problems.

To my family, thank you for your patience, for stepping up when things were difficult, and for giving me space when I needed to prioritize academics over the fun stuff. I can't wait to see what comes next.

ABSTRACT

Simulation-based practices are widely utilized in medical education and are known to be a safe and effective way to train and assess learners, improve provider confidence and competency, and improve patient safety. Competency-based initiatives are being more broadly utilized to assess learner proficiency in health professions education. Recent publication of competencies expected of new graduate physician assistants, and updated accreditation requirements which include assessment of learner competencies in non-knowledge based domains, have led to the creation of this simulation-based summative assessment of learner competency in communication and patient care skills for Physician Assistant students.

The purpose of this quantitative study was to identify if this simulation assessment had appropriate construct validity and rater consistency, and to identify if correlation existed between learner performance on the simulation exam and in required Supervised Clinical Training Experiences for measures of communication skills and patient care skills.

While raters for the simulation assessment had minimal variability, measures of internal consistency did not achieve suitable thresholds for patient care skills. Communication skills assessment was able to achieve the minimum suitable threshold for internal consistency with minor revisions. No correlation was noted between exam performance for communication skills or patient care skills and clinical practice ratings. Several key areas exist which may explain these results including the rating scale for the

simulation exam which utilized checklists and not global rating scales, faculty raters with broad and diverse clinical backgrounds, observation-related factors on the part of the student, and the high-complexity and multidimensional nature of provider-patient interactions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAMC	Association of American Medical Colleges
AAPA	American Academy of Physician Assistants
ABIM	American Board of Internal Medicine
ABMS	American Board of Medical Specialties
ACGME	Accreditation Council for Graduate Medical Education
ARC-PA	Accreditation Review Commission for the Education of Physician Assistants
BH	Behavioral Health
CBE	Competency Based Education
CBME	Competency Based Medical Education
CISL	Center for Immersive and Simulation based Learning
EM	Emergency Medicine
EPA	Entrustable Professional Activities
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
IM	Internal Medicine
IPE	Interprofessional Education
MSPA	Masters of Science in Physician Assistant Studies
NBME	National Board of Medical Examiners
NCCPA	National Commission on Certification of Physician Assistants
OSCE	Observed Structured Clinical Exam
PA	Physician Assistant
PAEA	Physician Assistant Education Association
PANCE	Physician Assistant National Certification Exam

PANRE	Physician Assistant National Recertification Exam
PBL	Problem Based Learning
PC	Primary Care
SCPE	Supervised Clinical Practice Experience
SP	Standardized Patient
USMLE	United States Medical Licensing Exam
WH	Women’s Health

CHAPTER ONE: INTRODUCTION

Competency-based practices have been utilized and studied in healthcare education since the 1970s (McGahie, 1978) and have support from key stakeholders in the medical and healthcare education community. While it is imperative that healthcare providers achieve a certain level of competency prior to practicing clinically, identifying how to assess competency can be a challenge. Simulation-based assessment has been implemented in certain sectors of medical education and has been shown to be an effective tool for skill development and both formative and summative assessment. The physician assistant education community has recently developed a set of new graduate competencies, but no established and validated means of assessing competency across domains currently exists. This represents an opportunity to develop a simulation-based exam for physician assistant learners to assure that appropriate levels of competency have been achieved prior to workforce entry.

Background

Physician assistants (PA) are medical providers who diagnose and treat illness, perform medical procedures, prescribe medical therapies including prescriptions, and work collaboratively with others in the healthcare team, including physicians, nurses, and other allied health professionals. The profession was founded in the mid 1960's at Duke University and provided expedited medical training to four Navy corpsman. Since that time, the profession has grown to include 238 accredited training programs and 131,000 certified providers (AAPA, 2019). Physician Assistants work in primary care settings and subspecialty practice in the United States and several other countries, including the United Kingdom, the Netherlands, Ghana, and Canada (AAPA, n.d.). New training programs continue to receive provisional accreditation and numbers of PAs in training and in practice are projected to increase in the coming years.

Physician Assistant Education

Physician assistant curriculum varies somewhat at the programmatic level, but all accredited PA programs must adhere to the standards set forth by the Accreditation Review Commission for the Education of Physician Assistants (ARC-PA). According to recent data collected by the Physician Assistant Education Association, PA programs average 27 months in length and graduates of physician assistant programs are graduate-level prepared with at least a master's degree (AAPA, 2019). Training includes rigorous didactic curriculum, which spans all areas of medicine, clinical procedural training, and supervised clinical practice experiences, averaging two thousand hours, across medical and surgical settings caring for patients across the lifespan.

Physician Assistant Professional Organizations

The Physician Assistant Educational Association (PAEA) is the national organization, which represents PA programs and sets strategic plans, facilitates research, and supports faculty development and professional advocacy. The National Commission on Certification of Physician Assistants (NCCPA) is the organization that administers the Physician Assistant National Certifying Exam (PANCE) and Physician Assistant National Recertifying Exam (PANRE) and oversees that new graduate and certified PAs have met the requirements to obtain and retain certification for professional practice. The

Accreditation Review Commission on Education for the Physician Assistants (ARC-PA) defines educational standards and evaluates PA programs.

Physician Assistant Competencies

In 2018, the Physician Assistant Education Association developed the Core Competencies for New Physician Assistant Graduates (Physician Assistant Education Association, 2018). Prior to 2018, there had been several iterations of recommended competencies for practicing physician assistants, but none spoke specifically to the expectations or skills necessary for workforce entry. This document was prepared in a standardization effort so that all programs training PAs can work toward common alignment. At the PAEA Stakeholder Summit 2016, employers indicated that new graduates should possess not only medical knowledge and clinical skills, but they will need to know how to synthesize and incorporate interpersonal and communications skills. Stakeholders identified "the need for greater emphasis on critical thinking, empathy, and communication skills" (Physician Assistant Education Association, 2018, p.6). The domains identified are as follows:

Patient-centered practice knowledge

Society and population health

Health literacy and communication

Interprofessional collaborative practice and leadership

Professional and legal aspects of health care

Health care finance and systems

The committee also identified cultural humility and self-assessment and ongoing professional development as foundational skills that pertain to each of the above competencies.

Statement of the Problem

The PAEA and key stakeholders have recently established competencies required for workforce entry of new graduates to the physician assistant profession. The ARC-PA has indicated that programs must adopt competency standards for workforce entry of program graduates. The NCCPA PANCE exam only assesses a single domain of competency, medical knowledge, on the current certification examination. PA training programs are left to establish and validate their measures of learner readiness for workforce entry, and per accreditation standards, should align with the domains outlined in the new graduate competencies. This represents an opportunity for the creation of a competency-based assessment tool for use in summative assessment of PA students.

Purpose of the Study

The purpose of this quantitative, correlational pilot study will be to analyze the use of simulation as an assessment for competency-based summative evaluation of physician assistant students. Assessing competency domains other than knowledge is currently completed via a survey of clinical preceptors from supervised clinical practice experiences (SCPE). Given that clinical practice experiences and impressions of evaluators are inherently variable and may have poor inter-rater reliability, it is crucial to identify a reliable and valid manner of assessing learners in competency domains beyond that of medical knowledge. A standardized simulation assessment could be used to ensure that learners have more broadly met competency standards prior to graduation and entry into clinical practice. This would also ensure that each learner is assessed in a comparable manner.

Research Questions

1. Does the simulation-based assessment have sufficient construct validity and criterion related validity to be used as a high-stakes summative evaluative tool?
    a. Does each station of the simulation exam perform with sufficient internal reliability?
    b. Within stations, is there sufficient reliability in response consistency among raters?
2. To what extent does performance on the simulation-based summative evaluation correlate with supervised clinical practice measures for communication skills?
    a. Does this level of correlation support the use of simulation-based competency evaluation for co-assessment of communication skills?
    b. Do the ratings for specific clinical practice specialty areas impact the level of correlation with the simulation exam scores for communication skills?
3. To what extent does performance on the simulation based summative evaluation correlate with supervised clinical practice measures for patient care skills?
    a. Does this level of correlation support the use of simulation-based competency evaluation for co-assessment of patient care skills?
    b. Do the ratings for specific clinical practice specialty areas impact the level of correlation with the simulation exam scores for patient care skills?

Research Design

Quantitative research methodology will be utilized to analyze the data from this study. Data collection methods will include Likert-scale survey data and exam scores reported numerically and as such, a variety of quantitative analyses can be performed. Correlation analysis via calculation of Pearson correlation coefficient and reliability analysis of assessment components will be conducted.

Limitations and Delimitations

There are several limitations to this study based on the deployment to a single cohort at one training program. Sample size will be limited to the enrollment size of a single cohort of students (n=27). Also, admissions criteria can vary from one program to another, which may impact generalization of the study results more broadly.

Due to the time and space limitations for conducting the simulation assessment, eight unique faculty evaluators will be involved in the scoring process, which introduces the potential for inter-rater reliability challenges. The assessment will also occur over two days, and while there is a strict honor code in place, the possibility remains that information regarding the content of the assessment could be shared between students. This could potentially impact the integrity of the exam between student groups.

Approaching graduation from the program, it is expected that all learners will have achieved the appropriate level of competency to perform adequately on a summative assessment. In the analysis of the data, there is a possibility that statistical significance may not be reached due to similar performance ratings across learners in the cohort. Delimitations will include the exclusion of non-physician assistant student learners from within the institution, as well as learners from outside of the institution. This exclusion will allow for a more specific assessment of students from the home institution for the purposes of program and curricular evaluation.

CHAPTER TWO: LITERATURE REVIEW

The quantitative study proposed will utilize simulation as a competency-based assessment tool for physician assistant learners. This comprehensive review of the literature will focus on the following topics central to this study. Learning theories related to healthcare education and simulation will be reviewed. Simulation use for both skill acquisition and assessment of healthcare learners must be explored as this is the basis for creation of the assessment for this study. Competency-based assessment in medical education and how this relates to the PA new graduate competencies will be discussed more specifically.

Learning Theories and Simulation

In a review of the medical literature on simulation, McGahie (2018) advocates that simulation-based mastery learning occurs not by virtue of a single theory, but by a convergence of behavioral, constructivist, and social cognitive principles. Exploring each of these in the context of medical education and simulation-based initiates will help to shape the foundation of this study.

Constructivism and Experiential Learning

Constructivism is a theory of learning in which the act of learning is based on a process that connects new knowledge to pre-existing knowledge (Dennick, 2016). Fundamental principles of constructivism include the following tenets: students learn best when learning is active, reflective, and centered around reasoning and processes. The learner creates meaning within context from lessons, instead of being a passive recipient of knowledge transfer (Ertmer, 2013).

These principles align directly with simulation-based learning, as students are actively participating in situations that engage their clinical judgment and problem-solving ability. The simulation-based formats of interactive case-based learning, standardized patient work, and high-fidelity simulation all rely on these concepts. Standardized patient encounters and high-fidelity team training scenarios are also frequently accompanied by debriefing sessions. These facilitated discussions include opportunities for feedback and self-reflection.

One can also include the more specific concept of experiential learning within this discussion. Fundamental principles of experiential learning focus on the scaffolding of advanced concepts on existing knowledge, as well as active learning principles. Kolb's (2014) exploration of experiential learning fits well in this context. Kolb's educational model consisting of a "holistic integrative perspective on learning that combines experience, perception, cognition, and behavior" (p. 31), is a common practice in clinical education. The scaffolding of new and increasingly complex knowledge or skills upon an existing knowledge base to improve performance and expertise, as described by Dennick (2016), is common practice in medical education. This aligns with the apprenticeship style model, where learners gain increasing levels of responsibility over time and with demonstrated competency.

Additionally, Kolb proposed that learning takes place in a cycle with episodic experience, reflection, conceptualization, and experimentation. This is reflective of medical education simulation models of revisiting concepts, receiving and integrating feedback, and analysis of findings in progressively complex ways in order to solve problems. This is also representative of psychomotor skill acquisition required for the development of procedural competency. Barsuk, McGaghie, Cohen, O'Leary, and Wayne (2009) demonstrated that simulation-based training, which used experiential learning principles for procedural skill training, resulted in improved performance and fewer procedural complications.

Critics of constructivism and experiential learning, such as Kirshner (2006), argue that, particularly with novices, expecting learners to sift through massive amounts of information to establish solutions to complex problems, with minimum guidance, may be counterproductive. Taylor and Hamdy (2013) suggest that there should be a threshold level of knowledge in place before the introduction of experiential practices to give students an appropriate framework to allow for scaffolding to occur. Considering the report from the National Academies of Science, Engineering and Medicine (2018, p.33), distinctions exist between novice and expert learners in both their general abilities as well as their problem-solving strategies (pattern recognition, organization, and interpretive skill). In these situations, the curriculum must be carefully designed to allow for more structured experiences with less variability until basic proficiency is established. Following the introduction of basic concepts, extending to the application and more abstract and complex reasoning, will deepen understanding. A purely experiential curriculum in medical education would also pose significant limitations due to the vast quantities of factual knowledge, which must be delivered in a somewhat fixed timeframe.

Conversely, strict constructivists may reject structured fact-based instruction in favor of extensive practical work. While this may be suitable for those who already have a firm grasp of factual knowledge, true novices may benefit from a combination of early traditional instruction, followed by application of these facts experientially, as can be accomplished with simulation.

Complexity Theory

Complexity theory examines how learning emerges from the convergence of numerous external factors, including material, social, and settings (Fenwick & Dahlgren, 2005). Central to this theory is the concept of the distinction between competency and capability. Fraser and Greenhalgh (2001) define capability as the "extent to which individuals can adapt to change, generate new knowledge, and continue to improve their performance" (p.799). This is a more dynamic and fluid concept of application of knowledge than a simple recall of facts, which may be present in competency.

The following are key factors related to complexity theory as identified by Fraser and Greenhalgh (2001):

Neither the system nor its external environment is, or ever will be, constant

Individuals within a system are independent and creative decision-makers

Uncertainty and paradox are inherent within the system

Problems that cannot be solved can nevertheless be "moved forward"

Effective solutions can emerge from minimum specification

Small changes can have big effects

Behaviour exhibits patterns (that can be termed "attractors")

Change is more easily adopted when it taps into attractor patterns. (p.800)

The concept of emergence, “non-linear dynamics of internal interactions among a quantity of diverse elements, such as diverse ways of thinking and acting, or diverse information” (Fenwick & Dahlgren, 2015, p. 362) is essential for training clinicians that must apply clinical concepts in varied circumstances. As medicine is continually evolving, and individual patient interactions and situations are dynamic and unpredictable, utilizing the foundations of complexity theory to develop medical education interventions seems to be a natural fit. The non-linear nature of clinical management and inherent variability in daily practice environments is well suited for the integration of process-oriented learning methods of complexity theory.

Levels of complexity should also be carefully considered when designing simulation-based learning activities. As Haji, Cheung, Woods, Regehr, de Ribaupierre, and Dubrowski (2016) identified, when novice learners are involved, excessively complex circumstances can reduce the quality of task performance. While it is important for skills to be reproducible in patient care contexts, care when developing educational interventions should be taken to allow novices gradual increases in complexity when possible to optimize performance.

For medical learners, high fidelity simulation, in particular, lends itself to teaching how to respond to variable and dynamic circumstances. When designing scenarios for simulation, the educator must also consider that learners may choose to make decisions that are atypical or not a part of the scenario algorithm, and even if learners do not follow the path of the specifically intended concept, there are still opportunities to learn. This can also provide a wealth of learning opportunity through debriefing, both in discussing how and why decisions were made and to tease out the context of problem-solving strategies.

Problem-based Learning

Problem-based learning (PBL) is the student-centered practice to present learners with a complex applied problem to solve through facilitated discussion, often in a small group setting.

The fundamental principles of problem-based learning are identified by Duch, Groh, & Allen (2001) are outlined as follows:

> Think critically and be able to analyze and solve complex, real-world problems,
> Find, evaluate, and use appropriate learning resources,
> Work cooperatively in teams and small groups,
> Demonstrate versatile and effective communication skills, both verbal and written,
> Use content knowledge and intellectual skills acquired at the university to become continual learners. (p.6)

This process of collaborative work and utilization of both intellectual acumen as well as communication skills aligns well with the medical model of training. Regarding simulation-based training initiatives, this most closely pairs with interactive case-based models of instruction, and perhaps high-fidelity simulation in a team training type scenario.

In interactive case-based scenarios, learners are presented with a complex or challenging case and work through most likely diagnoses based on their prior knowledge while identifying areas to research further. These can either be completed in a small group with facilitated discussion or via digital case study software programs. Computer-based clinical problem solving has been well received both as a teaching and assessment tool in medical education and is even included in the United States Medical Licensing Exam. Feldman et al. (2008) showed that student performance on a computer-based case study program correlated with performance on other commonly used evaluative tools such as the standardized pediatrics exam and clinical performance as rated by supervisors.

High-fidelity simulation for team training could also allow for principles of problem-based learning. While these exercises will often limit a participant's ability to conduct self-directed research in real-time, collaboration with the team is encouraged, which can provide alternative perspectives and additional knowledge. At the same time, high fidelity team training emphasizes the importance of communication skills, which are crucial to PBL (Weinstock & Halamek, 2008).

Learning Theory Summary

Medical trainees must learn a vast quantity of content in a relatively brief period of time and must not merely acquire factual knowledge but be able to apply knowledge in varied and complex environments. Ensuring that learners possess the ability to recall and apply knowledge poses a challenge with regard to curriculum design and confirming readiness for clinical practice after educational programming. Simulation-based learning has been proposed as a way to provide learners with opportunities to apply knowledge and actively engage in clinical problem-solving.

Simulation in Healthcare Education

Simulation is a technique initially pioneered in the commercial aviation industry, to amplify or recreate realistic circumstances so that learners may experience them in an authentic way (Gaba, 2004). Since the 1960s, mannequin-based simulators and more elaborate simulator devices have been developed for use in healthcare education (Cooper & Taqueti, 2004). Various modalities of simulation allow for the deliberate practice of a variety of skills and challenging experiences in an effort to improve technical acumen and critical thinking in low frequency, high-stakes circumstances that can occur in medicine. Utilization of simulation as a training modality also reduces risk to patients (Ziv, 2003).

Efforts to ensure that physicians not only have sufficient factual knowledge to practice, but competency in interpersonal and communication skills, and patient care have also led the United States Medical Licensing Exam (USMLE) to include the use of simulation for professional assessment (Boulet, 2008). A series of Standardized Patient (SP) encounters, which assess patient interview and physical exam skills, and oral and written communication is now a key component of high-stakes assessment in undergraduate medical education.

Coerver, Multak, Marquardt and Larson (2017), found that utilization of simulation and standardized patients is quite common in physician assistant education as well. Standardized patients are also commonly utilized for physician assistant learner simulation-based assessments (Coplan, 2008). Additionally, PA programs report using simulation for formative and summative assessments with rates as high as 83% (Coerver, 2017).

Simulation Modalities

Cooper (2004) defines five broad categories of simulation in the healthcare space: verbal (role play), standardized patient (actor), partial task-trainers (part models, virtual reality), computer patient (virtual world), electronic patient (replica of the clinical setting, interactive mannequin). Each of these has a role in healthcare education and how and when they are utilized will vary based on the type of learner and the learning objective of the exercise.

