2017

An Examination of the Impact of Simulation and Multimedia Instruction on Central Venous Catheterization

Lord, Jason

UNIVERSITY OF CALGARY

An Examination of the Impact of Simulation and Multimedia Instruction on Central Venous Catheterization

by

Jason Allan Lord

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

MARCH, 2017

Abstract

Dependable assessment tools are essential for Competency Based Medical Education (CBME). Competence in central venous catheterization (CVC) is a key objective to be learned by trainees. Tools to assess technical competency include checklists, critical error tools, Objective Structured Assessment for Technical Skills (OSATS) tools and the Ottawa Surgical Competency Operating evaluation (O-SCORE) tool. This study examined the impact of a simulation-based educational intervention on resident knowledge and performance of CVC. It also compared the dependability of the scores derived from the four assessment tools. Junior residents completing their first ICU rotation in Calgary participated in the study. The control group received didactic instruction. The intervention group received simulation-based teaching and an online multimedia educational module. No observed differences between groups were identified in any of the assessment measures. Global rating scales such as the OSATS or O-SCORE tools outperformed checklists or critical error tools when assessing competence for this procedure.

Acknowledgements

First and foremost, thank you to Dr. Jocelyn Lockyer for agreeing to supervise the completion of this project. I am deeply grateful for all your efforts and guidance.

Thank you to Dr. Maria Palacios Mackay for helping me to better understand the methods used in this study. It has enabled me to grow as a researcher in the field of medical education.

Thank you to Dr. Dan Zuege for agreeing to sit on my committee and participate in the data collection and for being both an academic and clinical mentor.

Thank you to Dr. Amanda Roze des Ordons for dedicating hours of time analyzing the videos and helping to collect data for this study.

Thank you to Dr. John Norcini for reviewing the results of the generalizability study and to Dr. Chip Doig for providing the opportunity and time to complete this project.

Finally, thank you Fiona for your tireless support and encouragement in helping me to get this done.

Dedication

To Fiona:

Your love and support have always been there. I could not have done it without you.

To my mom and dad:

Thank you for providing me the love, guidance and opportunity to succeed. I am thinking of you always.

Table of Contents

Abstract ii
Acknowledgements iii
Dedication iv
Table of Contents v
List of Tables viii
List of Figures and Illustrations ix
List of Symbols, Abbreviations and Nomenclature x

CHAPTER ONE: INTRODUCTION 1
1.1 Background 1
1.2 The research team 2
1.3 Study goal 3

CHAPTER TWO: LITERATURE REVIEW 4
2.1 Competency based medical education in Canada 4
2.2 Simulation-based procedural skills training 5
2.3 Central venous catheterization 7
2.4 Acquiring procedural competence 8
2.5 Assessment tools for procedural competence 10
2.6 The validity argument 12
2.7 Research questions 14
2.7.1 Resident knowledge pertaining to CVC (Questions 1a and 1b) 14
2.7.2 Resident competence performing CVC (Question 2) 14
2.7.3 Resident comfort and satisfaction regarding their educational experience (Questions 3a, 3b & 4) 14
2.7.4 Dependability of performance scores (Question 5) 15

CHAPTER THREE: METHODS 16
3.1 Basic study design 16
3.2 Methods 19
3.2.1 Setting 19
3.2.2 Participants 19
3.3 Procedure 20
3.4 Outcome measures 22
3.4.1 Research questions 1a & 1b: Resident knowledge pertaining to CVC 22
3.4.2 Research question 2: Resident competence performing CVC 23
3.4.2.1 Competence tool #1 – Checklist 23
3.4.2.2 Competence tool #2 – Critical error tool 24
3.4.2.3 Competence tool #3 – Objective Structured Assessment for Technical Skills (OSATS) 25
3.4.2.4 Competence tool #4 – Modified Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) 25
3.4.3 Research questions 3a, 3b & 4: Resident comfort and satisfaction regarding their educational experience 26
3.4.4 Research question 5: Dependability of the assessment tools 27

3.5 Data analysis 28
3.6 Validity evidence 30
3.7 Ethics approval 30

CHAPTER FOUR: RESULTS 31
4.1 Participants' demographic characteristics and prior CVC experience 31
4.2 Knowledge acquisition and change 34
4.2.1 Knowledge acquisition 34
4.2.2 Change in knowledge 35
4.3 Competency achievement 35
4.3.1 Checklist 35
4.3.2 Critical error tool 37
4.3.3 OSATS 38
4.3.4 O-SCORE 38
4.3.5 Time to cannulation and number of attempts 39
4.4 Exit survey results 40
4.4.1 Survey responses 40
4.4.2 Between groups differences 41
4.4.3 Change in comfort 42
4.5 Dependability of the performance scores 43
4.5.1 Variance component analysis 43
4.5.2 Dependability coefficient calculation 46
4.5.3 Decision study 46
4.6 Validity evidence 48

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS 51
5.1 Thesis questions 51
5.1.1 Question 1a: Does an educational intervention using a multimedia teaching module, together with simulation based teaching, improve junior residents' knowledge regarding the indications, the contraindications, the technique and the potential complications related to CVC as compared to a control group receiving the current educational format during an ICU rotation at our institution? 51
5.1.2 Question 1b: Does an ICU rotation at our institution improve junior residents' knowledge related to CVC regardless of the educational intervention? 52
5.1.3 Question 2: Is there a difference in performance of CVC between junior residents trained using a multimedia teaching module, together with simulation based teaching, and those receiving the standard educational format during an ICU rotation at our institution? 53
5.1.4 Question 3: Is there a difference in self-reported comfort with CVC between junior residents trained using a multimedia teaching module, together with simulation based teaching, and those receiving the standard educational format during an ICU rotation at our institution? 54
5.1.5 Question 4: Is there a difference in self-reported satisfaction with the overall educational experience between residents using a multimedia teaching module, together with simulation based instruction using a

procedural task trainer, and those receiving the standard educational format following an ICU rotation at our institution? 55
5.1.6 Question 5: What is the dependability of the scores derived from the performance assessment tools used in this study for decisions regarding resident competency in CVC? 57
5.2 Implications for education 60
5.3 Limitations 64
5.4 Future directions 66
5.5 Conclusions...66

REFERENCES 68

APPENDIX A: PRE-TEST WRITTEN EXAMINATION 76

APPENDIX B: POST-TEST WRITTEN EXAMINATION 78

APPENDIX C: CHECKLIST TOOL 80

APPENDIX D: CRITICAL ERROR TOOL 81

APPENDIX E: OSATS TOOL 82

APPENDIX F: O-SCORE EVALUATION TOOL 83

APPENDIX G: ROTATION EXIT SURVEY 84

APPENDIX H: VARIANCE COMPONENT RESULTS 86
H.1. Right IJ checklist variance component results 86
H.2. Right SC checklist variance component results 86
H.3. Right IJ critical error tool variance component results 87
H.4. Right SC critical error tool variance component results 87
H.5. Right IJ OSATS tool variance component results 88
H.6. Right SC OSATS tool variance component results 88
H.7. Right IJ O-SCORE tool variance component results 89
H.8. Right SC O-SCORE tool variance component results 89

List of Tables

Table 3.1 Summary of research questions, tests applied and data analysis 29

Table 4.1 Baseline demographic information by study group 33

Table 4.2 Frequency of self-reported prior CVC experience by study group 33

Table 4.3 Between group comparison on written examination passing scores 34

Table 4.4 Between group comparison on frequency of residents with passing scores at the internal jugular catheterization site for all performance tools 36

Table 4.5 Between group comparison on frequency of residents with passing scores at the subclavian catheterization site for all performance tools 37

Table 4.6 Frequency of critical errors by catheterization site 38

Table 4.8 Summary descriptives for survey responses 41

Table 4.9 Subscale survey responses - Comparisons between groups 42

Table 4.11 Pre-rotation written examination variance components 44

Table 4.12 Post-rotation written examination variance component 44

Table 4.13 Percentage of variance explained for each component by performance assessment tool – Internal jugular site 45

Table 4.14 Percentage of variance explained for each component by performance assessment tool – Subclavian site 45

Table 4.15 Dependability coefficients and standard errors of measurement (SEM) for all assessment tools 46

Table 4.16 Summary – Decision study 48

Table 4.17 Specific evidence to support our validity argument 50

List of Figures and Illustrations

Figure 3.1. Study design layout 18

Figure 4.1 Study enrollment 32

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
CBME	Competency Based Medical Education
Cohen's *d*	Cohen's Effect Size
CVC	Central Venous Catheterization
D	Decision Study
df	Degrees of Freedom
e	Unmeasured Error
EPA	Enabling Professional Activity
FM	Family Medicine
GRS	Global Rating Scale
G-Theory	Generalizability Theory
ICU	Intensive Care Unit
i	Items
Max	Maximum
Min	Minimum
OSATS	Objective Structured Assessment for Technical Skills
OSCE	Objective Structured Clinical Examination
O-SCORE	Ottawa Surgical Competency Operating Room Evaluation
p	Participants
PGY	Post Graduate Year
r	Raters
RC	Royal College of Physicians and Surgeons of Canada
n	Sample Size
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences
Φ	Dependability Coefficient
X^2	Chi-Squared

Chapter One: Introduction

1.1 Background

Procedural skills training in medicine has traditionally relied on an apprenticeship model referred to as “see one, do one, teach one”. In this model, demonstration is followed by supervised performance and then progression of the newly trained learner into the teaching role.(1, 2) It is the educational mandate of residency programs to provide sufficient training opportunities to enable their residents to become competent in clinical procedures. Frank et al. defines clinical competence as “an observable ability of a health professional, integrating multiple components such as knowledge, skills, values and attitudes”.(3) As such the importance of procedural skills training in the broader context of clinical competency has been recognized. However, to learn these skills, residents must have access to adequate practice opportunities, as well as practitioners who are qualified to teach and assess their performance.

Central venous catheterization (CVC) is a required procedural skill for medical practitioners caring for acutely-ill patients.(4) As such, it is a training requirement for a variety of residency specialties. The procedure involves the insertion of a multi-lumen catheter into a large, deeply located vein at the neck (internal jugular), chest (subclavian) or groin (femoral).(5) These catheters are often utilized to administer medications or fluids which would injure a smaller, more peripherally located vein.(5) Other indications for their placement include central venous blood sampling or pressure measurements, rapid intravenous volume administration, the introduction of transvenous pacemakers or pulmonary artery catheters and the need to obtain venous access in an unstable patient when peripheral sites are not accessible.(4, 5)

Unfortunately, the current educational format at many institutions inadequately trains residents to develop competence in procedures such as CVC, deemed necessary for their clinical careers. Opportunities to "see" some of the rarer, but potentially life-saving, clinical procedures during clinical rotations are often limited.(6) Inconsistent and inadequate exposure means residents lack opportunities to develop competence to "do" these procedures, let alone "teach them.(7) Furthermore, many of these procedures are performed in life-threatening clinical scenarios, where stress and time constraints relegate the learner to the role of passive observer. Other limitations imposed by work hour restrictions, insufficient faculty supervision and patient reluctance to act as teaching participants further contribute to the inadequate experience of many trainees.(8, 9) Such suboptimal training shapes residents into clinicians who feel uncomfortable performing procedures due to inadequate preparation and experience.(10-12) Given that the number of opportunities for clinical procedures varies and are often limited, there is a clear need to develop safe, alternative teaching strategies to help learners acquire these skills during training rather than hoping to gain exposure and experience by chance.

1.2 The research team

A team of researchers conducted the study. It was led by the student submitting this thesis, and included two other critical care physicians who are also Faculty in the Department of Critical Care Medicine, Cumming School of Medicine, University of Calgary.

1.3 Study goal

Residents currently learn to perform CVC in the Calgary Zone hospitals in an opportunistic fashion. They are taught by various instructors, using differing educational methods and differing techniques. Many training programs rely heavily on the intensive care rotation to provide their residents with a consistent experience to gain the necessary training to competently perform this skill. The primary goal of this study is to explore whether the introduction of a multimedia learning module, combined with simulation based teaching, improves the knowledge, performance, and educational experience in junior residents learning to perform CVC during their intensive care rotation at the University of Calgary. A secondary goal of this study is to determine whether the scores derived from the assessment tools developed for this study are dependable in assessing resident competence in performing this procedure. We (the team) intend to use the results of this study to develop a CVC assessment tool that can be utilized in our intensive care units as we transition to Competency Based Medical Education (CBME).

This thesis is organized into five chapters. The second chapter (Literature Review) reviews the existing literature and provides a rationale for the research questions contained in this thesis. The third chapter (Methods) describes the study participants, data collection instruments and statistical analyses performed to answer the research questions. The fourth chapter (Results) presents the data from the study. The final chapter (Discussion) contextually reviews the results of each of the research questions. It explains how the results build upon the existing literature and demonstrates the educational implications of the study. Finally, it highlights some limitations of the work and suggests areas for future research.

Chapter Two: Literature Review

This chapter reviews the existing literature relevant to the research questions in this thesis. It introduces the reader to the changes that are likely to occur in residency education as a consequence of the national implementation of competency based medical education (CBME) by the Royal College of Physicians and Surgeons of Canada (RC). It reviews the evidence for simulation-based procedural skill teaching. It identifies the importance of repeated practice and mastery in skills training, with a focus on central venous catheterization. It describes some of the commonly used scoring tools to assess procedural competence. Finally, it demonstrates the importance of considering validity and dependability when developing and utilizing scoring tools to judge performance. This chapter concludes by outlining the research questions for this thesis.

2.1 Competency based medical education in Canada

The framework of residency education in Canada is changing. With the introduction of "Competence By Design", the RC proposes to replace the existing time-based training platform with one that is specifically competency-based.(13) This CBME structure will promote greater learner-centeredness while emphasizing abilities and focusing on outcomes.(13) In this framework, competencies will be structured as Entrustable Professional Activities (EPAs) containing individual milestones to be demonstrated by the learner. These milestones (such as CVC) are defined as discrete, observable tasks of a discipline that an individual can be trusted to perform without direct supervision once sufficient competence has been demonstrated.(14) As EPAs are introduced into residency curricula, medical educators will be responsible for creating assessment tools to determine competence.(15) A variety of tools have been reported in the literature to assess performance, including a variety of task specific checklists, entrustability scales and global rating tools.(16-20) Currently, no specific tool has been identified that can serve to adequately assess resident competency in CVC within the context of a CBME training platform.

