Title
A Study of Relationships Among Three Assessment Methods for Nurse Anesthetists

Author
Gabriel, Nicholas

Publication Date
2013

Peer reviewed|Thesis/dissertation

A Study of Relationships Among

Three Assessment Methods for Nurse Anesthetists

by

Nicholas Gabriel, CRNA

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Nursing

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Acknowledgements

This dissertation would not be possible if not for the continued sacrifices over the years of my best friend, partner, and beloved wife of 31 years, Beth. She supported me throughout my 23 years of active duty in the Air Force and managed to move two small boys across the country alone on two separate occasions without me. These last six years of sacrifice have been the greatest of all, and I fear I will not be able to repay my debt to her. I wish to extol the patient virtue of Beth who brought timely words of encouragement to continue on this endeavor. May God bless and keep her.

To: Mary Blegen, RN, PhD, FAAN: my advisor and dissertation chair. She took on this topic of simulation when others had no idea what I was talking about. My thanks for her guidance and direction.

To: Susan Chapman, RN, PhD: Dissertation committee member, who gave light to theory and enabled me to drive this dissertation by theory. It is a foundation I will keep centerpiece in my future endeavors in simulation research.

To: Bernadette Henrich, PhD, CRNA, my outside expert, dissertation committee member, and dear friend. I could not have accomplished this task without your assessment tool which you so graciously allowed me to use for this research. Your passion and energy for simulation and its future in our profession of nurse anesthesia has encouraged me to continue in this endeavor long after I have completed this dissertation.

A Study of Relationships Among Three Assessment Methods for Nurse Anesthetists

Nicholas Gabriel, CRNA

Abstract

The current measures of competency for nurse anesthetists for recertification are continuing education units for each biennial recertification cycle, and records of current practice and state licensure. Recently, the National Board of Certification and Recertification of Nurse Anesthetists (NBCRNA) adopted new standards that include a written examination every 8 years, an increase of 40 continuing education units per year, and completion of four core competency modules each four-year recertification cycle. However, little is known about the validity of these competency measures.

The purpose of this study was to determine relationships between written examination scores, self-assessment scores, and performance scores in a simulated environment. Eighteen nurse anesthetists from three hospitals completed the written exam, self-assessment, and 8 scenarios in the simulation lab.

The mean score on the 30 item written exam was 67.22%, SD = 11.42. There were no significant differences in scores between groups CRNA employed at different hospitals or of differing age and experience. The mean percentage score of the eight scenarios was 77.28%, SD = 7.35, with a range of scores from 64.50-89.00%. The only statistically significant correlation among the three competency measure was a negative correlation between the written examination and total performance scores ($r = -.407$, $p = .094$). No statistically significant correlations were found between the competency measures and age, years of experience, workplace, and prior exposure to simulation. Self-assessments

were completed before and after taking the multiple-choice exam and the simulation performance tests; scores were lower after the tests but still correlated at .496($p = .036$).

These results bring attention to the need to address the relationship between knowledge and performance. Utilizing performance assessments that have been validated and deemed reliable will help to improve practice standards. This in turn will lead to greater safety in anesthesia patient care.

Table of Contents

List of Tables

Page

List of Figures

CHAPTER I: INTRODUCTION TO THE PROBLEM

Introduction

In the past decade and a half, there has been a strong focus on patient safety and demands for better methods to determine the ongoing qualifications of healthcare professionals to improve patient safety (Burns, 2009). States boards of nursing, medicine and other credentialing organizations are searching for better means to assess continuing competency of healthcare professionals. Continuing competency now encompasses safe practice, improvement in skills, currency of knowledge, and competent care since initial licensure (National Council of States Boards of Nursing, 2005).

Assessment of competence at completion of training is different from competence assessment of performance in ongoing practice. Healthcare professionals start their career as qualified to provide patient care after successful passage of state licensure boards and specialty certification; and have generally been deemed competent to care for patients throughout their career until proven otherwise. This assumption of continuing competence is now being questioned. Studies have shown that physicians performance deteriorated over time and reports of poor performance have been published (Mahmood, 2010). A recent review of the literature demonstrated that 6-12% of physicians fail to meet competency standards. The extraordinary explosion of knowledge and technology has challenged healthcare professionals to keep pace and has added to the urgency to finding new and innovative methods to assess competency on a continuing basis (Hays et al., 2002).

Debate has intensified over the usage of the terms competence and performance. Many definitions have been suggested, and the terms have been used interchangeably in the literature. For the purposes of this completed study, competence is the possession of the knowledge, skills, and attitudes necessary to provide safe patient care. Performance is the application of competence in actual patient care. Repeated assessment of performance is necessary to determining how healthcare professionals obtain and apply new knowledge and skills throughout their career. This summative assessment of performance focuses on concepts of knowledge, clinical reasoning, skills, patient care, interpersonal and communication skills (Boulet, et al., 2008).

Nursing has also begun to address the continuing competency issue. The Advisory Panel for the National Council of State Boards of Nursing is developing outlines for competence assessments (National Council of State Boards of Nursing, 2011). There are many methods for assessing continued competency considered for inclusion: self-assessments, professional portfolios, mandated continuing education programs, peer review programs, reexamination, certification and maintenance of certification, and practice experience (Burns, 2009). Since no one method constitutes competence, a suggested approach is triangulation, which is the mix of assessment tools or methods to gain a more complete picture of the competence of the individual (Fotheringham, 2010).

Anesthesia is a unique healthcare profession in that both nurses and physicians are licensed and certified by specialty board examination to practice anesthesia. For anesthesiologists, board certification is valid for ten years. Recertification of anesthesiologists, a component of which is written exam, has shown a decline in pass rates with time since initial testing. Reasons for this decline are not known (Rhodes, 2007). The purpose of recertification by exam has been to assess knowledge only. Other evidence has also demonstrated a relationship between declines in practice performance and the duration of practice (Choudry, Fletcher, & Soumerai, 2005). In response, the American Board of Medical Specialties developed the Maintenance of Certification to assess more frequently the core issues surrounding professionalism and address the performance assessment of physician practice in all specialties. Assessment includes a written exam to measure cognitive expertise, and assessment of performance (Steinbrook, 2005).

The six core competencies underlying the Maintenance of Certification include medical knowledge, patient care, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice. These competencies were translated into four measureable components: evidence of professional standing, commitment to lifelong learning and self-assessment, evidence of cognitive expertise, and practice performance assessment. Maintenance of Certification of Anesthesia now requires simulation to assess a physician's clinical and teamwork skills in managing critical events (Anesthesiologists, 2011).

Nurse anesthetists have provided anesthesia care for over 100 years. The current measures of competency for nurse anesthetists for recertification are a minimum of 40 continuing education units for each biennial recertification cycle, and records of current practice and state licensure ("National Board on Certification of Registered Nurse Anesthetists," 2010). Recertification by written exam or other measures of competence are not required. The National Board on Certification and Recertification of Nurse Anesthetists (NBCRNA) convened a taskforce to study improvements and enhancements to the recertification process, including the use of simulation for performance assessment (Anesthetists, 2011).

The conclusions of the taskforce resulted in the NBCRNA adopting the Continued Certification Program (CPC) as the new standard for recertification of nurse anesthetists beginning in January 2016. These standards include an increase of 40 continuing education units per year in order to recertify; a recertification written examination every 8 years, starting in 2024; completion of four core competency modules each four-year recertification cycle. The four core competencies include airway management techniques, applied clinical pharmacology, human physiology and pathophysiology, and anesthesia technology. For continuing education units, 15 units of assessed learning per year are required of the total 40 units. These assessed units will be more defined to include simulation as an assessment component (NBRCRNA, 2011).

Simulation has become an increasingly popular mechanism for learning and assessing elements of competence (Sinz, 2007). It is recognized as a medium to assess management of rare clinical events that are otherwise not readily observable in the clinical environment (Tetzlaff, 2009). Simulation-based assessment can provide the platform for assessing skills and behaviors that are deemed crucial to anesthesia practice, especially the ability to recognize and respond to changes in patient condition, problem-solving, crisis and communication management (Tetzlaff, 2009).

Statement of the Problem

Currently, nurse anesthetists are not required, as are their physician counterparts, to recertify by written exam or performance assessment in a simulated environment. A review of the literature finds no studies of nurse anesthetists that included as a component of assessment a written exam or self-assessments. Only two studies have included nurse anesthetists for measuring performance in a simulated environment (Gaba et al., 1998; Henrichs et al., 2009). While a few studies have examined the correlation of different assessment tools in the field of anesthesiology, (Schwid, et al., 2002; Weller, et al., 2005) no studies have been published regarding the relationships between written exams, self-assessments, or performance assessment in a simulated environment of practicing nurse anesthetists. Without knowledge of these assessment approaches, it is difficult to devise future approaches for maintaining safe anesthesia practice.

Purposes and Significance of the Study

This study examined the relationships between written exam scores, self-assessment scores, and performance scores in a simulated environment of nurse anesthetists. First, this study examined written exam scores of practicing nurse anesthetists with various years of practice and different practice settings. Second, this study examined performance scores in a simulated environment using a performance assessment tool. Third, the participant completed a self-assessment tool prior to simulation performance evaluation, and after the evaluation. Fourth, the relationships among these three measures were examined. By examining these relationships, a better understanding of how to implement mixed assessment tools for measuring performance of nurse anesthetists was explored.

The specific aims of this study were:

to examine performance scores of nurse anesthetists on three different assessment methods:

Self-assessment

a. Written knowledge examination
b. Simulated environment using an assessment tool utilized in a prior study

(1) to determine the relationships among knowledge, self-assessment, and performance.

To meet the study aims, a written multiple-choice exam was developed using an expert validity approach through Delphi technique and a pre-test to determine reliability. Performance assessment was measured in a recognized simulation center at UC Davis Medical Center utilizing a performance assessment tool developed by Henrichs and colleagues (2009). A self-assessment tool using a Likert scale modified from a visual analog scale used for reporting self-confidence was utilized (Elgie, Sapien, Fullerton, & Moore, 2010). By using this multi-method assessment format this study determined if these three assessment methods provide similar competency and performance measures of nurse anesthetists.

Setting and Subjects

This study was conducted in the Virtual Center for Patient Care, which is the simulation center for the University of California at Davis Medical Center in Sacramento, California. The sample of participants included nurse anesthetists at UC Davis Medical Center, Kaiser Permanente Hospital North, and Kaiser Permanente Hospital South in

Sacramento, California. The targeted recruitment was 60 participants; however, only 18 volunteers completed the study. Each participant consented to all parts of the study, consented to being videotaped during the simulation exercise, and completed a self-assessment prior to and after their performance session in the simulation center and completed a multiple choice written examination. Each participant completed eight simulated anesthesia scenarios lasting five minutes each. A debriefing of the simulation session was offered to each participant after all participants have completed their study session and data collected. At the completion of the study, their exam scores and simulation scores were provided to each participant along with the study findings.

CHAPTER II: LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

Literature Review

This literature review focuses on the use of written examinations and performance assessment, multi-method assessment, and performance assessment using a simulated environment. This review explores assessments in the healthcare disciplines and specifically to the discipline of anesthesia. Finally, the theoretical framework that forms the base of this research study is explained.

Performance assessment in simulation is defined as "measurements of behavior and product based on settings designed to mimic real-life conditions in which specific knowledge or skills are actually applied" (Palm, 2008, p. 5). Performance measures ideally are able to quantify the degree to which competent care in simulation sessions translates into competence in "real" clinical care settings. While written tests and certification exams are used to demonstrate the "knows-how", simulation has emerged as a vehicle and setting in which individuals and teams can demonstrate the "shows-how". It is at this level that providers can demonstrate their clinical performance.

The current trend for recertification of physicians in all specialties, including anesthesia, is with Maintenance of Certification. This process requires physicians to demonstrate, every ten years, competency in the specialty via self assessment, written exam, specific continuing education, and assessment of practice performance (Rhodes, 2007). These processes are new to the recertification process and are a work in progress.

State Boards of Nursing, State Boards of Medicine, and specialty certification processes in both nursing and medicine require successful passage of written examinations for licensure and certification. This was once considered a one-time event and was valid for the career of the individual healthcare professional (Rhodes, 2007). With the advent of the Maintenance of Certification by the American Board of Medical Specialties, the assessment of cognitive expertise is no longer considered to be a one-time event but a continuous assessment for physicians. This evidence is now provided via a written exam taken at ten-year intervals. Nursing and nursing specialties, as of date, do not require the written exam as a part of re-licensure or recertification.

Comparisons of Assessment Methods

Several studies have been conducted to compare written exam scores to skills performance in pediatric and adult life support courses, nursing student education, intensive care medical residents, pulmonary fellows, and first year internal medicine residents. Some of these studies did not report the quantitative results and are difficult to include in summaries. A systematic literature review performed by the 2005 international consensus conference on cardiopulmonary resuscitation found a weak correlation between written examination of basic life support and skills performance (Resuscitation, 2005). Studies on pediatric life support by Nadel and colleagues (2000) and White, Shugerman, and Brownlee (1998) reported poor correlations between written exam scores and skills performance. Their results were based on percentage scores of written exams and performance scores. Correlations and statistical significance levels were not reported.

In the domain of nursing, there is a prevalent view of knowledge and competence being equal (Ericcson, Whyte, & Ward, 2007). Self-assessment of competency was reported in various studies (Rieman and Gordon, 2007; Sandie and Heindel, 1999). The authors came to several conclusions: those who scored higher on knowledge tests were deemed as competent and having met standards despite no measurement of actual performance; knowledge equated to performance; and, nurses that are more knowledgeable perform at a higher level. Whyte and colleagues (2009) conducted a study to measure the knowledge and performance of experience and novice critical care nurses in a simulated environment. They found no statistically significant correlation between knowledge and clinical performance. Specific statistics for the correlations were not reported.

Crawford and Colt (2004) conducted their study to determine a correlation between written exam on bronchoscopy theory and performance scores of pulmonary fellows in a simulated environment. They found no statistically significant relationship between technical skill and theoretical knowledge. Significant levels and statistics were not reported.

Quantitative Comparisons Among Assessment Methods

Several studies reporting correlations among assessment methods were found in the literature. The results of these studies are presented in Table 1. These studies were placed into the following categories of comparison: written exam and performance score, and performance score and self-assessment. Performance scores include performance in simulation labs or with the objective structured clinical examination (OSCE). Written exams included exams for advanced cardiac and pediatric life support, medical school exams, national board exams, oral board exams, and the United States Medical Licensing

Exam (USMLE), Steps One and Two. Self-assessment included visual analog scales and questionnaires.

The majority of correlational studies reviewed compared written exams to performance scores. Performance scores utilizing a simulation lab were prominent. These results showed a low to moderate correlation between written exam scores and performance scores for two studies: a study on ACLS by Rodgers and colleagues (2010)($r=0.194$), and one on anesthesia by Morgan and Cleeve-Hogg (2001)($r=0.19$, $p<.05$). Three studies in anesthesia found moderate correlations: Morgan and colleagues (2000) ($r=0.43$, $p<.05$), Savoldelli and colleagues (2006) ($r=0.52$, 0.53, $p<.05$), and Schwid and colleagues (2002) ($r=0.44$-0.49, $p<.01$). The discipline of nursing was studied by Hauber and colleagues (2010) ($r=0.542$, $p<.05$), and Elgie and colleagues (2010) ($r=0.37$). These results showed a moderate correlation. Written exam scores and performance scores with medical students and the United States Medical Licensing Examination (USMLE) Steps One and Two were conducted by Rifkin and Rifkin (2005)($r=0.20$,$p=0.27$, $r=0.09$, $p<.001$) and Simon and colleagues (2007)($r=0.395$, $p<.001$). These studies demonstrated one low and one moderate correlation between written exam scores and performance scores. Studies comparing written exams and performance scores with medical students and residents found moderate correlations: Nunnink and colleagues (2010) ($r=0.31$), Waldmann and colleagues (2008) ($r=0.36$, $p<.005$), and Hull and colleagues (1995) ($r=0.36$, $p<.01$). Mudumbai and colleagues (2012) found low to moderate correlations ($r = 0.14 - 0.49$) between anesthesia residents in-training examination scores and performance scores in a simulated environment. No significant statistical levels were reported.

Results from these 13 studies comparing written examination scores and performance scores found all (21) correlations to be positive; those with higher scores on examinations had high performance scores, although only 10 of these were reported to be statistically significant. The correlations ranged between 0.09 and 0.54.

Three studies were conducted to examine correlations between performance scores and self-assessment. Performance scores were elicited from simulator labs and an animal lab, and self-assessment scores were obtained from review of performance via videotape, analog scale, and a five point Likert scale. Ward and colleagues (2003) found a moderate correlation($r=0.50$, $p<.01$), but, Weller and colleagues (2005) ($r=0.321$, $p<.01$), and Elgie and colleagues (2010) ($r=.064$) found low correlations between performance scores and self-assessment scores.

Blanch-Hartigan (2011) did a meta-analysis of 30 studies of medical students' self-assessment of performance and criterion scores. Criterion scores included faculty evaluation ($r=0.16$), test scores ($r=0.37$, $p<.01$), and objective structured clinical exam/standard patient (0.19), and showed an overall low correlation ($r=0.21$, $p<.001$). She found that self-assessment accuracy was significantly higher when self-assessment was compared to test scores and grade point average than faculty evaluations or the evaluation of performance with a standardized patient. These findings are similar to those presented in Table 1.