Low fidelity (less realistic) task training devices are used for simple procedural training such as for airway management, placement of intravenous lines, or urinary catheters. High-fidelity (authentic) simulation exercises are used to replicate experiences such as operating room emergencies, or patient resuscitation attempts for a multi-disciplinary team (Halamek, 2000; Lighthall, Barr, Howard, Gellar, Sowb, Bertacini, & Gaba, 2003). Virtual reality-based trainers may are utilized for learning surgical or procedural skills, modeling difficult conversations, or exploring 3-dimensional complex anatomy (Grantcharov et al., 2004; Aggarwal et al., 2009; Maresky et al., 2019). Standardized patients (actors) are utilized to replicate challenging patient encounters and hone communication skills. Across specialties and modalities, simulation is well received by trainees.

Incorporating problem-based learning exercises following the introduction of the material gives learners the opportunity for scaffolding, which is key to a more in-depth understanding of the material. Additional experiences in either the simulation lab with standardized patients or simulated patients (mannequins) in the spirit of complexity theory and experiential learning can provide richer opportunities to integrate various sources of data in an applied context. Synthesis of data in the applied context that can be achieved in these learning experiences provides learners with the opportunity to problem-solve and develop management strategies without compromising patient safety. Skilled facilitated debriefing following these encounters incorporates principles of self-reflection, analysis, and integration of feedback from experiential learning models and problem-based learning.

Simulation integration and utilization in PA and MD education

While, anecdotally, simulation was thought to be widely used across medical education, in 2011, the Association of American Medical Colleges (AAMC) commissioned an exploration of the use of simulation initiatives in undergraduate and post-graduate medical education. The Physician Assistant Education Association subsequently followed this path to gain a more comprehensive understanding of how simulation was being utilized. A summary of these reports follows.

Association of American Medical Colleges report on Simulation

In 2011, the Association of American Medical Colleges (AAMC) along with the Society for Simulation in Healthcare, the Association of Standardized Patient Educators, and the American Association of Colleges of Nursing conducted a survey of member programs regarding their utilization of simulation for education and assessment. The summary report by Passiment, Huang, and Sacks (2012) indicates broad and extensive usage of simulation activities in both physician training programs and postgraduate medical education. For the purposes of this survey, simulation included mannequin-based, physical models (task trainers), standardized patients (actors) or computer-based programs. A total of 133 medical schools and 263 teaching hospitals were invited to participate and the response rate was 68% for medical schools (n=90) and 24% for teaching hospitals (n=64).

Of respondents, all 90 medical schools indicate that they use some form of simulation each year of undergraduate medical training and all 64 teaching hospitals report utilizing simulation at some point during the four years of undergraduate training. The most common content areas taught with simulation were emergency medicine, obstetrics-gynecology, internal medicine, pediatrics, surgery, and anesthesiology. For delivery of preclinical content, 84% of programs utilize simulation in some way, and clinical skills, clinical medicine, and physical diagnosis the most common domains covered. During the clinical phase of training (clerkship), 95% of medical schools, and 68% of teaching hospitals incorporate simulation with internal medicine, emergency medicine, pediatrics, and anesthesiology most commonly represented. A wide variety of simulation modalities are utilized for training, with mannequins, task trainers and standardized patients (actors) all represented at 84% of medical schools.

When post-graduate (residency) training is considered, rates of simulation use are at approximately 90% for both teaching hospitals and medical schools for the first 3 years of residency. These rates decline for training programs that extend to four or five years in length. For subspecialty physician training, critical care medicine, pulmonology, cardiology, neonatology, and gastroenterology most commonly utilize simulation training.

The integration of interprofessional educational (IPE) experiences are also common with 93% of medical schools and 84% of teaching hospitals reporting participation in simulation initiatives. Nurses, emergency medical technician/paramedics, pharmacists, physician assistants, and respiratory therapists are the most common non-physician providers included in IPE activities.

The Accreditation Council for Graduate Medical Education (ACGME) core competencies for medical school graduates were also explored with regard to simulation. These competencies based on health care quality goals were developed in the late 1990's and fundamentally changed the way physician trainees are evaluated (Swing, 2007). The general competencies of medical knowledge, patient care, interpersonal communication skills, professionalism, practice-based learning, and system-based practice were included, and four additional domains important to clinical practice (psychomotor tasks, leadership, team training, and critical thinking) were also assessed. High rates of simulation use for educational exercises in most domains are reported, and simulation is also utilized as an assessment tool across many domains.

Simulation initiatives can result in increased cost of medical education services. These costs include staffing, administrative expenses, and equipment cost. While they vary widely, expenditures in excess of $750,000 per year are reported in over one-third of medical schools.

In summary, simulation is being used extensively for physician training in the United States. A variety of modalities are commonly used for training across different domains of practice and medical specialties. For post-graduate physician training, specialties that include procedural skill training, or patient resuscitation are more likely to incorporate simulation.

Physician Assistant Education and Simulation Use

In 2014, a national survey of physician assistant programs was conducted by Coerver, Multak, Marquardt and Larson (2017) to assess the utilization of simulation-based medical education. Of 177 programs contacted, there was a 35.6% response rate (n=63) which is somewhat limiting for a comprehensive overview but can still provide valuable insight about simulation use broadly across PA programs.

Of responding programs, 96% report some use of simulation-based teaching or assessment. Standardized patient use was utilized by 93% of programs, followed by mannequins (83%), task trainers (77%) and hybrid simulations (55%). Cardiology and pulmonology skills were the most frequently addressed clinical areas with 97% and 82% of programs reporting use. Both formative and summative assessments are conducted with simulation at 83% of programs responding. Interprofessional education and training is conducted via simulation at 72% of responding programs with nurses, medical students and pharmacists most commonly included. Communication skills and team training are most often taught in this context.

Simulation for Assessment in Healthcare Education

Observed Structured Clinical Exams (OSCE) are performance-based assessments in the simulation environment which were first described by Harden et al in 1975. Standardization of clinical or performance scenarios were thought to increase the validity and reliability of the assessment of performance. These can be used for both formative and summative assessment for learners of all levels. Considering Miller's pyramid of assessment (1990), utilization of OSCEs falls within the 'show's how' level of performance assessment as opposed to simply recalling facts as on a multiple-choice assessment.

In 1999, the USMLE integrated a digital case-based assessment to the Step 3 medical boards exam to evaluate clinical reasoning, application and synthesis of medical knowledge. Subsequently, in 2004, a Clinical Skills assessment was incorporated into the Step 2 medical boards exam. Successful performance in both digital case-based, and SP based high stakes assessments are a requirement for medical licensure in the United States.

As indicated by Coerver, Multak, Marquardt and Larson (2017), utilization of simulation and standardized patients is quite common in physician assistant education. Standardized patients are also commonly utilized for physician assistant learner simulation-based assessments (Coplan, 2008). Additionally, PA programs report using simulation for formative and summative assessments with rates as high as 83% (Coerver, 2017).

While utilization of simulation in physician assistant education is broad, there are currently no specific guidelines or accreditation standards in place to guide physician assistant programs in the development or administration of simulation assessments. OSCEs, while prevalent, are program specific, non-standardized, and may not be validated. Evidence-based guidance regarding implementation of simulation programs, both for formative and summative assessment, would be beneficial to educators as well as agencies providing oversight, and would provide consistent measures of learner competency. Additionally, alignment of simulation-based assessments with professional competencies required for workforce entry would ensure professional practice readiness in clinical environments for new graduates.

Competency-based Medical Education and Assessment

Competency-based practices have been utilized and studied in healthcare education since the 1970s (McGahie, 1978) but only recently have attempts been made to more clearly standardize and define these practices. Frank, Mungroo, Ahmad, Wang, De Rossi, & Horsley (2010) conducted a systematic review of the literature and qualitative methodological approach in order to develop a definition of competency-based education. An initial search of the medical and education literature yielded 173 sources from the United States, the United Kingdom, Canada, and Australia, which were deemed appropriate for subsequent analysis. Resulting from this analysis, four major themes (organizing framework, rationale, contrast with time, and implementing CBE), and six sub-themes (outcomes defined, a curriculum of competencies, demonstrable, assessment, learner-centered and societal needs) were identified. Ultimately, the definition of Competency-based education that arose from Frank et al. (2010) is as follows:

> Competency-based education (CBE) is an approach to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organized around competencies derived from an analysis of societal and patient needs. It de-emphasizes time-based training and promises greater accountability, flexibility, and learner-centeredness. (p. 638)

When developing the Core Competencies for New Physician Assistant Graduates, the PAEA (2018) defined competency as a "specific skill, knowledge or ability that is both observable and measurable." (p. 4) Meretoja and Koponen (2012) in discussing competency for the nursing profession define competence as "an underlying characteristic of an individual that is directly related to various quantifiable aspects of effective job performance." (p. 415)

To synthesize the critical features of each of these definitions, for the purposes of this study, competency is defined as specific knowledge, skills, and abilities that are observable, or measurable, to assure acquisition.

Competency-based Practices in Medical Education

Competency-based practices have been utilized for both formative and summative purposes in medical education and have support from key stakeholders such as the Association of American Medical Colleges (AAMC), Accreditation Council for Graduate Medical Education (ACGME), the American Board of Medical Specialties (ABMS), and the United States Medical Licensing Exam (USMLE). This is evidenced by adoption of demonstrable competencies in both USMLE Step 2 and Step 3 of the medical licensing exam, and the AAMC's adoption of Entrustable Professional Activities (EPAs) for entering medical residency (AAMC, 2014), and the adoption of the ABMS Maintenance of Certification standards (Hawkins, Lipner, Ham,Wagner & Holmboe, 2013).

The AAMC developed competency standards in four primary areas: curriculum development, assessment of competency, the path to entrustment, and faculty development. Entrustable professional activities (EPAs) reflect key skills that all medical students must possess prior to residency, regardless of the intended medical specialty. EPAs are defined as "units of professional practice, defined as tasks or responsibilities that trainees are entrusted to perform unsupervised once they have attained sufficient specific competence" (AAMC, 2013, p. 2). Entrustable professional activities are essentially competencies in context; that is, an integration of the competencies that allow one to perform professional activities in the clinical setting. That they are observable, measurable in outcome, and are independently executable makes them ideal for assessment tools in medical education.

Competency Frameworks

A competency framework is defined by Juneja (n.d.) as a "comprehensive structure which describes different competencies with its specific set of behavioral indicators and measurement criteria." Development of a competency framework involves the following steps

Define the purpose and performance objectives of a job or position

Identify the competencies and behaviours that predict and describe superior performance in the job

Validate selected competencies

Implement/integrate competencies (Sanghi, 2016, p. 91)

As such, when developing competency frameworks, one must carefully consider what objectives lay at the endpoint of the process. Clear and specific competencies that are predictive of high-quality work should be identified and validated prior to attempts to implement programs or assess individuals.

Concerning medical education specifically, Van Melle et al. (2019) identified the need for a framework to describe and evaluate competency-based medical education (CBME). The process by which their framework is developed is outlined to follow. The first step in framework development was to explore the core components and best practices of CBME by exploring the literature. Stakeholders were identified, with representation from medical organizations, educators, assessment specialists, and researchers. Consensus was established surrounding key components of competency by the Delphi method. A draft of competencies was developed based on stakeholder responses to surveys and focus groups. Five key components of CBME were identified: outcome competencies, sequenced progression, tailored learning experiences, competency-focused instruction, and programmatic assessment. Each of these components was then further explored to include practices, principles, and conceptual frameworks.

Most relevant to the discussion of competency-based assessment are the core competencies related to outcomes. When outcomes are explored related to contexts of professional practice, Van Melle et al. (2019) reports that "required outcome competencies are based on a profile of graduate and/or practice-based abilities” (p.1005). Considering the context of principles, “specification of learning outcomes promotes focus and accountability” (p.1005). Theoretical frameworks most applicable to CBME include those of backward design, job-task analysis, social accountability, and outcomes-based education.

Competency-based Assessment

When considering the standard evaluative tools and strategies utilized in competency-based assessment, it is helpful first to consider the domains of competency that are commonly referenced. While these vary from one organization to another, competencies common to many stakeholder and oversight groups in medicine include medical knowledge, communication skills, patient care skills, professionalism, and practice-based learning/performance improvement.

Medical Knowledge

Medical knowledge and clinical diagnostic reasoning are most commonly assessed with examinations. Various examinations are utilized to establish competency for healthcare students and professionals. For physicians, these include the three USMLE exams, the National Board of Medical Examiners (NBME) subject exams, and various certification and re-certification exams associated with medical specialty boards.

For physician assistants, several validated examinations exist to determine competency in the realm of medical knowledge. The NCCPA oversees the PANCE exam for graduates seeking certification, and the PANRE for recertification. The PAEA offers End of Rotation Exams (analogous to the NBME subject exams) and the End of Curriculum Exam as a summative assessment of medical knowledge.

Studies exist to support that high performance on exams of medical knowledge are associated with clinical competency as rated by supervisors (Shea, Norcini & Kimball 1993), professionalism as it relates to disciplinary action (Papadakis, Arnold, Blank, Holmboe & Lipner, 2008), and clinical outcomes (Norcini, Lipner & Kimball, 2002).

Communication Skills and Professionalism

Communication skills are a cornerstone of patient care and are of crucial importance when assessing provider competency. There are several approaches to the assessment of communication reported in the literature. Survey data can be collected from patients as with the validated American Board of Internal Medicine (ABIM) Patient Assessment survey (Abadel and Hattab, 2014). Surveys can also be conducted among peers, and co-workers, or via self-assessment. Violato, Marini, Toews, Lockyer, and

Fidler (1997) found that when survey data related to communication skills were collected from peers, coworkers, and patients, a reliable assessment was achieved.

Standardized patient encounters can also be used to assess communication skills. Chang, Mann, Sommer, Fallar, Weinberg, and Friedman (2017) found that SP assessment of provider communication skills had good inter-rater reliability and correlated with Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey data obtained from patients. Participation in the SP communication exercise also improved the confidence of providers related to their communication skills with patients.

Patient Care Skills

Assessment of patient care related skills may vary based on what stage of the educational process the learner is in at the time of evaluation. Expectations for a practicing provider with many years of clinical experience may be quite different than for that of a trainee who has not yet completed their education or achieved licensure.
For licensed professionals, patient care may be indirectly assessed through fulfilling obligations for the maintenance of licensure. When applying for a renewal license, any pending or resolved medical malpractice claims or other disciplinary action must be disclosed to both state medical boards and national accreditation groups.

For trainees, other evaluative frameworks have been developed to more directly measure competency in the patient care domain. Pangaro (1999) introduced the RIME (Reporter, Interpreter, Manager, Educator) framework to describe trainee progress throughout clinical education. Tolsgaard, Arendrup, Lindhardt, Hillingsø, Stoltenberg, & Ringsted (2012) showed that the RIME framework demonstrated good construct validity and interrater reliability when used to assess competency during standardized patient encounters. Statistically significant differences in performance were noted between trainee cohorts as they progress to higher levels of education.

Surveys are frequently conducted to assess learner performance during structured clinical practice experiences (SCPE). Observed Structured Clinical Exams (OSCE) are competency-based assessments to measure clinical performance in simulated settings (Khan, Gaunt, Ramachandran, and Pushkar, 2013). Since all trainees will experience the same clinical scenarios and be evaluated with the same tools, this is a more fair and equitable assessment method than relying on clinical practice evaluations alone. In OSCEs, students will proceed through multiple stations where different skills will be demonstrated and assessed, such as history taking and physical exam skills, clinical documentation, communications, procedural skills, or clinical reasoning.

There are two primary means of evaluation for OSCEs. Analytical measures such as checklists may be binary (yes/no, completed/did not complete) or may include quality measures related to the level of performance. Holistic measures or global ratings assess the quality of the encounter as a whole and are typically measured on a rating scale. Turner and Dankowski (2008) suggest that global rating scales may yield superior results that have better inter-station reliability and construct validity.

Practice-based Learning and Performance Improvement

Performance improvement measures are most often related to those who are currently practicing and will engage in continuing education activities. Licensing and certification agencies have varied requirements for continuing education requirements, and these activities are logged and reported based on recertification cycles.

Recertification exams can ensure that providers have ongoing competency concerning medical knowledge.

Self-reported performance improvement data, which includes a review of survey data from patients and chart review, as well as provider self-assessment, has been proposed by the American Board of Internal Medicine for the evaluation of performance and practice improvement (Duffy et al., 2008). When considering self-assessment as a measure of competency, it is essential to assess for reliability. In a systematic review of physician self-assessment measures compared to external measures of competency, Daves, Mazmanian, Fordis, Van Harrision, Thorpe, and Perrier (2006) found that physicians do not accurately self-assess. This was true across domains of assessment, level of training, and clinical specialty.

In summary, a wide variety of simulation-based platforms and assessments exist in the medical education sector. These have been widely adopted to assess medical students and graduates in domains beyond that of medical knowledge. The physician assistant education community should consider adoption of such assessments in parallel with the AAMC and USMLE to standardize assessment of additional domains of competency.

Summary

The healthcare education sector is becoming increasingly focused on assuring not just a minimum level of knowledge, but the broad competency of graduates and clinicians. Traditional benchmarks for programmatic completion such as standardized knowledge assessments do not assess domains beyond clinical knowledge that are critical to clinical practice. In exploring alternative assessment techniques applicable to additional domains, simulation has emerged as a promising method of assessment. Simulation has been adopted for both formative and summative practices for healthcare learners and reflects alignment with principles of problem-based learning, and experiential learning by applying and synthesizing prior academic experiences to their performance assessment in a simulated environment. That learners will actively engage with the simulation environment and both recall and apply clinical judgement and demonstrate reasoning allows for a robust assessment of clinical skills. The additional benefit of multiple learners being assessed for the same clinical scenario also offers the advantage of improved reliability over learner evaluation in variable clinical practice environments.

As physician assistant training programs attempt to fulfill accreditation requirements ensuring that graduates are competent across broad domains, educators must identify reliable and valid means of competency assessment. The graduate and post graduate medical education community has adopted simulation as a valid and reliable method to assess competency in domains beyond knowledge at the institutional level, as well as for national certifying organizations. While physician assistant education often parallels what has been adopted by the physician education community, no studies exist to evaluate the use of such simulation-based competency assessments for physician assistant learners. This study will pilot a simulation-based competency assessment for physician assistant students. Correlating student performance on the simulation assessment to their clinical practice performance will provide insight to the reliability of utilizing simulation to demonstrate learner competency.

CHAPTER THREE: METHODOLOGY

This chapter will outline the research methodology, data collection and analysis, and assessment design. The goal of this study was to establish the construct validity of simulation as a competency-based summative assessment for physician assistant learners. More specifically, competencies related to communication and patient care skills in learners at the conclusion of their education were the basis for this research study.

Research Questions

1) Does the simulation-based assessment have sufficient construct validity and criterion related validity to be used as a high-stakes summative evaluative tool?
   a) Does each station of the simulation exam perform with sufficient internal reliability?
   b) Within stations, is there sufficient reliability in response consistency among raters?
2) To what extent does performance on the simulation-based summative evaluation correlate with supervised clinical practice measures for communication skills?
   a) Does this level of correlation support the use of simulation-based competency evaluation for co-assessment of communication skills?
   b) Do the ratings for specific clinical practice specialty areas impact the level of correlation with the simulation exam scores for communication skills?
3) To what extent does performance on the simulation based summative evaluation correlate with supervised clinical practice measures for patient care skills?

a) Does this level of correlation support the use of simulation-based competency evaluation for co-assessment of patient care skills?

b) Do the ratings for specific clinical practice specialty areas impact the level of correlation with the simulation exam scores for patient care skills?

Research Design

This research was conducted as a cohort based, quantitative, correlational study to explore the extent of relationships between performance on supervised clinical training experiences, and simulation-based summative assessments at the conclusion of a physician assistant training program. Correlation was selected for use in this study to compare performance between observed clinical experiences and standardized simulation encounters to determine the internal reliability of simulation based assessment of competency. Additionally, factors related to the simulation exam such as interrater and interstation reliability were analyzed, and internal consistency was evaluated to determine suitability for use as a high stakes evaluative tool.