2.2 Simulation-based procedural skills training

In concert with Bloom's taxonomy, the North American medical educational system has traditionally focused on teaching higher level cognitive skills such as problem solving, synthesis and analysis.(21) Affective and psychomotor domains of medical education were largely neglected until the 1999 Institute of Medicine report "To Err is Human" highlighted the diagnostic and cognitive roles in medical error and emphasized the role of simulation in procedural skills training.(22) Subsequently, national licensing and governing bodies have moved to accreditation as a strategy to improve physician competence and patient safety.(23)

Support for simulation in skills acquisition and crisis management originated in other professional domains, including aviation, business and the military.(24-26) In medicine, research in the fields of obstetrics and anesthesia pioneered the use of simulation to improve performance while reducing patient risk.(27) Using realistic interactive patient simulators, anesthetists borrowed concepts from non-medical industries to train residents in technical skills and behavioral performances including leadership, communication, and teamwork.(2, 28)

There are a number of advantages to simulation-based education. Simulators allow trainees to "see one" then to "do one" and finally to "teach one" in a controlled environment. Focused, repeated deliberate practice of procedural skills can be performed in a safe, controlled environment with formal teaching and supervision prior to performing the procedure on an actual patient.(29-34) Simulations can be further manipulated to suit the learning objectives, and can be utilized to allow exposure to scenarios or procedures that are more rarely encountered in clinical practice.(35) Basic tasks can be introduced using lower fidelity training instruments such as multimedia teaching modules and procedural task trainers. More complex tasks can be taught with newer simulation modalities such as computerized simulators, and virtual-reality enhanced environments. Furthermore, simulations can be tailored to allow for standardized training or manipulated to demonstrate important consequences of trainee interventions or decisions.(35) This controlled practice environment helps learners expand their knowledge of and reinforces appropriate techniques, which in turn leads to increased confidence and improved performance.(36) Additionally, the simulated environment allows for the elimination of anchor bias. Anchor bias occurs when raters are influenced by their knowledge of the past performance of the learner.(37) In the simulated environment, where video recording is available, the identification of the learner can be blinded from the rater to eliminate this influence.

Most initial studies using simulators as training tools concentrated on scenario development and curriculum integration but did not provide evidence for improved learner performance or patient safety.(38, 39) Fortunately, as simulators have become more ubiquitous, evidence of improved performance in procedural skills training is becoming apparent. In addition to performance improvements in the simulated environment, a growing body of evidence demonstrates translation of performance from the simulated arena to the clinical setting. Improvements in resident performance as well as self-perceived confidence have been reported in a variety of study settings including clinical and procedural skills training and team-based resuscitation.(30, 40-46)

2.3 Central venous catheterization

Given its common role in the management of unstable patients, competence in CVC has been identified as a key objective to be learned by trainees in a variety of specialties.(47-51) Nearly three million central lines are inserted annually in the United States.(52) There are an estimated 15 million central line days per year in North American intensive care unit (ICU) populations.(53, 54) The reported incidence of mechanical and infectious complications related to this procedure has been reported to range from 15 percent to as high as 26 percent, and complication rates are inversely related to operator experience.(55-58) These complications directly impact patient morbidity, mortality and system costs related to prolonged hospitalization, underscoring the need for adequate training.(59) The opportunity to learn CVC in a controlled setting, which supports repeated and deliberate practice, seems an ideal solution to help address this issue.

Despite its apparent advantages, simulation based training has been criticized as potentially expensive, time consuming and resource intensive.(60, 61) Scientific evidence on newer and costlier educational tools is necessary, due to the increasing cost of new technology and the pressures to demonstrate correlations between simulation based teaching and improved patient safety.(62, 63) An expanding body of literature pertaining to simulation based central line insertion training demonstrates improved performance and self-confidence as well as increased satisfaction with the simulated teaching environment.(16, 64-68) Furthermore, patient safety seems to be enhanced through decreased average number of needle passes, improved barrier protection practices, decreased infectious complications, protocol adherence and decreased rates of pneumothorax.(64, 69-71)

2.4 Acquiring procedural competence

Procedural competence and decreased incidence of complications are associated with operator experience.(57, 72) Unfortunately, the exact number of opportunities required for competency in CVC remains unclear. In an attempt to accelerate the learning curve, opportunities for adequate exposure can be enhanced with simulation. The simulated learning environment provides opportunities to practice technique, refine behavioural skills and integrate safety procedures in a safe and controlled environment.(73) Simulation based practice has been shown to improve CVC insertion skills and resident confidence while decreasing patient complications and catheter-related bloodstream infections.(10, 64, 74, 75)

Several theories have been proposed to define a structured approach to skill acquisition and procedural competence. These models suggest that competence is not acquired through simple practice or unstructured repetition of an activity. Rather, competence is a function of improved accuracy, speed and performance resulting from structured, repeated practice. The ‘deliberate practice’ model of skill acquisition stipulates that specific conditions must exist to enable competence. These conditions include: 1) highly motivated learners who exert effort to improve performance; 2) well defined learning objectives or tasks; 3) appropriate levels of task difficulty; 4) opportunities for focused repetitive practice; 5) rigorous, reliable measurement strategies; 6) immediate, informative feedback and knowledge of results of performance; 7) monitoring, error correction and more deliberate practice; 8) further evaluation and performance to the expected minimal outcome; and 9) advancement to increasingly difficult levels or tasks.(76, 77)

The 'mastery learning' model of medical education extends this argument to a higher level of proficiency. This model suggests that all learners should ultimately accomplish educational objectives with little or no variation in outcome. In essence, it is a more stringent form of deliberate practice and competency-based education. It consists of seven complimentary features: 1) baseline or diagnostic assessment; 2) clear learning objectives sequenced as units in increasing difficulty; 3) engagement in sustained educational activities (such as deliberate practice) focused on reaching the objectives; 4) a fixed minimum passing standard; 5) formative assessment with specific feedback to gauge completion; 6) advancement to the next educational unit given measured achievement at or above the mastery standard; and 7) continued practice or study until the mastery standard is achieved.(73) Implementation of this teaching approach has demonstrated improved performance and skill acquisition as compared to traditional teaching methods.(30) A systematic review of mastery-based teaching in simulation skills training demonstrated improved skill acquisition, improved communication, decreased complications, reduced health care costs and improved patient outcomes. As McGaghie *et al*. state, “the Barsuk *et al*. study is clearly a wake-up call for all of us who were trained in the era of *see one, do one, teach one*”.(73)

While increased exposure to procedures is associated with improved learner confidence in their abilities, studies have consistently failed to identify an association between resident confidence and clinical competence.(75, 78-81) This disconnect reinforces the importance of ensuring that educational strategies targeting procedural competence are implemented in a structured way with clear learning objectives, opportunities for repeated practice and rigorous assessment strategies.

2.5 Assessment tools for procedural competence

Determining competence in clinical performance can only be achieved if scores derived from the assessment tool are dependable and valid. A variety of scoring tools to assess procedural competence have been reported in the literature. Common instruments include itemized checklists, scoring tools targeting critical errors, global rating scales and combinations thereof.

Checklists are often focused on technical aspects of procedural competence and allow identification specific observable, dichotomous behaviours.(82, 83) Since they provide an objective measure of performance and are intuitive to use, it has been assumed that scores derived from these instruments are more reliable from those derived from global assessment tools.(84) For these reasons, they are widely reported in the literature for the assessment of procedural competency.(16, 70) Unfortunately, although checklists are able to identify and measure successful completion of specific steps related to a procedure, they typically fail to adequately assess other important cognitive domains related to procedural competency such as ‘infection control’, ‘safety’ or ‘team collaboration’.(85, 86) In clinical skills assessment, checklists are less sensitive than global assessment tools in discriminating between learners with varying levels of expertise.(87, 88) These limitations suggest that checklists may reward thoroughness at the expense of behaviours that more adequately assess performance.(89) For the assessment of procedural skill competence, checklist tools often sacrifice specificity for sensitivity. Low checklist scores are typically associated with procedural incompetence, while high checklist scores do not necessarily ensure competence.(17, 86, 90) Furthermore, itemized checklists may fail to appropriately account for procedural errors that are committed by the trainee during the procedure.(86) Consequently, the impact of these errors may minimally contribute to the assessment process and determination of competence. Studies utilizing error-focused assessment tools for procedural skills have demonstrated greater accurate than traditional checklists at determining competence.(17, 91)

Global rating scales, such as the Objective Structured Assessment of Technical Skills (OSATS) or Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) tools, define performance on a global behaviour scale, or set of sub-scales.(82) Improved reliability of scores and discrimination of expertise of performance has been demonstrated with global rating scales (GRS).(17, 20, 86) Scores derived from tools using a GRS format have demonstrated higher reliability and validity than those from tools using a checklist format in Objective Structured Clinical Examination (OSCE) scenarios.(92) A recent meta-analysis comparing checklist and GRS scoring tools in simulation-based procedural assessment identified high inter-rater reliability using both scale types.(93) However, inter-item and inter-station reliabilities favoured GRS-type tools and allow for identification of important deviations from desired performance such as critical errors.(93) Behaviourally anchored GRS such as widely used OSATS and recently developed O-SCORE tools avoid central tendencies by setting standards in such a way as to identify safe independent performance, rather than comparing performance to an arbitrary or peer-reference standard.(20, 31) However, judgments made using GRS are subjective; therefore, it is important that evidence of validity and reliability be provided to defend the decisions that are made while using these tools.

2.6 The validity argument

The development and application of appropriate assessment tools is essential for a competency-based training environment to be successful. Identification and testing of these tools must occur, to ensure that judgments regarding performance are valid and defensible. As emphasized by Downing, 'without evidence of validity, assessments in medical education have little or no intrinsic meaning'.(94) Validation of an assessment tool requires the collection and demonstration of evidence to defend the use of scores to make decisions. Traditional validity theory proposed that validity evidence could be categorized into various 'types' (e.g. face, criterion, content). Recently this has been challenged. Theories developed by Messick, and more recently Kane, abandon this traditional classification of validity.(95, 96) Rather they argue that all validity supports the broader context of construct validity. Unlike Messick's theory, Kane's approach allows for prioritization and organization of the validity evidence to support a proposed validity argument.(97) Cook *et al*. organize Kane's framework of validity into four definable categories (*scoring, generalization, extrapolation and implications*).(97)

Evidence from these categories can be demonstrated, organized and prioritized to create a thorough and comprehensive validity argument to support and justify decisions made by using a particular tool within a particular context. To strengthen this argument, the development of an assessment tool should include evidence from each of these categories. In this framework, validity evidence related to *scoring* supports the assumption that scores derived using a proposed assessment tool are fair, accurate and reproducible. *Generalization* provides evidence that sampled items from the assessment process are representative of all possible items within the universe of interest. *Extrapolation* extends the argument and provides evidence that the measured scores from the 'test-world' environment accurately predict performance in the real-world. Finally, evidence related to the *implications* of performance helps define the consequences of assessment, both on the learner as well as others, impacted by the decisions. Our study examines the usefulness of a variety of tools in the assessment of central venous catheterization.

Kane's theory recommends that an examination of the evidence for validity begin with a clear statement of the proposed use of the assessment scores. The upcoming transition to CBME necessitates that we develop tools with adequate psychometric properties to judge performance and assess resident competence in procedures such as CVC. This study will help determine which assessment tools are more dependable and therefore should be utilized at our institution to assess performance in central line insertion. Scores derived from the chosen assessment tool will be used to determine competency in the procedure.

2.7 Research questions

2.7.1 Resident knowledge pertaining to CVC (Questions 1a and 1b)

1a. Does an educational intervention using a multimedia learning module, together with simulation based teaching, improve junior residents' knowledge regarding the indications, the contraindications, the technique and the potential complications related to CVC as compared to a control group receiving the current educational format during an intensive care unit (ICU) rotation at our institution?

1b. Does an ICU rotation at our institution improve junior residents' knowledge related to CVC regardless of the educational intervention?

2.7.2 Resident competence performing CVC (Question 2)

2. Is there a difference in performance of CVC between junior residents trained using a multimedia learning module, together with simulation based teaching, and those receiving the standard educational format during an ICU rotation at our institution?

2.7.3 Resident comfort and satisfaction regarding their educational experience (Questions 3a, 3b & 4)

3a. Is there a difference in self-reported comfort with CVC between junior residents trained using a multimedia learning module, together with simulation based teaching, and those receiving the standard educational format during an ICU rotation at our institution?

3b. Does an ICU rotation at our institution improve junior residents' comfort related to CVC regardless of the educational intervention?

4. Is there a difference in self-reported satisfaction with the overall educational experience between residents using a multimedia learning module, together with simulation based instruction using a procedural task trainer, and those receiving the standard educational format following an ICU rotation at our institution?

2.7.4 Dependability of performance scores (Question 5)

5. What is the dependability of the scores derived from the performance assessment tools used in this study for decisions regarding resident competency in CVC?