Simulation and Performance Assessment in Anesthesia

The discipline of anesthesia has been a pioneer in the use of simulation with educating and assessing anesthesia residents over the last decade. The concept of performance assessment for anesthesia providers (anesthesiologists and nurse anesthetists) is a new endeavor of which a paucity of studies has been conducted. For clarity, simulation is a process that attempts to emulate a system, environment, or circumstance. The goal of simulation is to effectively emulate said system, environment, or circumstance sufficiently to meet the objectives for which the simulation was designed (Society for Simulation in Healthcare, 2010).

Studies with anesthesia residents have demonstrated that simulation is a useful environment for assessing performance. Schwid et al. (2001) conducted one study with computer screen-based simulation and found an overall improvement in performance of anesthesia residents on a mannequin-based simulator who first managed anesthetic problems on a computer-based simulator.

Table 1. Correlational Studies

Study	Written Exam And Performance Score	Performance ScoreAnd Self Assessment
Elgie et al., (2010) n = 52 school nurses	r=0.37 no p value	p=.064 no r value
Hauber et al., (2010) n = 15 nursing students	r=0.542 p<.05	
Hull et al., (1995) n = 438 internal medicine clerks	r=0.36 p<.01	
Morgan and Cleave-Hogg (2001) n = 24 medical students	r=0.19 p<.05	
Morgan et al., (2000) n = 140 medical students	R=0.43 P<.05	
Mudumbai et al., (2012) n = 12 anesthesia residents	r=0.14-0.49 No p value	
Nunnick et al., (2010) n = 45 intensive care residents	r=0.31 No p value	
Rifkin and Rifkin (2005) n = 17 medical residents	Step 1: r=0.20 p=0.27 Step 2: r=0.09 p=0.001	
Rodgers et al., (2010) n = 34 nursing students	r=0.194 No p value	
Savoldelli et al., (2006) n = 20 anesthesia residents	Resuscitation r=0.52 p<.05 Trauma r=0.53 p<.05	
Schwid et al., (2002) n = 99 anesthesia residents	r=0.44-0.49 p<.01	
Simon et al., (2007) n = 390 medical students	r=0.395 p<0.001	
Waldmann et al., (2008) n = 147 medical students	r= 0.36 p<.005	
Ward et al., (2003) n = 26 surgical residents		r=0.50 p<.01
Weller et al., (2005) n = 21 anesthesia residents		r=0.321 p<.01

Boulet and colleagues (2003) found significant variations in performance of acute care skills amongst medical students and first year residents with their six-scenario assessment tool. Murray and colleagues (2004) used the same assessment tool using six scenarios but developed four different scoring methods. Results demonstrated that senior residents did better across scenarios than first year residents. The third study was conducted by Murray and colleagues (2005) comparing performance assessment of junior anesthesia residents, senior anesthesia residents, and final year nurse anesthesia students. The same number of scenarios and scoring methods in the previous study (2004) were used. Senior anesthesia residents performed significantly better than junior residents or nurse anesthesia students. Junior residents and nurse anesthesia students performed at a similar level. Limitations included small sample size and number of scenarios.

Modifications to the assessment tool were studied in a more robust design by Murray and colleagues (2007). Eight of twelve scenarios were utilized. Results demonstrated significant differences in performance scores between novice 1^{st} year anesthesia residents and 3^{rd} year residents and practicing anesthesiologists. There was no significant difference in scores between advanced 1^{st} year, 2^{nd} year, 3^{rd} year anesthesia residents and practicing anesthesiologists. The anesthesiologists had no prior exposure to the simulation center or prior experience to a simulation environment.

A recent study with this modified assessment tool was done by Henrichs and colleagues (2009). Their study examined practicing nurse anesthetists and anesthesiologists using the same format of the assessment tool as Murray et al. (2007). A significant group effect demonstrated that overall, the anesthesiologists had higher scores across eight scenarios than did the nurse anesthetists. Limitations included small sample size and practice setting of the nurse anesthetists and anesthesiologists in the study. The limitations of practice setting were participants were from a very limited geographic location.

A study using the assessment tool by Murray et al. (2007) was conducted by Mudumbai, Gaba, Boulet, Howard, & Davies (2012). Using six five-minute scenarios from the assessment tool to measure performance and one created 30-minute scenario to measure non-technical skills were used. Twelve graduating 3rd year anesthesia residents were recruited. Results of the simulation performance were compared to residency examination scores and clinical evaluations by anesthesia faculty and operating room nursing staff. The examination scores were only correlated to the one 30 minute scenario. Results demonstrated a positive correlation between the six 5 minute scenarios and clinical evaluations ($r = 0.18$). Overall performance for the one 30 minute scenario correlated positively with written examination scores ($r = 0.19$). Statistical significance was not reported. Limitations included a small number of participants, selection from only one institution, and the use of only key action scores and not timing or order of actions.

Despite the small number of studies on performance assessment of anesthesiologists using a simulated environment, the American Society of Anesthesiologists have instituted a new recertification format under the Maintenance of Certification for Anesthesiologists (MOCA) under the guidelines based on the six competencies of the American Board of Medical Specialties (Steinbrook, 2005). This new format includes a written examination and a six-hour simulation session assessment covering three common anesthesia events. After two years of this requirement, a post course survey demonstrated that participants felt the content was relevant to their practice (99%) and what they learned changed their practice (94%) (McIvor, Burden, Weinger, & Steadman, 2012). It is not certain whether other medical disciplines will develop simulation scenarios to help in performance assessment of their specialty.

Henrichs's study (2009) is the only published study to measure performance assessment of practicing nurse anesthetists. It is clear that more studies need to be conducted to address this shortcoming. It is unclear how nurse anesthetists will be recertified in the future, but pressure from outside agencies will demand a better process to assess continued competency and performance as other disciplines develop criteria for maintenance of certification for all practicing health care providers.

In summary, studies of performance assessment using simulation in the discipline of anesthesia have evolved from students to practicing providers. Studies have found overall improvement in resident's performance during their residency training and have shown that performance varied across scenarios and levels of experience. Limitations of number of scenarios used and small sample sizes demonstrate the issues that require further research and development of more robust assessment tools and recruitment of subjects.

Assessment Tools

In the discipline of anesthesia, only a few performance assessment tools for simulation have been developed and used since the incorporation of simulation in training and continued competency of anesthesia providers. These tools are comprised of scenarios that can occur under anesthesia and can potentially have an impact on morbidity and mortality.

Schwid and colleagues (2002) used four scenarios and grading forms on 99 anesthesia residents from 10 different institutions to assess performance in simulation. An expert panel of 32 anesthesiologists was used to determine content validity. Two different scoring systems, a short and long form, were used. Three raters had an inter-rater reliability of 0.94-0.96. Construct validity was determined by the progressive scoring commensurate with level of training. Internal consistency of assessment was Cronbach's alpha = 0.71-0.76 for both the long and short form scoring system.

Weller and colleagues (2003) developed an assessment tool using three anesthesia scenarios and a global rating scale to measure performance. The global rating scale was developed by a panel of anesthesiologists who agreed on the tasks for generic management of situations in an operating room. This generic rating scale looked at knowledge and behavior. A total of 28 videotapes from a crisis resource management course were reviewed by three raters. Inter-rater reliability was determined to be 0.79-0.85. Validity was determined by the use of the expert panel and the content of the scenarios based on human factors and accepted principles of crisis resource management.

Weller and colleagues (2005) then used the same three scenarios in a follow-up study to assess psychometric properties. Four raters were used to score the videotapes from 22 anesthesia residents. Inter-rater reliability was 0.73. A generalizability coefficient of 0.58 was calculated. The same global rating scale was used as in the previous study. They drew two conclusions: (a) raters need to have training, practice and feedback to have an acceptable inter-rater reliability, and (b) for an acceptable generalizability coefficient of 0.80, a total of 12-15 scenarios are needed. Validity was addressed by the use of the assessment tool and scoring system from the previous study.

Morgan and Cleeve-Hogg (2000) studied medical students to determine the validity and reliability of an anesthesia performance assessment tool using simulation. Six scenarios were developed and a 25-point checklist was created. Each student participated in one scenario only; a total of 140 students participated. Each student was randomly assigned to a scenario. Inter-rater reliability for the paired raters was 0.86. Validity was based on content of scenarios that was based on learning objectives the committee had developed.

Savoldelli and colleagues (2006) developed two scenarios, resuscitation and trauma, based on oral board examination criteria, to assess performance of final year anesthesia residents in simulation. Scores were a nine-item scale developed by the oral examination board. Each item was rated on a five-point scale. A total of four raters was used, and the inter-rater reliability was 0.93-0.98. Validity was based on the established criteria of the oral examination board.

Boulet and colleagues (2003) developed 10 acute care scenarios based on the USMLE step 3. A scoring checklist was developed of expected actions. Forty medical students participated in 6 of the scenarios and were scored by 4 raters. The ten scenarios were placed into one of two sets, with two scenarios common to each set. Inter-rater reliability was 0.95-0.97. Generalizability coefficient was 0.74 for set 1, and 0.53 for set 2.

Content validity was established by using curriculum from the USMLE step 3 evaluations and a review by four faculty members.

Murray and colleagues (2004, 2005, and 2007) created scenarios based on perioperative events that could be replicated in simulation. They were developed by faculty and simulation staff. The initial assessment tool consisted of 6 scenarios (2004), and 28 anesthesia residents participated in the study. A time-based key action-scoring checklist, along with a key action scoring and global rating scoring checklist was created during scenario development for 6 raters to score the simulation assessments. Generalizability coefficient for the global rating score was 0.72, and the checklist scoring method was 0.59. Inter-rater reliability was not reported. Content validity was determined by faculty and staff development of the assessment tool and scoring system. In a follow-up study, Murray and colleagues (2005) used the same scenario assessment tool on 42 anesthesia residents and 15 nurse anesthesia students. Six raters scored the assessments using the same 3 scoring methods in the previous study. Inter-rater reliability was not reported. Generalizability coefficient for the three scoring methods was 0.51-0.71, with the global rating scoring highest.

In 2007, Murray expanded the assessment tool to 12 scenarios based on review of faculty. These 12 scenarios were topics covered by curriculum content outline for anesthesia residents. Participants were 64 anesthesia residents and 35 practicing anesthesiologists, who participated in 8 of the 12 scenarios. Key action scores were used only. Two primary raters were used. Inter-rater reliability was reported as 0.91. The generalizability coefficient was 0.56. Validity was established by content used in earlier studies and faculty review and input for the additional 4 scenarios.

Henrichs and colleagues (2009) used Murray's (2007) assessment tool and key action scoring method on 26 practicing nurse anesthetists and compared their scores to the 35 anesthesiologists in the previous study (Murray, 2007). The nurse anesthetists participated in 8 of the 12 scenarios. Inter-rater reliability for the two raters was 0.88. The generalizability coefficient was 0.80. Content validity was based on the prior use and review of the anesthesia assessment tool.

One final study done by Mudumbai and colleagues (2012) used Murray's anesthesia performance assessment tool with twelve 3rd year anesthesia residents. Only 6 of the scenarios were used and one long scenario was created to assess nontechnical skills. Two faculty raters were enlisted to score the videotapes. Interrater reliability for the 6 short scenarios was 0.84, and for the one long scenario, 0.88. The generalizability coefficient was not calculated.

Of the six performance assessment tools created, only two (Weller & Murray) have been used more than once. The number of scenarios ranged from three to twelve, and in one study, the participants only performed in one scenario. While still in its early evolution, performance assessment in anesthesia has shown promise.

Several issues remain. First, inter-rater reliability is important to the process of assessing performance from an objective and subjective view, depending on the scoring method used. The debate of using checklists, timing of key actions, and global rating scores continues. The six performance assessment tools used one or a combination of all of the scoring methods. Inter-rater reliability does not appear to have a significant effect on scores.

Second, the number of scenarios appears to have an effect on the performance assessment. Weller and colleagues (2005) found that for the generalizability coefficient to be acceptable (0.80), 12-15 scenarios are needed to adequately measure performance. This is based on the finding that participant by scenario accounts for more variance in scores than rater or scenario alone.

Third, one of the major limitations with assessing anesthesia providers is participation in studies. Sample sizes rarely meet power requirements. Recruitment techniques to enhance participation need to be explored. Surveys to determine methods to encourage participation are needed.

Fourth, current performance assessment studies need to be replicated to determine validity and reliability of the performance assessment tool. Re-exploring content validity, scoring methods, and reliability are important to determining if the tool measures what it is supposed to measure. This is a dynamic process that requires constant re-evaluation. Movement towards performance assessment for recertification will require that the tools used are indeed valid and reliable.

Theoretical Framework

This study used a theoretical framework developed on the concept of competence and competency. The competency model is based on the work of Miller (1990). His framework for clinical assessment of physicians changed the methods by which medical students and residents were evaluated during their education and training (Epstein and Hundert, 2002). His model has evolved and has been modified to more fully delineate the differences between performance and competence (Rethans et al., 2002). This revision focuses on a model for assessing practicing physicians.

Concept of Competence/Competency

The major concept of this study was competence/competency. It has various definitions, depending on the source and context. According to Webster's Dictionary Online (2011), competence is the ability to do something successfully and efficiently; a quality of being adequately or well qualified both physically and intellectually. Competence is a behavior or series of actions that can be assessed, observed, or demonstrated (Bradley & Huseman, 2003). According to the Joint Commission, competency is "a determination of an individual's skills, knowledge, and capability to meet defined expectations" (Joint Commission, 2006, p.234).

To date, there is no official theoretical or operational definition of competency for healthcare providers. The confusion of the definitions of competence and competency further add to the problem. For the purposes of this study, competence and competency were used interchangeably to provide a clearer concept of competence and how this concept relates to the measurement of performance. Specific competencies have been identified as being obtained through skills, knowledge, and performance roles that are assessed by specific criterion (Axley, 2008). Healthcare, especially nursing, has identified competency with technical skill that can be measured. However, nursing involves more abstract skills such as attitudes, critical thinking skills, motivation, and self-assessment (Bradshaw, 1998). Methods to measure and evaluate these aspects of the concept are complex and evolving (Axley, 2008).

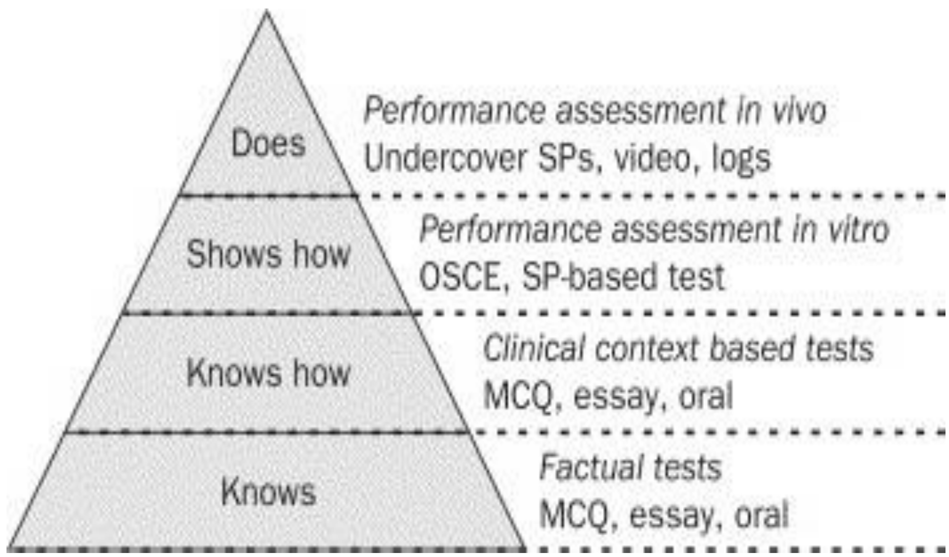
While the evidence reviewed in the previous section is mixed, there does seem to be weak correlations between written examinations and clinical performance. This evidence is based on medical students, residents, and participants in advanced life support courses. Studies with practicing physicians or advanced practice nurses have not been conducted to examine correlations between written examinations, clinical evaluations, and performance assessments for recertification. Maintenance of Certification for medical specialties is being changed to include these methods of assessments.

Miller's Model of Competence

In a landmark article published in Academic Medicine, George Miller (1990) proposed a four-tier framework for clinical assessment of medical students and residents. He designated these tiers or levels as 'knows', 'knows how', 'shows how', and 'does', as shown in Figure 1. 'Knows' is a measurement of knowledge which is usually assessed through tests of knowledge. There are three common formats of knowledge testing: multiple choice questions, essay questions, and short answer questions. The multiple-choice question is considered the gold standard in knowledge assessment (Bashook, 2005). They are frequently used by licensing and certifying specialty boards of medicine and nursing.

'Knows how' is the next level that requires the analysis and interpretation of knowledge and translating the results into a patient care management plan. It is at this level that Miller felt that competence was derived from the skills, judgment, and knowledge of an individual (Miller, 1990). Miller states that modified essay questions and patient management problems are used to assess this level. He stated concerns, however, with the patient management problem format due to varying reliability of raters and the weighting of items scored in the assessment. Miller endorsed the use of assessment tools, but was concerned about the careful design of scoring instruments, checklists, and rating methods.