Quantitative Design

Quantitative design was pertinent in this case because all the data collected and analyzed was either assessment data reported in exam scores, or clinical performance survey data reported by Likert-scale. Descriptive analyses as well as additional statistical analysis including correlational analyses, ANOVA, principle component analysis for constructive validity, and reliability analyses were conducted (Table 2). This research was deemed to be exempt by the Boise State University Institutional Review Board (Appendix D).

Participants

The study population was obtained by convenience sampling of a single cohort of physician assistant students enrolled in the Stanford School of Medicine Masters of Science in Physician Assistant (MSPA) Studies Program. The MSPA program is a highly competitive professional training program with an acceptance rate of less than 5%. Training is 30-months in duration and comprised of five academic quarters of didactic training followed by twelve months of supervised clinical practice experiences. Assessment data from one entire cohort of senior students was evaluated in this research. This data was obtained from the first cohort of students approaching program completion which is comprised of 27 students (22 women and 5 men) with an average of 28.9 years (24.9-38 years of age).

Summative Assessment Activities

As required for accreditation by ARC-PA, program summative activities must include assessment of students across competency domains within the final 4 months of matriculation. In order to meet accreditation standards and document learner competency, program leadership requires satisfactory completion of several unique assessment measures during this time period (Table1).

Table 1 Summary of Program Summative Activities

Assessment Tool	Competency Domain	Data Collection Period	Number of scores collected	Utilization in this Study
End of Curriculum Exam (MCQ)	Medical Knowledge	Month 30	1	Not utilized for analysis
SP Evaluation of Student for Simulation Exam (Appendix A)	Communication Skills	Month 27	4	Research question 1 and 2, reference instrument
Faculty Checklists for Simulation Exam (Appendix B)	Patient Care Skills	Month 27	4	Research question 1 and 3, reference instrument
Supervised Clinical Practice Experience (Appendix C)	Communication Skills, Patient Care Skills	Every four weeks from Month 18-27	9	Research question 2, 3, target instrument

Simulation Assessment

The entire cohort of senior students took a simulation based summative assessment during month 27 of training as a part of the program summative evaluation activities. The study population consisted of the first cohort of students for the MSPA program and in order to fulfill accreditation requirements for summative assessment of graduates, simulation was proposed by program leadership as a means to evaluate students in a standardized manner for domains other than medical knowledge. The simulation exam mirrored the existing assessment structure for learners in the doctor of medicine program at the sponsoring institution.

Simulation encounters (Appendices A and B) took place at the Stanford School of Medicine Goodman Center for Immersive and Simulation-Based Learning (CISL), a 28,000 square foot, state-of-the-art simulation facility that supports immersive learning and assessment activities for the school of medicine. Simulated patient exam rooms equipped with video capture capability were utilized for the exam. Two different camera angles were captured to allow for optimal observation of physical exam skills, and cameras were adjusted by a technician for video data capture. Standardized patients (SPs), trained actors who portray patients with specific medical conditions and concerns, engaged with students during simulation encounters, as well as assessed student communication skills. Students proceeded through four encounters, each twenty minutes in length, with standardized patients in simulated exam rooms. Due to the time and space limitations of the simulation lab, this assessment took place over two days in January of 2020.

The simulation encounters were designed to reflect the Core Competencies for New Physician Assistant Graduates (Physician Assistant Education Association, 2018), as well as the NCCPA content blueprint (National Commission on the Certification of Physician Assistants, 2019), were scripted, and had a fixed time for students to obtain a focused history, perform an appropriate physical exam and provide any necessary patient education or counseling. Standardized patient scripts and assessment documents were developed by members of the MSPA program faculty collaboratively with the Medical Director of the Standardized Patient program at CISL (Appendices A and B).

Physician assistant program faculty completed a checklist reflecting patient care skill competency while viewing the encounter via video feed. Scoring occurred in real time while watching the video feed of the encounter from a different room at CISL. All faculty raters received training regarding the clinical performance evaluative tools prior to the testing event. Video review capability was available, if needed, for clarification and verification purposes.

Standardized patients completed a student communication skills checklist (Appendix A) immediately following each encounter, which included individual items related to performance as well as a global rating scale. Training related to completion of the rating scale occurred for all standardized patients prior to the testing event.

Data Collection

A variety of quantitative data was collected and analyzed for this study. Over the course of the simulation exam eight unique evaluations of each student were conducted. Standardized patients completed an evaluation related to communication skills at each of four simulation stations (Appendix A). Learners had faculty rater scores which reflected history taking, physical exam evaluation for Patient Care Skills at each of four simulation stations (Appendix B).

Standardized Patient Communication Evaluations

A 10-item evaluation, with three rating options (Agree, Somewhat, Disagree) and a global rating scale (Yes, No, Undecided) which reflects professionalism and communication skills was completed by four individual SPs for each learner. (Appendix A) Prior to the testing event, SPs received training on how to rate learners and utilize the evaluative tool. Standardized patients scored students immediately following each encounter.

Faculty Evaluations of Patient Care Skills

Faculty raters viewed student encounters from a different room via live video feed. Recordings were available for immediate review, upon request, for additional verification. Skill checklists with three rating options (Done, Done Incorrectly, and Not Done) were completed by a trained member of the faculty for each student encounter. Checklists varied slightly by case, were between twenty-eight and thirty-four items in length (Appendix B) and were completed in real-time while watching student encounters on video. Each member of the faculty was a rater for only one of the four clinical scenarios utilized for the assessment.

Preceptor Evaluation of Student Survey

All students routinely had clinical performance survey data collected at the conclusion of each supervised clinical training experience as a required component of assessment for the clinical phase of training. Students completed one 4 week-long SCPE in each of the following settings: Pediatrics, Women’s Health, Psychiatry/Behavioral

Health, Surgery, and Emergency Medicine. Two 4 week-long SCPEs were completed for Internal Medicine and Primary Care. At the conclusion of each of the nine clinical training experiences, a student performance evaluation completed by the supervising healthcare provider.

These surveys assessed student performance in the competency domains of patient care and medical knowledge, practice-based learning and improvement, professionalism, communication skills, and systems-based practice. Surveys were administered via secure email and rated on a 5-point Likert scale (Excellent, Very Good, Satisfactory, Poor, Unacceptable, N/A). (Appendix C) For the purposes of this study, eleven survey items related to patient care skills, and nine items are related to communication skills were used for analysis.

Data Storage

Data was collected and managed using REDCap electronic data capture tools. REDCap (Research Electronic Data Capture) is a web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. (Harris, Taylor, Thielke, Payne, Gonzalez and Conde, 2009).

Data collected from Preceptor Evaluation of Students, checklist scores from the simulation exam, and ratings data from standardized patients were uploaded and stored in REDCap. Video data from the standardized patient encounters will remain stored securely in the CISL video database but was not utilized for analysis in this study.

Data Analysis

Data analysis was performed using SPSS software version 26 (Chicago, IL). Descriptive analyses of demographics and assessment ratings were summarized using means and standard deviations for continuous variables and proportions for categorical variables. Standardized patient evaluations of student communication skills and patient care skills were scored and analyzed as continuous variables. Clinical performance assessment checklists for each standardized patient encounter were scored as continuous variables. A two-sided p value of <0.05 and an effect size of >0.8 was considered significant, and moderate correlation of $r=0.5-0.7$, and strong correlation were reflected by $r>0.7$.

Histograms were created with raw data from each assessment and were evaluated for normality. Items with high variability or outliers beyond 2 standard deviations beyond the mean were considered for exclusion from subsequent analysis.

Data collected from the simulation exam was analyzed separately for Patient Care Skills and Communication. The rating scale for correctly completed skills scored a 5, incorrectly completed skills scored a 3, and missing items scored a 1. The sum of ratings for each content area evaluated was established and utilized for subsequent analysis. Rater (both faculty and SP) reliability was reviewed for each of the four clinical scenarios. Additionally, exam performance was analyzed for statistically significant performance differences from day one to day two of exam administration. Data collected from simulation assessment for Patient Care Skills- History, and Patient Care Skills-Physical Exam, and Communication were analyzed separately for construct validity via principle component analysis.

SCPE survey data were treated as continuous variables. The sum of 5-point Likert-scale ratings for measures of Patient Care Skills-History, Patient Care Skills-Physical Exam and Communication were calculated for each clinical practice experience. Survey items with response rates less than 50% were excluded from subsequent analysis.

Table 2 Analyses for Research Questions

Research Question	Related Data	Statistical Tests	Anticipated Results
Simulation assessment reliability and validity	Simulation assessment scores organized by rater, case and assessment date	Descriptive statistics Cronbach's alpha for internal consistency One-way ANOVA post-hoc to identify outlier evaluators Principle component analysis for construct validity	Cronbach's alpha >0.7 indicating good internal consistency One-way ANOVA p>0.05 indicating minimal variance between raters or assessment dates PCA shows a high constructive validity

Correlation between SCPE evaluation and Simulation exam performance for communication skills	SCPE communication scores Simulation exam communication scores	Descriptive statistics Pearson's correlation	Pearson's correlation of r>0.7 indicating strong positive correlation
Correlation between SCPE evaluation and Simulation exam for patient care skills	SCPE Patient Care Skill scores Simulation exam Patient Care Skill scores	Descriptive statistics Pearson's correlation	Pearson's correlation of r>0.7 indicating strong positive correlation

Note. Summary of planned statistical analysis for each of the research questions in this study.

Support for Analysis Practices

Wallenstein, Heron, Santen, Shayne & Ander (2010) conducted a cohort based correlational study of new graduate physicians with performance on a structured simulation exam and ratings of clinical performance across competency domains. They found that performance on structured simulation assessment did correlate with clinical performance in domains of patient care skills, knowledge and practice-based learning but not in professionalism, communication skills, or systems-based practice. In spite of their small sample size (N=18), Wallenstein et al. found that sufficient power existed to detect a correlation of 0.6 (one-tailed alpha of 0.05 and beta 0.2). The proposed study will be similarly structured with use of a single cohort of learners (but a slightly larger sample size), simulation assessment based on competency domains and analysis via correlation with clinical practice evaluation.

In their evaluation of simulation assessment for physician competencies in post-graduate training, Jefferies, Simmons, Tabak, Mcilroy, Lee, Roukema, & Skidmore (2007) established observed structured simulation assessment as a valid and reliable tool for assessing competency across multiple domains of practice. Correlation analysis was utilized to measure alignment between expert ratings in the clinical setting, and simulation assessment measures across different competency domains. Similarly, the proposed study will analyze for correlation of ratings of SPs and of clinical evaluators on communication skills measures. Similar analyses will be conducted for measures of patient care skill with clinical performance ratings and simulation-assessment scores of faculty raters.

When taking sample size into consideration, this analysis is being approached from the lens of an internal pilot study. Ryan (2013, p.21) states that internal pilot studies such as this may be conducted in order to determine estimated parameters for later use in determining sample sizes for larger studies and to establish generalizability. Johansen & Brooks (2010) state that a sample size of 24-30 participants is permissible for valid analysis in pilot studies. On their analysis of increasing sample size on the impact of confidence intervals for correlation analysis, Johansen and Brooks (2010, p.397) found that once the sample size exceeds 24-30, there is a flattening of the curve regardless of the level of correlation (Image 1). This effect would suggest that significantly larger sample size is unlikely to have a significant impact on the correlation once the sample size exceeds 24-30.

Additionally, Bonett & Wright (2000) indicate that Pearson correlation analysis may be used in sample sizes greater than 25. This is similarly reported by Bujang & Baharum, (2016) who identify that to get a statistically significant result ($p<0.05$) and sufficient power (80%) would require a minimum sample size of 29 to detect correlation of $r=0.5$, and minimum sample size of 13 to detect correlation of $r=0.7$. If this study achieves the goal of identifying high correlation ($r>0.7$), a sample size of 27 should have sufficient power. Considering an alpha of 0.05, power >0.8 will be considered sufficient as was the case for Wallenstein et al (2010).

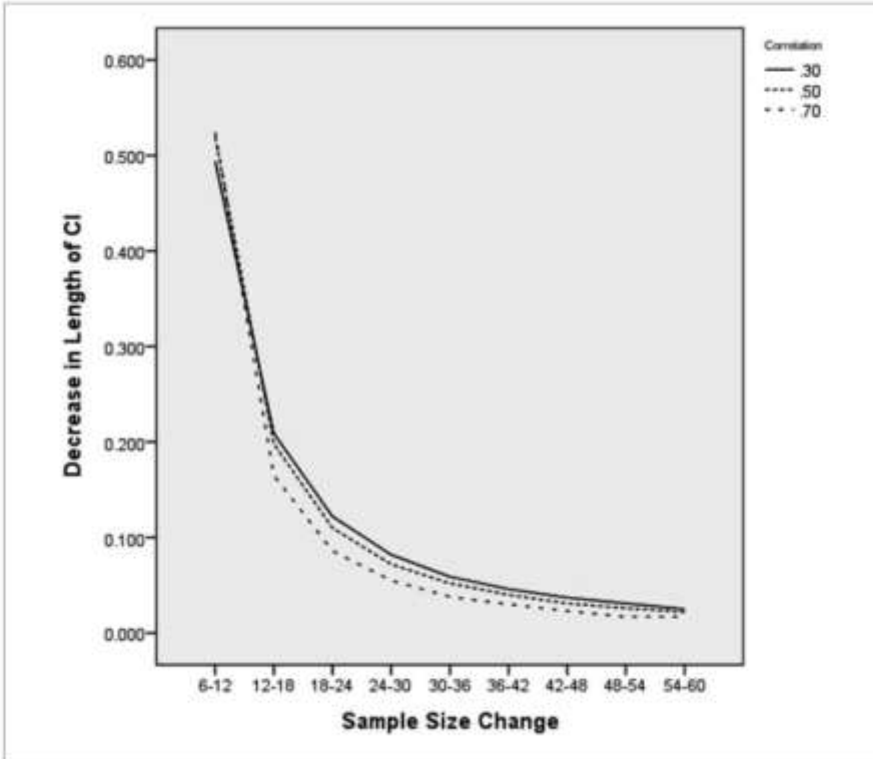

Figure 1 Impact of Increasing Sample Size on Length of Confidence Intervals in Correlation Analysis

Note. Image reprinted from Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement*, *70*(3), 394-400. https://doi-org.stanford.idm.oclc.org/10.1177/0013164409355692

To establish reliability of the simulation exam as a competency-based assessment tool, analysis will be conducted in the manner of Jefferies et al. (2007) with calculation of Cronbach's alpha for interstation reliability. Pell, Fuller, Homer & Roberts (2010) identify Cronbach's alpha >0.7 as a suitable measure of internal consistency for high stakes assessment of standardized patient encounters. Pell et al. (2007) also advocate for the use of one-way ANOVA when assessing interrater and interstation variance. For the proposed study, it is important to establish internal consistency reliability and criterion validity across the examination scores from simulation assessment and SCPE evaluation ratings. This statistical analysis will be performed for and interrater and interstation reliability with Cronbach alpha calculations as well as one-way ANOVA.

Assumptions

Assumptions of the simulation exam include that all of the standardized patients follow the script provided and answer student questions correctly. To mitigate variability, all SPs undergo training prior to assessment deployment and any questions or clarifications can be addressed by members of the faculty.

This simulation assessment scoring also assumes that all faculty are scoring skills similarly on video review. To mitigate variability in this area, all faculty received training about scoring and any questions about scoring are addressed in real time and verified with video review and/or clarification with SPs.

Delimitations and Limitations

Several significant limitations exist with this study. The sample size is currently limited to a single cohort of students, which is a fixed number (n=27). Due to the time and space limitations for conducting the simulation assessment, twelve unique faculty evaluators, and nine standardized patients were involved in the scoring process. The assessment occurred over two days, and while there is a strict honor code in place regarding exam integrity, the possibility remains that information regarding the content of the assessment will be shared between students.

Limitations also exist with relation to clinical practice environments which can be assessed in the simulation lab. For the purposes of this exam, the simulation lab facilities that replicate outpatient clinic offices were utilized and all standardized patients were adults. Women's health presenting complaints, pediatric aged patients, and performance in surgical settings were not assessed.

Expected Results and Impact

Upon correlation analysis, I would anticipate that there would be a strong positive correlation($r>0.7$) between student performance on SCPE evaluations and simulation assessment scores for both communication skills and patient care skills. This would indicate that performance on the simulation assessment is reflective of performance in actual clinical environments, and reflective of competency in the domains of interest. High levels of correlation would support the further use and expansion of simulation-based assessments of competency.

Correlation scores of $r<0.7$ may reflect the limitations of the exam and evaluative settings and patient types available in the simulation lab (no women's health, pediatrics, or surgery) which are captured in the clinical practice evaluations. The impact of variation by scope of clinical practice will be evaluated. Lower than anticipated positive correlation may also be an issue if there are reliability issues with the assessment stations or raters. Minimal variability may be noted in the simulation assessment scores which could complicate analysis. As this is a summative exam, one would assume that all learners will have achieved the minimum competency by the end of the program and as such there may be insufficient score variation for these data to draw conclusions of significance.

Regarding assessment reliability and validity, I anticipate that there will be good interrater reliability ($p>0.05$) as pre-briefing and training of the faculty and standardized patients should be quite consistent. On one-way ANOVA, both interstation and interrater analyses should reach statistical significance ($p>0.05$).

Considering interstation reliability, it is my hope that this will also be strong across all four stations, however each station reflects different specialties of medicine and individual students may find certain subtopics knowledge to be more complex and challenging.

Impacts of this simulation assessment being valid, reliable and with strong correlation to other measures of competency are important to establish if this is to be considered for more widespread use. Future study should include utilization of these assessment tools at other institutions to see if the results are similar. While all programs have differences in educational programming, all are held to the same set of accreditation standards and all learners must meet the same basic demonstration of competency for graduation and clinical practice. Additionally, expansion of the simulation exam to include more varied patient types and clinical settings, as well as designing scenarios which capture other competency domains would be beneficial.

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

The purpose of this quantitative, correlational study was to analyze the construct validity, criterion related validity and reliability of a simulation-based summative assessment for physician assistant students, and to establish if there is correlation between the simulation exam and clinical practice evaluations. More specifically, this study addressed the non-knowledge-based competencies of patient care skills and communication skills. What follows are the analytic procedures and results of the quantitative analysis.

Sample Demographics

Twenty-seven students representing one entire cohort of the Stanford MSPA Program were recruited via convenience sampling. This cohort consisted of twenty-two women, and five men with an average age of 28.9 years (24.9-38.0 years of age).

Simulation Assessment Analysis

Data from the four-station simulation assessment which consisted of Standardized Patient Checklists and Faculty Checklists (Appendix A and B) had no missing values. Scores were tabulated for each simulation scenario in categories of patient care skills (history-taking and physical exam skills) and communication skills. The assessment was conducted over two days where 11 students took the exam day one, and the remaining 16 took the exam day two. Score distributions demonstrated normality (figures 2-5).