Chapter Three: Methods

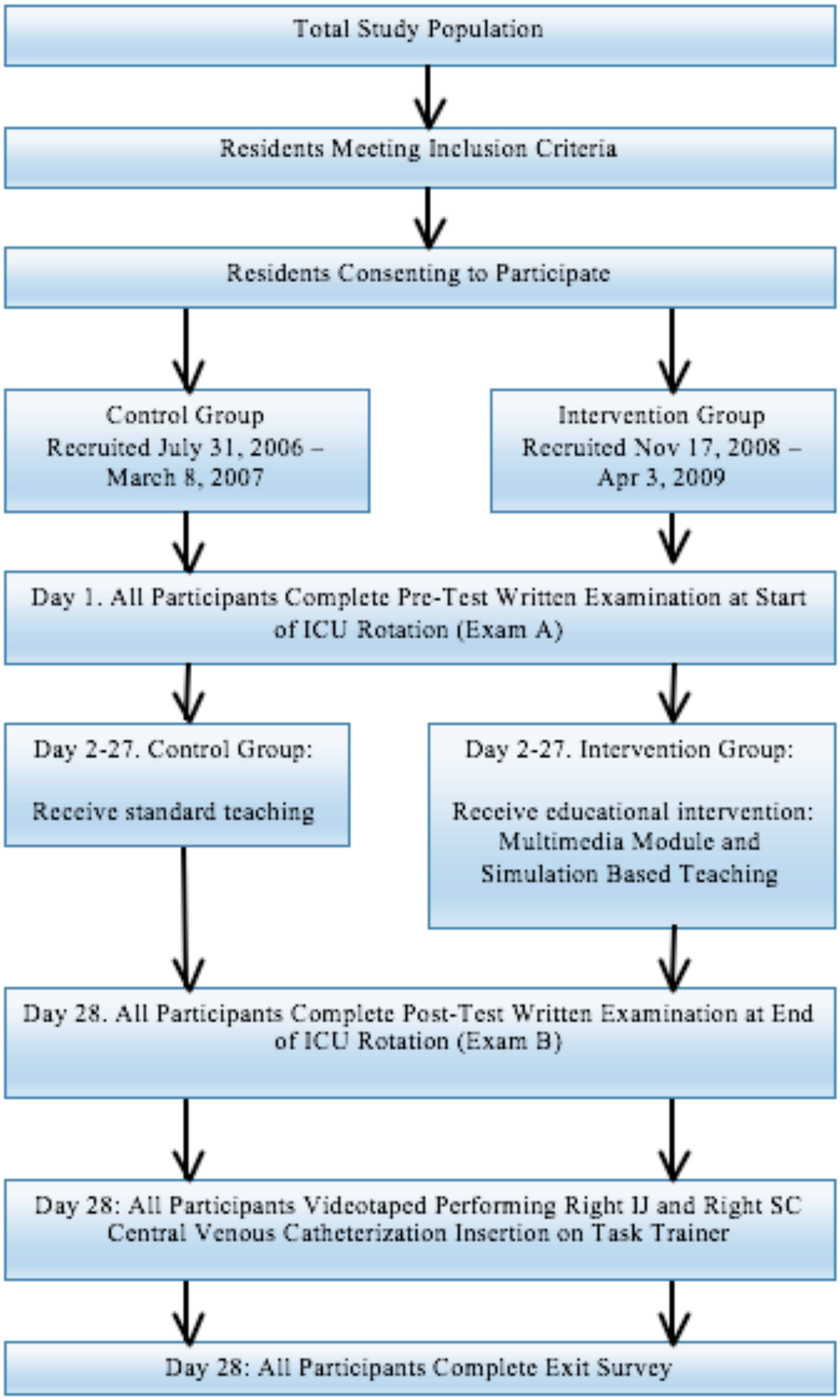
This chapter describes the study design and methods used for data collection and analysis. It begins with a general overview of the study design, setting and participants; presents a review of the procedures; provides the details for the outcome measures, and summarizes how they were administered and analyzed to answer the research questions.

3.1 Basic study design

With the purpose of improving CVC instruction of our medical residents and based on a review of the relevant medical education literature, there was a Departmental decision to adopt a new method of CVC instruction based on simulation and multimedia technology. This decision implied discarding the traditional teaching modality, and switching all the residents simultaneously to the new teaching method once the instructional materials had been fully developed. Under these circumstances it was impossible to assess the impact of the new method by conducting a truly experimental design involving two groups, which could have been randomly assigned to the old and new CVC instructional methods. The only option we had was to collect baseline data before implementing the new teaching modality; i.e.: enrolling residents in a "control' group while they were still learning with the traditional CVC instructional method, and subsequently enrolling residents in an "intervention' group once the new teaching method had been implemented.

Based on the above, we used a quasi-experimental between-group pre-test post-test design to assess the impact of the simulation-based educational program on the knowledge and confidence regarding central line insertion for junior residents rotating through their first ICU rotation in our Department. We assessed competence in CVC performance at the completion of the rotation, and examined the validity and dependability of the competence scores derived from several assessment tools used to judge competence in CVC. See Fig. 3.1 for a graphical layout of the study design

Figure 3.1. Study design layout



3.2 Methods

3.2.1 Setting

The study was conducted within the ICU system of the Calgary Zone of Alberta Health Services (formerly Calgary Health Region). At the time of this study, this zone served a population of approximately 1.14 million people. Its catchment territory included southern Alberta as well as parts of British Columbia and Saskatchewan. The region contained three multisystem ICUs containing a total of 48 beds. Due to limitations related to the number of video cameras available for study, participants were enrolled from two regional ICUs at the Peter Lougheed Centre and Foothills Medical Centre.

3.2.2 Participants

Residents from nearly all postgraduate training programs at the University of Calgary rotated through one of these intensive care units during their first three years of residency training. Prior to the initiation of the study, based on the annual clinical schedule for the ICUs, we expected to recruit between 25 and 30 residents into each group. To be included in the study, participants had to be residents commencing their first intensive care unit rotation. Medical students, senior residents, and residents who had previously completed an ICU rotation, were absent on the first day of the rotation or were scheduled as elective rotations from outside centres, were excluded from the study.

All junior residents who met the inclusion criteria at the Peter Lougheed Centre or Foothills Medical Centre were invited to participate in the study. Written informed consent was required prior to enrollment. The first group of students enrolled in the study prior to the implementation of the new teaching method served as "controls" and provided baseline data for further comparisons. A second group of students enrolled upon implementation of the new instructional method constituted the "intervention" group. Due to the time required to develop and implement the multimedia learning module, a delay occurred between the completion of enrolment and data collection for the control group and initiation of enrollment and data collection for the intervention group.

3.3 Procedure

Participants included in the control group completed a short-answer written exam prior to the start of their clinical rotation in the ICU to assess for baseline knowledge related to CVC (Appendix A). During their rotation, they received usual instruction on the procedure, which included a didactic educational session at the start of their rotation and opportunities for CVC with supervision that arose during their rotation. At the end of their clinical rotation, the residents completed a post-rotation short-answer written examination to assess for changes in knowledge related to CVC (Appendix B). To demonstrate their proficiency in CVC, they were asked to perform the procedure on a task trainer mannequin at the internal jugular (IJ) and the subclavian site. Their performance was videotaped and subsequently scored by three independent raters using four assessment tools specifically designed for this purpose (Appendices C-F). Finally, the residents completed an exit survey to rate their comfort with the procedure and evaluate their educational experience (Appendix G).

The intervention group received a one-hour hands-on, simulation-based teaching session demonstrating CVC on the first day of their ICU. At the time of the study, ultrasound guidance was not routinely utilized for CVC at our institution. For this reason, the residents were taught to perform the procedure without the use of ultrasound. This simulation-based teaching session took place prior to patient contact. These residents were also introduced to a newly developed multimedia learning module. This module was designed to provide residents with a standardized approach to CVC, which supplemented the hands-on learning by integrating narrated text together with images and video for each step of the procedure. The module was available to all study participants through the regional internal web server and via an external link. Residents were encouraged to utilize this module during their rotation to refine their procedural technique.

The assessment measures for the intervention group were similar to the control group. Residents in the intervention group completed a pre- and post-rotation short-answer written exam, were videotaped performing CVC at the internal jugular and subclavian site, and completed the exit survey. The residents were supervised during their written examination to ensure the assessment process was rigorously conducted. The video recorded data were digitally transferred to compact disc for analysis. All of the research records were securely stored in a locked cabinet at an office at the Peter Lougheed Centre. All data were kept in a secured office inaccessible to others. Electronic records were kept on a password protected computer with data encryption. Only the investigator of the study had access to this data.

3.4 Outcome measures

The research questions for this study were addressed using several outcome measures. This section presents the links between the research study questions and the methods used to address them.

3.4.1 Research questions 1a & 1b: Resident knowledge pertaining to CVC

Two 15 question short-answer written exams (to be completed pre-and post-rotation) were developed to assess resident cognitive knowledge pertaining to the indications, contraindications, technique and complications related to central venous catheter insertion (Appendices A & B). They were created by two medical experts and developed to be similar in content, structure and level of difficulty. Content contained in these exams was explicitly reviewed during the educational intervention and was clearly described in the multimedia module.

Three expert raters established explicit criteria relating to the answer key and the relative weighting of scores. Each question in the written exam was weighted equally for the purpose of assessment, regardless of the number of responses required for a given question. Following exam completion by the residents, the raters felt that one question from each of the exams was unclear and it was removed from the exam scoring. The total maximum score for each of the exams was 14. A clearly defined pass/fail criterion (cut-score) was developed by the raters for each exam using the modified Angoff method.(98) The written exams were scored by three independent expert raters who were blinded to residents' identity and group assignment. Three raters were chosen to allow for a majority consensus in the event of scoring discrepancies. All resident exams were scored independently by all three raters.

3.4.2 Research question 2: Resident competence performing CVC

To assess competence in performing CVC, the residents were videotaped performing right-sided internal jugular and right-sided subclavian CVC on the task trainers at the end of their rotation. Due to practical issues related to manpower, room availability and conflicts with other resident academic priorities of their rotation, we were unable to assess pre-rotation competence in CVC.

The performance of each resident was reviewed by three raters, using four competency-based assessment tools. The raters were blinded to residents' identity and group assignment. All residents were scored by all raters using all four tools. The assessment tools were created *a-priori* by the raters and piloted on three randomly selected video recordings to ensure that they were appropriate for data collection. The relevant content, answer key, relative weighting of scores, and pass/fail criterion was defined by the expert opinion of the raters. Cut-scores were calculated by the raters for each of the competence assessment tools. Central line insertion was performed by the raters at the internal jugular and subclavian site on the task trainers to verify authenticity of the assessment environment.

3.4.2.1 Competence tool #1 – Checklist

An itemized checklist was designed to assess performance of nine key steps related to the CVC (Appendix C). Each step in the checklist was weighted equally. All residents received a mark for each step they performed correctly. The maximum score possible for this tool was nine.

The checklist was used by one rater to measure time to cannulation and number of passes with the needle for each resident. Since these measurements are discrete and objectively quantified, they were assessed by a single rater. Time to cannulation (measured in seconds) was defined as the time from which the 14g steel needle or angiocatheter entered the skin, until the time that the central venous catheter was fully inserted into the vein. A needle pass was defined as a movement where the resident advanced the needle towards the vein. An additional pass was counted if the resident withdrew the needle back to the level of the skin after missing the target vein and then performed another attempt at localizing the vein. The total time to cannulation and number of needle passes was recorded for all residents.

3.4.2.2 Competence tool #2 – Critical error tool

A critical error tool was designed to identify significant errors that occur during the procedure (Appendix D). The raters defined a total of six types of potential critical errors, which were weighted equally. Residents were scored for each critical error that was made during their CVC. The maximum score possible for this tool (worst score) was six.

3.4.2.3 Competence tool #3 – Objective Structured Assessment for Technical Skills (OSATS)

The OSATS is a global rating scale widely used in the assessment of technical skills. (17, 18, 86, 99) A modified OSATS (Appendix E) was developed to assess five distinct domains specifically related to the technical components of CVC. Competency in each domain was scored using a 4-point Likert scale. All domains were weighted equally. The residents' score on the tool was defined as the total score for all domains. The maximum score possible for this tool was 20, which represented optimum technical competency.

3.4.2.4 Competence tool #4 – Modified Ottawa Surgical Competency Operating Room Evaluation (O-SCORE)

The final scoring tool used in this study was a modified O-SCORE (Appendix F).(14, 20) This was a 5-point global-rating scale used to assess performance along a competence continuum. The behaviours anchored in the scale mirrored practical decisions regarding graded supervision and progression of independence for procedural skills training in a CBME framework. The maximum score possible for this tool was five.

3.4.3 Research questions 3a, 3b & 4: Resident comfort and satisfaction regarding their educational experience

An exit survey was developed to assess resident self-perceived comfort with CVC, and satisfaction with their educational experience during the ICU rotation (Appendix G). This survey included six items designed to capture standard demographic information such as residency training specialty, year of training, gender, age, and educational experiences related to CVC prior to their ICU rotation. In addition, a 5-point Likert scale (1=*Strongly Disagree* to 5=*Strongly Agree*) composed of nine items aggregated into 4 sub-scales was developed to measure:

1) Resident self-perceived comfort with performing CVC before their rotation (*subscale items 1 and 2*); minimum/maximum scores possible were 2/10, where two indicated minimum, and 10 maximum comfort with the procedure.
2) Resident self-perceived comfort with performing CVC after their rotation (*subscale items 3 and 4);* minimum/maximum scores possible were 2/10, where two indicated minimum, and 10 maximum comfort with the procedure.
3) Resident perception of knowledge gained in the rotation (*subscale items 6, 7, and 8*); minimum/maximum scores possible were 3/15, where 3 indicated minimum and 15 maximum knowledge gain.
4) Quality of the rotation (*subscale items 5 and 9*); minimum/maximum scores possible were 2/10, where 2 indicated poor quality of the rotation, and 10 good quality of the rotation.

3.4.4 Research question 5: Dependability of the assessment tools

Generalizability theory (G-theory) was used to determine the dependability of the residents' scores on the written examinations as well as the four competency assessment tools.(100) G-theory was used to estimate the relative contribution of resident performance (our measure of interest for each test) to the test scores, as compared to the contribution from measurement error. Our objects of interest were the participants (p), and the potential sources of measurement error included the raters (r); the items (i); and the interactions between raters-items, raters-participants, item-participants, and items-raters-participants. A two-facet, fully-crossed design was used; i.e.: each participant (p) was examined by each rater (r) on each exam item (i).

The test scores were used in an absolute manner (i.e.: a cut-score based on a standard was used for decision making); therefore, a Dependability coefficient (Φ) was calculated. This coefficient ranges from 0 to 1.0 and provides a measure of the extent to which score consistency is affected by absolute error (100). A Φ greater than 0.8 is generally considered adequate for high-stakes assessment such as licensing examinations.(101) For the purpose of our assessment, we elected to accept a lower threshold of 0.7, which is generally considered adequate for lower-stakes resident assessments such as that performed in this study.(102, 103) The G-Study was followed by a Decision study (D). This enabled us to estimate changes in the dependability of the scores as a result of increasing/decreasing the number of raters, and/or exam items; thereby, maximizing the dependability of the results obtained under a variety of conditions.