Figure 1. Miller's Triangle of Competence Assessment



'Shows how' is the third tier or level of the pyramid. Miller refers to this level as the assessing performance level (Rethans, et al., 2002). He based his model on the assumption that competence predicts performance. This level is generally assessed with limited direct observation on a ward or clinic by clinical teachers. Miller felt the most effective method to assess this level was with the simulated clinical encounter utilizing the standardized patient. The trained standardized patient is limited to use for history taking and physical exam, communication skills assessment, and assessing ethnic and cultural differences (Miller, 1990).

The fourth level of the pyramid is 'does'. This level relates to whether the individual can actually function independently in clinical practice. However, to assess a healthcare provider at this level requires a direct observation in the clinical environment. Assessing performance of rare occurrence/ high stakes critical events is difficult at best.

Rethans et al. (2002) proposed a modification of Miller's model of competence. He felt one of the limitations of Miller's model was an assumption of competence predicting performance. Rethans felt that other factors complicated the assumption, such as external factors (time pressure, day of week, mood of the patient) and system influences (clinic facility, government regulation, policies). While Miller's model worked well in an educational setting, Rethans felt that the model was not useful for assessment in actual practice. Miller's model was more static than flexible, and flexibility is important for measuring performance by different methods.

In the modified model, called the Cambridge Model for Performance and Competence, Rethans et al. (2002) inverted the model, emphasized competence (shows how), and performance (does). Their contention was that not all problems of performance are related to competence. Systems and individual influences contribute to acceptable performance or substandard performance. Therefore, performance assessment should be done first in the actual clinical environment by various screening methods to determine if performance standards are met or not. Rethans concludes that addressing the systemic and individual influences first would be more advantageous and less costly than addressing the competence component. Competence could be assessed by written examinations, oral exams, standardized patients, and simulation of clinical and practical skills.

While competence-based assessments of healthcare providers are being tested currently for validity and reliability, performance-based assessments are evolving and require the same rigorous testing for reliability. Rethans et al. (2002) describes performance as "a product of competence combined with the influences of factors related to the system (e.g. facilities, practice time)" (Rethans, et al., 2002, p. 908).

Hays et al. (2002) complemented Rethans Cambridge Model for Performance and Competence by further defining competence and performance constructs. He states that performance assessment is healthcare provider's application of knowledge, skills, and attitudes towards their patients in the clinical environment as opposed to a simulated environment. Limitations of performance assessment are that most applications used are adapted from competence assessment. Hays felt that performance assessment needed its own domains and assessment tools. His concern is its intended use and how poor performers will be addressed.

A list of performance constructs or components include clinical expertise, communication, collaboration, management, personal development, education, professional attributes, and personal health. These constructs could be accessed through direct observation in practice, video observation in practice, covert simulated patients, surveys, interviews, and medical record and data analysis. These constructs reflect the six core competencies of the American Board of Medical Specialties (Steinbrook, 2005).

Theoretical Application and Research Hypotheses

The concept of competence has its early origins in the legal profession and psychology (Axley, 2008). While there is no standardized definition of competence across disciplines, nursing and medicine have begun to address the defining of and implementing of the concept as part of the continuing education and certification processes (Epstein and Hundert, 2002). Debate continues over the interchangeable use of the terms performance assessment and competence assessment. While the ideal performance assessment will occur in clinical settings in which a healthcare provider practices, certain skills required during low frequency/ high stakes events will continue to be problematic if these skills (knowledge, skills, critical thinking, teamwork, communication, etc) are a part of the performance assessment. Waiting for a cardiac arrest to occur in a primary care clinic to assess a nurse's performance could take a long time.

Assessing such performance in a simulated setting is more practical and timely. While the debate rages over the proper use and application of the concepts of performance assessment and competence assessment, this study will assess performance of practicing nurse anesthetists in a simulated environment of low occurrence/high stakes events that could more commonly occur in the operating room environment. Since competence is a part of the concept of performance (shows how and does), the theoretical framework of Miller's model of competence ,with the added emphasis of the Cambridge Model for Performance and Competence, helps to define and clarify the research question at hand.

Based on the theoretical framework in this study, hypotheses were developed to help explore the relationships among written exam scores, self-assessment scores, and performance scores in a simulated environment. Based on the Cambridge model, it was expected that performance scores of nurse anesthetists would be associated with their written exam scores, and their self-assessment scores. These associations are based on the idea that competency assessment measures the fundamental concepts of knows, knows how, shows how, and does. The knows is measured by the written examination. The knows how, shows how, and does, can be measured in the clinical or simulation environment. However, the strength of these associations was open to question. This study has contributed important information to support the development of better methods to assess competence for recertification of nurse anesthetists. The current method of recertification is being challenged in light of the anesthesiologists move to the Certification of Maintenance program. The use of simulation as a learning platform in healthcare education of all disciplines, including nursing and medicine, has now crossed over into the continuing education and certification of various healthcare disciplines.

The lack of studies supporting this move to use simulation to assess performance by the American Board of Medical Specialties has not deterred them from implementing its application. Nurse anesthetists use simulation to teach and train nurse anesthesia students in their educational program. While little is known about the impact of simulation on certification or recertification of nurse anesthetists, the need to study these phenomena is imperative. The knowledge gained from such studies would help drive the validity and reliability of how to properly implement such a program in the domain of nurse anesthesia.

Research done so far has focused on medical students, residents, their written exam scores, and performance in simulation. These students and residents are in a learning environment and whose knowledge and performance is being routinely evaluated. Poor to weak correlations have been found between written exam scores and performance. No study has been done with practicing nurse anesthetists or healthcare providers.

Miller in 1990 developed and published his concept of competence. This concept was before the use of simulation came into practice. Rethans in 2002 modified this competency model to include performance measurements including simulation. The knows is measured by the gold standard multiple choice written examination. Rethans' model focuses on the shows how and does. This can be accomplished by written exams, oral exams, and simulation of clinical and practical skills.

Since the early 2000's, the American Board of Medical Specialties mandated the new concept of life-long learning into the Maintenance of Certification, a method to recertify physicians of all medical specialities. This method includes written examinations and performance measurements specific to each specialty. The American Society of Anesthesiology has chosen the use of simulation as a component of their recertification. The Association of Nurse Anesthetists in 2011 adopted similar content to their Continued Professional Certification without adopting simulation.

What we do not know are correlations between knowledge, self-assessment, and performance assessment. While performance assessment can be measured in a multitude of ways, simulation has taken hold in the academic settings and is making headway on the healthcare provider side. We do not know if knowledge and performance correlate into clinical practice and ultimately its effect on patient safety. Using Rethans Cambridge Model components of knows and shows how, I propose to determine if there are correlations between a written examination, self-assessment, and performance assessment in a simulated environment for nurse anesthetists. This knowledge will help to determine the appropriate path in the future for recertification as to the content of the written examination and the validity and reliability of a performance assessment tool.

CHAPTER III: METHODS

Research Design

This study examined the relationships among different measures of performance of nurse anesthetists. First, this study examined written exam scores of practicing nurse anesthetists with various years of practice and different practice settings. Second, this study examined performance scores in a simulated environment using a performance assessment tool. Third, the participants completed a self-assessment tool prior to simulation performance evaluation, and after the evaluation. Fourth, the relationships among these three measures were examined. By examining these relationships, a better understanding of how to implement mixed assessment tools for measuring performance of nurse anesthetists was explored.

This study examined the relationships between performance assessment, written examination scores, and self-assessment scores using a correlational design. The major limitation of correlation design is that correlation does not prove causation: it only shows that two or more variables are related in some systematic way. Since it was not the intent of this study to prove or disprove a causal relationship between the variables, this limitation did not impact this study.

Setting and Sample

The study took place in the simulation lab at the Center for Virtual Care at the University of California at Davis Medical Center, Sacramento, California. This center is a state of the art simulation lab that provides a briefing room, a simulation room recreating an operating room environment, and a control room. The simulation room has a current anesthesia machine and anesthesia cart stocked with commonly used equipment and medications. The mannequin is from METI Corporation. A videotaping system with multi-view cameras is incorporated into the simulation room. The scenarios were programmed into the system so that each participant was exposed to the same parameters for performance assessment. Each participant session, including the administration of the written exam, self-assessment tool, orientation to the simulation lab, and paperwork requirements (consent, demographic data) took approximately two hours to complete.

The sample of nurse anesthetists was recruited from four area hospitals in the Sacramento, California area. The hospitals are: University of California at Davis Medical Center with 24 CRNA's, Kaiser Permanente North Sacramento with 68 CRNA's, Kaiser South Sacramento with 42 CRNA's, and David Grant USAF Medical Center, Travis Air Force Base, Fairfield, California with 13 CRNA's. A recruiting poster/letter was placed in each facility as well as direct contact made with the Chief Nurse Anesthetist of each facility (See Appendix B). A total of sixty nurse anesthetists was the recruitment goal from all facilities. Selection was to be made on a first come basis until a total of sixty was to be reached. After participants were recruited, a schedule was presented to each participant for convenient self-scheduling of study date participation.

Prior to the start of recruitment, approval was obtained from the Institutional Review Board (IRB) from the University of California at Davis Medical Center and then the Committee on Human Research (CHR) at the University of California at San Francisco (See Appendix A). The original protocol submitted to the UC Davis Medical Center IRB included a clinical evaluation of each nurse anesthetist by their respective supervisor. The UC Davis Medical Center IRB had concerns about the nature of the confidentiality of a practicing provider's clinical evaluation to be shared with the principal investigator. Of special concern was that participants from UC Davis Medical Center worked in the same institution as the principal investigator, and that such information may interfere with participation and informed consent. The clinical evaluation component was removed from the study.

Informed consent was obtained prior to the start of the study day for the participant (See Appendix D). The consent included consent to videotape for purposes of scoring the simulation sessions, and information about how the videotapes and scores from the written exam, and self-assessment would be collected and stored for anonymity and security. Each participant was given a number and all three assessment tools (including the scores from the videotapes) were identified only by number.

The participant information, videotapes, written exam scores, and self-assessment scores were stored and secured in a locked file cabinet in the Anesthesia Department Office in the Patient Support Services Building, located adjacent to the University of California at Davis Medical Center. Only the two raters and the researcher have access to the identity of the participants. Each participant was assigned a number determined by sequence of schedule in the simulation lab and all data from each were identified by the assigned number. After the study was completed, each participant was provided an opportunity to come back to the simulation lab to access their written scores, self-assessment scores, and performance assessment scores during the simulation sessions. When all participants have had an opportunity to review the results, the individual scores and videotapes will be kept on file for one year and then will be destroyed.

Data Collection Tools and Procedures

Measurement Tools

This study utilized 3 different tools and a demographic questionnaire to gather data. Participants were recruited and scheduled for the study in the Center for Virtual Care at UC Davis Medical Center. Data collected in order was demographic data, pre-test self-assessment, written examination, performance assessment in the simulation lab, and post-test self-assessment.

Demographic Data

Demographic data collected included age, current practice setting (Kaiser or UC Davis Medical Center), years in practice since graduation, and prior exposure to simulation in training or practice (See Appendix C). Originally, sex of the participant was included in the demographic data form but was removed as a potential identifier by request of the UC Davis Medical Center IRB.

Self-Assessment

The next data collection tool was a self-assessment tool. This self-assessment tool was completed by the participant prior to the written examination and again after the end of the performance assessment in the simulation lab (See Appendix E & F). This self-assessment form, extracted from the mini-clinical evaluation exercise (Mini-CEX) form, is commonly used to evaluate medical residents on multiple encounters over time in their clinical performance (American Board of Internal Medicine, 2011). This specific clinical evaluation was used by faculty to rate the overall clinical competence of residents. This evaluation included demonstration of judgment, synthesis, caring, effectiveness, and efficiency. A nine point rating scale is used: Unsatisfactory was measured on a scale of 1-3; satisfactory was measured on a scale of 4-6, and superior on a scale of 7-9.

1 2 3	4 5 6	7 8 9
Unsatisfactory	**Satisfactory**	**Superior**

Participants were asked to circle the number that best reflects their clinical practice. This scale was chosen due to its common use in determining clinical competence of medical residents. This particular measurement of the Mini-CEX was found to be an adequate measurement scale for self-assessment for this study. This form is used and approved for medical student and resident evaluation by the American Board of Internal Medicine. Weller et al., (2009) studied the use of this form on 61 anesthesia residents. Results demonstrated a Cronbach's alpha = 0.95.

Written Knowledge Test

The next collection tool was the multiple choice written examination containing 30 clinically based questions (See Appendix G). To compose the written exam for this study, test questions were obtained from two anesthesia board review books (Hall & Chantigian, 2010; Dershwitz & Walz, 2006). The authors state that a panel of experts in the field of anesthesia reviewed these questions to prepare students for their board exams (Hall & Chantigian, 2010; Dershwitz & Walz, 2006).

Fifty questions were extracted from these review books for the exam. The content validity of the questions was established by the fact that nurse anesthesia students and anesthesiology residents for their upcoming written board examination use these review books. The written board examination for nurse anesthetists' exam must be successfully passed in order to become board certified and practice anesthesia in the United States (NBCRNA, 2010).

The fifty questions were then put together into a multiple-choice random format for a pilot study to determine the validity and reliability of the exam. The exam was given to twelve volunteer board certified nurse anesthetists at the annual meeting of the American Association of Nurse Anesthetists in Boston, Massachusetts, in August 2011. For confidentiality purposes, no identifying information was collected, and the written examination was not timed. All twelve pilot study exams were returned completed. Scores were recorded as percent correct. The mean score for the pilot exam of 50 questions was 55.58%. This exam was given to the participants in the main study and was completed after the pre-test self-assessment and prior to the start of the performance assessment portion of the study in the simulation lab.

First, each of the fifty questions was fitted to one of the four core competencies. Second, a Kuder-Richardson-20 reliability analysis was performed on the fifty questions. Results demonstrated a low Cronbach's alpha = .471. To improve internal reliability, a) questions that were specialty related (cardiac, neuro, pediatric, etc) were removed due to the fact that the NBCRNA proposed guidelines for the written examination would not contain questions related to anesthesia specialties (NBCRNA, 2011); b) questions that asked similar content in a different fashion were deleted; and c) questions with extremes scores (greater than 80% correct or greater than 80% incorrect) were removed as outliers. The remaining 30 questions were randomly ordered into the final written examination product for use in this study.

Performance Assessment Tool

The anesthesia performance assessment tool was utilized to measure performance assessment in a simulated environment. The original tool contained twelve clinical anesthesia scenarios, each representing an intraoperative emergency event likely to be experienced by a nurse anesthetist over the course of their professional career. Those twelve scenarios were bronchospasm, anaphylaxis, ventricular tachycardia, myocardial ischemia, right main stem intubation, tension pneumothorax, malignant hyperthermia, blocked endotracheal tube, total spinal, loss of oxygen pipeline, hyperkalemia, and acute hemorrhage.

This instrument/tool was utilized for the first time to assess practicing nurse anesthetists by Henrichs and colleagues (2009). Eight of the twelve scenarios were randomly used to assess 26 nurse anesthetists. From that study, the 4 scenarios in which the nurse anesthetists scored highest and the 4 scenarios in which they scored lowest were selected for this study. Those scenarios included bronchospasm, acute hemorrhage, right main stem intubation, hyperkalemia, tension pneumothorax, total spinal, loss of oxygen pipeline, and malignant hyperthermia. Each participant was given the same 8 scenarios in the same order (See Appendix H & J).

These scenarios are low-occurrence high-stakes events that can occur during an anesthetic either in an operating room or other locations where anesthesia is delivered to patients. Each scenario has 5 to 7 key action scores, which the participant is expected to accomplish in a 5 minute time period (See Appendix H). Each scenario has a unique patient description of their medical history, medications, allergies, type of surgical procedure, and type of anesthetic to be given. Participants were expected to recognize and intervene appropriately for each scripted intraoperative event during a five-minute period.

Scoring is done by percent completion of key action scores. Participants are videotaped for scoring by experienced raters. Scoring discrepancies between the two raters were resolved by the principal investigator.

Inter-rater Reliability

For scoring of the videotapes recorded during the simulation scenarios, two experienced nurse anesthetists were recruited. These individuals are currently employed as full-time practicing nurse anesthetists at UC Davis Medical Center. Years of experience were 5 years and 20 years, respectively. Both were well acquainted with simulation and had experience in the simulation lab. Both raters were oriented to the simulation lab at UC Davis Medical Center, and a comprehensive description of the study per IRB protocol was discussed. Both raters then were taken through the same eight scenarios that the study participants would experience. They were both given two videotapes of individuals who accomplished the same scenarios for practice scoring to determine issues or concerns with the scoring process. The practice scores were then placed into SPSS software for analysis. A total of 96 key actions were scored. A total of 5 discrepancies were found in the scoring requiring an intervention on the part of the primary investigator to resolve. The inter-rater reliability was determined to be .858. Issues with camera angles and interpretation of some key actions were the common theme of discrepancy. These issues were discussed with the raters to more finely tune their rating skills and accuracy of scoring. It was determined that the reliability was acceptable and that the study could start. In previous studies using the same performance assessment tool, inter-rater reliability for Murray and colleagues (2007) was $r = .91$, and Henrichs and colleagues (2009) $r = .88$. Both studies utilized a two-rater system.