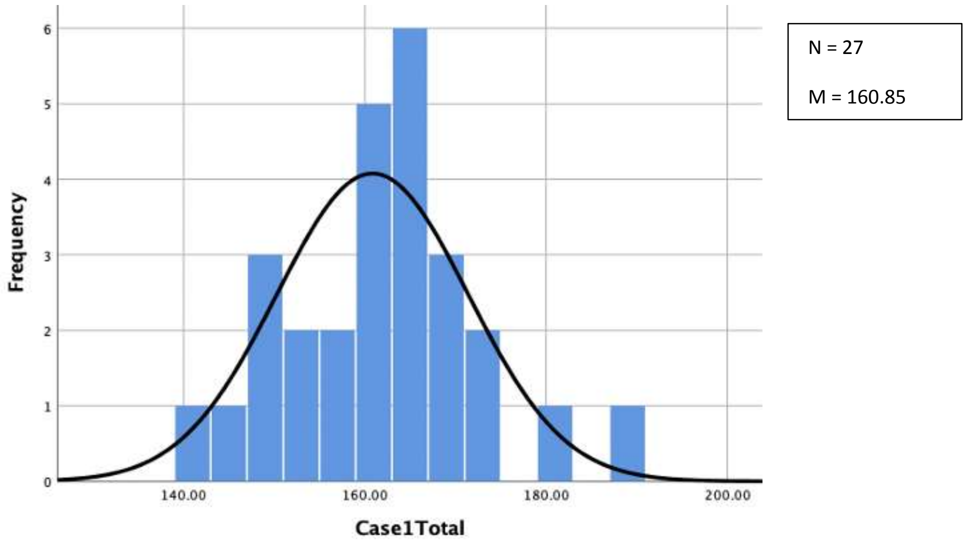

Figure 2 **Histogram for Simulation Exam Scores for Case 1**

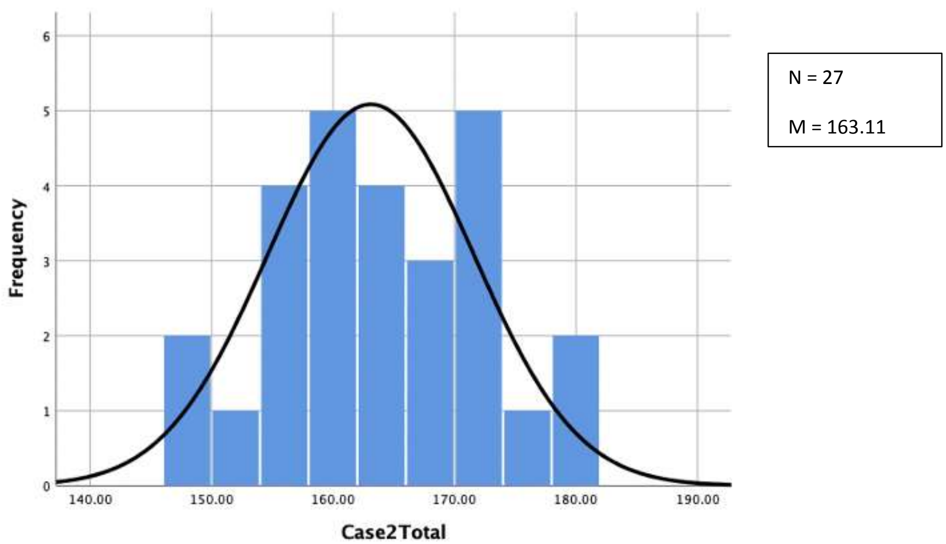

Figure 3 **Histogram for Simulation Exam Scores for Case 2**

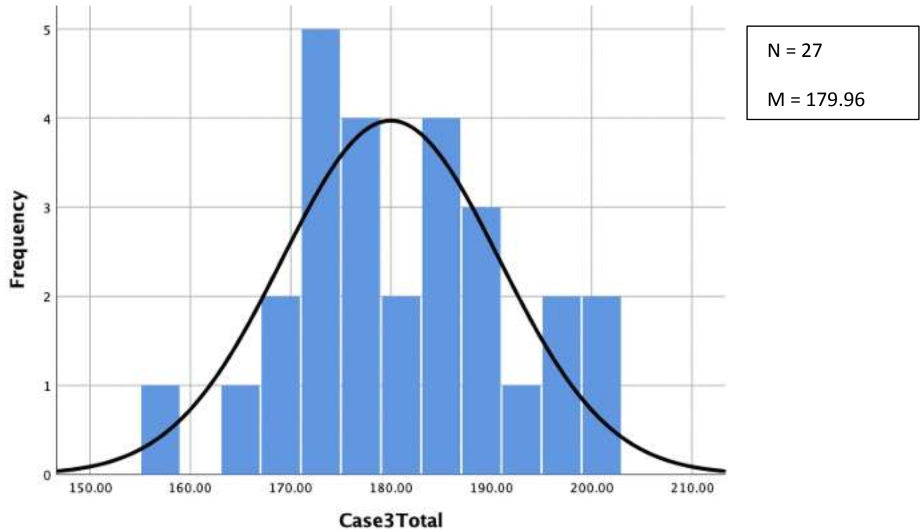

Figure 4 Histogram for Simulation Exam Scores- Case 3

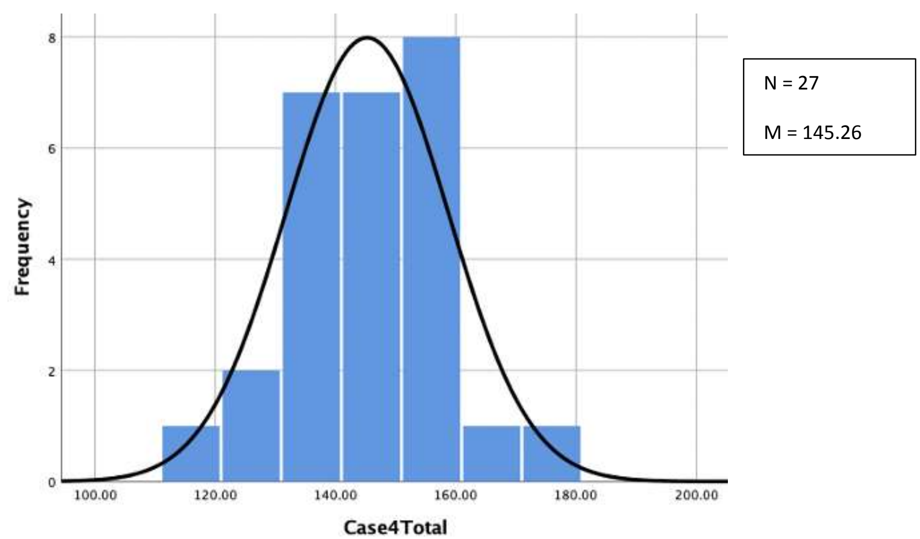

Figure 5 Histogram for Simulation Exam Scores- Case 4

Table 3 **Descriptive statistics for Simulation Exam Communication Scores**

	N	Minimum	Maximum	Mean	Std. Deviation
Communication Case 1	27	39	53	47.52	4.42
Communication Case 2	27	43	55	51.9	2.90
Communication Case 3	27	37	53	47.67	4.93
Communication Case 4	27	39	55	48.26	4.27
Valid N (listwise)	27				

Note. Score distributions demonstrated normality.

Table 4 **Descriptive statistics for Simulation Exam Patient Care Skills Scores**

	N	Minimum	Maximum	Mean	Std. Deviation
Patient Care Skills Case 1	27	90	136	113.33	10.00
Patient Care Skills Case 2	27	93	127	111.15	8.47
Patient Care Skills Case 3	27	108	154	132.30	10.90
Patient Care Skills Case 4	27	77	131	97.00	11.54
Valid N (listwise)	27				

Note. Score distributions demonstrated normality.

The Communication Scores were further analyzed by ANOVA to assess for statistically significant differences between assessment dates (table 5), and raters (table 6). When considering differences with exam and student performance from one testing date to another, a statistically significant difference was noted between scores on exam day one and two for the communication scores for case 2 ($p=0.035$) (Table 5). It is notable, however, that scores decreased on the second day of the exam which would not be suggestive of potentially problematic academic dishonesty due cases being utilized on two separate days.

Table 5 **ANOVA for communication scores between testing dates**

		Sum of Squares	df	Mean Square	F	Sig.
Case 1	Between Groups	2.10	1	2.10	.104	.750
	Within Groups	506.64	25	20.27	4.986	
	Total	508.74	26			
Case 2	Between Groups	36.42	1	36.42	4.99	*0.035
	Within Groups	182.55	25	7.30		
	Total	218.97	26			
Case 3	Between Groups	53.46	1	53.46	2.31	.141
	Within Groups	578.55	25	23.14		
	Total	632.00	26			
Case 4	Between Groups	.71	1	.71	.037	.848
	Within Groups	472.48	25	18.90		
	Total	473.19	26			

Note. Case 2 had a statistically significant communication score difference between day 1 and 2 of the assessment (p=0.035). Subsequent review showed lower scores for the later date which would not be suggestive of academic dishonesty contributing to differences.

Analysis of variation between SP raters for communication skills revealed no statistically significant differences for any of the four cases in the assessment (table 6). This is a reassuring finding and indicative that the Standardized Patient Checklist (Appendix A) and preparatory training of raters was sufficient to mitigate interrater reliability challenges.

Table 6 **ANOVA for communication scores by SP raters**

		Sum of Squares	df	Mean Square	F	Sig.
Case 1	Between Groups	79.31	2	39.65	2.22	.131
	Within Groups	429.43	24	17.89		
	Total	508.74	26			
Case 2	Between Groups	33.86	2	16.93	2.20	.133
	Within Groups	185.10	24	7.71		
	Total	218.97	26			
Case 3	Between Groups	72.86	2	36.43	1.56	.230
	Within Groups	559.14	24	23.30		
	Total	632.00	26			
Case 4	Between Groups	12.69	2	6.34	.331	.722
	Within Groups	460.50	24	19.19		
	Total	473.19	26			

Note. No statistically significant differences were noted for communication ratings by SPs between cases indicating that there were no outlier evaluators.

When patient care skill ratings were analyzed for variance, there was no statistically significant difference in patient care skill scores between testing dates (table 7). Administration of the exam over multiple testing dates did not reflect evidence of academic dishonesty.

Table 7 **ANOVA for patient care skill scores between testing dates**

		Sum of Squares	df	Mean Square	F	Sig.
Case 1	Between Groups	33.00	1	33.0	.32	.576
	Within Groups	2567.00	25	102.680		
	Total	2600.00	26			
Case 2	Between Groups	106.68	1	106.68	1.52	.229
	Within Groups	1756.73	25	70.27		
	Total	1863.41	26			
Case 3	Between Groups	53.88	1	53.88	.44	.511
	Within Groups	3031.75	25	121.27		
	Total	3085.63	26			
Case 4	Between Groups	414.82	1	414.82	3.40	.077
	Within Groups	3049.18	25	121.97		
	Total	3464.00	26			

Note. No statistically significant differences were noted for patient care skill scores between testing dates. These results suggest that administration of the exam over several days is not problematic from an assessment integrity perspective.

Considering variability among faculty raters, there was a statistically significant rating difference for Case 4 ($p=0.04$) as noted on table 8. This distinction can be attributed to a single outlier faculty rater. It is also notable that case 4 was the only case which had four different faculty raters compared to two or three for other cases in the assessment. Subsequent analysis for case 4 attributes this variability largely to history taking scoring from a single rater ($p<.001$) and may indicate challenges with either interpretation of the rating tool, pre-briefing of faculty, or rater bias related to scoring specific items.

Table 8 ANOVA for patient care skill scores between faculty raters

		Sum of Squares	df	Mean Square	F	Sig.
Case 1	Between Groups	205.41	2	102.71	1.03	.372
	Within Groups	2394.59	24	99.775		
	Total	2600.00	26			
Case 2	Between Groups	152.62	1	152.62	2.23	.148
	Within Groups	1710.80	25	68.43		
	Total	1863.41	26			
Case 3	Between Groups	137.25	2	68.62	.559	.579
	Within Groups	2948.38	24	122.85		
	Total	3085.63	26			
Case 4	Between Groups	1031.00	3	343.67	3.25	*.040
	Within Groups	2433.00	23	105.78		
	Total	3464.00	26			

Note. A statistically significant difference was found for faculty scoring in case 4 (p=0.04). Subsequent review attributed this to a single faculty rater outlier.

When conducting reliability analysis for internal consistency of the communication scores for the simulation exam, Cronbach's alpha of 0.591 was calculated (table 9). The minimum acceptable value for internal consistency was a Cronbach's alpha of 0.7, and as such, optimization of the rating scale should be considered prior to redeployment.

Since each case had distinctive clinical content, Patient Care Skills were evaluated for internal consistency on a case by case basis. Cronbach's alpha for Patient Care Skills scores (history taking and physical exam skills in aggregate) for the four cases in the simulation exam were 0.461, 0.1, 0.494, and 0.43 respectively (table 9). Poor internal consistency was noted for all cases, but particularly for case 2.

Table 9 **Reliability statistics for the simulation assessment components**

		Cronbach's Alpha	N of items
Communication	Cases 1-4	.591	11
Patient Care Skills (History taking + physical exam skills)	Case 1	.461	28
	Case 2	.100	34
	Case 3	.494	34
	Case 4	.430	29
History taking	Cases 1-4	.512	44
Physical Exam Skills	Cases 1-4	.502	60

Note. All measures for internal consistency fell below acceptable thresholds when Cronbach's alpha was calculated indicating poor reliability of the assessment instruments.

Principle component analysis was then conducted to assess construct validity. The communication checklist component of the simulation exam should be measuring a single construct; communication skills of the student being assessed. Review of the principle component analysis and extraction data (table 10) indicates that validity of the communication assessment would be improved by removing items 7 and 9. When single component extraction was conducted, the items on this checklist can explain only 27.465% of variance which falls below the acceptable minimum threshold of 50% indicating that multiple constructs are being assessed.

Table 10 **Principal Component Analysis for communication assessment**

	Initial	Extraction	Cronbach's alpha if item deleted	Corrected Item-Total Correlation
Comm Item 1	1	.494	.534	.540
Comm Item 2	1	.624	.513	.437
Comm Item 3	1	.361	.537	.408
Comm Item 4	1	.206	.565	.270
Comm Item 5	1	.099	.573	.243
Comm Item 6	1	.093	.591	.112
Comm Item 7	1	.016	.639	*-.163
Comm Item 8	1	.746	.439	.668
Comm Item 9	1	.048	.664	*-.135
Comm Item 10	1	.144	.555	.346
Comm Item 11	1	.190	.559	.308

Note. Items 7 and 9 on the Communications rating scale were found to have negative values on item total correlation indicating that they may not contribute to reliability of the rating scale and should be considered for exclusion.

Table 11 **Total Variance Explained for communication assessment**

	Initial Eigenvalues			**Extraction sums of squared loadings**		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.021	27.465	27.465	3.021	27.465	27.465
2	1.668	15.162	42.626			
3	1.361	12.370	54.997			
4	1.138	10.344	65.340			
5	.874	7.947	73.287			
6	.773	7.025	80.312			
7	.625	5.682	85.994			
8	.588	5.341	91.336			
9	.452	4.106	95.442			
10	.257	2.336	97.778			
11	.244	2.222	100.000			

If communication items 7 and 9 are both excluded, the Cronbach's alpha for the communication rating scale is improved to .709 which is within the lower acceptable range for internal consistency. Removal of items 7 and 9 also improves the total variance explained from 27.465% to 33.073%, but this remains reflective of the presence of multiple constructs measured by this component of the assessment.

Turning attention to the measures of patient care skills, raw data was recoded to reflect a score of zero for incorrectly executed and missing items, and a score of five for correctly completed items. The decision to recode in this way was made on the assumption that in actual patient care situations, an incorrectly performed skill will not yield clinically useful information to the clinician. Items with no variability among examinees were omitted from subsequent analysis. Omitted items are summarized in table 12.

Table 12 **Patient care skills assessment items omitted from analysis due to lack of variability**

Case	Item Type	Item number	Description	Rationale for omission
1	History	1	How did the pain begin?	All received full credit
		2	How long has the pain lasted?	All received full credit
	Exam skills	10	Discussed differential diagnosis	All received full credit
		11	Addressed concerns re: medical history	All received full credit
		16	Explained exams in advance	All received full credit
2	History	1	How long have you been short of breath?	All received full credit
	Exam skills	9	Discussed differential diagnosis	All received full credit
		15	Explained exams in advance	All received full credit
3	History	2	Describe the location of the pain	All received full credit

		4	Describe the duration of pain	All received full credit
	Exam skills	10	Discussed differential diagnosis	All received full credit
		16	Hand hygiene	All received full credit
		17	Explained exams in advance	All received full credit
4	History	1	Describe the quality of pain	All received full credit
	Exam skills	16	Hand hygiene	All received full credit
		17	Explained exams in advance	All received full credit
		18	Discussed findings of physical exam	All received full credit

Construct validity was then assessed separately for History taking (table 13 and 14) Exam Skills (table 15 and 16). For the History taking components of the assessment, reliability was suboptimal with many low values for extraction. Aggregate construct validity for history taking components of the assessment was .512 which falls below acceptable thresholds (table 9), and there were no specific items that could be identified for exclusion that would substantively improve the performance to an acceptable value (table 13). When considering that history taking should be reducible to a single target construct, these checklists could only explain for 11.359% of respondent variance (table 14) which is suboptimal.

Table 13 **Factor Analysis for Patient Care Skill – History Taking (cases 1-4)**

	Initial	Extraction	Cronbach's alpha if item deleted
History Case 1 Item 3	1	.409	.499
History Case 1 Item 4	1	.022	.517
History Case 1 Item 5	1	.045	.516
History Case 1 Item 6	1	2.991E-5	.512
History Case 1 Item 7	1	.308	.509
History Case 1 Item 8	1	.017	.495
History Case 1 Item 9	1	.099	.499
History Case 1 Item 10	1	.061	.523
History Case 1 Item 11	1	.462	.499
History Case 2 Item 2	1	.280	.518
History Case 2 Item3	1	.082	.500
History Case 2 Item4	1	.133	.491
History Case 2 Item5	1	.032	.522

History Case 2 Item6	1	.011	.505
History Case 2 Item7	1	.011	.494
History Case 2 Item8	1	.016	.506
History Case 2 Item9	1	.401	.509
History Case 2 Item 10	1	.020	.513
History Case 2 Item 11	1	.053	.505
History Case 2 Item 12	1	.014	.490
History Case 2 Item 13	1	.014	.532
History Case 3 Item 1	1	.069	.522
History Case 3 Item 3	1	.053	.508
History Case 3 Item 5	1	.027	.551
History Case 3 Item 6	1	.060	.514
History Case 3 Item 7	1	.000	.514
History Case 3 Item 8	1	.020	.539
History Case 3 Item 9	1	.254	.520
History Case 3 Item 10	1	.411	.497

History Case 3 Item 11	1	.288	.493
History Case 3 Item 12	1	.003	.520
History Case 3 Item 13	1	.028	.463
History Case 3 Item 14	1	.007	.500
History Case 3 Item 15	1	.009	.510
History Case 4 Item 2	1	.065	.492
History Case 4 Item 3	1	.003	.492
History Case 4 Item 4	1	.156	.511
History Case 4 Item 5	1	.107	.500
History Case 4 Item 6	1	.018	.494
History Case 4 Item 7	1	.130	.506
History Case 4 Item 8	1	.095	.455
History Case 4 Item 9	1	.262	.474
History Case 4 Item 10	1	.316	.500
History Case 4 Item 11	1	.125	.509

Table 14 Total Variance Explained for Patient Care Skill- History Taking (Cases 1-4)

	Initial Eigenvalues			**Extraction of Sums Squared Loadings**		
Comp	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.998	11.359	11.359	4.998	11.359	11.359
2	39.89	9.066	20.425			
3	3.739	8.498	28.923			
4	3.466	7.877	36.800			
5	3.079	6.998	43.797			
6	2.853	6.484	50.282			
7	2.449	5.567	55.848			
8	2.404	5.464	61.201			
9	2.151	4.889	66.201			
10	1.775	4.035	70.236			
11	1.701	3.865	74.101			

12	1.580	3.590	77.691
13	1.486	3.378	81.069
14	1.366	3.105	84.174
15	1.208	2.746	86.920
16	.969	2.201	89.121
17	.896	2.037	91.158
18	.792	1.799	92.957
19	.740	1.683	94.640
20	.647	1.471	96.111
21	.466	1.059	97.170
22	.414	.941	98.111
23	.366	.832	98.943
24	.198	.450	99.393
25	.151	.344	99.736
26	.116	.264	100.000
27	1.536E-15	3.490E-15	100.000

28	1.112E-15	2.527E-15	100.000
29	7.113E-16	1.616E-15	100.000
30	4.923E-16	1.119E-15	100.000
31	3.763E-16	8.551E-16	100.000
32	2.320E-16	5.272E-16	100.000
33	1.645E-16	3.740E-16	100.000
34	3.125E-17	7.306E-17	100.000
35	-1.932E-17	-4.392E-17	100.000
36	-1.830E-16	-4.160E-16	100.000
37	-2.671E-16	-6.070E-16	100.000
38	-4.002E-16	-9.094E-16	100.000
39	-5.596E-16	-1.272E-15	100.000
40	-7.515E-16	-1.708E-15	100.000
41	-8.917E-16	-2.027E15	100.000
42	-1.013E-15	-2.302E-15	100.000
43	-1.202E-15	-2.731E-15	100.000

44	-1.707E-15	-3.880E-15	100.000

Factor analysis of physical exam skill items also revealed a large proportion items with very low extraction value (table 15), and an aggregate Cronbach's alpha of .502 (table 9). While there are several items that could be considered for exclusion to improve construct validity, none would substantially improve the performance of the physical exam portion of the assessment (table 15). In analysis of the ability of the assessment of physical exam skills to reflect a single construct, the checklists could only explain 10.545% of respondent variability (table 16). These measures indicate suboptimal construct validity for the simulation assessment for the measurement of patient care skills, for both history taking and physical exam skills.