3.5 Data analysis

The data were analyzed using SPSS® statistical software for descriptive and inferential statistics. Scores from the written examinations and the performance assessment tools were treated as continuous data. See Table 3.1 for a summary outlining how the tools were utilized to address each research question and how the data were analyzed.

Demographic data from the survey was used to portrait the sample. Descriptive statistics included mode scores, means, standard deviations (SDs), frequencies and percentages. The choice of descriptors was dependent on the type of data (e.g.: categorical vs. continuous). Repeated measures analyses were used to compare pre- vs. post-rotation results. Independent samples *t*-tests and Chi Squared (X^2) analyses were conducted to test for between-group differences where appropriate. Paired samples *t*-tests were performed to assess the change pre-post ICU rotation for all residents, regardless of their group. Generalizability theory (G-theory) was used to determine the dependability of the scores from the assessment tools. A Decision study (D-study) was performed to predict changes in dependability under a variety of conditions.

Table 3.1 Summary of research questions, tests applied and data analysis

Research Question	Domain	Data	Statistical Analyses
1a	Between group knowledge differences	Post-rotation written examinations	• Independent samples *t*-test • G-study (Φ)
1b	Individual differences in knowledge pre-post ICU rotation	Pre-vs post-rotation written examinations	• Paired samples *t*-test • Cohen's *d*
2	Performance assessment of CVC insertion	Checklist	• Independent samples *t*-test • X^2 analysis • G-study (Φ)
		Critical error tool	• Independent samples t-test • X^2 analysis • G-study (Φ)
		OSATS	• Independent samples *t*-test • X^2 analysis • G-study (Φ)
		O-SCORE	• Independent samples *t*-test • X^2 analysis • G-study (Φ)
		Differences in number of needle attempts between groups	• Independent samples *t*-test
		Differences in time to cannulation between groups	• Independent samples *t*-test
3 & 4	Resident satisfaction and comfort	Survey scores	• Descriptive analysis • Independent samples *t*-test

			• Paired samples *t*-test • Cohen's *d*
5	Dependability of performance tools	Performance assessment tools	• G-study (Φ) • Decision-study (D)

3.6 Validity evidence

As we transition to CBME, tools with good psychometric properties must be developed to ensure valid and reliable decisions about resident performance are made. Our work also included the development of an argument for validity as the results of the study were intended to provide information to help design an assessment process for central line insertion in junior residents rotating through the ICUs at our institutions. To support this argument, the study was planned to provide evidence related to each of the four validity categories defined by Kane.(96, 97) These categories include evidence for *scoring, extrapolation, generalization and implication.*

3.7 Ethics approval

This study was reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board. Study ID number REB16-1035.

Chapter Four: Results

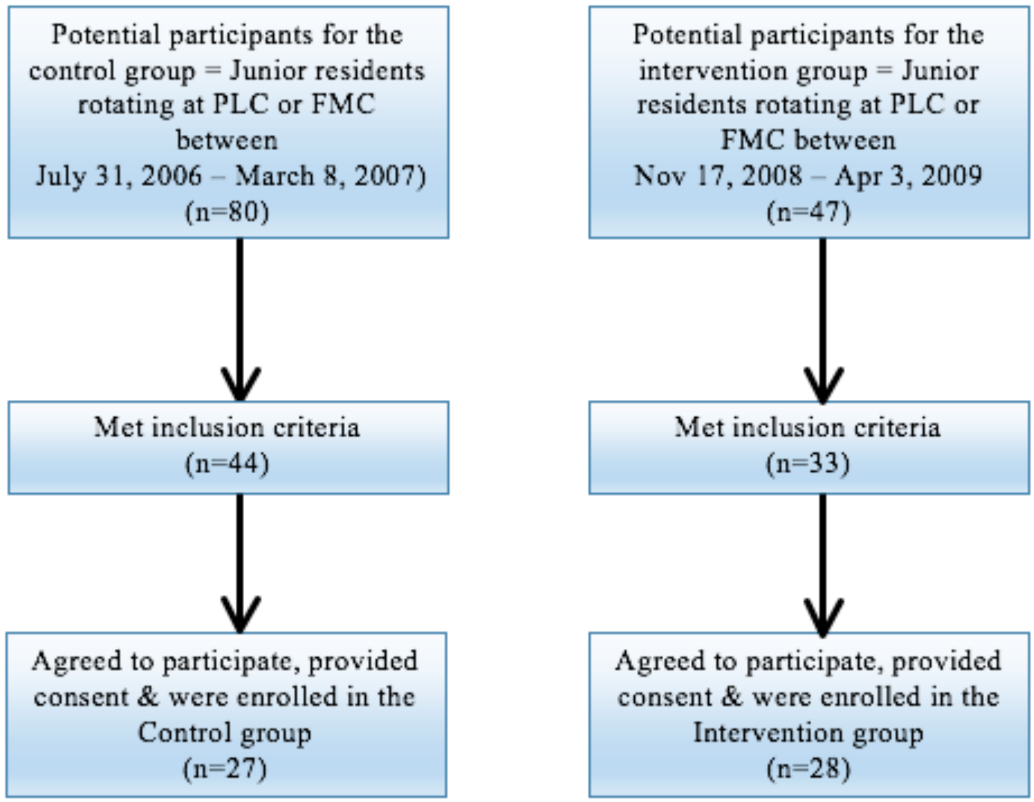
This chapter presents the results of this study in 5 sections: 1) Participants' demographic characteristics and prior central venous catheterization experience; 2) Knowledge acquisition and change; 3) Competency achievement; 4) Exit Survey Results (Residents' self-perceived comfort and confidence with central venous catheterization, perceived gain in knowledge, and satisfaction with the rotation); and 5) Dependability of the performance scores.

4.1 Participants' demographic characteristics and prior CVC experience

Eighty junior residents rotated through the ICUs at either the Peter Lougheed Centre or Foothills Medical Centre within the Calgary Zone of Alberta Health Services between July 31, 2006 and March 9, 2007. Forty-four of these residents met our inclusion criteria and were invited to participate in the study. Twenty-seven residents agreed to participate and provided informed consent. These 27 residents comprise the control group.

Forty-seven residents rotated through the intensive care units at either the Peter Lougheed Centre or Foothills Medical Centre within the Calgary Zone of Alberta Health Services between November 17, 2008 and April 3, 2009. Thirty-three of those residents met our inclusion criteria and were invited to participate in the study. Twenty-eight residents agreed to participate and provided informed consent. These 28 residents comprise the intervention group. A total of 55 residents participated in the study (Figure 4.1).

Figure 4.1 Study enrollment



The frequencies (percentages) of residents in the different training programs, years of training and gender for each of the study groups are presented in Table 4.1. A summary of participants' self-reported prior central venous catheter insertion experience is presented in Table 4.2.

Table 4.1 Baseline demographic information by study group

Demographic Information	Control (n=27) Frequency (%)	Intervention (n=28) Frequency (%)	Total (n=55) Frequency (%)
Training Program			
FM Urban	9 (33.3)	11 (39.3)	20 (36)
Internal Medicine	4 (14.8)	4 (14.3)	8 (14)
FM Rural	4 (14.8)	3 (10.7)	7 (13)
Anesthesia	3 (11.1)	2 (7.1)	5 (9)
Orthopedic Surgery	0 (0.0)	4 (14.3)	4 (7)
General Surgery	1 (3.7)	2 (7.1)	3 (5)
Otolaryngology	1 (3.7)	1 (3.6)	2 (4)
Neurology	2 (7.4)	0 (0.0)	2 (4)
Emergency Medicine	1 (3.7)	0 (0.0)	1 (2)
Neurosurgery	1 (3.7)	0 (0.0)	1 (2)
Missing	1 (3.7)	0 (0.0)	1 (2)
Other	0 (0.0)	1 (3.6)	1 (2)
Year of Training			
PGY 1	0 (0.0)	5 (17.9)	5 (9)
PGY 2	23 (85.2)	22 (78.6)	45 (82)
PGY 3	3 (11.1)	1 (3.6)	4 (7)
Missing	1 (3.7)	0 (0.0)	1 (2)
Gender			
Female	13 (48.1)	16 (57.1)	29 (53)
Male	13 (48.1)	12 (42.9)	25 (45)
Missing	1 (3.7)	0 (0.0)	1 (2)

Table 4.2 Frequency of self-reported prior CVC experience by study group

Number of CVCs Done Prior to the Rotation	Control (n=27) Frequency (%)	Intervention (n=28) Frequency (%)	Total (n=55) Frequency (%)
0	17 (63.0)	15 (53.6)	32 (58)
1-5	5 (18.5)	8 (28.6)	13 (25)
6-10	1 (3.7)	3 (10.7)	4 (7)
>10	3 (11.1)	1 (3.6)	4 (7)
Did not respond	1 (3.7)	1 (3.6)	2 (4)

4.2 Knowledge acquisition and change

This section presents the differences in mean knowledge scores between the two groups pre-and post-intervention, and changes in knowledge for all participants pre-post intervention.

4.2.1 Knowledge acquisition

The mean (SD) scores on the pre-rotation written examination were 6.9 (2.3) in the control group, and 7.6 (2.6) in the intervention group. This difference was not statistically significant, $t(52) = -1.028$, p= .31. The mean (SD) scores on the post-rotation written examination were 8.6 (1.7) in the control group, and 8.8 (1.6) in the intervention group. This difference was not statistically significant, $t(52) = 0.474$, $p = .64$.

Cut-scores for pre/post rotation written exams were 5.6 and 5.6 respectively. There were no statistically significant differences in the proportion of residents who passed the knowledge examination either pre-or post-rotation between the study groups. (Table 4.3).

Table 4.3 Between group comparison on written examination passing scores

Written Examination	Cut-Score	Control Group (%)	Intervention Group (%)	*p*
Pre-Rotation (n=27)	5.53	17 (63%)	21 (78%)	.23*
Post-Rotation (n=26)	5.60	25 (96%)	26 (93%)	.53**

*$X^2(1, N = 54) = 1.42, p = .23$
** $X^2(1, N = 54) = 0.28, p = .53$

4.2.2 Change in knowledge

Since there was no significant difference in knowledge acquisition pertaining to CVC between the two groups before or after the rotation the 'change in knowledge' analysis was done for all residents together, regardless of their study group. The mean scores pre-post rotation scores on the written examination of all residents were compared using paired samples t-tests. The overall (n=52) mean (SD) scores pre and post rotation on the knowledge written exam were 7.2 (2.5); and 8.7 (1.7) respectively. The increase in knowledge scores post-rotation was statistically significant, $t(52) = -4.46$, $p < .001$. To provide a quantitative measurement of the magnitude of the effect of the intervention, a Cohen's *d* effect size was calculated. Effect sizes were interpreted as small (<0.5), moderate (.6-.8) and large (>.9).(104) The Cohen's *d* effect size was moderate (.66).

4.3 Competency achievement

This section presents the results of resident performance on internal jugular and subclavian CVC for each of the four performance assessment tools (checklist, critical error tool, OSATS tool, and O-SCORE tool), including time to catheterization and number of attempts made by each resident to perform CVC at each catheterization site. The differences in performance scores between groups are presented separately for each tool.

4.3.1 Checklist

At the internal jugular catheterization site, the mean (SD) scores were 7.2 (1.5) in the control group, and 7.2 (1.8) in the intervention group. This difference was not statistically significant, $t(51) = 0.52$, $p = .89$. At the subclavian site, the mean (SD) scores were 7.3 (2.1) in the control group, and 8.0 (1.4) in the intervention group. This difference was not statistically significant, $t(52) = -1.37$, $p = .18$.

A cut-score of 4.4 was determined for this checklist tool. There was no statistically significant difference between groups in the proportion of residents who passed using any of the competency assessment tools at either the internal jugular (Table 4.4) or the subclavian catheterization sites (Table 4.5).

Table 4.4 Between group comparison on frequency of residents with passing scores at the internal jugular catheterization site for all performance tools

Tool	Cut Score/Max Score Possible	Control Group Frequency (%)	Intervention Group Frequency (%)	*p*
Checklist	4.40/9	24 (92)	24 (89)	.52*
Critical Error Tool	0/6	6 (23)	4 (15)	.50**
OSATS	15/20	10 (38)	7 (27)	.33***
O-SCORE	5/5	3 (12)	2 (8)	.48****

*$X^2(1, N = 53) = .181$, $p = .52$
**$X(1, N = 53) = .591$, $p = .50$
***$X^2(1, N = 53) = .955$, $p = .33$
****$X^2(1, N = 53) = .265$, $p = .48$

Table 4.5 Between group comparison on frequency of residents with passing scores at the subclavian catheterization site for all performance tools

Tool	**Cut Score/Max Score Possible**	**Control Group Frequency (%)**	**Intervention Group Frequency (%)**	***p***
Checklist	4.40/9	23 (88)	28 (100)	.11*
Critical Error Tool	0/6	5 (19)	8 (31)	.42**
OSATS	15/20	13 (52)	19 (68)	.24***
O-SCORE	5/5	2 (8)	4 (14)	.39****

*X^2 (1, N = 53) = 3.421, p = .11
**X^2 (1, N = 53) = .644, p = .42
***X^2 (1, N = 53) = 1.388, p = .24
****X^2 (1, N = 53) = .520, p = .39

4.3.2 Critical error tool

At the internal jugular catheterization site, the mean (SD) scores were 1.1 (1.2) in the control group, and 1.3 (1.2) in the intervention group. This difference was not statistically significant, $t(51) = -0.68$, $p = .50$. At the subclavian site, the mean (SD) scores were 1.0 (1.2) in the control group, and 0.8 (0.8) in the intervention group. This difference was not statistically significant, $t(52) = 1.02$, $p = .31$.

Ideally, no critical errors should occur; therefore, the cut-score for this tool was determined to be 0. There was no statistically significant difference between groups in the proportion of residents committing at least one critical error at either site (Tables 4.4 & 4.5). The proportion of critical errors from each category, committed at each catheterization site, is summarized in Table 4.6.