Procedure

Recruitment began in earnest by using approved recruitment flyers approved by the IRB and sent to the Chief Nurse Anesthetists at UC Davis Medical Center, Kaiser Medical Center Sacramento North, Kaiser Medical Center Sacramento South, and Travis Air Force Base, Fairfield, California. Follow-up emails to the respective chiefs were sent and questions answered. Phone call follow-up was initiated at their convenience and further questions answered. No response was obtained, however, from the Chief Nurse Anesthetist at Travis Air Force Base. A personal visit confirmed no interest in participation. Further contact was abandoned.

The simulation center was prepared and dress rehearsal was done three times to ensure a smooth process and debugging. At the time, a new simulation center was completed and moved in to. However, it did not meet the needs of this study due to lack of video and software support, so the study continued in the old simulation center in the main hospital. The recording capabilities in the old simulation center were marginal at best and so it was determined that a stand-alone video camera on a tripod would be used to videotape the participants.

The participants contacted the principal investigator and were scheduled on a first-come first-served basis. The participants were met at the front of the Center for Virtual

Care (simulation lab) and were taken to the conference room. The study was explained and the informed consent was completed. A demographics data sheet was completed, as well as the pre-test self-assessment. The written examination was then administered with no time limit. During that time, the principal investigator completed preparation of the simulation lab. Deliberately scheduled after hours and weekends, the principal investigator was the only individual in the Center for Virtual Care to ensure anonymity and confidentiality. The doors to the Center for Virtual Care were locked.

After completion of the written examination, the participant was taken into the simulation lab for orientation to the lab, mannequin, anesthesia machine, equipment, and anesthesia cart. Standardized medications in syringes appropriately labeled were on top of the anesthesia cart (syringes filled with normal saline) and additional equipment needed such as stethoscope, ambu-bag, and large bore I.V. catheters for insertion into the chest if decompression were needed as part of the therapy for a given scenario (tension pneumothorax). Every item needed for successful completion of key actions was available. Mannequin capabilities and limitations were demonstrated, rules of engagement were discussed, such as each scenario was five minutes and would not result in complete resolution of the problem encountered. Questions during the scenario were encouraged, and requests for information such as confirmation of breath sounds or lab results were to be provided only if asked.

A practice scenario was then done to give the participant a “feel” for what to expect. The practice scenario was a routine induction of general anesthesia using the “routine medications” normally used and intubation of the mannequin. The scenario entailed a hypotensive episode after intubation and resolution once the participant identified the problem and treated the episode with known therapeutic modalities. After the practice scenario, further questions were entertained and answered.

The participants were given 8 scenarios, in order, each lasting 5 minutes. Each scenario was videotaped for rater scoring. A written patient history and description of the type of anesthetic and surgical procedure was given prior to each scenario. Each participant was given 5 minutes to review the written description and ask any clarifying questions prior to the start of each scenario. Each scenario began with the participant going into an operating room to give a break to a nurse anesthetist who already started the anesthetic and was waiting for the surgeon to start the procedure. Approximately thirty seconds into the room, the scenario began and was timed. Videotaping began as soon as the participant entered the room. At the end of five minutes, the scenario was stopped regardless of where the participant was in the scenario, and was taken to the conference room and given the next patient history and surgical procedure to be performed. This occurred for each of the eight scenarios. After each scenario, the simulation lab was reset for the next scenario while the participant was taken to the conference room with the door closed to review the next written patient history.

At the end of the eighth scenario, the participant went to the conference room and completed the post-test self-assessment. Once this was accomplished, further questions were answered about setting up a time after data collection to allow the participant to see the videotape and written examination scores if they so chose. Reinforcement of confidentiality of the participant's results and the need of the participant to keep confidential the integrity of the study and data collected. The participant was then thanked for their participation and was given a fifty-dollar gift card. Two videotapes were produced using the eight scenarios designated for the performance evaluation scores in the simulation lab.

The videotapes were scored by two nurse anesthetist raters not participating in the study. A signed statement of confidentiality by each rater was obtained and filed. The raters are two experienced nurse anesthetists who volunteered to review and score the videotapes of the simulation scenarios. Each rater was oriented to the study contents and the simulation lab. Each rater completed the same eight scenarios as the participants in order to understand the scoring system and expectations. Several videotapes of previous performances using the same scenarios were given to the raters to practice scoring and to determine inter-rater reliability prior to scoring of actual participants. Issues and problems with the scoring were addressed prior to the first participant entry into the study. During the study scoring discrepancies between the two raters were resolved by the principal investigator.

Data Analyses

To analyze the data, this study used SPSS version 21 program. The analyses described the responses to each assessment method, psychometric analyses of each tool, and the relationship between the variables. Means, frequencies, inter-item reliabilities, generalizability analysis, correlations, paired sample t-tests, and one-way analysis of variance were used. Due to the small sample size and the exploratory nature of this study, Spearman's correlation coefficients and a p value of 0.10 was used to minimize type 2 error.

A univariate descriptive analysis of demographic data and each study variable was conducted. The categorical variables of practice setting and prior exposure to simulation (yes/no) were extracted from the demographics form completed by each participant. The continuous variables of age and years in practice were entered as well. Data were analyzed to detect any systematic differences across practice sites, years of experience, and prior exposure to simulation as a student or provider.

The written examination was calculated as a percentage score of correct answers. Mean scores within group and between groups with standard deviation were calculated. Correlations with performance scores, self-assessment scores, and written exam scores were calculated. Internal consistency was analyzed using the Kuder Richardson approach to provide an estimate of reliability of content.

The mean score of all participants self-assessment was calculated. A mean score of pre-post self-assessment was calculated. Correlation between self-assessment and other study variables was calculated and analyzed. The difference between the self assessment score prior to and after the performance in simulation lab was analyzed using the paired sample t-test (repeated measures).

Performance scores were a percent score based on the number of key actions accomplished during each simulation scenario. The number of key actions ranged from three to six, depending on the scenario. A total score for each participant was the average score across the eight scenarios. Descriptive statistics such as mean and standard deviation were calculated to determine performance by scenario. One-way between-groups ANOVA was conducted to analyze any potential difference in practice setting and performance scores. Reliability was assessed using inter-rater reliability and by calculating Cronbach's alpha; the generalizability coefficient was used based on the eight scenarios and two raters. To provide scenario discrimination, each partipants' total key action score was correlated with each individual scenario score. This was done to assess if performance on an individual score can predict overall performance by the participant.

Generalizability Coefficient

The generalizability coefficient is used to estimate reliability and measurement fidelity of instruments (Mushquash & O'Connor, 2006). Known as G-theory, it use is to help provide estimates of variance between sources. In this study, the sources are person, scenario, and rater. Improvements in reliability can be obtained by this method by altering or changing the number of persons, scenarios, or raters. Specific statistical software has been created outside of SPSS to help estimate the generalizability coefficient. The acceptable criterion for G-theory reliability is .80. For this study, EduG software version 6.1-e (Educan, Inc.) was obtained.

Summary

This study addressed the question of the relationships among performance scores, written examination scores, and self-assessment scores of nurse anesthetists. To answer this question, this study had two aims: to examine performance scores of nurse anesthetists in a simulated environment; and to determine the relationships among performance scores and knowledge, and self-assessment scores. From the analyses, the findings of this study will contribute to the body of knowledge that can be used to help determine the best methods of recertification of nurse anesthetists in the future.

Chapter IV

Results

Purposes and Aims of the Study

This study examined three measures of competence to improve understanding of performance assessment. The specific aims of this study were:

1. to examine performance scores of nurse anesthetists on three different assessment methods:
    a. self-assessment
    b. Written knowledge examination
    c. Simulated environment using an assessment tool utilized in a prior study
2. To determine the relationships among knowledge, self-assessment, and performance.

Sample

Data were collected on eighteen individuals who volunteered to participate in this study. These CRNA's practiced at either the University of California at Davis Medical Center or Kaiser Permanente sites in Sacramento. Recruitment flyers were sent out to reach an estimated 143 practicing nurse anesthetists in the Sacramento, California area. The estimated response rate was 12.6%.

Demographic Data

The mean age of the participants was 43.90, with a range from age 30 to age 64 (See Table 2). The mean years of experience as a nurse anesthetist were 8.5 years, with a range of between 2 and 20 years of experience. Participants with prior exposure to simulation as a student or practicing nurse anesthetist comprised 72% of the total number of participants, while 28% had not been exposed to a simulation lab or environment prior to this study.

Table 2. Demographic Data

Mean Age	43.90 (30-64)
Workplace: Kaiser	5
Workplace: UCDMC	13
Mean Years of Experience	8.5 (2-20)
Prior Exposure to Simulation Yes(No)	72% (28%)

Self-Assessment Scores

Participants were asked to rate themselves on their clinical performance over the last year on a likert scale with a rating from 1 to 9: a rating of 1-3 reflected unsatisfactory; a rating of 4-6 was satisfactory; and a rating of 7-9 was superior. After the simulation session of eight scenarios, the participant was asked to rate themselves again using the same rating scale. The mean score for pre-assessment was 6.50, with a range of 6-7(See Table 3). Half of all participants rated themselves 6, while the other half rated themselves 7. The mean score for the post-assessment score was 5.39, with a range of 3-7. Approximately 33% of participants rated themselves 7, 25% rated themselves 6, and 44% rated themselves 5 or less. A correlation between pre and post scores was done and demonstrated a correlation of .496, with a significance level of $p = .036$. A paired samples t-test was done and results demonstrated a statistically significant decrease in self-assessment scores after the written exam and simulation session $p = .003$ (two-tailed). The mean decrease in self-assessment scores was 1.11 with a confidence interval ranging from .431 to 1.79. The eta-squared statistic representing effect size for paired samples t-test was .522, indicating a large effect size.

Table 3. Self-Assessment Scores

Self-Assessment	Mean	Standard Deviation	Range
Pre Self-Assessment	6.50	.514	6-7
Post Self-Assessment	5.39	1.54	3-7

Written Examination Scores

The 30 item written exam was administered to the participant after informed consent was obtained and the pre-testing self-assessment form was completed. A time limit was not established. Results demonstrated a mean score overall 67.22, SD=11.42, with a score range of 40-80. Assessing the differences between groups, an independent sample t-test was conducted (UCDMC M=67.54, SD = 12.79; Kaiser M= 66.40, SD = 7.89; $t(16) = .18$, $p=.22$). There was no significant difference in scores between participants from UCDMC and Kaiser. Reliability analysis utilizing Kuder-Richardson 20 demonstrated an alpha = .60. This is an improvement over the pilot study 50 question examination with an alpha = .47.

Table 4. Written Examination Scores

Written Examination	Mean Score %	Range %
Original Pilot Study (50 item)	55.58	32-66
Revised Pilot Study (30 item)	52.17	40-60
Study Exam (30 item)	67.22	40-80

Simulation Performance Scores

Each participant had their performance assessed over eight different scenarios. The total mean percentage score of the eight scenarios for the eighteen participants was 77.28%, SD = 7.35, with a range of scores from 64.50 – 89.00%. The mean percentage scores for each scenario were: bronchospasm: 80%, acute hemorrhage: 81%, right main stem intubation: 74%, hyperkalemia: 60%, tension pneumothorax: 80%, total spinal: 80%, oxygen pipeline loss: 79%, and malignant hyperthermia: 83%.

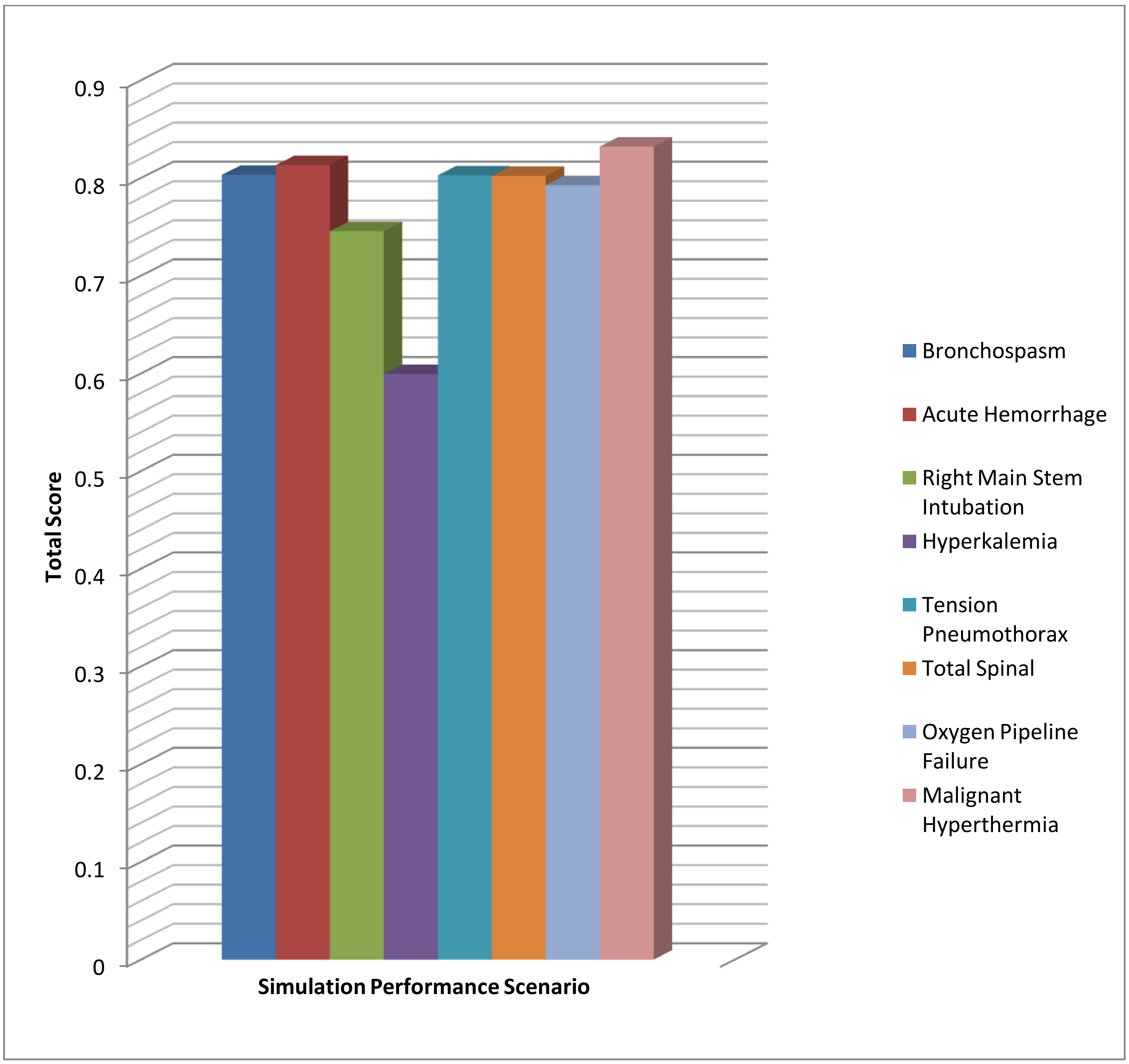
Inter-rater Reliability

Post-study data demonstrated an inter-rater reliability of .844, based on 48 key actions per participant, eighteen total participants, for a total of 864 key actions scored via videotape. A total of 68 discrepancies that required intervention by the principle investigator were found and resolved.

Table 5. Inter-rater Reliability-Simulation Performance

Rating Method	Key Action Scores	Inter-rater Reliability
Pre-study Practice	96	.858
Study	864	.844

Figure 2. Percentage of Key Action Scores Accomplished



Scores for individual scenarios were correlated with the total score to describe internal consistency. A moderate to strong correlation to total was found for five of the eight scenarios: (acute hemorrhage r = .217; right main stem r = .508; Hyperkalemia r = .429; tension pneumothorax r = .646; malignant hyperthermia r = .612) and negative to low correlation in three (bronchospasm r = -.137; total spinal r = .043; and pipeline oxygen loss r = .044). When the three low to negative correlation scenarios are taken out of the model, the results show slightly stronger correlations with the total score: acute hemorrhage r = .359, right mainstem intubation r = .464, hyperkalemia r = .559, tension pneumothorax r = .665, and malignant hyperthermia r = .564. A Cronbach's alpha analysis of the eight individual scenarios demonstrated a negative reliability (inter-item consistency) estimate of -.234.

Another approach to describing the reliability and validity of this tool is the Generalizability Coefficient. Data from each person for each scenario, and the rater score for each individual scenario was placed into the software. The generalizability coefficient for this study was 0.52. When looking for the source of variance and the relative error variance, 91.5% of all variance could be attributed to person x scenario. Rater x person x scenario accounted for 8.5% of all relative error variance. The acceptable criterion for G-theory reliability is .80. Murray et. al (2007) used the same instrument with anesthesiologists and anesthesia residents and had a reliability estimate of 0.56. Hendrichs et. al (2009) study using the same instrument had a reliability estimate of .80. However, her study randomly used all twelve scenarios and her study population was of anesthesiologists and nurse anesthetists, totaling 61 participants.

Therefore, this studies' reliability coefficient reflects the limitations of the number of scenarios, and more importantly, the number of participants.