Table 15 Factor Analysis for Patient Care Skills – Physical Exam (cases 1-4)

	Initial	**Extraction**	**Cronbach's alpha if item is deleted**
Case 1 Skill 1	1	.086	.526
Case 1 Skill 2	1	.004	.493
Case 1 Skill 3	1	.282	.459
Case 1 Skill 4	1	.244	.469
Case 1 Skill 5	1	.047	.507
Case 1 Skill 6	1	.270	.511
Case 1 Skill 7	1	.153	.525
Case 1 Skill 8	1	.008	.503
Case 1 Skill 9	1	.012	.499
Case 1 Skill 12	1	.010	.501
Case 1 Skill 13	1	.001	.508
Case 1 Skill 14	1	.003	.508
Case 1 Skill 15	1	.016	.501

Case 1 Skill 17	1	.200	.521
Case 2 Skill 1	1	.024	.486
Case 2 Skill 2	1	.087	.513
Case 2 Skill 3	1	3.765E-5	.497
Case 2 Skill 4	1	.381	.475
Case 2 Skill 5	1	.037	.501
Case 2 Skill 6	1	.125	.470
Case 2 Skill 7	1	4.791E-6	.495
Case 2 Skill 8	1	.252	.484
Case 2 Skill 10	1	.006	.510
Case 2 Skill 11	1	.020	.506
Case 2 Skill 12	1	.012	.515
Case 2 Skill 13	1	.124	.534
Case 2 Skill 14	1	.176	.484
Case 2 Skill 16	1	.335	.468
Case 3 Skill 1	1	.118	.505

Case 3 Skill 2	1	.001	.469
Case 3 Skill 3	1	.419	.476
Case 3 Skill 4	1	.506	.463
Case 3 Skill 5	1	.046	.509
Case 3 Skill 6	1	.060	.484
Case 3 Skill 7	1	.104	.493
Case 3 Skill 8	1	.055	.493
Case 3 Skill 9	1	.009	.506
Case 3 Skill 11	1	.082	.507
Case 3 Skill 12	1	.222	.500
Case 3 Skill 13	1	.078	.502
Case 3 Skill 14	1	.217	.521
Case 3 Skill 15	1	.003	.497
Case 3 Skill 18	1	.003	.509
Case 3 Skill 19	1	.251	.481
Case 4 Skill 1	1	.031	.505

Case 4 Skill 2	1	.082	.502
Case 4 Skill 3	1	.066	.512
Case 4 Skill 4	1	.080	.478
Case 4 Skill 5	1	.003	.475
Case 4 Skill 6	1	.001	.489
Case 4 Skill 7	1	.039	.528
Case 4 Skill 8	1	.301	.458
Case 4 Skill 9	1	.080	.476
Case 4 Skill 10	1	.098	.473
Case 4 Skill 11	1	.019	.481
Case 4 Skill 12	1	.123	.518
Case 4 Skill 13	1	.099	.494
Case 4 Skill 14	1	.105	.539
Case 4 Skill 15	1	.005	.497

Table 16 **Total Variance Explained for Patient Care Skills- physical exam skills (cases 1-4)**

	Initial Eigenvalues			Extraction of Sums Squared Loadings		
Comp.	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.222	10.545	10.545	6.222	10.545	10.545
2	5.537	9.384	19.929			
3	4.402	7.160	27.390			
4	4.225	7.160	34.550			
5	3.825	6.482	41.032			
6	3.710	6.288	47.321			
7	3.271	5.544	52.865			
8	3.241	5.492	58.357			
9	2.759	4.676	63.033			
10	2.562	4.342	67.375			
11	2.393	4.056	71.430			

12	2.080	3.526	74.957
13	1.861	3.154	78.111
14	1.763	2.987	81.098
15	1.683	2.852	83.950
16	1.505	2.550	86.500
17	1.303	2.209	88.710
18	1.160	1.967	90.676
19	1.051	1.782	92.458
20	1.013	1.716	94.174
21	.771	1.307	95.481
22	.635	1.076	96.558
23	.609	1.033	97.591
24	.563	.954	98.545
25	.455	.771	99.316
26	.403	.684	100.000
27	1.426E-15	2.416E-15	100.000

28	1.277E-15	2.164E-15	100.000
29	1.163E-15	1.972E-15	100.000
30	9.956E-16	1.687E-15	100.000
31	9.223E-16	1.563E-15	100.000
32	8.138E-16	1.379E-15	100.000
33	6.991E-16	1.185E-15	100.000
34	6.208E-16	1.052E-15	100.000
35	5.603E-16	9.497E-16	100.000
36	4.705E-16	7.975E-16	100.000
37	4.312E-16	7.308E-16	100.000
38	3.316E-16	5.621E-16	100.000
39	1.590E-16	2.696E-16	100.000
40	1.380E-16	2.340E-16	100.000
41	4.756E-17	8.060E-17	100.000
42	-2.922E-17	-4.953E-17	100.000
43	-1.559E-16	-2642E-16	100.000

44	-1.740E-16	-2.949E-16	100.000
45	-2.463E-16	-4.175E-16	100.000
46	-2.744E-16	-4.650E-16	100.000
47	-3.234E-16	-5.482E-16	100.000
48	-3.872E-16	-6.562E-16	100.000
49	-4.425E-16	-7.500E-16	100.000
50	-4.897E-16	-8.300E-16	100.000
51	-5.330E-16	-9.034E-16	100.000
52	-6.918E-16	-1.173E-15	100.000
53	-7.467E-16	-1.266E-15	100.000
54	-8.419E-16	-1.427E-15	100.000
55	-9.328E-16	-1.581E-15	100.000
56	-1.063E-15	-1.802E-15	100.000
57	-1.188E-15	-2.014E-15	100.000
58	-1.371E-15	-2.324E-15	100.000
59	-1.875E-15	-3.178E-15	100.000

In an effort to more precisely identify individual assessment items that should be considered for exclusion, history-taking and exams skills were then analyzed individually for each case of the assessment. Following the removal of items with no variability (table 12), reliability statistics were recalculated and principle component analysis was conducted. Items with negative values on item total statistics were then omitted and reliability statistics were recalculated. The summary of recalculated construct validity is presented in table 16. While reliability was improved in all cases, none of the four cases reached the minimum threshold of Cronbach's alpha >0.7.

Table 16 **Stepwise recalculation of internal consistency for Patient Care Skills**

		Cronbach's alpha		
		All Items Included	No respondent variability removed	Items unrelated to construct removed
Case 1	History	.393	.402	.499
	N	11	9	7
	Exam Skills	.153	.155	.523
	N	17	15	9
	All PCS	.461		.605
	N	28		16
Case 2	History	.405	.407	.548
	N	13	12	9
	Exam Skills	-.012	-.012	-.565
	N	17	15	6

	All PCS	.100		.470
	N	34		15
Case 3	History	-.127	-.128	.431
	N	15	13	5
	Exam Skills	.491	.472	.599
	N	19	16	13
	All PCS	.494		.544
	N	34		18
Case 4	History	.521	.526	.637
	N	11	10	8
	Exam Skills	.114	.116	.502
	N	18	15	9

All PCS	.430	.611
N	29	17

Note. Cases from the simulation exam were analyzed individually for patient care skills construct validity. Cronbach's alpha was recalculated after removing items with no variability and again when items with negative values on corrected item-total correlation were excluded. Internal consistency remained below acceptable thresholds for all cases.

Supervised Clinical Practice Experience Analysis

Supervised clinical practice experience survey data was collected for nine clinical experiences: Primary Care (PC) 1 and 2, Internal Medicine (IM) 1 and 2, Emergency Medicine (EM), Pediatrics (Peds), Women's Health (WH) and Psychiatry/Behavioral Medicine (BH). For the purposes of this study, nine items related to communication skills and eleven items related to patient care skills were collected analyzed (Appendix C). In the few cases where an isolated survey data point was omitted, the mean value for the other collected data points for that item was utilized. Patient Care Skills item 5 was reported as not applicable in 42% of responses (n=102) and so it was excluded from subsequent analyses. Of note, this item referred to procedural skill competency and this skill is not pertinent in all clinical settings. Descriptive statistics for SCPE communication and patient care skills are reported in table 17 by clinical specialty.

Table 17 Descriptive Statistics for SCPE ratings in Communication and Patient Care Skills by specialty

		IM1	IM2	PC1	PC2	EM	Peds	WH	BH	Surg
Comm	N	27	27	27	27	27	27	27	27	27
	Min	27	15	25	27	27	15	31	35	27
	Max	45	45	45	45	45	45	45	45	45
	Mean	40.89	40.44	41.25	42.54	41.81	40.01	40.83	39.44	37.61
	SD	3.945	6.750	5.399	4.102	5.249	6.680	4.427	3.154	5.153
Patient Care Skills	Min	28	30	29.5	30	30	30	32	33	30
	Max	50	50	50	50	48	50	50	46	50
	Mean	40.07	39.90	43.26	44.66	43.26	43.21	42.62	39.85	39.57
	SD	6.63	4.91	6.01	5.71	4.82	6.67	6.30	3.19	6.01

Note. SCPE rating scores demonstrated normality across all settings.

For each student, the sum of all communication ratings from the nine required SCPE were calculated and correlated with the sum of communication scores from the 4 stations of the simulation assessment. On correlation analysis of communication ratings, no statistically significant correlation was noted between the simulation assessment scores and the SCPE ratings (fig. 6). Statistically significant correlation was not identified when all SCPE were included in aggregate, when only the most similar SCPE to the simulation exam were evaluated (primary care, internal medicine, and emergency medicine) (table 18) and when each SCPE was analyzed individually (table 19).

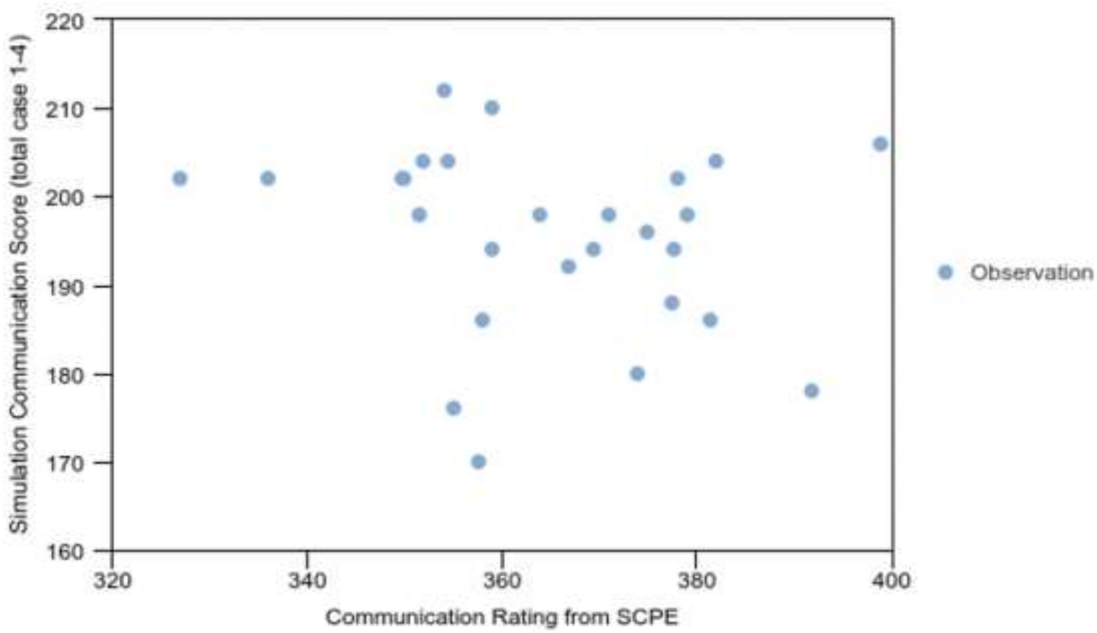

Figure 6. Scatterplot of Communication Scores from simulation exam (cases 1-4) and SCPE Communication ratings (all settings)

Table 18 Correlation between simulation exam communication scores and SCPE communication ratings

		Simulation Exam	SCPE All Settings	SCPE similar settings
Sim Exam all values	Pearson Correlation	1	-.207	-.191
	Sig. (2-tailed)		.299	.339
	N	27	27	27
Sim Exam Items 7,9 excluded	Pearson Correlation		-.166	-.174
	Sig. (2-tailed)		.407	.386
	N	27	27	27

Note. No statistical significance was identified between simulation assessment performance and SCPE performance for measures of communication when all items were included in the analysis, and when low performing items (7, and 9) were omitted.

Table 19 **Correlation between simulation exam communication scores and SCPE ratings by specific clinical practice area**

		Sim	IM1	IM2	PC1	PC2	EM	Peds	WH	BH	Surg
Sim Exam	Pearson Correlation	1	-.287	-.238	.170	-.257	.154	-.078	-.052	.189	-.229
	Sig. (2-tailed)		.147	.233	.395	.195	.443	.701	.798	.345	.251
	N	27	27	27	27	27	27	27	27	27	27

Note. No statistical significance was identified for correlation between simulation exam performance and measures of communication skills in any specific clinical setting.

The sum of patient care skill scores from the simulation exam cases one, three and four were calculated. Case 2 was excluded from subsequent analysis due to extremely poor reliability. The sum of patient care skill scores from nine required SCPE was calculated and utilized for correlation analysis. When patient care skills from the simulation exam and SCPE evaluations were analyzed with correlation, no statistically significant correlations were noted for SCPE taken in aggregate (fig. 7) or when the most similar clinical experiences (primary care, internal medicine, and emergency medicine) were reviewed (table 20). When correlation was analyzed between the simulation exam and SCPE ratings by specific clinical practice area, there was a single statistically significant finding. Patient care skill ratings in women's health were noted to have a moderately negative correlation with scores from the simulation exam ($r = -.436$, $p = .023$).

Table 20 Correlation between simulation exam Patient Care Skill scores (case 1, 3, 4) and SCPE Patient Care Skill ratings

		Simulation Exam	SCPE All Settings	SCPE similar settings
Simulation Exam	Pearson Correlation	1	-.121	.122
	Sig. (2-tailed)		.549	.545
	N	27	27	27

Note. No statistical significance was identified between simulation assessment performance and SCPE performance for measures of communication.

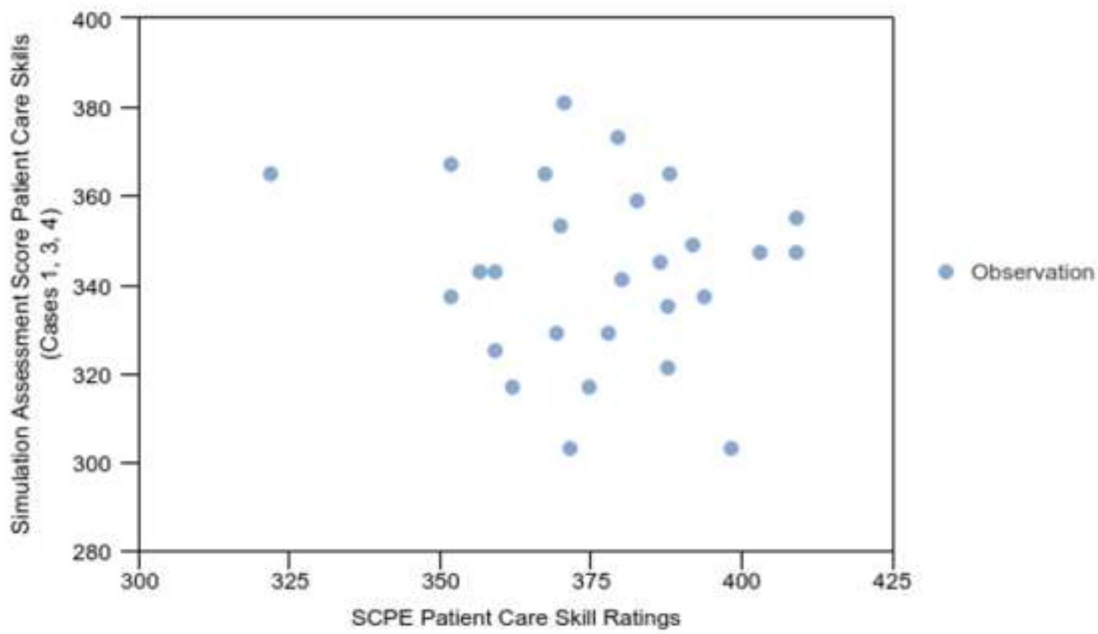

Figure 7. **Scatterplot of Patient Care Skill Scores for Simulation Exam (cases 1, 3, 4) and SCPE Patient Care Skill Ratings (all settings)**

Table 21 **Correlation between simulation exam Patient Care Skill Scores (case 1, 3, 4) and SCPE Patient Care Skill ratings by clinical practice area**

		Sim	IM1	IM2	PC1	PC2	EM	Peds	WH	BH	Surg
Sim Exam	Pearson Correlation	1	.187	-.146	.313	.049	-.242	-.022	-.436	-.243	-.036
	Sig. (2-tailed)		.351	.468	.112	.807	.224	.913	.023*	.223	.857
	N	27	27	27	27	27	27	27	27	27	27

Note. Statistical significance was identified for negative correlation between simulation exam performance and measures of patient care skills in women's health. No other statistically significant correlations were noted for correlation between simulation exam performance and measures of patient care skills in any specific clinical setting.