Table 4.6 Frequency of critical errors by catheterization site

Critical Error Type	Internal Jugular Frequency (%)	Subclavian Frequency (%)
Mishandling of Needle	23 (43.4)	18 (33.3)
Inadequate Sterile Technique	17 (32.1)	10 (18.6)
Inadquate Flushing	7 (13.2)	5 (9.3)
Mishandling of Wire	5 (9.4)	3 (5.6)

4.3.3 OSATS

At the internal jugular site, the mean (SD) scores were 13.6 (4.6) in the control group, and 12.8 (3.9) in the intervention group. This difference was not statistically significant, $t(51) = -0.69$, $p = .49$. At the subclavian site, the mean (SD) scores were 14.9 (3.4) in the control group, and 15.9 (3.4) in the intervention group. This difference was not statistically significant, $t(52) = -1.07$, $p = .29$.

A cut-score of 15 was determined for this assessment tool. There was no statistically significant difference between groups on the proportion of residents who passed at either the internal jugular (Table 4.4) or the subclavian catheterization site (Table 4.5).

4.3.4 O-SCORE

At the internal jugular site, the mean (SD) scores were 3.2 (1.3) in the control group, and 3.0 (1.1) in the intervention group. This difference was not statistically significant, $t(51) = 0.59$, $p = .56$. At the subclavian site, the mean (SD) scores were 3.5 (1.1) in the control group, and 3.8 (1.0) in the intervention group. This difference was not statistically significant, $t(52) = -0.88$, $p = .38$.

Since a score of 5 on this tool represented optimum competence in central venous catheterization, the cut-score for the O-SCORE tool was determined to be 5. There was no statistically significant difference between groups in the proportion of residents who passed at either the internal jugular (Table 4.4) or the subclavian catheterization site (Table 4.5).

4.3.5 Time to cannulation and number of attempts

The time to cannulation and number of attempts for cannulation were not measured for residents who were unable to cannulate the vessel and/or requested assistance from the instructor to locate the target vessel during the procedure. Forty-two residents were included for analysis at the internal jugular site, and forty-nine residents for the subclavian catheterization site.

The mean (SD) time to cannulation (measured in seconds) and mean (SD) number of attempts for cannulation for the study groups at each catheterization site are summarized in Table 4.7.

Table 4.7 Between group comparison of time to insertion and number of attempts at cannulation by catheterization site

Site	Control Group *M* (SD)	Intervention Group *M* (SD)	*p*
Internal Jugular			
Time (s)	285.04 (214.47)	318.56 (176.96)	.59*
# Attempts	6.22 (11.74)	13 (52)	.87**
Subclavian			
Time	238.57 (105.97)	190.46 (88.84)	.09***
# Attempts	5.13 (5.22)	3.00 (6.24)	.20****

* t (40) = .610, p=.593
** t (39) = -1.64, p=.87
*** t (47) = 1.728, p=.09
**** t (49) = -1.30, p=.20

4.4 Exit survey results

This section presents the results of the resident exit survey. Differences in survey scores between groups are presented for each of the individual survey items as well as the 4 predefined subscales; 1. Residents' self-perceived comfort with performing CVC before and 2. Residents' self-perceived comfort with performing CVC after their ICU rotation; 3. Perception of knowledge gained in the rotation; and 4. Perception of the quality of the rotation.

4.4.1 Survey responses

Fifty-four residents (98.18%) completed the exit survey. The overall (n=54) mean (SD) scores, the range of scores and the minimum/maximum score for the nine survey items are presented in Table 4.8.

Table 4.8 Summary descriptives for survey responses

Survey Question	Mean (SD)	Range	Min	Max
Prior to this rotation, I felt comfortable inserting a central venous catheter *under direct* supervision	2.78 (1.53)	4.00	1.00	5.00
Prior to this rotation, I felt comfortable inserting a central venous catheter *without* direct supervision	1.89 (1.24)	4.00	1.00	5.00
Following this ICU rotation, I feel comfortable inserting a central venous catheter *under direct* supervision	4.69 (58)	2.00	3.00	5.00
Following this ICU rotation, I feel comfortable inserting a central venous catheter *without* direct supervision	3.85 (1.11)	4.00	1.00	5.00
I feel my training in central venous catheter insertion during this rotation was adequate	3.94 (1.05)	4.00	1.00	5.00
I understand the indications for central venous catheter insertion	4.04 (.80)	3.00	2.00	5.00
I understand the contra-indications for central venous catheter insertion	3.50 (1.00)	3.00	2.00	5.00
I understand the potential complications of central venous catheter insertion	3.93 (.80)	3.00	2.00	5.00
I achieved the clinical objectives pertaining to central venous catheter insertion during this rotation	3.50 (1.11)	4.00	1.00	5.00

4.4.2 Between groups differences

Table 4.9 presents a summary of comparisons between groups in survey responses relating to comfort with CVC pre- and post-rotation, their perceived knowledge gained during the rotation and their perceived quality of the rotation. There were no significant differences between the groups in residents' self-perceived comfort with CVC insertion before or after the rotation. There was a statistically significant difference in perceived knowledge gain between the groups; the residents in the intervention group perceived greater knowledge acquisition as a result of the rotation than those in the control group. Likewise, the residents in the intervention group rated significantly higher the quality of the rotation than those in the control group.

Table 4.9 Subscale survey responses - Comparisons between groups

Survey Subscale	Control Group Mean (SD) n = 26	Intervention Group Mean (SD) n = 28	*p*
Comfort with CVC Insertion Pre-Rotation	4.34* (2.76)	4.96 (2.47)	.39
Comfort with CVC Insertion Post-Rotation	8.42* (1.75)	8.64 (1.34)	.61
Perceived CVC Knowledge Gained during Rotation	10.63** (2.47)	12.12 (2.23)	.02
Perceived Quality of Rotation	6.71** (2.19)	8.10 (1.45)	.01

*Maximum score possible = 10 points
**Maximum score possible = 15 points

4.4.3 Change in comfort

Since there was no significant difference in comfort pertaining to CVC between the two groups before or after the rotation the 'change in comfort' analysis was done for all residents together, regardless of their study group. The mean scores pre-post rotation scores on the written examination of all residents were compared using paired samples t-tests. The overall (n=54) mean (SD) scores pre-post rotation on the comfort performing

CVC without direct supervision pre- and post-rotation were 1.9 (1.2); and 3.9 (1.1) respectively. The increase in comfort scores post-rotation was statistically significant, $t(53) = -10.95$, $p < .001$. To provide a quantitative measurement of the magnitude of the effect of this intervention, a Cohen's *d* effect size was calculated. The Cohen's *d* effect size was large (1.74). Stated another way, a total of 14.9% of all residents responded that they agreed or strongly agreed with the statement that they felt comfortable inserting a CVC without direct supervision prior to the rotation. After the rotation, 72.2% of residents felt comfortable inserting a CVC without direct supervision.

4.5 Dependability of the performance scores

This section presents the results of the variance component analyses and generalizability study of the assessment scores (knowledge and competence) as well as the results of the Decision study.

4.5.1 Variance component analysis

The relative proportion of variance explained by each of the variance components was calculated for each of the assessment tools. Sources of variance in this study were participants, raters, items, and all the interactions between them. Tables 4.11 and 4.12 summarize the sources of variance for the pre- and post-rotation written examinations. The largest source of variance for the both written examinations was the participant-by-item interaction (pre-rotation = 62.81%; post-rotation = 58.86%).

Table 4.10 Pre-rotation written examination variance components

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	53	1.297	0.022	11.06
Raters (*r*)	2	.075	0.000	0.00
Items *(i)*	13	5.917	0.034	17.09
p x *r*	106	.014	0.000	0.00
p x *i*	689	.392	0.125	62.81
r x *i*	26	.052	0.001	0.50
pri, e	1378		0.017	8.54
Total				100.00

Table 4.11 Post-rotation written examination variance component

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	53	.624	0.007	4.00
Raters (*r*)	2	.095	0.000	0.00
Items *(i)*	13	7.023	0.041	23.43
p x *r*	106	.025	0.000	0.00
p x *i*	689	.331	0.103	58.86
r x *i*	26	.101	0.001	0.57
pri, e	1378	.023	0.023	13.14
Total				100.00

Similar calculations were performed for each of the performance tools at each anatomical site (Appendix H). A summary of the variance components from each assessment tool at the internal jugular site is provided in Table 4.13. The percentage of variance explained for each component by assessment tool at the subclavian site is summarized in Table 4.14.

Table 4.12 Percentage of variance explained for each component by performance assessment tool – Internal jugular site

Source of Variation	Checklist (%)	Critical Error (%)	OSATS (%)	O-SCORE (%)
Participants (*p*)	11.36	13.30	58.54	65.49
Raters (*r*)	1.14	1.60	1.49	2.99
Items *(i)*	6.25	11.70	0.00	N/A
p x *r*	1.70	1.59	16.16	N/A
p x *i*	26.14	21.81	4.02	N/A
r x *i*	0.57	1.06	1.12	N/A
pri, e	52.84	48.94	18.67	N/A
pr, e				31.52
Total				100.00

Table 4.13 Percentage of variance explained for each component by performance assessment tool – Subclavian site

Source of Variation	Checklist (%)	Critical Error (%)	OSATS (%)	O-SCORE (%)
Participants (*p*)	12.59	14.79	42.82	54.20
Raters (*r*)	0.74	0.70	1.74	2.48
Items *(i)*	2.22	5.63	0.00	N/A
p x *r*	3.71	0.00	26.50	N/A
p x *i*	27.41	19.72	4.17	N/A
r x *i*	2.96	2.82	0.81	N/A
pri, e	50.37	56.34	23.96	N/A
pr, e				43.32
Total				100.00

4.5.2 Dependability coefficient calculation

A summary of the calculated Dependability coefficients (Φ) and standard errors of measurement (SEM) for each of the assessment tools is summarized in Table 4.15.

Table 4.14 Dependability coefficients and standard errors of measurement (SEM) for all assessment tools

Assessment Tool	Φ	SEM
Pre-Rotation Written Exam	.65	.11
Post-Rotation Written Exam	.39	.10
Checklist Tool		
Internal Jugular	.64	.11
Subclavian	.65	.10
Critical Error Tool		
Internal Jugular	.59	.13
Subclavian	.66	.10
OSATS Tool		
Internal Jugular	.88	.29
Sublavian	.78	.32
O-SCORE Tool		
Internal Jugular	.85	.45
Subclavian	.78	.50

4.5.3 Decision study

The Decision study was performed to identify potential strategies for improving the dependability of the scores from each of the assessment tools, while optimizing efficiency of the testing environment. For this study, we set a Dependability coefficient (Φ) threshold of 0.7, which is considered acceptable for low stakes examinations such as the one in this study.(101) The impact of increasing/decreasing the number of raters or items on each of the assessment tools is summarized in Table 4.16. To meet this threshold for the pre-rotation written examination, we determined that we would need to add a substantial number of additional items to our examination. Increasing the number of raters would have no meaningful effect on Φ. Since Φ for the post-rotation written examination was only .39, we would require substantially more items or raters to elevate Φ above our threshold.

Both the checklist and critical error tool could be improved to meet our threshold, but would require a substantial increase in the number of items. For the internal jugular site, this would mean adding at least four items to the checklist and more than four items to the critical error tool. For the subclavian site, it would require adding at least three items to the checklist and two to the critical error tool.

Because Φ for the OSATS and O-SCORE tools exceeded the acceptable dependability threshold, we looked to identify ways to improve efficiency while maintaining Φ within acceptable limits. For the IJ site, the Decision study (D) revealed that we could maintain our dependability threshold for the OSATS tool by eliminating two raters or alternatively, by decreasing the number of items to as little as one item. For the Subclavian site, we could achieve a similar Φ for the O-SCORE, OSATS or the O-SCORE tool by removing one rater.

Table 4.15 Summary – Decision study

Exam	Φ	-2 Raters	-1 Rater	+ 1 Rater	+1 Item	+2 Items	+3 Items
Pre-Rotation Written Exam	.65	.65	.65	.65	.67	.68	.69
Post-Rotation Written Exam	.39	.37	.39	.40	.41	.42	.45
Checklist Tool - IJ	.64	.48	.59	.66	.66	.67	.69
Checklist Tool - SC	.65	.48	.60	.68	.67	.68	.70
Critical Error Tool - IJ	.59	.44	.54	.62	.62	.64	.67
Critical Error Tool - SC	.66	.50	.61	.68	.69	.72	.74
OSATS Tool - IJ	.88	.72	.83	.90	.88	.89	.89
OSATS Tool – SC	.78	.56	.71	.82	.79	.79	.80
O-SCORE Tool - IJ	.85	.65	.79	.88	-	-	-
O-SCORE Tool – SC	.78	.54	.70	.83	-	-	-

4.6 Validity evidence

Evidence for *scoring* validity was provided by a number of ways. The written exams were created by two medical experts and were developed to be similar in content, structure and level of difficulty. Three expert raters established explicit criteria relating to the answer key and the relative weighting of scores on these written exams. The exams' content was reviewed with the residents during the educational intervention and they were oriented to the multimedia module to prepare for their assessment. All residents were informed of the assessment process for each step of the study. Additionally, the four competency assessment tools were created *a-priori* by three expert raters and piloted on several randomly selected video recordings to ensure they were appropriate for data collection. For each tool, expert opinion from the raters was utilized to define the relevant content, create the answer key, and identify the relative weighting of scores. All residents were scored by all raters using all four performance assessment tools. Furthermore, the quality assurance processes outlined above were adhered to throughout the study.

Evidence of *extrapolation* validity was provided by requiring the three expert raters to independently perform CVC at the internal jugular and subclavian site on the task trainers to verify authenticity of the assessment environment. In addition, surrogate measures of competence (time to cannulation, number of attempts) were analyzed to further assess resident performance, our construct of interest.

Evidence of *generalization* validity was provided through the G-study to assess the dependability of the scores derived from the written and performance assessment tools.

Evidence of *implications* validity was provided through the development of distinct pass/fail criterion for each of the assessment tools. This process was established in an attempt to align the methodology of the study with our validity argument. See Table 4.17 for a summary of the evidence to support our validity argument.