A one-way between-groups analysis of variance was conducted to explore the impact of workplace, age, experience and prior exposure to simulation on simulation scores. There was no statistically significant difference found for: workplace, ($F(1, 17) = 1.491$, $p = .240$), age ($F(1, 17) = .789$, $p = .661$), experience ($F(1.17) = 1.751$, $p = .279$), and prior exposure to simulation ($F(1, 17) = .193$, $p = .666$.

Table 6. Key Action Scores

Individual Scenarios N = 144	No. of Key Actions in Scenario	Key Action Score Mean ± SD	Scenario Discrimination Individual - Total Score
Bronchospasm	5	4.00 ± 0.84	-.137
Acute Hemorrhage	6	4.89 ± 1.52	.217
Right Main Stem Intubation	6	4.50 ± 1.50	.508
Hyperkalemia	6	3.61 ± 1.91	.429
Tension Pneumothorax	6	4.83 ± 1.04	.646
Total Spinal	6	4.83 ± 0.92	.043
Oxygen Pipeline Failure	6	4.78 ± 0.94	.044
Malignant Hyperthermia	7	5.83 ± 1.69	.612

Correlations

The last research question for this study was to determine if a correlation existed among the three methods of assessing competency: a written examination, self-assessment, and performance scores in a simulated environment. A statistically significant, but negative correlation was found between the written examination and total performance scores ($r = -.407$, $p = .094$). That participants scored higher on the written examination scored lower on the performance exam. Overall self-assessment scores were not correlated at a statistically significant level with total performance scores: pre-test self-assessment ($r = .289$, $p = .244$), post-test self-assessment ($r = .313$, $p = .205$). When the three negative to low correlated scenarios were removed from the model, correlations did not improve pre-test self-assessment ($r = .162$, $p = .521$) and post-test self-assessment ($r = .322$, $p = .192$). Correlations between written exam scores and self-assessment scores were not statistically significant: pre-test self-assessment ($r = .185$, $p = .464$), post-test self-assessment ($r = -.203$, $p = .419$) (see Table 7).

Correlations were done to determine the strength and relationship of the variables age, workplace, years of experience, and prior exposure to any simulation experience to each of the assessment measures. First, the variable age had no statistical significance with the written examination ($r= -.278$, $p = .264$), pre-test self-assessment ($r = -.322$, $p = .192$), post-test self-assessment ($r = -.174$, $p = .490$), and simulation scores ($r = .038$, $p = .882$) (see Table 7).

Next, the variable workplace had no statistical significance with the written exam ($r = -.218$, $p = .385$), pre-test self-assessment ($r = -.124$, $p = .624$), post-test self-assessment ($r = .148$, $p = .559$), and simulation scores ($r = .239$, $p = .339$). (see Table 7).

Since the number of participants from UCDMC was so much larger, further description of the differences by work site were done. UCDMC participants included younger and less experienced CRNA's than Kaiser and had more prior exposure to simulation. Therefore, partial correlations were performed to determine any influence age or years of experience may have had on workplace and simulation scores. Workplace and simulation scores continued to be nonsignificant controlling for age (r = .290, p = .258), and controlling for years of experience (r = .277, p = .282).

Years of experience was analyzed to determine a relationship between the written examination, self-assessment (pre/post), and simulation scores. No correlation was found between years of experience and written examination (r = -.014, p = .955); years of experience and pre-test self-assessment (r = .107, p = .671), years of experience and post-test self-assessment (r = .154, p = .542), and years of experience and simulation scores (r = .238, p = .342) (see Table 7).

Prior exposure to simulation was analyzed to determine a relationship between the written examination, self-assessment (pre/post), and simulation scores. No correlations were found between prior exposure to simulation and written exam scores (r = -.218, p = .385), prior exposure to simulation and pre-test self-assessment (r = -.124, p = .624), prior exposure to simulation and post-test self-assessment (r = -.098, p = .698), and prior exposure to simulation and simulation scores (r = .024, p = .925). (see Table 7).

A closer examination was done to determine if there were correlations between individual scenarios and the written exam, pre-test self-assessment, and post-test self-assessment. The following results were found. Statistically significant negative correlations were found between the written examination and tension pneumothorax r = -.583, p = .011; pre-test self-assessment and right main stem intubation scenario: r = -.389,

$p = .101$; and age and total spinal: $r = -.421$, $p = .082$. That is participants who scored higher on the written exam score lower in performance scores on the tension pneumothorax scenario; participants who rated themselves higher on the pre-test self-assessment had a lower performance score on the right main stem intubation scenario; and participants who were younger in age had a lower performance score in the total spinal scenario. Statistically significant correlations were found between total performance scores and hyperkalemia: $r = .426$, $p = .078$; total performance scores and tension pneumothorax: $r = .635$, $p = .005$; total performance scores and malignant hyperthermia: $r = .601$, $p = .008$; and workplace and right main stem intubation: $r = .532$, $p = .023$. That is participants who scored higher on overall performance scores scored higher in the scenarios hyperkalemia, tension pneumothorax, and malignant hyperthermia; participants from Kaiser Permanente sites scored higher on right main stem intubation than participants from UC Davis Medical Center.

Table 7. Correlations

Variable	Written Examination	Pre-Test Self-Assessment	Post-Test Self Assessment	Simulation Performance Scores	Bronchospasm	Acute Hemorrhage	Right Main Stem	Hyperkalemia	Tension Pneumo	Total Spinal	Oxygen Pipeline	Malignant Hyperthermia
Written Examination					r = -.106	r = -.020	r = -.351	r = .074	r =-.583**	r = .279	r = -.054	r = -.202
Pre-Test Self-Assessment	r = .185				r = -.034	r = .290	r = -.389*	r = .143	r = -.022	r = -.045	r = .202	r = .057
Post-Test Self Assessment	r = -.203	r = 496**			r = 0.00	r = -.078	r = -.120	r = .365	r = .248	r = -.146	r = -.064	r = -.028
Simulation Scores	r = -.407*	r = .289	r = .313		r = -.087	r = .282	r = .320	r = .426*	r =.635**	r = -.041	r = .143	r = .601**
Age	r = - .278	r = -.322	r = -.174	r = .038	r = .346	r = -.075	r = .041	r = -.119	r = .206	r = -.421*	r = -.120	r = .117
Years of Experience	r = -.014	r = .107	r = .154	r = .238	r = .394	r = .131	r = -.080	r = -.070	r = .141	r = -.260	r = -.015	r = .105
Workplace	r = -.218	r = -.124	r = .148	r = .239	r = -.163	r = .039	r =.532**	r = -.171	r = .349	r = -.038	r = -.318	r = .368
Prior Exposure To Simulation	r = -.218	r = -.124	r = -.098	r = .024	r = -.289	r = -.220	r = .260	r = .086	r = .000	r = .276	r = -.119	r = .241

Note: *p<.10 **p<.05 all correlations are Spearman's rho

Chapter V

Discussion and Conclusions

The purpose of this doctoral dissertation was to determine whether different methods of assessing competence would produce similar results. Despite very assertive recruitment, only eighteen practicing nurse anesthetists volunteered to participate in this study. Five of the participants came from the Kaiser Sacramento facilities, and thirteen came from UC Davis Medical Center. No participants came from Travis Air Force Base. There were no other major issues with data collection or participation in the simulation performance.

Their average age was 44 years old, with 8.5 years of experience, and nearly three quarters had been in a simulation lab before, either as a student or as a practicing nurse anesthetist. The average score on the written examination was 67%, the mean score for the pre-test self-assessment was 6.5, and for the post-test self-assessment was 5.39. The mean score for simulation performance was 77.28, and the inter-rater reliability was .844. The generalizability coefficient was .522.

To answer the research question: is there a correlation between written examination scores, self-assessment scores, and simulation scores, bivariate correlations were performed. The only statistically significant correlation found among the different assessment measures was negative; participants with higher written examination scores had lower performance scores. Previous study findings demonstrated low correlation between performance scores and self-assessment scores (Elgie et. al, 2010; Blanch-Hartigan, 2011) to moderate correlations (Weller, et. al, 2003; Ward et. al, 2003). The results of this study demonstrate no correlations compared to previous published studies. Another study using these scenarios (Mudumbai et. al, 2012) also found that the scores from simulation were only moderately and sometimes negatively (statistical significance not reported) related to written exams or clinical assessments.

Previous studies that looked at performance assessments and written examination scores showed low correlations (Morgan and Cleev-Hogg, 2001; Rifkin and Rifkin, 2005) to moderate correlations (Hauber et. al, 2010; Hull et. al, 1995; Morgan et. al, 2000; Savoldelli et. al, 2006; Schwid et. al, 2002; Simon et. al. 2007; Waldmann et. al, 2008). Other studies did not report a statistically significant value.

Additional correlations were also analyzed. The variables age, workplace, years of experience, and prior exposure to simulation were not correlated at a statistically significant level with written examination, self-assessment, and performance scores. A meta-analysis by Choudry (2005) of age, years of experience, and clinical experience reported that in 32 of 62 studies physician performance decreased with increasing years of experience. This study found no correlation between age, years of experience and performance.

Workplace had no correlations with the written examination and self-assessment or simulation scores. Age and years of experience did not have an effect on workplace and simulation scores. Statistically, there was no difference seen in the performance scores between Kaiser and UC Davis Medical Center.

The last of these variables, prior exposure to simulation, also showed no correlations with simulation scores. Two potential reasons for this lack of statistically significant correlations are small sample size and lack of reliability of measures. These are considered below, as is the possibility that each of these measures assesses a different aspect of competence and, therefore, cannot actually be compared.

Written Examination

The mean score was 67, with a range of 40-80. A reliability analysis using Kuder-Richardson demonstrated an alpha = .603. Limited items on an examination are usually the source of error (Nunnally & Bernstein, 1994). Limits of this examination include test length and content validity. Written examinations in the literature are limited to students or residents in training with required written examinations that lead to board certification. No studies have been found that correlate recertification exam scores with performance assessment of practicing healthcare providers. It is not known at this time the type and context of questions that the NBCRNA will focus on for the written examination. The questions on this examination focused on clinical application of anesthesia and not rote knowledge of basic sciences and pharmacology.

Self-Assessment

Self-assessment was done twice, a pre-test and a post-test score to determine if the written examination and performance in simulation had an effect on self-assessment. The pre-test self-assessment of 6.5, and the post-test self-assessment of 5.39; a statistically significant decrease. It appears that the written exam and/or the performance in simulation significantly changed the participant’s self-assessments. Since the post-test was given after both the written examination and the simulation assessment, it would be difficult to extrapolate whether only one or both of the variables changed the participant’s self-assessment. A meta-analysis by Blanch-Hartigan and colleagues (2011) on medical students showed that self-assessment was less accurate if taken before the written exam or performance assessment, and that the specific measurement had an effect on the accuracy of the self-assessment. Self-assessment accuracy was better on written tests than on performance assessment. Ward and colleagues (2003) had senior surgical residents perform a self-assessment at three intervals: after performance on a model, after self-observation of their videotaped performance, and after viewing four videotapes of benchmark performances. These were compared to expert evaluations of their performance. Results showed increased accuracy of self-assessment after self-observation of their videotaped performance. While the research protocol for this study did not allow the participants to view their performance prior to completing the post-test self-assessment, the decrease in self-assessment scores is consistent with previous work.

Simulation Performance

All participants started and finished all aspects of this study, including the performance assessment in the simulation lab. Eight scenarios of low occurrence high stakes issues that can occur under anesthesia were utilized. The eight scenarios were a part of an instrument with a total of twelve scenarios. The mean average score for all eight scenarios was 77.28. Total scores ranged from 64.50 to 89. A moderate correlation between individual scenario scores and the total score was determined for the scenarios acute hemorrhage, right main stem intubation, hyperkalemia, tension pneumothorax, and malignant hyperthermia. However, negative to low correlations were found for bronchospasm, total spinal, and pipeline oxygen loss. When the three negative to low correlation scenarios are removed from the model, the remaining five show moderate to large correlations to the total score. This suggests that there is poor internal reliability among multiple scenarios assessing performance. The Cronbach's alpha of these eight scenarios was negative, further indicating poor reliability.

Two previous studies used this performance tool on anesthesiologists and nurse anesthetists. Murray and colleagues (2007) reported mean scores of anesthesiologists and anesthesia residents ranging from 27-100%, depending on the scenario. Henrichs and colleagues (2009) reported mean score of 66.6% for anesthesiologists and 59.9% for nurse anesthetists. Of importance was the reliability estimate (G-study) calculated for this instrument. Previous reliability estimates ranged from 0.56 (Murray, 2007) to 0.80 (Henrichs, 2009). The reliability estimate for this study was 0.52. While adequate reliability is usually considered 0.80 or higher, this studies' lower reliability estimate can be partially explained by the low number of participants and using only eight of the twelve scenarios in the instrument. Both Henrichs and Murray used all twelve.

The findings of this study and the previous ones bring into doubt the reliability and validity of performance assessment in simulated settings as well as the reliability and validity of written examinations and self-assessments of competence. This is an extremely important issue given the movement in credentialing to periodic performance assessment. Further studies of these tools need to examine each one carefully. Specifically, improvement in the reliability estimate can be done by increasing the number of participants, increasing the number of items or scenarios, or improving the discrimination of the scoring actions. Further study is likely needed to ensure that the scenarios reflect current practice in each problem area.

Limitations

Several limitations for this study were identified. The most significant limitation was the number of total participants and the number of participation from each institution. Only eighteen participants, thirteen from UC Davis Medical Center, and five from Kaiser, out of a population of over one hundred nurse anesthetists in the Sacramento area, volunteered for this study. Every effort was maximized to reach out and contact the potential participants to explain the nature of the study and to reinforce confidentiality. Recruitment was conducted over a five and a half month period, more than adequate time to participate. Multiple reminders via email were sent during that time period. Limited feedback given for non-participation stemmed from discomfort with having their performance videotaped. Most nurse anesthetists who have graduated in the last eight years have had their performance videotaped in a simulation lab at their school as a part of their learning process. A survey in the future to further delineate reasons most chose not to participate might prove helpful to future studies.

The number of scenarios (eight) could also be identified as a limitation. The literature recommends between eight and twelve (Murray & Boulet, 2010) and upwards of fifteen (Weller, 2005). The rationale is that the more scenarios the better the measurement, since most participants will vary in their scoring scenario to scenario. A broad picture is necessary to come to reasonable conclusions of performance estimate. Increasing the number of participants and scenarios will improve the reliability estimate.

The Center for Virtual Care has been in existence for over a decade. New upgrades to equipment were not completed due to the creation of a new simulation center across the street from the Medical Center to the new Information Technology building. This new center became available late September 2012 and opened in October 2012. However, the audiovisual function was not completely installed. I decided to stay in the old simulation lab with its limitations. The main limitation was the audiovisual equipment reliability. The multi-camera suite had reliability issues and was not acceptable for the nature of this "one-shot" to collect data study. The previous simulation technician in charge of equipment left the Center for Virtual Care one month prior to starting the study, leaving no technical help available.

I decided to use a single camera from a less than optimal angle to videotape the participants. This limited the raters from multiple views of participant actions and did not have an overlay of the monitor screen with pertinent vital sign data that the participant was viewing. This limitation may have contributed to more disparities of rater scoring that required intervention. The inter-rater reliability was .844. While acceptable, it could potentially have been higher with better audiovisual presentation of participants.

The anesthesia machine was an older model, no longer in service for use with patients. This machine had various air leaks in its system requiring modification of normal air/oxygen flow levels to the mannequin. This in turn affected the audibility of breath sounds, which is already a major limitation on most mannequins. Auscultation with a stethoscope can become difficult in the middle of a scenario. To overcome this limitation, to ensure each participant had the accurate information to come to correct management of the scenario, the participant was instructed to verbally request if "breath sounds" were present when auscultating the mannequin chest, and to verify the quality of the breath sounds. While not ideal, it allowed each participant to have information vital to the scenarios. This type of modification from reality may diminish the process to evaluate performance and ability.

After review with the two raters, they requested that content validity may need to be revisited, especially the patient information given to the participants, (medical history, type of surgical procedure being done, type of anesthetic being given). A fresh look at key action scores to ensure they reflect accurate and current standards of practice should be considered, with appropriate modifications made. A new look at key action scores and the need to review the order of actions taken for future studies should be conducted

One final limitation of note from participants was the need to debrief after each scenario, or at least at the end of the simulation session. I chose not to do the debrief afterwards, which is a normal and major part of simulation learning, after consultation with the researchers of other studies including Bernadette Henrichs, one of the creators of the instrument used. Data integrity and security among participants and future participants was the rationale. It was a source of frustration on the part of the participants not to know what went right and what to improve upon. I have decided that in future studies with simulation to do a debrief and trust in the professional integrity of the practicing nurse anesthetist. This may have an effect on future recruitment of those who participated in this study.

Implications of Findings

The overall findings of this study demonstrate no significant correlations between the written examination, self-assessment, and performance assessment of nurse anesthetists. The variables age, years of experience, practice setting, and prior exposure to simulation showed no significant correlations. The reliability of the performance assessment tool questionable. These findings have implications in the areas of education, policy and practice, and research.