Summary

Considering research question number one, this particular simulation assessment can be revised to meet the minimum level of validity for utilization as a high-stakes assessment with regard to evaluation of communication skills, but not to assess patient care skills. SPs evaluations of communication were more reliable than faculty assessment of patient care skills for some exam cases. A single faculty rater outlier was identified for case 4 of the simulation exam, but otherwise, rater performance among faculty was consistent. Faculty-rater related factors should be reviewed to identify areas for clarity and optimization. There was largely no difference in student performance from day one to day two of the exam administration which suggests that there were not issues with integrity of the assessment materials or academic dishonesty.

Reliability analysis for the communication checklist approached the suitable threshold when it was administered, but with omission of two items, was improved to suitable minimum thresholds for use in a high-stakes assessment.

When checklists for patient care skills were analyzed for internal consistency for each case, results for case 2 were poor. Patient care skill rating scales for cases 1, 3, and 4 also did not achieve a suitable level of internal consistency, even with recalculation after the omission of problematic items. Specific assessment components responsible for poor reliability varied by case. These results indicate that substantive revision and additional analysis should be completed for patient care skill assessment.

Constructive validity analysis revealed suboptimal results for measures of communication, history taking and physical exam skills when conducted across the exam as a whole. Results for history taking and physical exam skills were particularly poor which may be due in part to the high complexity of the simulation exam scenarios that were assessed.

Considering research question two regarding the identification of correlation between simulation assessment and SCPE evaluations on student communication skills, no statistically significant positive correlations were identified. Failure to achieve meaningful correlation was observed for individual clinical experiences as well as for SCPE aggregate data.

With regard to research question number 3, related to identification of correlation between SCPE ratings for patient care skills and simulation assessment scores for patient care skills, there was no identified statistically significant positive correlation noted. Correlation was not identified for aggregate data for SCPE experiences or for individual clinical rotations. Due to the particularly poor performance of the patient care skill measures of the simulation exam, interpretation of correlation analysis may not be reliable for this domain of competency.

CHAPTER FIVE: DISCUSSION

Medical educators are tasked with ensuring that learners develop a broad-based set of skills that span multiple domains of performance including knowledge, effective communication, and technical skill performance. Assessing readiness for practice and competency of physician learners is achieved through varied evaluative tools which include examinations of factual knowledge, simulation lab assessments of skill, and survey data from a variety of sources (patients, peers, and supervising clinicians).

For physician assistant educators, the recently published Core Competencies for New Physician Assistant Graduates (PAEA, 2018) establishes a competency-based framework for PA student skill acquisition. Ensuring graduate competency in a range of domains has also been added to the most recent set of accreditation standards by ARC-PA (2019).

While medical schools have had valid and established assessment strategies for competency-based measures for a number of years, no such framework exits to support these initiatives for Physician Assistant programs. The purpose of this study was to pilot a simulation-based competency assessment for Physician Assistant learners approaching program completion. This assessment was created to assess learner competency in the domains of communication and patient care skills.

Research Question 1

Question 1 set out to determine the construct validity of components of the simulation assessment and to determine adequacy for utilization as a high-stakes assessment. A summary of recommendations to improve the performance of the simulation exam is compiled in table 22.

Communication

Standardized patients proved to be reliable raters of student communication skills and no outlier raters were identified. All stations of the simulation exam performed similarly with regard to communication ratings. This aligns with the findings of Chang, Mann, Sommer, Fallar, Weinberg, and Friedman (2017) which support the use of SP's in the evaluation of learner communication skills.

Analysis of the communication checklist for internal consistency fell just below the acceptable threshold but was able to meet the minimum accepted threshold with the omission of two items. Omitted items reflected the following: the student adapted to my level of understanding (item 7), and the student encouraged me to ask questions (item 9). Subjectivity of rater responses for item 7 may be problematic as this item may yield a variety of interpretations by raters based on their own knowledge or understanding of the disease processes, or their perceptions of what a typical lay person might know.

Regarding item 9 (the student encouraged me to ask questions), interactions for the simulation exam had a fixed time limit and soliciting and answering questions from the SP could have potentially utilized time needed for other key items accounted for on the assessment checklists. Time management considerations on the part of the student due to the structure of the exam may have influenced scoring for this item. Removal of items 7 and 9 is recommended for future use of this communication skills rating scale and these edits bring the reliability of the rating scale into the acceptable range.

While the omission of items 7 and 9 increases the internal consistency of the assessment component to the acceptable range, this portion of the simulation assessment remained without statistical significance to SCPE ratings regarding communication.

Patient Care Skills

Overall, use of the simulation exam for the evaluation of Patient Care Skills did not achieve acceptable thresholds for performance. When item total statistics and total variance explained by the exam were reviewed, the exam performed quite poorly related to constructive validity and reliability. Unlike with the communications ratings where there were clearly two problematic items on the assessment, there were no clearly identified items to delete which would have improved assessment performance, or there were a significant percentage of items that were not related to the construct. Since such a large number of items proved to be problematic history and physical exam components of the exam, no additional analyses or substantive revisions to the assessment were conducted. Analysis of the performance of the assessment will be evaluated for factors related to the exam administration, factors related to the raters, factors related to assessment development and content, and student-related factors.

Exam administration factors

Patient care skill scoring was consistent from day one to day 2 of the exam administration. There is no evidence that examination administration over multiple days negatively impacted the performance or integrity of the assessment. Due to the physical space limitations of the simulation lab, and time limitations of the standardized patient actors, eight faculty raters were required at all times. The introduction of additional raters for the examination has the potential to increase scoring variability or may inadvertently result in the inclusion of outlier raters.

Video review of the simulation encounters may also be reviewed to better understand student performance or faculty accuracy on scoring the exam. Recall that this exam was video recorded but was scored by faculty who watched the video feed of the encounter in real time. The footage of each assessment encounter was saved and remains available for review or auditing. Auditing of student performance on the video footage and comparing scoring on the faculty checklists may be indicated to ensure that scoring was accurate. If live scoring is found to be inaccurate, this presents an opportunity for faculty scoring to be completed exclusively via video review. Not being reliant on the specific time constraints of the exam, and utilizing video review in lieu of live scoring could also allow for the use of fewer raters and would reduce challenges related to scheduling for faculty.

Rater-related factors

When considering the performance of faculty raters in scoring patient care skills for this assessment, one outlier evaluator was identified. It is notable that each member of the faculty has a different clinical area of focus, varied teaching responsibilities within the program, and several different professional designations are represented (PA, MD, and NP). The outlier rater has extensive specialty clinical experience in the content area for the case that they observed but is not routinely involved in instruction of clinical skills for students. The variability in rater professional experiences may introduce bias or influence perceptions of student performance by the rater. For future simulation activities it may be advisable to select raters from a more specific pool of faculty, such as those who are primarily involved with the instruction for history-taking and physical exam skills course. Additionally, engaging faculty raters in a peer review process for the cases and scoring rubric may improve consistency of raters (Tavakol & Dennick, 2011).

Assessment development

Cases were developed collaboratively with input from the simulation center faculty and physician assistant faculty while considering guidance from stakeholder groups and accreditation agencies. The overall assessment design was modeled after a simulation assessment administered to the medical students annually but was reduced in scale from nine cases to a total of four cases due limitations related to budget and availability of the simulation lab space. Conducting a simulation assessment with a low station number can negatively impact reliability (Boulet, 2008) and post hoc analysis of this assessment indicates that this may have contributed to the overall performance of this exam.

While Khan, Gaunt, Ramachandran, and Pushkar (2013) advocate for piloting of the assessment as a training exercise for SPs and raters prior to the exam date, this was not feasible for this assessment from an expense or scheduling perspective. The inability to pilot and examine psychometrics prior to the exam administration may have played a role in the suboptimal validity and reliability of the assessment of patient care skills.

Internal consistency and construct validity measures for patient care skills were analyzed two ways. First, analysis was conducted for patient care skills (history-taking and physical exam skills in aggregate) for each case in the simulation exam (Tavakol, 2011). Each case assessed student ability to evaluate a patient with complaints related to specific organ systems. While all cases fell below the suitable threshold for internal consistency, case 2 (cardiac, pulmonary, infectious diseases) performed particularly poorly and should not be utilized for future exams.

There are several factors related to this clinical case which may have contributed to the poor performance. The working diagnosis in this case was a fungal infection that is relatively uncommon in the majority of primary care settings. It is notable that the primary author for this clinical case has specific expertise and training in infectious disease management, and so the author’s perception of the relative complexity of the case for non-specialist providers may have been underestimated. Recommendations for revision of the case would include modifying the working diagnosis to be more common and lower complexity such that a novice provider would be able to care for the patient with minimal collaboration. Incorporation of more robust peer review process for both the case script and the rating scale by generalist faculty members prior to deployment is also advisable (Tavakol & Dennick, 2012).

Case 1 (musculoskeletal, neurologic, psychiatric), case 3 (gastrointestinal, genitourinary) and case 4 (endocrine and chronic disease management) all fell below minimum acceptable thresholds for internal consistency with cases 1 and 4 reporting particularly poor reliability for physical exam skills and case 3 with very poor reliability for history taking. Given that specific deficiencies were noted in history taking for some cases, and physical exam skills for others, reliability may be optimized by uncoupling each of these skills into unique and more specific assessment stations (Khan, 2013). The highly complex nature of clinical cases and the reliance on many domains of performance in order to successful evaluate a patient, make construct validity difficult to achieve in even the most targeted cases (Barman, 2005).

When considering the poor results for construct validity, low values were calculated for total variance explained across the simulation assessment. These analyses were conducted with the assumption that categories of items on the assessment would be measuring a single construct. Due to the high complexity of the simulation assessment, assuming that any complex patient interaction can be reduced to a single construct may not have been a reasonable expectation. When interviewing and examining patients, for example, clinicians must begin with a broad sense of possible diagnoses across many body systems and the possibilities narrow over the course of the interview. A similar strategy is engaged with identifying which components of the physical exam are essential to establish a diagnosis. The highly complex interactions and need to synthesize large quantities of varied information in this process may make assessments of this type poorly suited to the narrow bounds and specificity of measuring the validity as it relates to a single construct.

In addition to a more robust peer review process with regard to case development and checklist development, there are other considerations for improvement of the assessment. Utilization of SPs to complete patient care skills is a consideration that may reduce the effect of faculty expertise and bias on ratings. The addition of a global rating scale for patient care skill assessment may provide a more accurate and nuanced evaluation of overall student performance (Hodges, 2003; Tolsgaard, 2012; Turner, 2008). Holistic ratings also allow the rater to capture egregious or dangerous actions and omissions that may not otherwise be accounted for on a checklist that captures only affirmative actions of the examinee (Boulet, 2008).

Another strategy to improve the validity and reliability of the assessment is with the addition of more points of data. Augmenting the assessment with either additional stations, additional components for rating on individual stations, or a larger sample size may be helpful (Tavakol, 2012; Khan, 2013). Turner (2008) estimates that to achieve reliability above 0.7, that simulation assessments would require at least six hours of testing and 10-12 stations. Increasing the length of the exam to six hours would represent a substantial increase in cost and personnel and may not be feasible with regard to scheduling at the simulation center.

Student-related factors

Since successful performance in a clinical encounter also requires a baseline level of subject matter knowledge, evaluation of knowledge deficiencies should include correlation with existing validated measures of student knowledge such as the end of curriculum exam. The addition of this post-hoc analysis would identify if poor simulation exam performance is related to learner knowledge deficits, or if challenges are more likely to be the result of the conditions or manner of scoring the simulation assessment (Turner, 2008).

Table 22 Summary of recommendations to improve simulation assessment quality

		Performance Challenge	**Statistical Test**	**Proposed Solutions**
Exam Administration	Multi-day assessment	None identified	ANOVA with no sig.	No changes
	SP Raters for Communication	None identified	ANOVA with no sig.	No changes
	Faculty Raters for Patient Care Skills	Outlier rater identified	ANOVA with one statistically sig outlier	Deliberate selection of raters from available faculty Improve pre-briefing of raters
	Exam scored live via video	Potential inaccuracy of raters scoring in real time		Consider grading assessment via video review, asynchronously,

				to allow for pause/rewind and verification
Communication rating	Checklist quality	Inadequate internal consistency	Cronbach's alpha .591	Remove items 7 and 9 to improve Cronbach's alpha
		Inadequate construct validity	Total variance explained 27.465%	Remove items 7 and 9 to improve total variance explained
Patient Care Skills	Case 1	Inadequate internal consistency	Cronbach's alpha .461	Add additional items for history and exam skills items Peer review and establish consensus for

				checklist items prior to redeployment Add global rating scale as primary means of assessment
	Case 2	Poor internal consistency	Cronbach's alpha .100	Consider creation of a new scenario to reflect a more common diagnosis
	Case 3	Inadequate internal consistency	Cronbach's alpha .494	Add additional items, particularly for history items Peer review and establish consensus for

				checklist items prior to redeployment Add global rating scale as primary means of assessment
	Case 4	Inadequate internal consistency	Cronbach's alpha .430	Add additional items, particularly related to exam skills Peer review and establish consensus for checklist items prior to redeployment Add global rating scale as primary means of assessment

	History taking	Inadequate construct validity	Total variance explained 11.359%	Consider restructuring the exam stations for more targeted skill assessment
	Physical exam skills	Inadequate construct validity	Total variance explained 10.545%	Consider restructuring the exam stations for more targeted skill assessment

The patient care skills assessment is highly complex and numerous factors related to assessment design and organization, complexity and variability of the subject matter, and rater characteristics contribute to challenges with assessment. The suboptimal performance of the patient care skills assessment is likely multifactorial.

Research Question 2

When Pearson correlation analysis was conducted for measures of communication skill among SCPE evaluations and the simulation exam scores, no statistically significant relationships were identified. Lack of meaningful correlation was noted when analysis was conducted for each individual SCPE, all SCPE assessed together, and when similar practice settings to the simulation exam were considered (PC, IM, EM). Standardized patients have been found to be suitable raters for measures of communication skills (Chang, 2017) and during this assessment there were not rater-related challenges associated with their scoring.

Communication skills can play a more or less significant role in patient care delivery based on sub-specialty of medicine. One might expect psychiatry or pediatric providers to be particularly attuned to the nuances of communication, whereas those practicing in surgical settings may rely less on patient communication in their daily work. As such, I suspected that alignment would vary by individual SCPE, but this was not the case. Addition of more assessment items on the rating scale, more stations, and creating stations that more closely reflect the range of clinical practice settings that students are likely to encounter may improve correlation for subsequent assessments (Calhoun, 2008).

The Hawthorne effect is another potential factor which may have impacted student performance on the simulation exam and negatively impacted correlation with clinical practice evaluation. The Hawthorne effect, or reactivity as it is also known, describes a phenomenon where participants positively alter their behavior when they know they are being observed (Boet, Sharma, Goldman and Reeves, 2012). As students are aware that they are being observed closely and scored on the simulation assessment, this may influence their conduct during the simulation assessment in a way that diverges from their behavior in true clinical settings. Paradis and Suskin (2016) identify assessment a circumstance that is particularly vulnerable to participant reactivity. The modification of student behavior in the exam setting could negatively impact the ability to identify correlation with SCPE ratings which are a result of four weeks of rater observations.

Research Question 3

When Pearson correlation analysis was conducted for measures of patient care skills among SCPE evaluations and the simulation scores, no statistically significant positive correlations were identified. Lack of correlation was identified when analysis was conducted for all SCPE scores as aggregate, individual SCPE and SCPEs that had the most commonality with the simulation exam (PC, IM, EM). While demonstration of specific patient care skills will vary from one clinical practice setting to another, there was no identified association with clinical practice areas most closely resembling the simulation exam scenarios.

The validity and reliability challenges associated with this simulation exam have been discussed in detail earlier in this chapter. As such, there is limited utility in discussing the assessment correlation with SCPE performance related to patient care skills. Until improvements in the simulation exam structure and scoring are conducted, the identification of correlation and any perceived impact cannot be meaningfully addressed. Following the quality improvement measures outlined above, and determination that the assessment has achieved sufficient reliability and internal consistency, identification of correlation may be revisited. Turner (2008) analyzed 33 studies where simulation exams were correlated to other typical assessment strategies for MD students (standardized testing, self-evaluation, clinical ratings, and course grades) and found a range of correlation coefficients from 0.10 to 1.00 with only 9 of 33 studies achieving a correlation coefficient of .70. Even with optimization of the assessment and rating scales, strong positive correlation may still not be achieved.

Limitations and Directions for Future Research

Several important limitations of this study exist. Sample size (n=27) is small with a female predominance, and the convenience sample reflected students from a single cohort at a single institution. As such, these results may not be reproducible. Increasing the sample size would improve statistical power but could be perceived as unethical with the current exam given the limitations of validity and reliability. Following modification of the assessment and re-piloting of the exam, if validity and reliability are improved, identifying additional institutions to administer the exam would be beneficial from both a sample size perspective and also to establish generalizability to PA learners more broadly.

It is clear that revision and optimization of the simulation exam should be carried out prior to re-deployment. The following items are recommendations for the approach to development and deployment of the assessment. First, when developing the clinical cases and rating scales, it may be beneficial to have a dedicated group of faculty who teach the course on patient care skills to create content and reach consensus prior to engagement with the simulation lab faculty. This could improve consistency, and alignment with PA curriculum and may be more reflective of clinical content that practicing PAs are likely to treat. These faculty members should also be used primarily as the raters for the exam.

Parsing out assessment stations to reflect more specific objectives is one consideration which may improve construct validity by removing confounding factors. For example, having a student complete a focused history for shortness of breath is likely to have better construct validity than a more comprehensive station where the student must incorporate and demonstrate history, physical exam and communication skills simultaneously. Assessing with a more targeted approach, however, is not as reflective of expectations or requirements of providers in the clinical setting and may not have utility for establishing learner competency appropriate for workforce entry.

Inclusion of a RIME rating as endorsed by Tolsgaard et al (2012) or a global rating scale (Turner, 2008) has been shown to be both reliable and allow for a more nuanced assessment of student performance. Global rating has been shown to be more valid and reliable for the assessment of more experienced learners when administered by expert raters. This style of rating may result in less modification of student behavior due to reactivity and a more authentic measure of student performance in the simulation setting. While global ratings could be captured for this assessment via retrospective video review, that students performed as if they were being scored via checklist may have impacted student assessment behavior and subsequent global scores. To pilot the impact of revised rating systems, students should be notified of any revisions to the scoring system prior to deployment. I would anticipate that utilization of a more authentic assessment of student performance would result in improved correlation with performance ratings from clinical settings.

Conclusions

This study explored the use of a simulation-based exam to assess physician assistant learner competency in the domains of patient care skills and communication. Through psychometric analysis of the examination components, several important factors were identified that may have negatively impacted the significance of the results. Faculty related factors such as variable clinical practice and teaching experiences, limited pre-briefing opportunities, and close modeling of this assessment off of medical student assessments may have reduced the validity of the patient care skills assessment. Validity for measures of communication skills approached acceptable values and can be optimized for future use with minor revisions. While no significant correlations were noted between the simulation exam and SCPE ratings, limitations related to the simulation exam performance and ratings scales may have impacted the quality of these results.

In addition to considering how to improve this assessment, clinician educators should also consider the feasibility of a large scale simulation exam and limitations related to the overall complexity of the subject matter. Achieving high construct validity may not be a feasible goal. Clinical encounters require students to obtain and synthesize large quantities of data across a variety of domains. Even the most carefully designed clinical case requires the student to utilize skills and knowledge from a variety of domains and clinical subspecialties thus reducing construct validity of the instrument. In each clinical encounter, for example, learners must have enough baseline medical knowledge to recognize that the differential diagnoses based on the presenting complaint could be represented by pathophysiology in several unique body systems. They must identify appropriate questions to ask the patient, while using language that the patient can understand, in order to solicit correct and useful responses. Physical exam skills need to be both comprehensive and targeted to both prove the working diagnosis, and disprove other possible diagnoses, all while maintaining rapport with the patient. They then must engage to answering patient questions and provide pertinent education while under strict time constraints. Capturing this variety of complex inter-related factors as a single, unique construct seems an unrealistic goal.