Table 4.16 Specific evidence to support our validity argument

Category of Validity Evidence	Evidence Provided
Scoring	Items and response option selection Scoring rubric Rater selection Rater training Data security and integrity
Extrapolation	Construct definition Authenticity of simulation fidelity Correlation with surrogate performance measure
Generalization	Generalizability analysis
Implication	Pass/Fail standard

Chapter Five: Discussion and Conclusions

This chapter discusses the results of our study in the context of the existing literature and describes some of the educational implications of our findings. In addition, the limitations of the study and some considerations for future research are proposed

5.1 Thesis questions

5.1.1 Question 1a: Does an educational intervention using a multimedia teaching module, together with simulation based teaching, improve junior residents' knowledge regarding the indications, the contraindications, the technique and the potential complications related to CVC as compared to a control group receiving the current educational format during an ICU rotation at our institution?

Resident knowledge pertaining to central venous catheterization was not improved by the educational intervention. Procedural competence is more than just technical ability. It requires skill-based knowledge acquisition. Learners must understand and demonstrate proficiency in the technical steps as well as the cognitive skills related to the procedure. They must understand the indications, contraindications and complications of the procedure. Educational strategies focusing instruction on cognitive skills together with procedural instruction have been shown to be more effective than those dedicated solely to procedural technique.(105) By enhancing knowledge acquisition and learner engagement, computer aided instruction has been shown to be effective in clinical skill acquisition, provided the cognitive load does not overwhelm the learner.(106, 107) To ensure that our educational intervention addressed cognitive domains, the simulation-based procedural instruction received by the intervention group was complemented by a multimedia learning module which addressed these issues. This online module incorporated instructional videos together with photographs and narrative text to provide specific instruction related to the cognitive issues pertaining to CVC. This strategy provided learners with a comprehensive educational process targeting both the technical and cognitive aspects of CVC.

It is possible that the lack of a difference in knowledge acquisition between groups relates to inadequate utilization of the multimedia module by the intervention group. Higher levels of interactivity and engagement are associated with increased time-on-task and have been shown to correlate with knowledge acquisition.(108) Since we were unable to track how frequently the residents accessed the learning module, we are uncertain whether the absence of a knowledge difference between groups simply reflects inadequate exposure of the intervention group to the module. Alternatively, the residents may have accessed the module, but it failed to engage them sufficiently to stimulate learning above and beyond that otherwise experienced during the rotation.

5.1.2 Question 1b: Does an ICU rotation at our institution improve junior residents' knowledge related to CVC regardless of the educational intervention?

Junior resident knowledge related to CVC was improved at the end of the ICU rotation, regardless of the educational intervention. Although opportunities for CVC during the rotation are variable, the skill is an essential task to be learned by many residents rotating through the ICU.(47-51) At our institution, residents are provided opportunities to perform the skill under guided and graded supervision when appropriate. Regardless of group assignment, it is possible that most residents spent a proportion of their time during the rotation learning about the procedure through standard means such as available textbooks or online resources to be better prepare themselves for opportunities to perform the procedure. Additionally, attending physicians and critical care fellows may have provided ad-hoc teaching sessions to the junior residents when time was available during the rotation.

5.1.3 Question 2: Is there a difference in performance of CVC between junior residents trained using a multimedia teaching module, together with simulation based teaching, and those receiving the standard educational format during an ICU rotation at our institution?

Because we did not collect pre-rotation baseline data related to procedural competence for CVC; therefore, we are unable to determine whether the intervention impacted resident competence. However, no difference was identified between the control and intervention groups in their CVC performance scores post-rotation, regardless of the anatomical site of catheterization or assessment tool utilized. This unexpected finding contradicts existing literature which suggests that simulation-based teaching is associated with improved performance.(30, 82) This finding may have resulted from the lack of exposure to repeated, structured practice in CVC during the rotation. Procedural competence is associated with experience.(30, 68) Since we did not track the number of central lines inserted during the rotation, we are unable to quantify the central line experiences received by the residents. Our intervention included a single simulation-based training experience and depended on further opportunities for supervised practice to reinforce procedural instruction. These additional learning opportunities may have been limited, or not occurred during the rotation. This unstructured educational process failed to meet the criteria specified by the deliberate practice or mastery models of skill acquisition. In the absence of sufficient opportunities for CVC, residents may not have received adequate feedback on their performance needed to acquire competence with more deliberate practice.(72)

Other measures of performance, such as time to cannulation and number of attempts, have been suggested as surrogates for competence.(32, 109) We did not identify a difference between groups in either measure.

Of particular concern is the number of critical errors made by the residents. At the end of the rotation, the mean number of critical errors per procedure at the internal jugular site was 1.1 by the control group and 1.3 by the intervention group. Seventy-seven percent of residents in the control group and 85% of residents in the intervention group committed at least one critical error at this site. Similarly, at the subclavian site, the mean number of critical errors by the control group was 1.0 and 0.8 by the intervention group. Eighty-one percent of residents in the control group and 79% of residents in the intervention group committed at least one critical error at this site. Mishandling of the needle and inadequate sterile technique were the most commonly observed errors. Mechanical and infectious complications related to CVC are reported to be as high as 26%, leading to patient morbidity, mortality and increased hospital costs.(55, 56, 59) In our study, the majority of our trainees were deemed "not competent" for independence in central line insertion by the raters at the end of their rotation. Since complication rates are inversely related to operator experience, the high complication rates in our study are not surprising.(57)

5.1.4 Question 3: Is there a difference in self-reported comfort with CVC between junior residents trained using a multimedia teaching module, together with simulation

based teaching, and those receiving the standard educational format during an ICU rotation at our institution?

There were no differences between the intervention and the control group related to resident comfort. However, overall resident comfort with CVC was statistically significantly improved for all residents at the end of the rotation, and the residents felt they would be equally comfortable performing the skill with or without direct supervision. Since comfort improves with procedural experience, this finding suggests that many residents perceived they had acquired sufficient CVC experience during the rotation to perform the procedure with limited, if any, supervision.(16) This finding is important, since resident comfort does not necessarily translate to competence.(75, 78, 79, 110) Since no difference in the gain in confidence was identified between the groups, we can assume that this confidence gain relates to other learning opportunities during the rotation. Without data to identify and quantify these other experiences, we are unable to determine whether this improved confidence relates to additional opportunities for central line placement that occurred during the rotation, additional teaching sessions provided by the senior residents or faculty, or self-directed learning.

5.1.5 Question 4: Is there a difference in self-reported satisfaction with the overall educational experience between residents using a multimedia teaching module, together with simulation based instruction using a procedural task trainer, and those receiving the standard educational format following an ICU rotation at our institution?

No difference was observed between groups in the proportion of residents who achieved their personal objectives for the rotation. Residents from a wide variety of training programs rotate through the Calgary intensive care units during their training. The personal learning objectives for these residents differ, based on their professional and individual needs. For many trainees, competence in CVC is a relatively minor learning objective for their ICU rotation. Other competing educational goals such as identifying and caring for critically-ill patients, learning how to lead resuscitations, and gaining experience in discussions related to end-of-life care may take priority. Since the distribution of residents from various training programs was similar between groups, targeting one specific procedural skill was likely insufficient to influence the perception of achievement of the residents' overall needs.

We found that residents who received the multimedia and simulation-based intervention reported an overall increased satisfaction with their educational experience. They were more satisfied with CVC knowledge gained during the rotation and with the quality of the rotation. They gave higher ratings of the adequacy of their training in CVC during the rotation and their understanding of the indications and contraindications for CVC. Improved satisfaction with simulation-based training is supported by the literature.(111-113) Furthermore, our intervention provided clearly defined learning objectives related to CVC, dedicated procedural teaching in a safe and engaging simulation environment, and an online module that specifically addressed the content related to the procedure. This structured approach may have been appreciated by the resident learners who received the simulation-based program. This finding is important, since satisfaction with the learning environment has been associated with improved performance, although this finding was not demonstrated in our study.(114)

5.1.6 Question 5: What is the dependability of the scores derived from the performance assessment tools used in this study for decisions regarding resident competency in CVC?

Ninety-two percent of the control group and 89% of the intervention group passed the cut-score for CVC at the internal jugular site using the checklist tool. However, using the OSATS tool, only 38% of the control group and 27% of the intervention group passed. Using the O-SCORE tool, only 12% of the control group and 8% of the intervention group passed. Similar findings were observed at the subclavian site. This discrepancy between scores using checklists and global rating scales is supported by the literature. Since the technical actions for procedures tend to be sequential and anticipated, itemized checklists are commonly used to ensure learners can competently perform all each necessary steps of the procedure.(70) However, scoring high on a checklist may simply reflect the learner's ability to 'tick' all the boxes for procedural performance, rather than demonstrating the behaviours necessary for competence.(17, 86, 90, 93) Global rating scales such as the OSATS and O-SCORE assess performance differently. These tools rely on subjective expert assessment, which has been shown to be accurate for judging performance, to measure overall competence using behaviourally-anchored scales.(87) The OSATS tool assesses performance against relevant constructs such as instrument handling, procedural flow and knowledge of the procedure, while the O-SCORE tool assesses performance along a continuum that reflects readiness for safe independent practice.(20) Since competence in procedural skills is more than completion of the steps of the procedure, assessment with these tools may better discriminate expertise and provide useful information regarding readiness for autonomy.

Dependability refers to the accuracy of generalizing from an observed test score to that person's average score under all the possible conditions.(100) Our findings suggest that scores derived from global rating tools such as the OSATS and O-SCORE are more dependable than those provided by either checklists or critical error tools. Similar findings for improved reliability with global rating scales are reported in the literature.(20, 86) Unlike the OSATS and O-SCORE tools, neither the checklist, nor the critical error tool achieved the pre-defined Dependability threshold of 0.7 in our study.

The G-study analysis helps to better determine why these tools performed differently. Unlike classical test theory types of reliability estimates (test-retest, interrater, internal consistency), where measurements of error are calculated separately and are not directly comparable to one another, generalizability theory enables us to simultaneously estimate measurement error from various sources and directly compare their relative contributions to the observed test scores.(100, 115) This analysis identified the relative proportions of variance from the participants as well as the different sources of measurement error contributing to the observed test scores.

For the checklist and critical error tools, substantial measurement error contributed to the overall variance in the observed test scores. At both CVC sites, measurement error resulting from the interaction between participant-by-rater-by-item contributed to nearly half the observed test scores. This variance reflects the inconsistency resulting from the three-way interaction between the participants, the raters, and the items, confounded with unmeasured sources of variation. Measurement error from the interaction between the participant-by-item was the second largest contributor to the observed test scores for these tools. This variance represents the inconsistency in the trainees' scores across the individual items contained in the scoring tools (equivalent to internal consistency in classical test theory).(116) Our findings support existing literature in several important ways. They suggest that successful performance in one construct of interest does not necessarily equate to success in other constructs, or overall competence.(117) They also reinforce our understanding that scores provided by checklist tools frequently suffer from lower internal consistency than scores from global rating scales.(93) Finally, the contribution to the observed test scores from resident performance (our object of interest) was relatively small. A similar pattern was identified using these tools at both anatomical sites. In summary, the scores obtained with the checklist and critical error tools contained substantial measurement error not related to resident performance. Consequently, the scores provided by these tools are less dependable.

Unlike the checklist or critical error tools, for the OSATS and O-SCORE tools the largest source of variance contributing to the observed test scores was derived from the participants themselves. For the OSATS tool, the second largest source of variance was measurement error derived from the three-way interaction between the participants, the raters, and the items, confounded with unmeasured sources of variation. The interaction between participant-by-rater was also substantial. This variance reflects the inconsistency of raters' evaluation of the participants (equivalent to inter-rater reliability in classical test theory). To make dependable judgements about performance, it is essential that interpretations of performance are similar between judges. Rater training has been shown to improve inter-rater reliability for the assessment of clinical skills.(118, 119) Despite our attempts to maximize agreement between raters by training them to use the developed tools on sampled videos, the relatively high measurement error attributed to this issue suggests that further training may be necessary.

Our Decision study supports the use of the OSATS or O-SCORE tool in the summative assessment of CVC competence. To improve the dependability of either the checklist or the critical error tool to meet our threshold, we would need to modify the scoring tools by either adding substantially more items, or increasing the number of raters. The items contained within both tools were developed by three expert raters and designed to capture the essential components of the procedure. Increasing the number of raters assessing performance would be both costly and inefficient. On the other hand, Φ for the OSATS and O-SCORE tools exceeded our threshold; thereby allowing us to predict ways to improve efficiency without losing dependability. We determined that we would be able to maintain the dependability for both tests even if we decreased the number of raters required for assessment.

5.2 Implications for education

The findings of this study provide useful information related to the assessment of central line insertion. Many critical care units in Canada are similarly structured to those in our study. Junior residents from numerous training programs rotate through the intensive care unit to acquire experience managing acutely ill patients. For many of these residents, it is expected that proficiency in CVC will occur during this rotation. This study demonstrates that at our institution, most trainees had not achieved sufficient competence for independence with central line insertion at the end of their one-month rotation, irrespective of a multimedia and simulation-based educational intervention at their ICU orientation. Since opportunities to receive supervised, repeated practice during the ICU rotation are variably present, it is unlikely that the ICU rotation alone will be able to provide an educational experience that meets criteria for acquiring competence.

The lack of impact by our multimedia and simulation-based education intervention on performance supports the need to adhere to principles of deliberate practice or mastery theory to achieve competence consistently for all trainees. Many of the defined conditions for success are present in our institution. We have motivated learners who are presented with defined learning objectives. They are given tasks at an appropriate level of difficulty and are assessed with reliable measurement strategies. However, without immediate, informative feedback on performance and opportunities for deliberate practice, they are unable to achieve the expected competence and advance to further levels or tasks. The notion of 'see one, do one, teach one' is no longer acceptable.(120) Training programs must develop organized, structured educational programs, with well delineated learning tasks, to ensure that the conditions for success are met. Our study provides evidence of a deficiency with our current curriculum. Namely, junior residents are not achieving competence in central line insertion during their ICU rotation. This finding may inform training program directors and educators to consider including alternative teaching strategies and repeated opportunities for practice such as dedicated simulation sessions to ensure that residents are receiving sufficient exposure and supervision to acquire competence in essential procedures such as CVC.(68) Relying on a basic simulation-based orientation, together with opportunistic experiences that occur during the ICU rotation is insufficient to attain competence in this skill.