Education

Recent changes to the recertification process of nurse anesthetists reflects a paradigm change which has occurred in all healthcare disciplines. The concept of life-long learning is now being integrated into this process. Schools of nurse anesthesia have incorporated simulation as a part of the learning process. Written examinations are a part of nurse anesthesia education. The material presented in simulation reflects course content as a part of standardized curriculum. This learning process will be incorporated into the life-long learning of nurse anesthetists once training is completed. A written examination every 8 years will now be standard. Various assessment methods are currently being developed by the NBCRNA. Anesthesiologists have already incorporated simulation as a part of their recertification process. The future of nurse anesthesia will undoubtedly include a performance assessment with simulation. This study has shown that much work is needed to incorporate a written examination reflecting clinical application of knowledge and a valid and reliable performance assessment tool to consistently measure performance. Continuing education requirements will need to focus on clinical application content and the availability of simulation to practice content that will be required for measuring performance, as is practiced by the airline industry. Continuity of education will be the new normal for continuing practice the nurse anesthesia discipline.

Policy and Practice

This study is the first to evaluate relationships between knowledge, self-assessment, and performance of nurse anesthetists. The current climate of healthcare has brought changes to measuring competency of all healthcare providers. The various governing bodies that regulate healthcare have dictated a change in evaluation of competence and performance. These changes reflect a growing concern for patient safety and patient outcomes. This study has shown that much work is still needed to define competence of knowledge and performance for the discipline of nurse anesthesia. Since the written examination is a new requirement for practicing nurse anesthetists, great care by the governing bodies needs to be taken to ensure that the written examination reflects current content necessary for the continuing practice of nurse anesthesia. Innovative methods to assist continuous learning need to be created and implemented, including online modules and content-focused assessments. While simulation is now integrated in schools of nurse anesthesia for students, practicing nurse anesthetists have little to no access at present to practice and rehearse low-occurring high stakes events that can effect patient morbidity and mortality in an intraoperative setting. Standardized scenarios and learning environments will need to be developed that reliably measure accepted performance and knowledge, and avenues to allow the practicing nurse anesthetist to increase their knowledge and performance in a safe learning environment. Other creative assessment methods, such as chart reviews, and evaluation of performance in an actual practice environment, needs to be addressed. If future practice and livelihood is based on knowledge and performance evaluation every eight years, the policy and practices of hospitals, institutions, and governing bodies, such as the NBCRNA, need to encourage and make available avenues to allow successful practice and maintaining of current knowledge for those high stakes examinations.

Research

While simulation has been used in healthcare for over 30 years, more research into validity and reliability of performance assessment needs to be done. Many instruments measuring performance have been created for every discipline, but few have been used more than once to address issues of validity and reliability. Therefore, research needs to focus on these pressing issues if both the written examination and performance assessment are incorporated into the recertification process. Some of these issues that have been found in this study are discussed below.

Performance Assessment

Assessment is becoming an integral part of recertification for medical specialties, and now with nurse anesthetists. The main goal is maintaining and improving performance. This performance can be measured in several ways. The multiple-choice written examination is still one of the most used tools for measuring knowledge, both didactic and in clinical decision-making. Simulation, on the other hand, is now being recognized as an effective method for formative assessment. Formative assessment in simulation includes feedback of strengths and weaknesses of the individual. Many of the low-frequency high-stakes events that can cause morbidity often go unrecognized in the real life setting of the operating room. With new advances and improvements in both mannequin capability and psychometric methods, simulation is slowly moving towards summative assessments. Summative assessments, once the domain of education to determine if a graduate was competent and ready to for independent practice, is now being considered to determine if a provider is competent or not. Besides formative assessments, simulation is now being used to help identify individual skills that need improvement, human factors problems, continuing medical education, and for assessment of providers with lapsed skills. Therefore, it is of great importance that the assessment of the individual must be at the ability level of the individual. Scenarios that give the best measure of the individual must be considered.

This brings forth the concerns of the quality of assessments. Reliability, how consistently the assessment measures the ability of an individual, is as important as identifying sources of measurement error. Issues with sampling and content must be addressed. The choice of raters and scenarios will have a direct influence on the accuracy of scores.

For assessment to be valid, the content must reflect the domains of practice, and be relevant to both knowledge and its application in practice. New ways to measure competencies must be developed and utilized. Currently, direct observation, chart reviews and audits are becoming more common. But, this has challenges as it is difficult to attribute actual patient outcomes to the performance of an individual practitioner. Anesthesia is beginning to utilize simulation to measure decision making and acute care skills in a shortened time frame in settings common to anesthesia, from the operating room to the recovery room. However, as shown in this study, the accuracy of simulation as a performance competence measure is uncertain.

It is becoming more clear that performance assessment will require multiple assessment methods in order to truly measure provider competence. Barriers include inadequate rater training, limited sampling of performance assessment, and contextual errors. If done improperly, the generalizability of the performance to other patient conditions or settings will be called into question. Performance assessment demands that the measurement difficulties be addressed. This includes subjective ratings, inadequate samplings of performances, whether by chart reviews or simulation, and to use a standardized format to ensure that the future measurements can be compared to prior performance measurements. This means that multiple measures are gathered from multiple assessments at several different time periods to be able to ensure reliability estimates.

Another issue to be addressed is the instrument used to measure performance. Whether checklists or key action scores, the validity of such measures by expert reviewers must be re-evaluated frequently to ensure they reflect best practice. If not, scores may not reflect true or intended ability. It will become important that the order and timing of actions be used more accurately reflect true ability. This will require the eventual combining of technical and non-technical skills to more accurately reflect what is required of healthcare providers in their daily practice.

Checklists can be subjective in how they are constructed. It is imperative to determine which actions are really important to the scenario. They also make timing and order of actions difficult to incorporate, thus requiring expert review. However, their use is common because in most acute care scenarios, there is general agreement on what are appropriate actions. They are also easy to score by qualified and trained raters, and if timing and order are important, the videotape can capture the time stamps of the scenario.

Another issue of validity is the paucity of evidence to demonstrate that expert review panels are actually composed of qualified members of the specialty, or that raters are properly trained as reviewers. This becomes an issue when certain actions scored are open to subjective interpretation. When well trained, raters have minimal impact on variance in scoring. Task sampling variance tends to dominate since usually participants tend to perform inconsistently from one task or scenario to the next. Thus, proper training of raters becomes paramount.

One final issue of validity is the scoring method chosen. If appropriate to the construct being measured, one should expect that individuals with more expertise would perform better across scenarios within simulation and across assessment approaches. Those who have trained with simulation should demonstrate improvement over time, and also improvement with additional training. They should be able to retain their skills over time. The strongest validity evidence will come with the ability to link simulation performance with actual patient care and outcomes. The ability to transfer the various skills learned in simulation to the actual patient care setting will transform the process into a valid and reliable measurement method.

Conclusion

Utilizing performance assessments that have been validated and deemed reliable will help to improve practice standards. This in turn will lead to greater safety in anesthesia patient care. By addressing the many challenges found in this study to increase the soundness of the psychometrically based performance evaluations, the profession of nurse anesthesia will continue to contribute significantly to the urgent matter of patient safety in our ever-changing health care environment.

References

American Society of Anesthesiologists (2011). Maintenance of Certification. Retrieved 01 April, 2011, from www.asahq.org

Axley, L. (2008). Competency: a concept analysis. *Nurs Forum, 43*(4), 214-222.

Bashook, P. G. (2005). Best practices for assessing competence and performance of the behavioral health workforce. *Administration and Policy in Mental Health, 32*(5/6), 563-592.

Blanch-Hartigan, D. (2011). Medical student's self-assessment of performance: Results from three meta-analyses. *Patient Educ Couns, 84*(1), 3-9.

Boulet, J. R., Murray, D., Kras, J., Woodhouse, J., McAllister, J., & Ziv, A. (2003). Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology, 99*(6), 1270-1280.

Boulet, J. R. (2008). Summative assessment in medicine: The promise of simulation for high-stakes evaluation. *Academic Emergency Medicine, 15*, 101701024.

Bradley, D., & Huseman, S. (2003). Validating competency at the bedside. *Journal for Nurses in Staff Development, 19*(4), 165-175.

Bradshaw, A. (1998). Defining "competency" in nursing (part II). *Journal of Clinical Nursing, 7*, 103-111.

Burns, B. (2009). Continuing Competency: What's ahead? *J Peinat Neonat Nurs, 23*(3), 218-227.

Byrne, A. J., & Jones, J. G. (1997). Responses to simulated anaesthetic emergencies by anaesthetists with different durations of clinical experience. *Br J Anaesth, 78*(5), 553-556.

Chopra, V., Gesink, B. J., de Jong, J., Bovill, J. G., Spierdijk, J., & Brand, R. (1994). Does training on an anaesthesia simulator lead to improvement in performance? *Br J Anaesth, 73*(3), 293-297.

Choudry, N. K., Fletcher, R.H., & Soumerai, S.B. (2005). The relationship between clinical experience and the quality of health care. *Ann Intern Med, 142*, 260-273.

Crawford, S. W., & Colt, H.G. (2004). Virtual reality and written assessments are of potential value to determine knowledge and skill in flexible bronchoscopy. *Respiration, 71*, 269-275.

Devitt, J. H., Kurrek, M. M., Cohen, M. M., Fish, K., Fish, P., Noel, A. G., et al. (1998). Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anesth Analg, 86*(6), 1160-1164.

Devitt, J. H., Kurrek, M. M., Cohen, M. M., & Cleave-Hogg, D. (2001). The validity of performance assessments using simulation. *Anesthesiology, 95*(1), 36-42.

Dictionary, W. s. (2011). Retrieved 10 June 2011, 2011, from www.websters-online-dictionary.org

Elgie, R., Sapien, R., Fullerton, L., & Moore, B. (2010). School nurse online emergency preparedness training: An analysis of knowledge, skills, and confidence. *The Journal of School Nursing, 26*(5), 368-376.

Epstein, R. M., & Hundert, E.M. (2002). Defining and assessing professional competence. *JAMA, 287*, 226-235.

Ericsson, K. A., Whyte IV, J., & Ward, P. (2007). Expert performance in nursing: reviewing research on expertise in nursing within the framework of the expert-performance approach. *Advances in Nursing Science, 30*(1), E58-E71.

Forrest, F. C., Taylor, M. A., Postlethwaite, K., & Aspinall, R. (2002). Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *Br J Anaesth, 88*(3), 338-344.

Fotheringham, D. (2010). Triangulation for the assessment of clinical nursing skills: A review of theory, use and methodology. *INternational Journal of Nursing Studies, 47*, 386-391.

Gaba, D. M., Howard, S. K., Flanagan, B., Smith, B. E., Fish, K. J., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology, 89*(1), 8-18.

Hauber, R. P., Cormier, E., & Whyte IV, J. (2010). An exploration of the relationship between knowledge and performance-related variables in high fidelity simulation: Designing instruction that promotes expertise in practice. *Nursing Education Perspective, 31*(4), 242-246.

Hull, A. L., Hodder, S., Berger, B., Ginsberg, D., Lindheim, N., Quan, J., & Kleinhenz, M.E. (1995). Validity of three clinical performance assessments of internal medicine clerks. *Academic Medicine, 70*(6), 517-522.

Hays, R. B., Davies, H.A., Beard, F.D., Caldon, L.F.M., Farmer, E.A., Finucane, P.M., et al. (2002). Selecting performance assessment methods for experienced physicians. *Medical Education, 36*, 910-917.

Henrichs, B. M., Avidan, M. S., Murray, D. J., Boulet, J. R., Kras, J., Krause, B., et al. (2009). Performance of certified registered nurse anesthetists and anesthesiologists in a simulation-based skills assessment. *Anesth Analg, 108*(1), 255-262.

Joint Commission. (2006). *Comprehensive accreditation manual for hospitals* (Vol. 2011). Oakbrook Terrace, IL.

Mahmood, T. (2010). Recertification and continuing professional development: The way ahead. *Best Practice & Research Clinical Obstetrics and Gynaecology, 24*, 807-818.

McIvor, W., Burden, A., Weinger, M. B., & Steadman, R. (2012). Simulation for maintenance of certification in anesthesiology: The first two years. *The Journal of continuing education in the health professions, 32*(4), 236-142.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9 suppl), s63-67.

Morgan, P. J., & Cleave-Hogg, D. (2000). Evaluation of medical students' performance using the anaesthesia simulator. *Med Educ, 34*(1), 42-45.

Morgan, P. J., Cleave-Hogg, D. M., Guest, C. B., & Herold, J. (2001). Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Can J Anaesth, 48*(3), 225-233.

Mudumbai, S. C., Gaba, D. M., Boulet, J. R., Howard, S. K., & Davies, M. F. (2012). External validation of simulation-based assessments with other performance measures of third-year anesthesiology residents. [Research Support, Non-U.S. Gov't

Validation Studies]. *Simulation in healthcare: journal of the Society for Simulation in Healthcare, 7*(2), 73-80.

Murray, D. J., Boulet, J. R., Kras, J. F., Woodhouse, J. A., Cox, T., & McAllister, J. D. (2004). Acute care skills in anesthesia practice: a simulation-based resident performance assessment. *Anesthesiology, 101*(5), 1084-1095.

Murray, D. J., Boulet, J. R., Kras, J. F., McAllister, J. D., & Cox, T. E. (2005). A simulation-based acute skills performance assessment for anesthesia training. *Anesth Analg, 101*(4), 1127-1134.

Murray, D. J., Boulet, J. R., Avidan, M., Kras, J. F., Henrichs, B., Woodhouse, J., et al. (2007). Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology, 107*(5), 705-713.

Mushquash, C., and O'Connor, B.P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavioral Research Methods,* 38(3), 542-7.

Nadel, F. M., Lavelle, J.M., Fein, J.A., Giardino, A.P., Decker, J.M., & Durbin, D.R. (2000). Assessing pediatric senior residents training in resuscitation: Fund of knowledge, technical skills, and perception of confidence. *Pediatric Emergency Care, 16*, 73-76.

National Board on Certification and Recertification of Nurse Anesthetists. (2010). Retrieved 10 Jan 2010, from www.nbcrna.com

National Board on Certification and Recertification of Nurse Anesthetists (2011). 2011 Criteria for Recertification. Retrieved 01 April, 2011, from www.nbcrna.com

Norman, G.R., & Streiner, D.L. (2008). Biostatistics: the bare essentials. Hamilton, Ontario: BC Decker, Inc.

Nunnink, L., Venkatesh, B., Bidhani, K., & Udy, A. (2010). A prospective comparison between written examination and either simulation-based or oral viva examination of intensive care trainees' procedural skills. *Anaesth Intensive Care, 38*(5), 876-882.

Nursing, N. C. o. S. B. o. (2005). Continued competence concept paper. Retrieved April 1, 2011, from www.ncbsn.org/919.htm

Nursing, N. C. o. S. B. o. (2011). Retrieved 01 April, 2011, from www.ncsbn.org

Palm, T. (2008). Performance assessment and authentic assessment: a conceptual analysis of the literature. *Practical Assessment Research & Evaluation, 13*(4), 1-11.

Resuscitation, I. L. C. o. (2005). Interdisciplinary topics. *Resuscitation, 67*, 305-314.

Rethans, F. F., Norcini, F.F., Baron-Maldonaldo, M., Blackmore, D., Jolly, B.C., et al. (2002). The relationship between competence and performance: Implications for assessing practice performance. *Medical Education, 36*, 901-909.

Rieman, M. T., & Gordon, M. (2007). Pain management competency evidenced by a survey of pediatric nurse's knowledge and attitudes. *Pediatric Nursing, 33*(4), 307-314.

Rhodes, R. S. (2007). Maintenance of certification. *The American Surgeon, 73*, 143-147.

Rifkin, W. D., & Rifkin, A. (2005). Correlation between housestaff performance on the united states medical licensing examination and standardized patient encounters. *The Mount Sinai Journal of Medicine, 72*(1), 47-49.

Rodgers, D. L., Bhanji, F., & McKee, B.R. (2010). Written evaluation is not a predictor for skills performance in an advanced cardiovascular life support course. *Resuscitation, 81*, 453-456.

Sandie, C. L., & Heindel, L.J. (1999). The knowledge and attitudes of nonanesthesia nurses regarding postoperative epidural analgesia. *AANA Journal, 67*(5), 455-460.

Savoldelli, G. L., Naik, V. N., Joo, H. S., Houston, P. L., Graham, M., Yee, B., et al. (2006). Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia residents. *Anesthesiology, 104*(3), 475-481.

Scavone, B. M., Sproviero, M. T., McCarthy, R. J., Wong, C. A., Sullivan, J. T., Siddall, V. J., et al. (2006). Development of an objective scoring system for measurement of resident performance on the human patient simulator. *Anesthesiology, 105*(2), 260-266.

Schwid, H. A., & O'Donnell, D. (1992). Anesthesiologists' management of simulated critical incidents. *Anesthesiology, 76*(4), 495-501.

Schwid, H. A., Rooke, G. A., Michalowski, P., & Ross, B. K. (2001). Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. *Teach Learn Med, 13*(2), 92-96.

Schwid, H. A., Rooke, G. A., Carline, J., Steadman, R. H., Murray, W. B., Olympio, M., et al. (2002). Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study. *Anesthesiology, 97*(6), 1434-1444.

Simon, S. R., Bui, A., Day, S., Berti, D., & Volkan, K. (2007). The relationship between second-year medical students' osce scores and usmle step 2 scores. *Journal of Evaluation in Clinical Practice, 13*, 901-905.