This broad-based and complex network of skills required of students is also applicable to communication assessment where concepts such as empathy, attentiveness, and ability to explain complex concepts in lay terms are all pertinent. It is unsurprising that construct validity was not achieved, and typical minimum thresholds of construct validity may not be an achievable goal for this style of assessment.

Additionally, the use of skill checklists alone as a measure of student skill, while useful for procedural skill training where there is a highly specific sequence of events, may not be appropriate for this more comprehensive and complex style of assessment. In the true clinical setting, performing accurately with efficiency is the goal. In assessments scored via checklists, student scores may benefit from including as many skills as possible that can be completed within the time limit, without penalty for inappropriate or extraneous skills. Moving to a global rating scale of performance scored by experienced faculty raters both disincentivizes students from casting a wide net in search of assessment points, and can capture more nuanced measures of student skill, efficiency and performance quality while capturing egregious omissions or dangerous practices. Utilization of global performance measures, as they encourage students to perform as they would in a true clinical setting, may also prove to have improved correlation with ratings from SCPE.

Simulation remains a widely used and accepted method for competency-based assessment in medical training. Revisions to the simulation exam cases and assessment measures with subsequent study is warranted as simulation continues to be both commonly utilized, and valuable for student learning and assessment for physician assistant students.

REFERENCES

Abadel, F. T., & Hattab, A. S. (2014). Patients' assessment of professionalism and communication skills of medical graduates. BMC Medical Education, 14(1), 28. https://doi.org/10.1186/1472-6920-14-28

Amin, Z., Chong, Y. S., & Khoo, H. E. (2006). Practical guide to medical student assessment. Singapore: World Scientific Publishing.

Accreditation standards for physician assistant education, 4th edition. (2010) Accreditation Review Commission for the Education of Physician Assistants. Retrieved from: http://www.arc-pa.org/wp-content/uploads/2018/06/Standards-4th-Ed-March-2018.pdf

Accreditation standards for physician assistant education, 5th edition (2019) Accreditation Review Commission for the Education of Physician Assistants. Retrieved from: http://www.arc-pa.org/wp-content/uploads/2019/09/Standards-5th-Ed-Sept-2019-fnl.pdf

AAPA (2019) What is a PA. Retrieved from: https://www.aapa.org/wp-content/uploads/2019/08/What_Is_A_PA_Infographic_LetterSize_July2019.pdf

AAPA (n.d.) PAs working abroad frequently asked questions. Retrieved from: https://www.aapa.org/career-central/career-advancement-and-transitions/international-practice/pas-working-abroad-frequently-asked-questions/

Association of American Medical Colleges. (2014) Core entrustable professional activities for entering residency: curriculum developers' guide. AAMC iCollaborative, Washington, DC. 2014. https://www.mededportal.org/icollaborative/resource/887

Barman, A. (2005). Critiques on the objective structured clinical examination. Annals-Academy of Medicine Singapore, 34(8), 478.

Barsuk, J. H., McGaghie, W. C., Cohen, E. R., O'Leary, K. J., & Wayne, D. B. (2009). Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. Critical Care Medicine, 37(10), 2697-2701.

Bateman, J., Allen, M., Samani, D., Kidd, J., & Davies, D. (2013). Virtual patient design: exploring what works and why. A grounded theory study. Medical Education, 47(6), 595-606. https://doi.org/10.1111/medu.12151

Boet, S., Sharma, S., Goldman, J., & Reeves, S. (2012). An overview of methods article de synthese: La recherche en éducation médicale: Un aperçu des méthodes. *Can J Anesth*, *59*, 159-70. https://doi.org/

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, *65*(1), 23-28.

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

Bok, H. G., de Jong, L. H., O'Neill, T., Maxey, C., & Hecker, K. G. (2018). Validity evidence for programmatic assessment in competency-based education. Perspectives on Medical Education, 7(6), 362-372. https://doi.org/10.1007/s40037-018-0481-2

Boyle, W. A., Murray, D. J., Beyatte, M. B., Knittel, J. G., Kerby, P. W., Woodhouse, J., & Boulet, J. R. (2018). Simulation-Based Assessment of Critical Care" Front-Line" Providers. Critical care medicine, 46(6), e516-e522. https://doi.org/10.1097/CCM.0000000000003073

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World*, *3*(1).

Chang, D., Mann, M., Sommer, T., Fallar, R., Weinberg, A., & Friedman, E. (2017). Using standardized patients to assess hospitalist communication skills. Journal of Hospital Medicine, 12(7), 562-566. https://doi.org/10.12788/jhm.2772

Coerver, D., Multak, N., Marquardt, A., & Larson, E. H. (2017). The use of simulation in physician assistant programs: A national survey. The Journal of Physician Assistant Education, 28(4), 175-181. https://doi.org/10.1097/JPA.0000000000000173

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Medical Teacher, 35(1), e867-e898. https://doi.org/10.3109/0142159X.2012.714886

Cox, C. W., & Gunderman, R. B. (2017). Andragogic approaches to continuing medical education. Academic Radiology, 24(10), 1325-1326. DOI: https://doi.org/10.1016/j.acra.2017.05.004

Cunnington, J., Neville, A., & Norman, G. (1997). The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE. *Advances in Health Sciences Education*, *1*(3), 227-233.

Daves, D. A., Mazmanian, P. E., Fordis, M., Van Harrision, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence. JAMA, 296, 1094-102. https://doi.org/10.1001/jama.296.9.1094

Dennick, R. (2016). Constructivism: Reflections on twenty five years teaching the constructivist approach in medical education. International Journal of Medical Education, 7, 200. https://doi.org/10.5116/ijme.5763.de11

Duch, B. J., Groh, S. E., & Allen, D. E. (2001). The power of problem-based learning: A practical" how to" for teaching undergraduate courses in any discipline. Stylus Publishing, LLC.

Duffy, F. D., Lynn, L. A., Didura, H., Hess, B., Caverzagie, K., Grosso, L., Lipner, R. & Holmboe, E. S. (2008). Self-assessment of practice performance: Development of the ABIM Practice Improvement Module (PIMSM). Journal of Continuing Education in the Health Professions, 28(1), 38-46. https://doi.org/10.1002/chp.154

Englander, R., Cameron, T., Ballard, A. J., Dodge, J., Bull, J., Aschenbrener, C. A. (2013). Toward a common taxonomy of competency domains for the health professions and competencies for physicians. Academic Medicine, 88(8), 1-6. https://doi.org/10.1097/ACM.0b013e31829a3b2b

Ertner, P.A. & Newby, T.J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. Performance Improvement Quarterly. 26 (43-71). https://doi.org/10.1111/j.1937-8327.1993.tb00605.x

Everhart, D., Sandeen, C., Seymour, D., & Yoshino, K. (2014). Clarifying competency based education terms. American Council on Education and Blackboard. Retrieved from: http://blog. blackboard. com/clarifyingcompetency-based-education-terms.

Fenwick, T., & Dahlgren, M. A. (2015). Towards socio-material approaches in simulation-based education: lessons from complexity theory. Medical Education, 49(4), 359-367. https://doi.org/10.1111/medu.12638

Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Medical Teacher, 29(7), 642-647. https://doi.org/10.1080/01421590701746983

Frank, J. R., Snell, L. S., Cate, O.T., Holmboe, E. S., Carraccio, C., Swing, S. R., Harris, P., Glasgow, N.J., Campbell, C., Dath, D., Harden, R., Iobst, W. Long, D., Mungroo, R., Richardson, D., Sherbino, J., Silver, I., Taber, S., Talbot, M., & Harris, K.(2010). Competency-based medical education: Theory to practice. Medical Teacher, 32(8), 638-645. https://doi.org/10.3109/0142159X.2010.501190

Frank, J. R., Mungroo, R., Ahmad, Y., Wang, M., De Rossi, S., & Horsley, T. (2010). Toward a definition of competency-based education in medicine: a systematic review of published definitions. Medical Teacher, 32(8), 631-637. https://doi.org/10.3109/0142159X.2010.500898

Feldman, M. J., Barnett, G. O., Link, D. A., Coleman, M. A., Lowe, J. A., & O'Rourke, E. J. (2006). Evaluation of the Clinical Assessment project: a computer-based multimedia tool to assess problem-solving ability in medical students. Pediatrics, 118(4), 1380-1387. https://doi.org/10.1542/peds.2006-0326

Fenwick, T., & Dahlgren, M. A. (2015). Towards socio-material approaches in simulation-based education: lessons from complexity theory. Medical Education, 49(4), 359-367. https://doi.org/10.1111/medu.12638

Fraser, S. W., & Greenhalgh, T. (2001). Coping with complexity: educating for capability. British Medical Journal, 323(7316), 799-803. https://doi.org/10.1136/bmj.323.7316.799

Gaba, D. M. (2004). The future vision of simulation in health care. BMJ Quality & Safety, 13(suppl 1), i2-i10. http://dx.doi.org/10.1136/qshc.2004.009878

Haji, F. A., Cheung, J. J., Woods, N., Regehr, G., de Ribaupierre, S., & Dubrowski, A. (2016). Thrive or overload? The effect of task complexity on novices' simulation-based learning. Medical Education, 50(9), 955-968. https://doi.org/10.1111/medu.13086

Halamek, L. P., Kaegi, D. M., Gaba, D. M., Sowb, Y. A., Smith, B. C., Smith, B. E., & Howard, S. K. (2000). Time for a new paradigm in pediatric medical education: teaching neonatal resuscitation in a simulated delivery room environment. Pediatrics, 106(4), e45-e45.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomed Informatics, 42(2):377-81. https://doi.org/10.1016/j.jbi.2008.08.010

Hawkins, R. E., Lipner, R. S., Ham, H. P., Wagner, R., & Holmboe, E. S. (2013). American Board of Medical Specialties Maintenance of Certification: Theory and evidence regarding the current framework. Journal of Continuing Education in the Health Professions, 33(S1), S7-S19. https://doi.org/10.1002/chp.21201

Holzinger, A., Kickmeier-Rust, M. D., Wassertheurer, S., & Hessinger, M. (2009). Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education, 52(2), 292-301. https://doi.org/10.1016/j.compedu.2008.08.008

Holzman, R. S., Cooper, J. B., Gaba, D. M., Philip, J. H., Small, S. D., & Feinstem, D. (1995). Anesthesia crisis resource management: real-life simulation training in operating room crises. Journal of Clinical Anesthesia, 7(8), 675-687.

Huang, G. C., Sacks, H., DeVita, M., Reynolds, R., Gammon, W., Saleh, M., & Cantrell, M. (2012). Characteristics of simulation activities at North American medical schools and teaching hospitals: an AAMC-SSH-ASPE-AACN collaboration. Simulation in Healthcare, 7(6), 329-333. https://doi.org/10.1097/SIH.0b013e318262007e

Jefferies, A., Simmons, B., Tabak, D., Mcilroy, J. H., Lee, K. S., Roukema, H., & Skidmore, M. (2007). Using an objective structured clinical examination (OSCE) to assess multiple physician competencies in postgraduate training. *Medical Teacher*, *29*(2-3), 183-191. https://doi-org.laneproxy.stanford.edu/10.1080/01421590701302290

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement*, *70*(3), 394-400. https://doi-org.stanford.idm.oclc.org/10.1177/0013164409355692

Juneja, P. (n.d.) Overview of competency framework and it's benefits. Retrieved November 7, 2019 from: https://www.managementstudyguide.com/competency-framework.htm

Kaakinen, J., & Arwood, E. (2009). Systematic review of nursing simulation literature for use of learning theory. International Journal of Nursing Education Scholarship, 6(1). https://doi.org/10.2202/1548-923X.1688

Khan, K. Z., Gaunt, K., Ramachandran, S., & Pushkar, P. (2013). The objective structured clinical examination (OSCE): AMEE guide no. 81. part II: Organisation & administration. Medical Teacher, 35(9), e1447-e1463. https://doi.org/10.3109/0142159X.2013.818635

Kobayashi, L., Dunbar-Viveiros, J. A., Devine, J., Jones, M. S., Overly, F. L., Gosbee, J. W., & Jay, G. D. (2012). Pilot-phase findings from high-fidelity in situ medical simulation investigation of emergency department procedural sedation. Simulation in Healthcare, 7(2), 81-94. https://doi.org/ 10.1097/SIH.0b013e31823b9923

Kolb, D. A. (2014). Experiential learning: Experience as the source of learning and development. Indianapolis, IN. FT press.

Lavoie, P., Michaud, C., Bélisle, M., Boyer, L., Gosselin, É., Grondin, M., & Pepin, J. (2018). Learning theories and tools for the assessment of core nursing competencies in simulation: A theoretical review. Journal of Advanced Nursing, 74(2), 239-250. https://doi.org/10.1111/jan.13416

Maresky, H. S., Oikonomou, A., Ali, I., Ditkofsky, N., Pakkal, M., & Ballyk, B. (2019). Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. Clinical Anatomy, 32(2), 238-243. https://doi.org/10.1002/ca.23292

McGaghie, W. C. (1978). Competency-Based Curriculum Development in Medical Education. An Introduction. Public Health Papers No. 68.

McGaghie, W. C., & Harris, I. B. (2018). Learning theory foundations of simulation-based mastery learning. Simulation in Healthcare, 13(3S), S15-S20. doi: 10.1097/SIH.0000000000000279

Meretoja, R., & Koponen, L. (2012). A systematic model to compare nurses' optimal and actual competencies in the clinical setting. Journal of Advanced Nursing, 68(2), 414-422. https://doi.org/10.1111/j.1365-2648.2011.05754.x

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. Academic Medicine, 65(9), S63-7.

Multak, N., Euliano, T., Gabrielli, A., & Layon, A. J. (2002). Human patient simulation: a preliminary report of an innovative training tool for physician assistant education. The Journal of Physician Assistant Education, 13(2), 103-105.

Naismith, L. M., & Cavalcanti, R. B. (2015). Validity of cognitive load measures in simulation-based training: a systematic review. Academic Medicine, 90(11), S24-S35. doi: 10.1097/ACM.0000000000000893

National Academies of Science, Engineering and Medicine. (2018). How people learn II: Learners, contexts, and cultures. Washington DC: The National Academies Press. National Commission on the Certification of Physician Assistants (2019) Content Blueprint for the Physician Assistant Certifying Exam. Retrieved from https://www.nccpa.net/pance-content-blueprint

Norcini, J. J., Lipner, R. S., & Kimball, H. R. (2002). Certifying examination performance and patient outcomes following acute myocardial infarction. Medical Education, 36(9), 853-859. https://doi.org/10.1046/j.1365-2923.2002.01293.x

Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. Academic Medicine, 74(11), 1203-7.

Papadakis, M. A., Arnold, G. K., Blank, L. L., Holmboe, E. S., & Lipner, R. S. (2008). Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. Annals of Internal Medicine, 148(11), 869-876. https://doi.org/10.7326/0003-4819-148-11-200806030-00009

Paradis, E., & Sutkin, G. (2017). Beyond a good story: from Hawthorne Effect to reactivity in health professions education research. Medical education, 51(1), 31-39. doi: 10.1111/medu.13122

Pell, G., Fuller, R., Homer, M., & Roberts, T. (2010). How to measure the quality of the OSCE: a review of metrics–AMEE guide no. 49. Medical Teacher, 32(10), 802-811.

Physician Assistant Education Association. (2018) Core Competencies for New Physician Assistant Graduates. Retrieved from https://paeaonline.org/wp-content/uploads/2018/09/core-competencies-for-pa-grads-20180919.pdf. Accessed October 12, 2018.

Ryan, T. P. (2013). Sample size determination and power. John Wiley & Sons.

Sanghi, S. (2016). The Handbook of Competency Mapping: Understanding, Designing and Implementing Competency Models in Organizations (Vol. Third edition). Thousand Oaks, California: Sage Publications Pvt. Ltd.

Schuwirth, L. W., & Van Der Vleuten, C. P. (2004). Changing education, changing assessment, changing research? Medical education, 38(8), 805-812.

https://doi.org/10.1111/j.1365-2929.2004.01851.x

Shea, J. A., Norcini, J. J., & Kimball, H. R. (1993). Relationships of ratings of clinical competence and ABIM scores to certification status. Academic Medicine, 68(10), S22-4.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*(1), 53-55.

Tavakol, M. & Dennick, R. (2012). Post-examination interpretation of objective test data: monitoring and improving the quality of high-stakes examinations: AMEE Guide No. 66. *Medical teacher*, *34*(3), e161-e175.

Taylor, D. C., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. Medical Teacher, 35(11), e1561-e1572. https://doi.org/10.3109/0142159X.2013.828153

Tolsgaard, M. G., Arendrup, H., Lindhardt, B. O., Hillingsø, J. G., Stoltenberg, M., & Ringsted, C. (2012). Construct validity of the reporter-interpreter-manager-educator structure for assessing students' patient encounter skills. Academic Medicine, 87(6), 799-806. https://doi.org/10.1097/ACM.0b013e318253c8be

Turner, J. L., & Dankoski, M. E. (2008). Objective structured clinical exams: a critical review. Family Medicine, 40(8), 574-8.

USMLE (September 2019). Content description and general information: Step 2 clinical skills. Retrieved from: https://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf

Van Melle, E., Frank, J. R., Holmboe, E. S., Dagnone, D., Stockley, D., Sherbino, J., & International Competency-based Medical Education Collaborators. (2019). A core components framework for evaluating implementation of competency-based medical education programs. Academic Medicine, 94(7), 1002-1009. https://doi.org/10.1097/ACM.0000000000002743

Violato, C., Marini, A., Toews, J., Lockyer, J., & Fidler, H. (1997). Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. Academic Medicine. *72*(10, Suppl 1), S82-S84. https://doi.org/10.1097/00001888-199710001-00028

Wallenstein, J., Heron, S., Santen, S., Shayne, P., & Ander, D. (2010). A core competency–based objective structured clinical examination (OSCE) can predict future resident performance. *Academic Emergency Medicine*, *17*, S67-S71.

Weinstock, P., & Halamek, L. P. (2008). Teamwork during resuscitation. Pediatric Clinics of North America, 55(4), 1011-1024. https://doi.org/10.1016/j.pcl.2008.04.001

Ziv, A., Wolpe, P. R., Small, S. D., & Glick, S. (2003). Simulation-based medical education: an ethical imperative. Academic Medicine, 78(8), 783-788. https://doi.org/10.1097/01.SIH.0000242724.08501.63

APPENDIX A

Standardized Patient Checklist for Simulation Exam: Communication Skills Score

STUDENT NAME: ________________________________

SP NAME: __________________________ DATE: _______________

OVERALL SATISFACTION

Based on my level of satisfaction with this encounter, I would return to see this student again.

YES UNDECIDED NO

Comment:

The student:

Made a personal connection during the visit.

- **Agree:** Was warm, went beyond medical issues at hand, conversed about personal background, interests, etc. (Acute cases: rapt attention to me.)
- **Somewhat:** Made an attempt to make a personal connection.
- **Disagree:** Gave me the impression s/he was only interested in me as a disease or a symptom.

Gave me an opportunity/time to talk.