The marked discrepancy between resident confidence and self-perceived ability, as compared to expert assessment of competence is concerning. However, this disconnect between self-perceived confidence and objective assessment of competence is supported in the literature.(110, 121) Educators should use caution when asking resident learners whether they feel comfortable performing procedures without supervision. Our study suggests that the majority of junior residents should not be allowed to perform CVC independently at our institution. Direct observation with feedback on performance should be considered mandatory for all learners, until proficiency is demonstrated. Since competence must be demonstrated before residents are entrusted to perform activities independently, it is expected that CBME will help mitigate these circumstances.

Although assessment of task performance is often focused on the specific motor skills related to the procedure, we found that the largest proportion of critical errors made by the residents included technical aspects related to the procedure itself such as mishandling of the needle, as well as errors relating to more general issues such as inadequate sterile technique. Educators should recognize that assessment tools, such as checklists, may not specifically target critical aspects of competence beyond technical proficiency.(86) Catheter related bloodstream infection is a common complication of CVC associated with significant morbidity and mortality.(74) Adherence to sterile technique is essential to minimize these complications. It is surprising that such a large proportion of residents failed to adhere to guidelines relating to sterile technique during central line insertion. This underscores the importance of ongoing education regarding the importance of maintaining sterility during clinical procedures.(74) We intend to use these results to improve resident education and better assess compliance with sterile technique policies in our ICU.

As educators prepare to transition to CBME, there is a need for reliable and valid assessments, made with task-appropriate tools. These tools must have adequate psychometric properties to assess procedural competence.(15) Furthermore, it is likely that increased observations of learner performance will be required to determine competency.(13, 15) Judgements of performance will be made using a variety of assessment tools, each with their own strengths and weakness. Assessments using proper tools, in the most efficient manner, may provide information necessary to make valid arguments about resident performance.(97) We will use the results of this study to develop an assessment tool for CVC that provides a summative assessment of competence, supported by formative feedback to our learners. To achieve this, we will integrate items from our checklist and critical error tool into the O-SCORE tool. The inclusion of these scales will ensure that residents are provided detailed information related to their performance, including the identification of critical mistakes requiring attention. Since the O-SCORE scale utilizes a behavioural assessment anchor (readiness for independence) that is similar to our current process for determining graded autonomy, this scale will be utilized to provide a summative assessment of competence.

Evidence for the validity of decisions using these tools should be clear. Our study proposed that a validity argument *a priori* is necessary to design assessment processes for central line insertion performed by junior residents at our institution. In this model, analysis of the scores derived from the four assessment tools was performed to determine which one would provide the most dependable scores for this situation. Kane's framework was helpful in defining the validity evidence required, namely scoring, extrapolation, generalization and implication.(96) We provided some evidence that the dependability of the scores provided by the OSATS tool or O-SCORE tools exceeded those from the checklist or the critical error tool.

5.3 Limitations

There are a number of important limitations to our study. First, this study includes a relatively small number of learners at two ICUs at our institution. Due to the design of this study, generalizations to other institutions or different learner groups is not possible.

Second, our educational intervention did not include ultrasound guidance for central line insertion. The use of ultrasound has been demonstrated to be associated with improved success rates and lower a lower incidence of complications.(71, 122) At the time of the study, ultrasound guidance was not routinely used at our institution. Furthermore, utilization of ultrasound is restricted to sites and departments who can afford the technology. This may not be readily available in rural or smaller centres and is often unavailable during emergency procedures such as those encountered on the ward. Clinical practitioners responsible for central venous catheterization must be able to perform this procedure with and without the assistance of ultrasonography, frequently under time constraints related to the nature of the medical emergency.

We did not track the online usage of the multimedia module by the residents. Nor did we assess performance by the residents prior to the start of the rotation. We did not monitor the number of opportunities for supervised, repeated practice of CVC by the residents during their rotation. As such, we are unable to quantify the total educational experiences of the learners.

Although attempts were made to capture all aspects of the procedure, the use of video-recordings of performance on low-fidelity mannequins limited our ability to observe and measure important aspects of CVC competence such as patient positioning, instillation of local anesthesia, suturing of the central line and confirmation of placement. Additionally, other non-technical skills were not assessed, such as donning and doffing of protective equipment or communication skills with the patient or allied health care staff. Despite orienting the residents to the mannequins, technical difficulties with the mannequins prevented some residents from successfully completing the procedure, occasionally requiring interference from the supervising study personnel.

Furthermore, the results of our study are context-specific. They are confined to the performance of junior residents in central line insertion, using landmark technique at the internal jugular or subclavian site on task-trainer mannequins, as judged by expert raters using the four specific assessment tools in this study. Generalization of the results beyond these parameters should be performed with caution.

Finally, the validity of our checklist, critical error tool and modified OSAT tool is uncertain. Although these tools were adapted from existing tools reported in the literature(16-18), with input from content experts, they had not been previously validated for use in the context of this study.

5.4 Future directions

Inclusion of valid and dependable tools to assess procedural competence will be necessary as we transition to CBME. This study will help to develop a scoring tool to assess CVC competence in the intensive care units at the University of Calgary. We intend to modify the behaviourally-anchored O-SCORE tool to include identification of critical errors and space for narrative feedback. In this way, the refined tool may be used to provide a global assessment of performance, together with formative feedback to help the residents recognize specific areas of deficiency in order to improve. A collection of assessments can be utilized to demonstrate improved proficiency over time, and evidence of readiness for independence. Furthermore, we plan to extrapolate these results to create similar scoring tools for other procedural skills frequently performed in the ICU, such as endotracheal intubation, chest tube insertion and arterial line insertion.

5.5 Conclusions

Central venous catheterization is an essential skill to be acquired during residency for many trainees. The results of this study did not demonstrate any impact on knowledge or performance in central line insertion after the introduction of a multimedia model and simulation-based education intervention.

Incorporation of assessment tools that provide dependable data is essential for determining competency in procedural skills during residency. Our results suggest that educators should integrate global rating scales such as the OSATS or O-SCORE tools, rather than checklists or critical error tools, to assess competency in central venous catheterization.

As our training program prepares for CBME implementation, these results will help us restructure our training program to more effectively teach procedural skills and assess competency in procedures such as central venous catheterization.

References

1. Norris TE, Cullison SW, Fihn SD. Teaching procedural skills. J Gen Intern Med. 1997;12 Suppl 2:S64-70.
2. Gorman PJ, Meier AH, Rawn C, Krummel TM. The future of medical education is no longer blood and guts, it is bits and bytes. Am J Surg. 2000;180(5):353-6.
3. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. Med Teach. 2010;32(8):631-7.
4. Taylor RW, Palagiri AV. Central venous catheterization. Crit Care Med. 2007;35(5):1390-6.
5. Graham AS, Ozment C, Tegtmeyer K, Lai S, Braner DA. Videos in clinical medicine. Central venous catheterization. N Engl J Med. 2007;356(21):e21.
6. Binstadt E, Donner S, Nelson J, Flottemesch T, Hegarty C. Simulator training improves fiber-optic intubation proficiency among emergency medicine residents. Acad Emerg Med. 2008;15(11):1211-4.
7. Ma IW, Teteris E, Roberts JM, Bacchus M. Who is teaching and supervising our junior residents' central venous catheterizations? BMC Med Educ. 2011;11:16.
8. Myers JS, Bellini LM, Morris JB, Graham D, Katz J, Potts JR, et al. Internal medicine and general surgery residents' attitudes about the ACGME duty hours regulations: a multicenter study. Acad Med. 2006;81(12):1052-8.
9. Santen SA, Hemphill RR, McDonald MF, Jo CO. Patients' willingness to allow residents to learn to practice medical procedures. Acad Med. 2004;79(2):144-7.
10. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit*. Critical Care Medicine. 2009;37(10):2697-701.
11. Card SE, Snell L, O'Brien B. Are Canadian General Internal Medicine training program graduates well prepared for their future careers? BMC Med Educ. 2006;6:56.
12. Huang GC, Smith CC, Gordon CE, Feller-Kopman DJ, Davis RB, Phillips RS, et al. Beyond the comfort zone: residents assess their comfort performing inpatient medical procedures. Am J Med. 2006;119(1):71 e17-24.
13. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638-45.
14. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability Scales: Outlining Their Usefulness for Competency-Based Clinical Assessment. Acad Med. 2016;91(2):186-90.
15. Fitzgerald JT, Burkhardt JC, Kasten SJ, Mullan PB, Santen SA, Sheets KJ, et al. Assessment challenges in competency-based education: A case study in health professions education. Med Teach. 2016;38(5):482-90.
16. Ma IW, Brindle ME, Ronksley PE, Lorenzetti DL, Sauve RS, Ghali WA. Use of simulation-based education to improve outcomes of central venous catheterization: a systematic review and meta-analysis. Acad Med. 2011;86(9):1137-47.

17. Walzak A, Bacchus M, Schaefer JP, Zarnke K, Glow J, Brass C, et al. Diagnosing technical competence in six bedside procedures: comparing checklists and a global rating scale in the assessment of resident performance. Acad Med. 2015;90(8):1100-8.
18. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-8.
19. MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. Continued Validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): Use in the Simulated Environment. Teach Learn Med. 2016;28(1):72-9.
20. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. Acad Med. 2012;87(10):1401-7.
21. Shapiro M, Simmons W. High fidelity medical simulation: a new paradigm in medical education. Med and Health. 2002;85(10):316-7.
22. Kohn LT, Corrigan J, Donaldson MS. To err is human : building a safer health system. Washington, D.C.: National Academy Press; 2000. xxi, 287 p. p.
23. Royal College of Physicians and Surgeons of Canada. Accreditation of Simulation Programs 2016 [Available from: http://www.royalcollege.ca/rcsite/cpd/accreditation-simulation-programs-e.
24. Goodman W. The world of simulators. Flight International Magazine. 1978;18:435.
25. Keys B, Wolfe J. THe role of management games and simulations in education and research. The Journal of Management. 1990;16:307-36.
26. Rolfe J, Staples K. Flight Simulation. Cambridge, England: Cambridge University Press; 1986.
27. Howard S, Gaba D, FIsh K, Yang G, Sarnquist F. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. Aviation Space Environ Med. 1992;63:763-70.
28. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. Simul Healthc. 2007;2(3):183-93.
29. Barsuk JH, Ahya SN, Cohen ER, McGaghie WC, Wayne DB. Mastery learning of temporary hemodialysis catheter insertion by nephrology fellows using simulation technology and deliberate practice. Am J Kidney Dis. 2009;54(1):70-6.
30. Barsuk JH, Cohen ER, Caprio T, McGaghie WC, Simuni T, Wayne DB. Simulation-based education with mastery learning improves residents' lumbar puncture skills. Neurology. 2012;79(2):132-7.
31. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. 2011;306(9):978-88.
32. Evans LV, Dodge KL, Shah TD, Kaplan LJ, Siegel MD, Moore CL, et al. Simulation training in central venous catheter insertion: improved performance in clinical practice. Acad Med. 2010;85(9):1462-9.
33. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. J Gen Intern Med. 2008;23 Suppl 1:46-9.

34. Sekiguchi H, Tokita JE, Minami T, Eisen LA, Mayo PH, Narasimhan M. A prerotational, simulation-based workshop improves the safety of central venous catheter insertion: results of a successful internal medicine house staff training program. Chest. 2011;140(3):652-8.
35. Smith CC, Huang GC, Newman LR, Clardy PF, Feller-Kopman D, Cho M, et al. Simulation training and its effect on long-term resident performance in central venous catheterization. Simul Healthc. 2010;5(3):146-51.
36. Euliano TY. Small group teaching: clinical correlation with a human patient simulator. Adv Physiol Educ. 2001;25(1-4):36-43.
37. Boet S, Sharma S, Goldman J, Reeves S. Review article: medical education research: an overview of methods. Can J Anaesth. 2012;59(2):159-70.
38. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg. 1999;177(2):167-70.
39. Curran VR, Aziz K, O'Young S, Bessell C. Evaluation of the effect of a computerized training simulator (ANAKIN) on the retention of neonatal resuscitation skills. Teach Learn Med. 2004;16(2):157-64.
40. Conroy SM, Bond WF, Pheasant KS, Ceccacci N. Competence and retention in performance of the lumbar puncture procedure in a task trainer model. Simul Healthc. 2010;5(3):133-8.
41. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg. 2004;91(2):146-50.
42. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg. 2008;32(2):182-8.
43. Nackman GB, Bermann M, Hammond J. Effective use of human simulators in surgical education. J Surg Res. 2003;115(2):214-8.
44. Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. Chest. 2008;133(1):56-61.
45. De Ponti R, Marazzi R, Ghiringhelli S, Salerno-Uriarte JA, Calkins H, Cheng A. Superiority of simulator-based training compared with conventional training methodologies in the performance of transseptal catheterization. J Am Coll Cardiol. 2011;58(4):359-63.
46. Draycott TJ, Crofts JF, Ash JP, Wilson LV, Yard E, Sibanda T, et al. Improving neonatal outcome through practical shoulder dystocia training. Obstet Gynecol. 2008;112(1):14-20.
47. Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Critical Care Medicine 2011 [Available from: http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/142_critical_care_int_med_2016.pdf?ver=2016-03-23-113132-843.
48. Royal College of Physicians and Surgeons of Canada. Objectives of Training in the Specialty of Internal Medicine 2011 [updated 2011. Available from: http://www.royalcollege.ca/rc/faces/oracle/webcenter/portalapp/pages/ibd.jspx;jsessionid