Sinz, E. H. (2007). Anesthesiology national CME program and ASA activities in simulation. *Anesthesiol Clin, 25*(2), 209-223.

Society for Simulation in Healthcare. (2010). Retrieved 08 June 2011, from www.ssih.org

Steinbrook, R. (2005). Renewing board certification. *N Engl J Med, 353*, 1994-1997.

Tetzlaff, J. E. (2009). Assessment of competence in anesthesiology. *Current Opinion in Anaesthesiology, 22*, 1-5.

Waldmann, U., Gulich, M.S., & Zeitler, H. (2008). Virtual patients for assessing medical students-important aspects when considering the introduction of a new assessment format. *Medical Teacher, 30*, 17-24.

Ward, M., MacRae, H., Schlachta, C., Mamazza, J., Poulin, E., Reznick, R., & Regehr, G. (2003). Resident self-assessment of operative performance. *The American Journal of Surgery, 185*, 521-524.

Weller, J. M., Robinson, B. J., Jolly, B., Watterson, L. M., Joseph, M., Bajenov, S., et al. (2005). Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. *Anaesthesia, 60*(3), 245-250.

Weller, J.M., Jolly, B., Misur, P., Merry, A.F., Jones, A., Crossley, J.G.M., Pedersen, K., & Smith, K. (2009). Mini-clinical evaluation exercise in anaesthesia training. *British Journal of Anaesthesia,* 103(4), 524-30.

White, J. R., Shugerman, R., & Brownlee, C. (1998). Performance of advanced resuscitaiion skills by pediatric housestaff. *Archives Pediatric Adolescent Medicine*, *152*, 1232-1235.

Whyte IV, J., Ward, P., & Eccles, D.W. (2009). The relationship between knowledge and clincial performance in novice and experienced critical care nurses. *Heart & Lung, 38*(6), 517-525.

Willams, B. W. (2006). The prevalence and special educational requirements of dyscompetent physicians. *J Cont Educ Health Prof, 26*, 173-191.

Appendix A

INSTITUTIONAL REVIEW BOARD
University of California, Davis
PROTECTION OF HUMAN SUBJECTS - DECLARATION / ASSURANCE OF IRB APPROVAL
The following research study has been determined to meet the definition of human subject's research as defined by Federal Regulations and UC Davis IRB Policy and has been reviewed by the IRB in accordance with the Common Rule and any other governing regulations:

Project Title
[303087-3] Nurse Anesthetist Performance Assessment
Principal Investigator
Nicholas Gabriel, CRNA, PhDc
School of Medicine
Protocol No.
303087-3
Approval Period
July 31, 2012 through
June 5, 2013
Risk Level
More than Minimal
Risk
Sponsor(s)
Status
New Project
Type of Review
Full Committee Review
Conditions / Comments:
As Principal Investigator for the above-referenced project, you assume certain responsibilities, including, but not limited to:

1. You will conduct the study according to the protocol approved by the IRB. As the PI you are ultimately responsible for the conduct of the research and the protection of rights and welfare of the human subjects. You will ensure, at all times, that you have the appropriate resources and facilities to conduct this study. You will ensure that all research personnel involved in the conduct of the study have been appropriately trained on the protection of human subjects, in addition to the study procedures.
2. Any unanticipated problems involving risks to participants or others will be reported within 5 days to the IRB or in accordance with IRB Standard Operating Procedures (SOPs).
3. Any changes in your research plan (including but not limited to advertisements) must be submitted to the IRB for review and approval prior to implementation of the change, except when necessary to eliminate immediate hazards to participants. Changes in approved research initiated without IRB approval to eliminate immediate hazards to the subject, are to be reported to the IRB in accordance with the SOP, "Reporting of Unanticipated Problems Involving Risks to Participants or Others."
4. Your protocol must be renewed prior to expiration of the study. Failure to submit renewal documents to the IRB Administration will result in a lapse in IRB approval or termination of the study by the IRB. All research involving human subjects must stop without on going IRB approval.
5. If you plan to collect protected health information, you are required to comply with HIPAA requirements.
6. Studies conducted at the CCRC must be reviewed and approved by the VA IRB and the Research & Development Committee prior to initiation of the study. Contact the VA Committee for submission requirements.
7. The UC Davis Health System requires that all investigational drugs be distributed through the UCDMC Pharmacy. You are required to provide a complete copy of the approved protocol to the Investigational Drug Service Pharmacy. A copy of the signed consent form must be submitted to the Pharmacy if investigational drugs are dispensed through the Outpatient Pharmacy.
8. For studies involving investigational drugs at Shriners Hospitals for Children Northern California, drugs must be distributed through Shriners Pharmacy. A copy of the signed consent form must be in the Pharmacy.

Name and Address of Institution
University of California, Davis
IRB Administration
CTSC Bldg, Suite 1400, Rm. 1429
2921 Stockton Blvd.
Sacramento, CA 95817
Institutional Administrator
Daniel Redline
Director, IRB Administration
daniel.redline@ucdmc.ucdavis.edu
Phone No. (916) 703-9151
Fax No. (916) 703-9160
This Assurance, on file with the Department of Health and Human Services, covers this activity:
FWA No: 00004557 Expiration Date: June 28, 2016

Appendix B

Nurse Anesthetist Performance Assessment Study

Be a part of an important study assessing performance of practicing CRNA's

The purpose of this doctoral dissertation study is to assess the performance of nurse anesthetists using four modalities, including simulation performance in a local simulation center

Practicing Certified Registered Nurse Anesthetists (CRNA's) at your institution are eligible to apply

This study will be conducted at UC Davis Medical Center, Sacramento, California, in the Center for Virtual Care, and will take four hours of your time in one session

Contact Nicholas Gabriel, CRNA, PhD$_c$ at

email: nicholas.gabriel@ucdmc.ucdavis.edu

for more information.

Appendix C

Nurse Anesthetist Performance Assessment Study

Demographics Form

Age: _______________

Institution ______________________________

Number of Years CRNA Practice ____________

Prior Exposure to Simulation (circle) YES NO

Appendix D

UNIVERSITY OF CALIFORNIA, DAVIS

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

STUDY TITLE: Performance Assessment of Nurse Anesthetists

This is a research study conducted by Nicholas W. Gabriel, CRNA, PhDc, who is a senior nurse anesthetist in the Department of Anesthesiology and Pain Medicine, and also a doctoral student at the University of California at San Francisco. Nicholas W. Gabriel will serve as the principal investigator (PI). Aaron E. Bair, M.D., Medical Director for the Center for Virtual Care at UC Davis Medical Center, will serve as the co-principal investigator (Co-PI). The Researcher has no outside financial interests in this research. This research is being conducted as a doctoral dissertation study. You have the right to know about the procedures, risks, and benefits of the research study.

Participating in research is your choice and voluntary. You have the right to know about the procedures, risks, and benefits of the research study. If you decide to take part, you can change your mind later and leave the study. No matter what decision you make, there will be no penalty to you.

If you decide to take part in this study, you can decide to stop at any time. We will tell you about new information or changes in the study that may affect your willingness to continue in the study. The Researcher may withdraw you from participating in this research if circumstances arise which warrant doing so, even if you would like to continue.

To participate in this study, you will need to give your written consent by signing this form.

ABOUT THIS RESEARCH STUDY

We hope to learn more about relationships among knowledge, self-assessment, and performance in simulation of nurse anesthetists, in order to contribute to the discussion of future changes in the re-certification process. About 45 practicing nurse anesthetists will take part in this study at UC Davis.

You are being asked to take part in this study because you are a practicing certified registered nurse anesthetist (CRNA) in the Sacramento region.

If you decide to participate in this study, you will be asked to do the following: You will be asked to come to the Center for Virtual Care at the University of California at Davis

Medical Center (UC Davis Medical Center), Sacramento, California, to participate in the study. You will be asked to complete a self-assessment form, complete a 30 question written examination, and participate in eight simulated anesthesia clinical scenarios in the simulation lab. The simulated scenarios will be videotaped for scoring purposes. Each scenario will last five minutes. Once the simulation portion is completed, you will be asked to complete another self-assessment form. The total time required will be four hours. Because this is a research study, the normal debriefing after completion of the eight simulated scenarios will not occur for reasons of study integrity. You will be given the opportunity once the study is completed to return for review of your results on the written exam, self-assessment forms, and viewing of the videotape of performance of the simulated eight anesthesia scenarios.

DISCOMFORT AND RISKS

There is minimal risk to participating in this study. However, you may experience some discomfort related to the assessments we are studying, including clinical and self-assessment, taking a written exam for which you did not prepare, and having your performance evaluated in a simulated environment.

While injury during this study is very unlikely, it is important that you promptly tell the Researcher if you believe that you have been injured because of taking part in this study. If you are injured as a result of being in this study, the University of California will provide necessary medical treatment. The costs of the treatment may be covered by University or the study sponsor or may be billed to your insurance company just like other medical costs. The University and the study sponsor do not normally provide any other form of compensation for injury. You do not lose any legal rights by signing this form.

There may be risks to your privacy. The results of your performance tests will not be shared with anyone, including your supervisors, employers, or fellow CRNA's. The videotapes will be scored by two CRNA's employed at UC Davis Medical Center. The Researchers will store study records and other information about you in a secure location and will grant access only to those with a need to know. However, just like with other personal information kept by your health care providers, your banks, and others, even these safeguards cannot guarantee absolute protection of the data. If private information gets into the wrong hands, it can cause harm.

Your participation in this study and your results will be kept confidential from your employer so as to minimize any potential adverse affect on your job evaluation conducted by your facility.

BENEFITS

There is no direct benefit from taking part in this research. However, you may benefit from the knowledge you get of your personal performance and assessment of strengths in knowledge and critical thinking skills. This may provide opportunities for improvement in the future. The information we get from this study may help us to increase understanding for assessing knowledge and performance of nurse anesthetists as part of the recertification process.

COSTS:

There is no cost to you beyond the time and effort required to complete the procedure(s) described above.

CONFIDENTIALITY

We will do our best to make sure that your personal information will be kept confidential. However, we cannot guarantee total privacy. Your personal information may be released if required by law.

To minimize the risks of a breach of confidentiality, we will assign each participant a research number for identification. Your personal information will be kept in a secured and locked file cabinet in the Department of Anesthesiology and Pain Medicine office. Only the researcher and the chief administrative assistant will have access to this information. All consent forms, completed written examinations, completed self-assessment forms, completed performance assessment score forms and videotapes of the performance assessment in the simulation lab will be kept in the same file cabinet as the personal information. During the study, the participant will be identified only by their assigned research number. The results of the study will be recorded for statistical analysis on a password-protected laptop computer owned by the researcher. This laptop will be kept at the home of the researcher at all times and locked securely in a file cabinet in the researcher's home office. Once the study is completed and data analyzed for the dissertation, the data will be transferred to a password-protected flash drive and placed in the locked file cabinet in the Department of Anesthesiology and Pain Medicine for one year to allow individual participants an opportunity to see their individual results and review the videotape of their performance in the simulation lab. At the end of the year, individual data will be destroyed and shredded.

Designated University officials, including the Institutional Review Board have the authority to review research records.

If information from the study is published or presented at scientific meetings, your name and other personal information will not be used.

WHAT OTHER CHOICES DO I HAVE IF I DO NOT TAKE PART IN THIS STUDY

If you choose not to take part in this study, your Alternative is to not participate.

DO I HAVE TO PARTICIPATE AND CAN I STOP BEING IN THE STUDY

Taking part in this study is your choice and completely voluntary. Alternatives to participating is to decline to participate and not provide informed consent. If you decide to take part in this study, you can decide to stop and withdraw at any time. Inform the principal investigator if you are thinking about stopping or decide to stop. The principal investigator may withdraw you from this research if circumstances arise that warrant doing so even if you would like to continue. We will tell you about new information or changes in the study that may affect your willingness to continue in the study.

COMPENSATION

Each participant who completes the study (one visit) will receive a $50.00 gift card for participation.

Participants who withdraw from the study for any reason will receive a pro-rated $10.00 gift card for participation. Parking will be paid for.

WHO CAN ANSWER MY QUESTIONS ABOUT THE STUDY?

If you have questions, please ask us. You can talk to the Researcher about any questions or concerns you have about this study at:

Nicholas W. Gabriel, CRNA, PhDc, at phone number (707) 494-8575 (cell number)

For questions about your rights while taking part in this study call the Institutional Review Board at (916) 703-9167 or write to IRB Administration, CTSC Building, Suite 1400, Room 1429, 2921 Stockton Blvd., Sacramento, CA 95817. Information to help you understand research is on-line at: www.research.ucdavis.edu/IRBAdmin.

My signature below indicates that I have decided to participate in this study as a research subject. have read and understand the information above. I understand that I will be given a signed and da copy of this consent form.

Signature of Subject	Date	Print Name of Subject
Signature of Person Obtaining Consent	Date	Print Name of Person Obtaining Consent

Appendix E

THE MINI-CEX FORM

The forms are conveniently designed in a slim packet of 10 duplicate forms (one for the resident, one for the program director) that easily fit into a coat pocket. Below is the description provided on the inside cover of the packet (left) and the mini-CEX form (right).

GUIDELINES FOR IMPLEMENTING THE MINI-CEX

The mini-clinical evaluation exercise (CEX) focuses on the core skills that residents demonstrate in patient encounters. It can be easily implemented by attending physicians as a routine, seamless evaluation of residents in any setting. The mini-CEX is a 15-20 minute observation or "snapshot" of a resident/patient interaction. Based on multiple encounters over time, this method provides a valid, reliable measure of residents' performance. *Attending physicians are encouraged to perform one mini-CEX per resident during the rotation.*

Settings to Conduct Mini-CEX:
In-patient services
(CCU/ICU, Ward)
Ambulatory
ED
Other including admission, discharge

Mini-CEX Evaluators:
Attending Physicians
Supervising Physicians
Chief Residents
Senior Residents

Forms and Rating Scale: Packet includes 10 forms; after completing form, provide "original" to program director and "copy" to resident. Nine point rating scale is used; *rating of 4 is defined as "marginal"* and conveys the expectation that with remediation the resident will meet the standards for Board certification.

DESCRIPTORS OF COMPETENCIES DEMONSTRATED DURING THE MINI-CEX

Medical Interviewing Skills: Facilitates patient's telling of story; effectively uses questions/directions to obtain accurate, adequate information needed; responds appropriately to affect, non-verbal cues.

Physical Examination Skills: Follows efficient, logical sequence; balances screening/diagnostic steps for problem; informs patient; sensitive to patient's comfort, modesty.

Humanistic Qualities/Professionalism: Shows respect, compassion, empathy, establishes trust; attends to patient's needs of comfort, modesty, confidentiality, information.

Clinical Judgment: Selectively orders/performs appropriate diagnostic studies, considers risks, benefits.

Counseling Skills: Explains rationale for test/treatment, obtains patient's consent, educates/counsels regarding management.

Organization/Efficiency: Prioritizes; is timely; succinct.

Overall Clinical Competence: Demonstrates judgment, synthesis, caring, effectiveness, efficiency.

If you have any questions, please call ABIM at 215.446.3568.

Mini-Clinical Evaluation Exercise (CEX)

Evaluator: ______________________ **Date:** __________

Resident: ______________________ ❍ R-1 ❍ R-2 ❍ R-3

Patient Problem/Dx: ______________________

Setting: ❍ Ambulatory ❍ In-patient ❍ ED ❍ Other __________

Patient: Age: ______ Sex: ______ ❍ New ❍ Follow-up

Complexity: ❍ Low ❍ Moderate ❍ High

Focus: ❍ Data Gathering ❍ Diagnosis ❍ Therapy ❍ Counseling

1. **Medical Interviewing Skills** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
2. **Physical Examination Skills** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
3. **Humanistic Qualities/Professionalism**
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
4. **Clinical Judgment** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
5. **Counseling Skills** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
6. **Organization/Efficiency** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR
7. **Overall Clinical Competence** (❍ Not observed)
   1 2 3 UNSATISFACTORY | 4 5 6 SATISFACTORY | 7 8 9 SUPERIOR

Mini-CEX Time: **Observing** ______ **Mins** **Providing Feedback:** ______ **Mins**

Evaluator Satisfaction with Mini-CEX
LOW 1 2 3 4 5 6 7 8 9 HIGH

Resident Satisfaction with Mini-CEX
LOW 1 2 3 4 5 6 7 8 9 HIGH

Comments: ______________________

______________________ ______________________
Resident Signature Evaluator Signature

Appendix F

Self-Assessment Tool

of Clinical Performance

Please rate your clinical performance by circling the number that best reflects your practice in your hospital during the past year.

1 2 3	**4 5 6**	**7 8 9**
Unsatisfactory	**Satisfactory**	**Superior**

Appendix G

Written Exam for CRNA's

Instructions: Circle the one best answer for exam multiple-choice question below.