- **Agree:** Invited me to speak. Encouraged me to tell my story. Asked open-ended questions. Used silence appropriately.
- **Somewhat:** I just answered the questions.
- **Disagree:** Did not give me the opportunity to speak. I felt rushed or interrupted.

Actively listened. Gave me undivided attention.

- **Agree:** Used body language that was open and encouraging – appropriate eye contact, body position. Let me know I was the student's focus.
- **Somewhat:** Made comfortable eye contact. Frequent use of notes, but still attentive.
- **Disagree:** Was not focused on me. Long unexplained pauses. Used closed body language. Focused solely on clipboard or notes. Positioned too close or too far away. Did not pay attention to my answers.

Summarized and/or clarified information.

- **Agree:** Followed up or clarified some of my answers or summarized what I said and allowed me to clarify if needed.
- **Somewhat:** Mostly echoed my answers.

○ **Disagree:** Never summarized nor verified what I was saying.

Treated me with respect.

○ **Agree:** Showed courtesy and consideration at all times. Did not talk down to me. I felt my concerns were taken seriously.
○ **Somewhat:** Was mostly respectful. There wasn't a pattern of disrespect.
○ **Disagree:** Their agenda was more important than mine. My problems were not important. Used humor inappropriately, talked down to me, belittled me.

Adapted to my level of understanding.

○ **Agree:** Spoke clearly in a way I could understand. All explanations and questions were clear.
○ **Somewhat:** Used a little jargon but explained with prompting.
○ **Disagree:** Used jargon without explaining or explanations were vague.

Verbally expressed empathy.

○ **Agree:** Offered comments to validate or acknowledge my feelings and concerns.
○ **Somewhat:** Minimally expressed empathy. Said the "right words," but it was strictly rote. Heard my concerns but didn't validate them.
○ **Disagree:** Made no verbal expressions of empathy.

Encouraged me to ask questions.

○ **Agree:** Asked if I had questions during the encounter.
○ **Disagree:** Didn't ask me if I had any questions.

Discussed assessment and explained rationale for next steps.

○ **Agree:** Assessment and rationale for next steps were clear.
○ **Somewhat:** Gave me a general understanding of the assessment and the rationale for next steps, but I would have appreciated more information/specifics.
○ **Disagree:** Failed to provide me with an assessment and/or any information about the rationale for next steps. The student ran out of time.

Elicited my perspective and concerns about the next steps.

○ **Agree:** Specifically asked how I felt about the next steps and addressed any concerns. Sought my approval/permission/ability to move forward with the next steps.
○ **Somewhat:** Simply asked if I was OK with the next steps.

- **Disagree:** Did not ask if I have any concerns about the next steps. Did not ask if I approved of/could follow the next steps. The student ran out of time.

SP COMMENTS:

Your comments are the patient's subjective personal feelings about the interpersonal skills of the student. What is important is the relationship that develops between the two of you. Do not hold it against the student if they don't finish the interview.

As the patient, I felt…

APPENDIX B

Faculty Checklists for Simulation Exam: Patient Care Skills Score

Case #1

HISTORY The student asked the patient:		
1. How the pain **began**.	☐ Yes	☐ No
2. **How long** the pain has lasted.	☐ Yes	☐ No
3. To **describe** the pain.	☐ Yes	☐ No
4. About any **radiation** of the pain.	☐ Yes	☐ No
5. About any **numbness**.	☐ Yes	☐ No
6. About any **weakness**.	☐ Yes	☐ No
7. What makes the current pain **better** AND **worse**.	☐ Yes	☐ No
8. About any **previous history** of back pain.	☐ Yes	☐ No
9. About any changes in **bladder** OR **bowel** functions	☐ Yes	☐ No
10. About the current status of my **prostate cancer**.	☐ Yes	☐ No
11. About any history of **trauma, weight loss, night sweats, steroid** or **IV drug use**. (Must ask at least one).	☐ Yes	☐ No

PHYSICAL EXAMINATION The student:			
12. **Felt my spine** and asked whether I experienced pain or tenderness while palpating over my spine AND the muscles on either side of my spine.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done
13. **Tapped** on both of my **ankles** AND both knees with a reflex hammer.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done
14. Asked me to **push down** against his/her hand with my foot (or big toe) [BOTH feet must be tested] OR asked me to walk on my toes.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done
15. Asked me to **pull up** against his/her hand with my foot (or big toe) [BOTH feet must be tested] OR asked me to walk on my heels.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done

16. **Touched my calves** OR feet in both legs in at least one place and **asked** about the touch.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
17. **Raised my RIGHT leg** while I was lying down AND asked about the location or quality of the pain I was experiencing with the maneuver.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
18. **Raised my LEFT leg** while I was lying down.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
19. Asked me to **walk** (normal gait).	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
20. Asked to perform a **rectal examination.**	☐ Done		☐ Not Done
	Notes:		

PATIENT EDUCATION AND COUNSELING

The student:

21. Told me that my pain might be being caused by **sciatica** ("pinched nerve") OR a **disk** problem ("slipped", "degenerated") OR **muscle strain**.	☐ Yes	☐ No
22. Addressed my concern that I am having a **recurrence** of prostate cancer.	☐ Yes	☐ No
23. Addressed my concern for better **pain management.**	☐ Yes	☐ No
24. Told me I did **not need imaging** at this time.	☐ Yes	☐ No

FUNDAMENTALS OF PHYSICAL EXAM BEHAVIOR

The student:

25. Introduced him/herself to patient (using last name, and title).	☐ Yes	☐ No
26. Washed his/her hands prior to the physical examination.	☐ Yes	☐ No
27. Explained in advance on at least one occasion what s/he would be doing during the physical exam.	☐ Yes	☐ No
28. Conveyed at least one aspect of the results of the physical exam to pt.	☐ Yes	☐ No

Case #2

HISTORY
The student asked the patient:

1. **How long** you have been feeling short of breath.	☐ Yes	☐ No
2. To **describe** the shortness of breath.	☐ Yes	☐ No
3. About the **progression** of the shortness of breath.	☐ Yes	☐ No
4. About what makes the shortness of breath **better AND worse**.	☐ Yes	☐ No
5. If I have a **cough** AND to **describe** the cough in detail.	☐ Yes	☐ No
6. If I have a **fever**.	☐ Yes	☐ No
7. If I have had any symptoms of **congestive heart failure**.	☐ Yes	☐ No
8. If I have any symptoms suggestive of **pulmonary embolus**.	☐ Yes	☐ No
9. If I have any **chest pain**.	☐ Yes	☐ No
10. If I have ever had any **lung problems**.	☐ Yes	☐ No
11. If I **smoke** (and/or how much I have smoked).	☐ Yes	☐ No
12. If I have any known non-pharmaceutical **allergies.**	☐ Yes	☐ No
13. About ONE **environmental exposure** or contact that I may have had.	☐ Yes	☐ No

PHYSICAL EXAMINATION
The student:

14. Felt **lymph nodes** on BOTH sides of my neck.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done
15. **Percussed** my lungs on 6 places on my back.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done
16. **Listened** to my lungs on 6 places on my **back**.	☐ Done Notes:	☐ Done Incorrectly	☐ Not Done

17. **Listened** to my lungs over my **chest** in two places.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
18. Listened to the lungs for ONE **adventitious sound** ("e", whisper, forced exhalation, "99")	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
19. Listened to my **heart** in four places.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		
20. Assessed for **swelling** in the lower legs.	☐ Done	☐ Done Incorrectly	☐ Not Done
	Notes:		

PATIENT EDUCATION AND COUNSELING

The student:

21. Explained the role of the physician assistant or their role on the team.	☐ Yes	☐ No
22. Indicated what they thought might be going on.	☐ Yes	☐ No
23. Informed me that they want to do ONE of the following: bloodwork or imaging.	☐ Yes	☐ No
24. Provided a specific plan to establish you with a primary care provider.	☐ Yes	☐ No
25. Told you that you should obtain a mammogram OR addressed your concern about cancer screening.	☐ Yes	☐ No

FUNDAMENTALS OF PHYSICAL EXAM BEHAVIOR

The student:

26. Introduced him/herself to patient (using last name and title).	☐ Yes	☐ No
27. Washed his/her hands prior to the physical examination.	☐ Yes	☐ No
28. Explained in advance on at least one occasion what s/he would be doing during the physical exam.	☐ Yes	☐ No
29. Conveyed at least one aspect of the results of the physical examination to patient.	☐ Yes	☐ No

Case #3

HISTORY The student asked the patient:		
1. To **describe** the pain you are currently having.	☐ Yes	☐ No
2. About the **location** of the pain.	☐ Yes	☐ No
3. About the **intensity** of the pain.	☐ Yes	☐ No
4. About the **duration** of the pain.	☐ Yes	☐ No
5. Whether the pain is **constant or intermittent.**	☐ Yes	☐ No
6. Whether the pain **radiates** to other parts of your body.	☐ Yes	☐ No
7. About what makes the pain **better** AND what makes it **worse**.	☐ Yes	☐ No
8. If you have felt **nauseated** OR vomited.	☐ Yes	☐ No
9. If you have vomited **blood** OR if you have had blood in the stool OR if you have had dark/tarry stools.	☐ Yes	☐ No
10. How much or how frequently you drink **alcohol**.	☐ Yes	☐ No
11. Asked about problematic alcohol use (impact on life, unsuccessful attempts to cut back, etc.)	☐ Yes	☐ No
12. Whether you have ever had, or been tested for, **gallstones.**	☐ Yes	☐ No
13. If you are **sexually active.**	☐ Yes	☐ No
14. When your last **menstrual period** was.	☐ Yes	☐ No
15. About any **medications** you are taking.	☐ Yes	☐ No

PHYSICAL EXAMINATION The student:	
16. Asked to check for **orthostatic** vital signs.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
17. **Auscultated** the abdomen for bowel sounds.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:

18. **Percussed** the abdomen.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
19. **Palpated** all four quadrants of the abdomen.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
20. Palpated the **liver.**	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
21. Assessed for **rebound tenderness**.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
22. Tapped on the lower back (at the **costovertebral angle**) on both sides to elicit tenderness.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
23. Listened to my **heart** in four places.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:	
24. Asked to perform a **pelvic exam.**	☐ Done ☐ Not Done Notes:	

PATIENT EDUCATION AND COUNSELING

The student:

25. Indicated what s/he thought might be going on (pancreas, gall-bladder, liver).	☐ Yes	☐ No
26. Discussed the importance of obtaining blood work/additional evaluation at this time.	☐ Yes	☐ No
27. Indicated that s/he would like to admit you to the hospital.	☐ Yes	☐ No
28. Addressed his/her concerns about my alcohol consumption.	☐ Yes	☐ No
29. Discussed safer sex practices.	☐ Yes	☐ No

FUNDAMENTALS OF PHYSICAL EXAM BEHAVIOR The student:		
30. Introduced him/herself to patient (using last name and title).	☐ Yes	☐ No
31. Washed his/her hands prior to the physical examination.	☐ Yes	☐ No
32. Explained in advance on at least one occasion what s/he would be doing during the physical exam.	☐ Yes	☐ No
33. Conveyed at least one aspect of the results of the physical examination to patient.	☐ Yes	☐ No
34. Appropriately draped patient during the physical exam.	☐ Yes	☐ No

Case #4

HISTORY The student asked the patient:		
1. To **describe** the quality of the pain you currently are having using an open ended question.	☐ Yes	☐ No
2. About the **intensity** of the pain.	☐ Yes	☐ No
3. about what makes the pain **better** AND **worse.**	☐ Yes	☐ No
4. If you have had any changes in, OR problems with the **sensation** in your feet.	☐ Yes	☐ No
5. If you have had any problems with the **motor function** of your feet OR legs.	☐ Yes	☐ No
6. If you **check** your **blood sugars** at home.	☐ Yes	☐ No
7. What your blood sugars have been running at home.	☐ Yes	☐ No
8. If I am following a "diabetic" diet.	☐ Yes	☐ No
9. If you are taking your **medications** as prescribed.	☐ Yes	☐ No
10. If you **check your feet** at home.	☐ Yes	☐ No
11. How frequently you **exercise.**	☐ Yes	☐ No

PHYSICAL EXAMINATION The student:	
12. Checked your **blood pressure** in one arm.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
13. Looked in both your **eyes** with an ophthalmoscope.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
14. **Listened to your heart** in four places.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:

15. Visually **inspected** the top AND bottom of BOTH of your **feet** AND in between your toes (without socks)	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
16. Felt the **pulses** in BOTH of your feet.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
17. Assessed for peripheral **edema** in BOTH legs.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
18. Placed a vibrating tuning fork on BOTH of your feet, and asked if you could feel the **vibration** OR moved your toe(s) on BOTH of your feet up and down and asked if you could tell whether your toe was up or down.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
19. Checked for **light touch sensation** on BOTH feet AND BOTH lower legs.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
20. Checked for **sharp touch sensation** on BOTH feet AND BOTH lower legs.	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
21. Assessed reflexes in BOTH **ankles.**	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:
22. Assessed reflexes in BOTH **knees.**	☐ Done ☐ Done Incorrectly ☐ Not Done Notes:

PATIENT EDUCATION AND COUNSELING The student:		
23. Indicated that patient's symptoms might be consistent with neuropathy/neurological/nerve damage in the feet.	☐ Yes	☐ No
24. Reinforced the importance of watching carbohydrate and fat intake OR eating meals regularly OR not gaining weight.	☐ Yes	☐ No
25. Offered patient assistance related to her visual impairment.	☐ Yes	☐ No

FUNDAMENTALS OF PHYSICAL EXAM BEHAVIOR The student:		
26. Introduced him/herself to patient (using last name and title).	☐ Yes	☐ No
27. Washed his/her hands prior to the physical examination.	☐ Yes	☐ No
28. Explained in advance on at least one occasion what s/he would be doing during the physical exam.	☐ Yes	☐ No
29. Conveyed at least one aspect of the results of the physical examination to patient.	☐ Yes	☐ No

APPENDIX C

Supervised Clinical Practice Experience Evaluation for Patient Care Skills and Communication Skills Ratings

Stanford MSPA Preceptor Evaluation Form

Student Name: ____________________ **Clerkship Setting:** ____________________ **Date:** ____________________

Competency Measure *Note: Poor or Unacceptable performances require comments in the space provided*	**Excellent** Superior in every way- top 10%	**Very Good** Exceeds in many areas- top 20%	**Satisfactory** Adequate	**Poor** Many deficiencies	**Unacceptable** Clearly inadequate: requires remediation	**N/A**
Patient Care Skills						
History Taking: Accurate and complete	☐	☐	☐	☐	☐	☐
Physical Exam: Required components present	☐	☐	☐	☐	☐	☐
Develops appropriate assessments and plans	☐	☐	☐	☐	☐	☐
Provides quality patient education	☐	☐	☐	☐	☐	☐
Competently performs medical and surgical procedures delineated by medical staff privileges- overall evaluation, when applicable	☐	☐	☐	☐	☐	☐
Demonstrates appropriate problem solving and critical thinking skills	☐	☐	☐	☐	☐	☐
Incorporates patient's cultural and personal values into evaluation and management plan	☐	☐	☐	☐	☐	☐
Demonstrates appropriate selection and interpretation of diagnostic tests	☐	☐	☐	☐	☐	☐
Appropriately integrates history, physical exam findings and diagnostic studies to formulate a differential diagnosis	☐	☐	☐	☐	☐	☐
Overall integration of clinical information into treatment planning	☐	☐	☐	☐	☐	☐
Demonstrates pharmacological knowledge/ appropriate ordering of therapeutics	☐	☐	☐	☐	☐	☐
Patient Care Skill Comments:	☐	☐	☐	☐	☐	☐

Competency Measure

Note: Poor or Unacceptable performances require comments in the space provided

Practice-Based Learning and Improvement	**Excellent** Superior in every way- top 10%	**Very Good** Exceeds in many areas- top 20%	**Satisfactory** Adequate	**Poor** Many deficiencies	**Unacceptable** Clearly inadequate: requires remediation	**N/A**
Applies evidence-based medicine to clinical decisions						
Identifies and engages in learning activities to address gaps in knowledge, skills, or attitudes	☐	☐	☐	☐	☐	☐
Actively seeks feedback and incorporates into daily practice	☐	☐	☐	☐	☐	☐
Professionalism						
Demonstrates understanding of the roles and responsibilities of a PA	☐	☐	☐	☐	☐	☐
Displays sensitivity and responsiveness to patients' culture, age, gender, and disabilities	☐	☐	☐	☐	☐	☐
Demonstrates compassion, integrity, and respect for others	☐	☐	☐	☐	☐	☐
Demonstrates respect for patient privacy and autonomy	☐	☐	☐	☐	☐	☐
Student is reliable and dependable	☐	☐	☐	☐	☐	☐
Communication Skills						
Demonstrates emotional resilience, adaptability, flexibility, tolerance of ambiguity and anxiety	☒	☐	☐	☐	☐	☐
Displays self-confidence in the clinical setting	☒	☐	☐	☐	☐	☐
Communications and behaviors with preceptors and staff are effective and appropriate	☒	☐	☐	☐	☐	☐

Competency Measure *Note: Poor or Unacceptable performances require comments in the space provided*	**Excellent** Superior in every way- top 10%	**Very Good** Exceeds in many areas- top 20%	**Satisfactory** Adequate	**Poor** Many deficiencies	**Unacceptable** Clearly inadequate: requires remediation	**N/A**
Demonstrates sensitivity, honesty, and compassion in patient interactions	☐	☐	☐	☐	☐	☐
Demonstrates effective listening, nonverbal, explanatory, interviewing, and writing skills to elicit and provide information	☐	☐	☐	☐	☐	☐
Responsiveness to feedback	☐	☐	☐	☐	☐	☐
Relatability to colleagues	☐	☐	☐	☐	☐	☐
Relatability to patients	☐	☐	☐	☐	☐	☐
Exhibits communication skills appropriate to patient level of understanding	☐	☐	☐	☐	☐	☐
Systems Based Practice						
Uses information technology resources to support patient care decisions and patient education						
Applies medical information and clinical data systems to provide more effective, efficient patient care	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐

APPENDIX D

BOISE STATE UNIVERSITY
RESEARCH AND ECONOMIC DEVELOPMENT

Date: January 15, 2020

To: Michele Toussaint cc: Brett Shelton

From: Office of Research Compliance (ORC)

Subject: SB-IRB Notification of Exemption - 101-SB19-267
Medical simulation as a competency-based assessment tool for physician assistant students

The Boise State University ORC has reviewed your protocol application and has determined that your research is exempt from further IRB review and supervision under 45 CFR 46.101(b).

Protocol Number: 101-SB19-267	Received: 12/13/2019	Review: Exempt
Expires: 1/14/2023	Approved: 1/15/2020	Category: 2

This exemption covers any research and data collected under your protocol as of the date of approval indicated above, unless terminated in writing by you, the Principal Investigator, or the Boise State University IRB. All amendments or changes (including personnel changes) to your approved protocol **<u>must</u>** be brought to the attention of the Office of Research Compliance for review and approval before they occur, as these modifications may change your exempt status. Complete and submit a Modification Form indicating any changes to your project.

Exempt protocols are set to expire after three years. Annual renewals are not required for exempt protocols. If the research project will continue beyond three years, a new application must be submitted for review. If the research project is completed before the expiration date, please notify our office by submitting a Final Report.

All forms are available on the ORC website at http://goo.gl/D2FYTV

Please direct any questions or concerns to ORC at 426-5401 or humansubjects@boisestate.edu.

Thank you and good luck with your research.

Office of Research Compliance

1910 University Drive Boise, Idaho 83725-1139
Phone (208) 426-5401 orc@boisestate.edu
This letter is an electronic communication from Boise State University