1. The likelihood of intraoperative awareness under general anesthesia is highest with the use of:
   a. Inadequate benzodiazepine dose
   b. High dose opioids
   c. Muscle relaxants
   d. MAC of 0.7% Sevoflurane

2. A patient is brought to the operating room for repair of an open fracture sustained from a fall
   from a window during a house fire. The patient was intubated at the scene and given 100% oxygen via ambu bag. The most reliable method for determining whether the patient has carbon monoxide poisoning while being ventilated with 100% FiO2 is:
   a. routine ABG
   b. pulse oximetry
   c. capnography
   d. EKG evidence of carbon monoxide-induced arrhythmias
   e. arterial carboxyhemoglobin level

3. A patient has had a total laryngectomy in the past. The patient now presents for Mastectomy and Axillary Node Dissection for the management of breast cancer. A reasonable method of managing this patients' airway during a general anesthetic include:
   a. inserting a low pressure cuffed endotracheal tube orally
   b. LMA insertion
   c. awake fiberoptic intubation
   d. inserting a reinforced cuffed endotracheal tube into the tracheostomy stoma

4. The first modality to be lost on the onset of spinal anesthesia is:
   a. touch
   b. motor
   c. temperature
   d. vibration
   e. pain

5. The most sensitive monitor for detection of intraoperative myocardial ischemia is:
   a. creatine phosphokinase levels
   b. EKG changes
   c. transesophageal echocardiography (TEE)
   d. troponin levels
   e. pulmonary artery catheter (PAC)

6. As the neurosurgeon manipulates tissue in the posterior fossa, there are sudden arrhythmias. The anesthetist should:
   a. lower the head
   b. administer lidocaine
   c. inform the neurosurgeon
   d. turn off the nitrous oxide

7. Once detected, the management of the patient with venous air embolism includes all of the following EXCEPT:
   a. inform the surgeon
   b. discontinue the nitrous oxide
   c. decrease elevation of the patients head
   d. control ventilation

8. The typical 2 year old child should be intubated with an endotracheal tube having an internal diameter of:
   a. 3mm
   b. 3.5mm
   c. 4.5mm
   d. 5.5mm
   e. 6.5mm

9. The patient with a full stomach is no longer at risk for aspiration:
   a. once fully relaxed with a muscle relaxant
   b. after the stomach has been decompressed with an nasogastric tube
   c. after proper placement of a cuffed endotracheal tube
   d. none of the above
   e.

10. You are asked to discontinue an epidural catheter on a 68-year-old patient who is 2 days post-sigmoid resection. On review of his medications, you see that he received a dose of low-molecular weight heparin (LMWH) 2 hours ago. You should:
    a. pull the catheter immediately
    b. wait 24 hours before discontinuing the catheter
    c. wait 12 hours before discontinuing the catheter
    d. give a unit of fresh frozen plasma, then discontinue the catheter
    e. remove the catheter after confirming that the activated partial thromboplastin time is normal.

11. The highest incidence of awareness intraoperatively occurs during which type of surgery?
    a. cardiac surgery
    b. obstetrical emergencies
    c. neurosurgery
    d. pediatric surgery
    e. trauma surgery

12. Once a fire is discovered during laser surgery, the FIRST thing the anesthetist must do is:
    a. extinguish the fire
    b. ventilate with 100% oxygen
    c. call for help
    d. stop ventilation
    e. remove and replace the endotracheal tube

13. Immediately after the uneventful induction of general anesthesia and placement of an LMA, you note gastric contents in the airway tubing. The first thing you should do is:

a. suction the patients airway
b. intubate the patient
c. give corticosteroids
d. wake the patient

14. Which of the following mechanisms is most frequently responsible for hypoxia in the recovery room?
    a. ventilation/perfusion mismatch
    b. hypoventilation
    c. hypoxic gas mixture
    d. intracardiac shunt
    e. hypercarbia
15. The most sensitive sign of malignant hyperthermia during general anesthesia is:
    a. tachycardia
    b. hypertension
    c. hypoxia
    d. increased end-tidal carbon dioxide

16. Allergic reactions occurring during the immediate perioperative period are most commonly attributable to the administration of:
    a. local anesthetics
    b. nondepolarizing muscle relaxants
    c. antibiotics
    d. opioids
    e. beta blockers

17. Hyperglycemia is more likely to occur in the diabetic surgical patient with which of the following diseases?
    a. renal disease
    b. rheumatoid arthritis requiring high dosage prednisone
    c. COPD
    d. manic-depressive disorder treated with lithium
    e. congestive heart failure

18. The nurse anesthetist is called to the emergency room to help manage a 3-year-old boy with high fever and upper airway obstruction. His mother states that earlier he complained of a sore throat and hoarseness. The patient is sitting erect and leaning forward, has inspiratory stridor, tachypnea, sterna retractions, and is drooling. Which of the following is the most appropriate management of airway obstruction in this patient?
    a. aerosolized racemic epinephrine
    b. awake tracheal intubation in the emergency room
    c. transfer to the OR and perform an awake tracheal intubation
    d. transfer to the OR, inhalation induction, tracheal intubation
    e. transfer to the OR, intravenous induction, paralysis with succinylcholine, and tracheal intubation

19. Which of the following opioids is unique in that it has both local anesthetic and narcotic properties?
    a. morphine
    b. demerol
    c. sufenta
    d. nalbuphine

20. Epidural use of which of the following opioids would result in the greatest incidence of delayed respiratory depression?
    a. sufentanil
    b. hydromorphone
    c. morphine
    d. meperidine
    e. fentanyl

21. Which is the most sensitive indicator of left ventricular myocardial ischemia?
    a. wall motion abnormalities on the echocardiogram
    b. ST segment changes in Lead 5 on the EKG
    c. appearances of V waves on the pulmonary capillary wedge pressure tracing
    d. decrease in cardiac output as measured by the thermodilution technique

22. The oxygen tanks on an anesthesia machine are:
   a. B Tanks
   b. D Tanks
   c. E Tanks
   d. G Tanks

23. Factors which increase the difference between arterial and measured end-tidal carbon dioxide include:
   a. mismatch of ventilation and profusion
   b. obesity
   c. low fresh gas flow rates
   d. low cardiac output

24. Nasopharyngeal airways are contraindicated in the patient with:
   a. nasal polyps
   b. the jaw wired shut
   c. low hematocrit
   d. patient is a mouth breather

25. Considering the oxygen cylinder supply source, which of the following statements is TRUE?
   a. anesthesia machine hold reserve D cylinders
   b. the hanger yoke assemblies that attach the cylinders to the anesthesia machine are equipped with a Pin Index Safety System to eliminate cylinder interchange
   c. the cylinder supply source is the primary gas source for the anesthesia machine
   d. the cylinder should be left open when the machine is in use in case of a pipeline failure

26. All of the following anesthetic techniques are associated with increased operating room contamination EXCEPT:
   a. failure to turn off gas flow at the end of the anesthetic
   b. filling of vaporizers
   c. use of low-volume high-pressure endotracheal tubes
   d. Jackson-Reese circuits

27. Problems associated with the bellows assembly include which of the following?
   a. Hyperventilation may occur if the ventilator relief valve is stuck in the closed position
   b. a bellows leak can lead to a change in FiO2
   c. hyperventilation may occur if the ventilator relief valve is incompetent
   d. a bellows leak can cause atelectasis if ventilators use a high-pressure driving gas

28. The rapid onset of the central nervous system (CNS) effects of most intravenous anesthetics is best explained by their:
   a. low hepatic extraction ratio
   b. small volume of distribution
   c. high lipid solubility
   d. slow elimination half-life

29. Rank the following induction agents in order of their degree of cardiovascular depression:
   a. propofol>etomidate>thiopental
   b. thiopental>propofol>etomidate
   c. propofol>thiopental>etomidate
   d. etomidate>thiopental>propofol
   e. thiopental>etomidate>propofol

30. Which physical characteristics of fentanyl best accounts for its rapid onset of clinical effect as well as its brief duration of action?
   a. high lipid solubility
   b. high degree of ionization
   c. relatively small molecular weight
   d. low hepatic clearance

Appendix H

Description of Events and Scoring Items for 8 Scenarios

Scenarios	Scoring items
Bronchospasm One minute after beginning the simulation, the oxygen saturation level decreases to 85% and heart rate increases to 120. The blood pressure remains at 105/60. Bilateral wheezing was detectable on auscultation. Elevated peak airway pressures were seen on the Bourdon gauge.	Listen to chest, increase inspired oxygen, state diagnosis, administer beta agonist/ epinephrine.
Acute hemorrhage One minute after the simulation starts the blood pressure begins to decrease to 85/50, heart rate increases to 115. One liter of "blood" is in the suction canister.	Ask about blood loss or evaluate for excessive blood loss (check suction canister), increase intravenous fluids. state diagnosis, request hemoglobin or hematocrit or blood product
Right bronchial intubation At the beginning of the simulation, the vital signs are stable except the oxygen saturation is 91%. There is no chest wall movement on the left side of the chest.	Auscultation or inspection of chest, increase inspired oxygen, state diagnosis, reposition endotracheal tube.
Hyperkalemia At the beginning of the simulation, the blood pressure is 170/90, the heart rate is 75. The heart rate increases, ventricular irritability increases, and peaked *T*-waves are evident on electrocardiogram.	Order or check electrolytes or arterial blood gas or potassium, state diagnosis, institute appropriate treatment.
Tension pneumothorax Pneumothorax is present from beginning of scenario, 60 s after beginning the scenario, the blood pressure decreases to 85/55, heart rate increases to 120, oxygen saturation continues to decrease to 85%.	Auscultation of chest, increase inspired oxygen, state diagnosis, relieve with needle, or place chest tube.
Total spinal The patient had a combined neuraxial/general anesthetic. Within 1 min of starting the simulation, the blood pressure starts decreasing to 60/40, heart rate decreases to 40.	Increase inspired oxygen, check blood pressure, turn-off agent, increase fluids, state diagnosis, give epinephrine.
Loss of pipeline oxygen Fifteen seconds after beginning the simulation, the pipeline oxygen is turned off. At 30 s, the alarm sounds. Vital signs remain stable. The first oxygen tank is empty.	State diagnosis, open oxygen tank #1, open oxygen tank #2.
Malignant hyperthermia Within 1 min of beginning the scenario, the heart rate and blood pressure increase to 115 and 180/90, respectively. Increased end-tidal carbon dioxide level.	State diagnosis, turn-off agent, call for Dantrolene or malignant hyperthermia cart.

.

Appendix I

Scoring Checklist for Raters

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Bronchospasm		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Listened to chest		
3. Increased inspired oxygen concentration		
4. Stopped antibiotic		
5. Administered beta agonist		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Acute Hemorrhage		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Checked suction canister or asked about blood loss		
3. Ordered hemoglobin/HCT		
4. Open fluids wide		
5. Treated hypotension with phenylephrine or ephedrine		
6. Inquired about and/or requested blood product		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Endotracheal Tube in Right Main Stem Bronchus		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Listened to chest		
3. Increased inspired oxygen concentration		
4. Pulled back ETT		
5. Re-expanded lung with increased tidal volumes or by hand ventilating after pulling back ETT		
6. Listened to chest again		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Hyperkalemia		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Hyperventilated patient		
3. Ordered or checked electrolytes/potassium		
4.Ordered arterial blood gases		
5. Gave 10% calcium chloride/gluconate IV and/or dextrose 50% IV + insulin IV		
6. Gave sodium bicarbonate IV		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Tension Pneumothorax		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Listened to chest		
3. Increased inspired oxygen concentration		
4. Hand ventilated patient		
5. Informed surgeon to stop CO_2 insufflation		
6. Inserted needle into chest to relieve tension pneumothorax		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Total Spinal		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Increased inspired oxygen concentration		
3. Establish airway		
4. Open fluids wide		
5. Decrease agent		
6. Administer epinephrine		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Loss of Pipeline Oxygen		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Checked function of oxygen flush		
3. Checked pipeline pressure gauge		
4. Checked oxygen flow meter		
5. Opened oxygen tank #1		
6. Opened oxygen tank #2		

Comments:

Please check “yes” or “no” if the subject carried out the task appropriately.

Scenario: Malignant Hyperthermia		
Task	***Yes***	***No***
1. Stated diagnosis		
2. Turned off inhaled agent		
3. Increased inspired oxygen concentration		
4. Begin hyperventilation		
5. Ordered dantrolene and administered		
6. Ordered ABGs		
7. Ordered electrolytes		

Comments:

Appendix J

Scenario # 1

A 28 year-old male, 80kg, NKDA, ASA II, with a history of bronchial asthma and a history of alcohol abuse is undergoing a laproscopic cholecystectomy. Anesthesia was induced with propofol 150mg, Fentanyl 150mcg, Lidocaine 100mg, and Rocuronium 50mg. Endotracheal tube placement was verified clinically and with capnography. Sevoflurane is at 1%. Ancef 1gm is infusing. Standard monitoring is being used. The patient has just been intubated and his abdomen is being prepped by the circulating OR nurse. You are relieving the CRNA and have received report.

Patient medications: albuterol inhaler prn; advair QD

Scenario # 2

A sixty-year-old male, 85 kg, NKDA, ASA II, with a history of hypertension, is undergoing a total left hip revision for the third time under general anesthesia. Induction was performed with Propofol 120mg, Fentanyl 150mcg, Lidocaine 100mg, and Rocuronium 50mg. Endotracheal intubation was accomplished without event using a #8 endotracheal tube. Pt has two 18g peripheral IV's. An arterial line has been established. Standard monitors are in place. Ancef 1gm IV is infusing. You have relieved the CRNA who started the case. Report has been received. Incision has been made.

Patient medications: lisinopril, hctz

Scenario # 3

A 22 year old male, 70kg, ASA I, allergy to penicillin (rash), with a history of sinusitis, presents for a laproscopic appendectomy. Rapid sequence induction was accomplished with Propofol 150mg, Fentanyl 150mcg, Lidocaine 100mg, and Succinylcholine 120mg. #8 endotracheal tube was placed and verified with capnography. Standard monitors are in place. Rocuronium 30mg IV has been given. Ancef 1gm IV is infusing. Patient has one 18g peripheral IV. Sevoflurane is at 1%. Patient's abdomen is being prepped by the circulating OR nurse. You have relieved the CRNA who started the case and have received report.

Scenario # 4

A 70 year old, 75 kg, ASA III, male patient with NKDA, and a history of stage III renal insufficiency, coronary heart disease, hypertension, diabetes type 2, is undergoing an exploratory laparotomy under general anesthesia. Induction was with Etomidate 14mg, Fentanyl 100mcg, Lidocaine 100mg, and Succinylcholine 120mg. #8 endotracheal tube was placed and confirmed with capnography. An arterial line was placed. Patient has two 18g peripheral IV's. Standard monitoring is in place. Sevoflurane is at 1%. Ancef 1gm IV is infusing. You relieve the CRNA who started the case. Report has been received. The OR circulating nurse has completed the abdominal prep.

Patient medications: Lisinopril, Norvasc, Metoprolol, glucophage.

Scenario # 5

A 25 year old female, 70kg, ASA II, NKDA, with a history of smoking presents for exploratory laparoscopy under general anesthesia. Induction was with Propofol 150mg, Fentanyl 200mg, Lidocaine 100mg, and Rocuronium 50mg. # 7 endotracheal tube was placed without event and confirmed by capnography. Patient has one 18g peripheral IV. Sevoflurane is at 1.2%. Ancef 1 gm IV is infusing. Standard non-invasive monitoring is in place. The abdomen has been prepped, and the surgeon has started the procedure. The pneumoperitoneum has started. Surgeon requests the patient in slight trendelenburg position. You relieve the CRNA who started the case. Report has been received.

Patient medication: birth control pills

Scenario # 6

A sixty year old male, 80kg, NKDA, ASA II, with a history of bladder cancer, hypertension, is undergoing bladder surgery under combined general/epidural anesthesia. A T11-12 epidural catheter was placed preop without event. 2% Lidocaine, 200mg was injected per epidural while abdomen was being prepped. The airway was secured with a #8 endotracheal tube after induction with Etomidate 14mg, Fentanyl 100mcg, Lidocaine 100mg, and Rocuronium 50mg.

Standard monitors are in place. Sevoflurane is at 1%. Ancef 1 gm IV is infusing. Surgeon has made incision. You relieve the CRNA who started the case. Report has been received.

Patient medications: Norvasc, HCTZ

Scenario # 7

A 55-year-old male, 100kg, ASA III, NKDA, with a history of significant COPD, asthma, smoker, and hypertension, is undergoing a right sided inguinal hernia repair under general anesthesia. Patient refuses regional anesthesia. Induction is uneventful with Etomidate 16mg, Lidocaine 100mg, Fentanyl 250 mcg, and Rocuronium 60mg. #8 endotrachael tube is placed and confirmed by capnography. Ancef 1 gm IV is infusing. The abdomen is being prepped by the OR circulator nurse. Sevoflurane is at 1%. Patient has one 18g peripheral IV. You relieve the CRNA who started the case. Report has been received.

Scenario # 8

A 16 year old male, 80 kg, ASA I, NKDA, with no prior medical history, presents for tonsillectomy under general anesthesia. Induction is Propofol 200mg, Fentanyl 200mcg, Lidocaine 100mg, and Succinylcholine 120mg. #7 endotracheal tube is placed and confirmed with capnography. Standard monitors are in place. One 18g peripheral IV is intact, and Ancef 1 gm IV is infusing. Sevoflurane is at 1%. You relieve the CRNA who started the case. Report has been received. The surgeon is about to begin.

Patient medications: None

Publishing Agreement

It is the policy of the University to encourage the distribution 0/ all theses, dissertations, and manuscripts. Copies 0/ all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of Califomia, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

O1 Aug 13

Date