=wD4t68tmdzOAbTpnuUsj5U1qr12cPu7l1YHoGNaZhAGW9uVUMqA4!1908427209?_afrLoop=9467776588024266&_afrWindowMode=0&_afrWindowId=null - !%40%40%3F_afrWindowId%3Dnull%26_afrLoop%3D9467776588024266%26_afrWindowMode%3D0%26_adf.ctrl-state%3Dkmaocqxik_4.
49. Royal College of Physicians and Surgeons of Canada. Objectives of Training in the Specialty of General Surgery 2010 [updated 2010 Available from: http://www.royalcollege.ca/rc/faces/oracle/webcenter/portalapp/pages/ibd.jspx?_adf.ctrl-state=kmaocqxik_4&_afrLoop=9467879085596266&_afrWindowMode=0&_afrWindowId=kmaocqxik_1 - !
50. Royal College of Physicians and Surgeons of Canada. Objectives of Training in the Subspecialty of Adult Critical Care Medicine 2014 [updated 2014. Available from: http://www.royalcollege.ca/rc/faces/oracle/webcenter/portalapp/pages/ibd.jspx?_adf.ctrl-state=kmaocqxik_4&_afrLoop=9467879085596266&_afrWindowMode=0&_afrWindowId=kmaocqxik_1 - !
51. Royal College of Physicians and Surgeons of Canada. Objectives of Training in the Specialty of Emergency Medicine 2014 [updated 2014. Available from: http://www.royalcollege.ca/rc/faces/oracle/webcenter/portalapp/pages/ibd.jspx?_adf.ctrl-state=kmaocqxik_4&_afrLoop=9467879085596266&_afrWindowMode=0&_afrWindowId=kmaocqxik_1 - !
52. Edgeworth J. Intravascular catheter infections. J Hosp Infect. 2009;73(4):323-30.
53. Mermel LA. Prevention of intravascular catheter-related infections. Ann Intern Med. 2000;132(5):391-402.
54. O'Grady NP, Alexander M, Burns LA, Dellinger EP, Garland J, Heard SO, et al. Guidelines for the prevention of intravascular catheter-related infections. Am J Infect Control. 2011;39(4 Suppl 1):S1-34.
55. Lemaster CH, Agrawal AT, Hou P, Schuur JD. Systematic review of emergency department central venous and arterial catheter infection. Int J Emerg Med. 2010;3(4):409-23.
56. Marik PE, Flemmer M, Harrison W. The risk of catheter-related bloodstream infection with femoral venous catheters as compared to subclavian and internal jugular venous catheters: a systematic review of the literature and meta-analysis. Crit Care Med. 2012;40(8):2479-85.
57. Sznajder JI, Zveibil FR, Bitterman H, Weiner P, Bursztein S. Central vein catheterization. Failure and complication rates by three percutaneous approaches. Arch Intern Med. 1986;146(2):259-61.
58. Deshpande KS, Hatem C, Ulrich HL, Currie BP, Aldrich TK, Bryan-Brown CW, et al. The incidence of infectious complications of central venous catheters at the subclavian, internal jugular, and femoral sites in an intensive care unit population. Critical Care Medicine. 2005;33(1):13-20.
59. Girard TD, Schectman JM. Ultrasound guidance during central venous catheterization: a survey of use by house staff physicians. J Crit Care. 2005;20(3):224-9.
60. Ogden PE, Cobbs LS, Howell MR, Sibbitt SJ, DiPette DJ. Clinical simulation: importance to the internal medicine educational mission. Am J Med. 2007;120(9):820-4.

61. Sherertz RJ, Ely EW, Westbrook DM, Gledhill KS, Streed SA, Kiger B, et al. Education of physicians-in-training can decrease the risk for vascular catheter infection. Ann Intern Med. 2000;132(8):641-8.
62. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, et al. Simulation technology for health care professional skills training and assessment. JAMA. 1999;282(9):861-6.
63. Maloney SH, T. Issues of cost-benefit and cost-effectiveness for simulation in health professions education. Advances in Simulation. 2016;1(13):1-6.
64. Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB. Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. J Hosp Med. 2009;4(7):397-403.
65. Britt RC, Novosel TJ, Britt LD, Sullivan M. The impact of central line simulation before the ICU experience. Am J Surg. 2009;197(4):533-6.
66. Dong Y, Suri HS, Cook DA, Kashani KB, Mullon JJ, Enders FT, et al. Simulation-based objective assessment discerns clinical proficiency in central line placement: a construct validation. Chest. 2010;137(5):1050-6.
67. Ault MJ, Rosen BT, Ault B. The use of tissue models for vascular access training. Phase I of the procedural patient safety initiative. J Gen Intern Med. 2006;21(5):514-7.
68. Alsaad AA, Bhide VY, Moss JL, Jr., Silvers SM, Johnson MM, Maniaci MJ. Central Line Proficiency Test Outcomes after Simulation Training vs. Traditional Training to Competence. Ann Am Thorac Soc. 2017.
69. Eisen LA, Narasimhan M, Berger JS, Mayo PH, Rosen MJ, Schneider RF. Mechanical complications of central venous catheters. J Intensive Care Med. 2006;21(1):40-6.
70. Evans LV, Dodge KL. Simulation and patient safety: evaluative checklists for central venous catheter insertion. Qual Saf Health Care. 2010;19 Suppl 3:i42-6.
71. Woo MY, Frank J, Lee AC, Thompson C, Cardinal P, Yeung M, et al. Effectiveness of a novel training program for emergency medicine residents in ultrasound-guided insertion of central venous catheters. CJEM. 2009;11(4):343-8.
72. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10 Suppl):S70-81.
73. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. Med Educ. 2014;48(4):375-85.
74. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. Arch Intern Med. 2009;169(15):1420-3.
75. Barr J, Graffeo CS. Procedural Experience and Confidence Among Graduating Medical Students. J Surg Educ. 2016;73(3):466-73.
76. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86(6):706-11.

77. Ericsson KA. Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. Acad Med. 2015;90(11):1471-86.
78. Katowa-Mukwato P, Banda S. Self-perceived versus objectively measured competence in performing clinical practical procedures by final year medical students. Int J Med Educ. 2016;7:122-9.
79. Kenny L, Teasdale R, Marsh M, McElnay P. Techniques of training in the management of tension pneumothorax: bridging the gap between confidence and competence. Ann Transl Med. 2016;4(12):233.
80. Augustine EM, Kahana M. Effect of procedure simulation workshops on resident procedural confidence and competence. J Grad Med Educ. 2012;4(4):479-85.
81. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296(9):1094-102.
82. Ma IW, Sharma N, Brindle ME, Caird J, McLaughlin K. Measuring competence in central venous catheterization: a systematic-review. Springerplus. 2014;3:33.
83. Rosenthal R, Gantert WA, Hamel C, Hahnloser D, Metzger J, Kocher T, et al. Assessment of construct validity of a virtual reality laparoscopy simulator. J Laparoendosc Adv Surg Tech A. 2007;17(4):407-13.
84. Cohen DS, Colliver JA, Robbs RS, Swartz MH. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. Adv Health Sci Educ Theory Pract. 1996;1(3):209-13.
85. McKinley RK, Strand J, Ward L, Gray T, Alun-Jones T, Miller H. Checklists for assessment and certification of clinical procedural skills omit essential competencies: a systematic review. Med Educ. 2008;42(4):338-49.
86. Ma IW, Zalunardo N, Pachev G, Beran T, Brown M, Hatala R, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. Adv Health Sci Educ Theory Pract. 2012;17(4):457-70.
87. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003;37(11):1012-6.
88. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med. 1998;73(9):993-7.
89. Norman G. Editorial--checklists vs. ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. Adv Health Sci Educ Theory Pract. 2005;10(1):1-3.
90. Gerard JM, Kessler DO, Braun C, Mehta R, Scalzo AJ, Auerbach M. Validation of global rating scale and checklist instruments for the infant lumbar puncture procedure. Simul Healthc. 2013;8(3):148-54.
91. Ma IW, Pugh D, Mema B, Brindle ME, Cooke L, Stromer JN. Use of an error-focused checklist to identify incompetence in lumbar puncture performances. Med Educ. 2015;49(10):1004-15.

92. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Acad Med. 1996;71(12):1363-5.
93. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ. 2015;49(2):161-73.
94. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003;37(9):830-7.
95. Messick S. Educational Measurement. 3rd ed. New York1989.
96. Kane M. Validation. 4th ed. Westport, CT: Praegar; 2006.
97. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560-75.
98. Norcini JJ. Setting standards on educational tests. Med Educ. 2003;37(5):464-9.
99. Wittler M, Hartman N, Manthey D, Hiestand B, Askew K. Video-augmented feedback for procedural performance. Med Teach. 2016;38(6):607-12.
100. Shavelson RJ, Webb NM. Generalizability theory : a primer. Newbury Park, Calif.: Sage Publications; 1991. xiii, 137 p. p.
101. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. Med Educ. 2002;36(10):972-8.
102. Lockyer J. Multisource feedback in the assessment of physician competencies. J Contin Educ Health Prof. 2003;23(1):4-12.
103. Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. BMJ. 2003;326(7388):546-8.
104. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates Inc.; 1988.
105. Kohls-Gatzoulis JA, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. Can J Surg. 2004;47(4):277-83.
106. Kalet AL, Song HS, Sarpel U, Schwartz R, Brenner J, Ark TK, et al. Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module. Med Teach. 2012;34(10):833-9.
107. Grover S, Currier PF, Elinoff JM, Katz JT, McMahon GT. Improving residents' knowledge of arterial and central line placement with a web-based curriculum. J Grad Med Educ. 2010;2(4):548-54.
108. Cook DA, Levinson AJ, Garside S. Time and learning efficiency in Internet-based learning: a systematic review and meta-analysis. Adv Health Sci Educ Theory Pract. 2010;15(5):755-70.
109. Carvalho P. Early introduction to central line placement: a curriculum for medical students. Med Educ. 2007;41(11):1098-9.
110. Tofil NM, Benner KW, Zinkan L, Alten J, Varisco BM, White ML. Pediatric intensive care simulation course: a new paradigm in teaching. J Grad Med Educ. 2011;3(1):81-7.
111. Lorello GR, Cook DA, Johnson RL, Brydges R. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. Br J Anaesth. 2014;112(2):231-45.

112. Lo BM, Devine AS, Evans DP, Byars DV, Lamm OY, Lee RJ, et al. Comparison of traditional versus high-fidelity simulation in the retention of ACLS knowledge. Resuscitation. 2011;82(11):1440-3.
113. Ten Eyck RP, Tews M, Ballester JM. Improved medical student satisfaction and test performance with a simulation-based emergency medicine curriculum: a randomized controlled trial. Ann Emerg Med. 2009;54(5):684-91.
114. Wayne SJ, Fortner SA, Kitzes JA, Timm C, Kalishman S. Cause or effect? The relationship between student perception of the medical school learning environment and academic performance on USMLE Step 1. Med Teach. 2013;35(5):376-80.
115. Lawson DM. Applying generalizability theory to high-stakes objective structured clinical examinations in a naturalistic environment. J Manipulative Physiol Ther. 2006;29(6):463-7.
116. Creswell JW. Research Design: Qualitative, Quantitative and Mixed Methods Approaches. 4th ed. USA: Sage; 2014.
117. Wimmers PF, Fung CC. The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. Med Educ. 2008;42(6):580-8.
118. Lou X, Lee R, Feins RH, Enter D, Hicks GL, Jr., Verrier ED, et al. Training less-experienced faculty improves reliability of skills assessment in cardiac surgery. J Thorac Cardiovasc Surg. 2014;148(6):2491-6 e1-2.
119. Evans LV, Morse JL, Hamann CJ, Osborne M, Lin Z, D'Onofrio G. The development of an independent rater system to assess residents' competence in invasive procedures. Acad Med. 2009;84(8):1135-43.
120. Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, et al. Beyond "see one, do one, teach one": toward a different training paradigm. Postgraduate Medical Journal. 2009;85(1003):244-9.
121. Lum ME, Galletly DC. Resuscitation skills of first year postgraduate doctors. N Z Med J. 1989;102(873):406-8.
122. Leung J, Duffy M, Finckh A. Real-time ultrasonographically-guided internal jugular vein catheterization in the emergency department increases success rates and reduces complications: a randomized, prospective study. Ann Emerg Med. 2006;48(5):540-7.

APPENDIX A: PRE-TEST WRITTEN EXAMINATION

Critical Care Medicine
Rotating Resident Written Exam (A)

1. List 5 indications for central venous catheter placement?
   i.________________________
   ii.________________________
   iii.________________________
   iv.________________________
   v.________________________

2. What barrier equipment must be worn by the physician inserting a central line?
   i.________________________
   ii.________________________
   iii.________________________
   iv.________________________

3. Following placement of a right-sided internal jugular catheter, where should the tip of the catheter be located on chest Xray?

____

____

4. What are 3 different ways to confirm venous placement of a central venous catheter?
   i.________________________
   ii.________________________
   iii.________________________

5. If peripheral access is unobtainable on a patient, what type of catheter should be centrally inserted for fluid resuscitation?

____

6. Approximately what percentage of central venous catheters are complicated by bacteremia?

____

7. What is the mean distance (cm) from the skin to the junction of the right atrium-SVC junction using the right internal jugular or right subclavian approach?

____

8. Identify 5 potential acute complications of a right internal jugular catheter insertion.
   i.________________________
   ii.________________________
   iii.________________________

iv.____
v.____

9. A patient being ventilated with pressure control ventilation becomes agitated following an attempted subclavian catheter insertion. Identify 3 physical exam signs that would suggest a tension pneumothorax.

i.____
ii.____
iii.____

10. A patient being ventilated with pressure control ventilation becomes agitated following an attempted subclavian catheter insertion. Identify 2 ventilator parameters that would suggest a tension pneumothorax.

i.____
ii.____

11. When inserting a right subclavian catheter, how should the bevel of the 14 gauge needle be oriented?

____

12. Approximately what percentage of subclavian catheters are complicated by pneumothorax?

____

13. Which anatomical site for central line placement has the most consistent anatomy?

____

14. Which site should be utilized for central venous catheter access in the trauma patient? List your reasons why.

____
____
____
____

15. Which anatomical site of central catheter insertion is associated with the highest rate of septic complications?

____

APPENDIX B: POST-TEST WRITTEN EXAMINATION

Critical Care Medicine
Rotating Resident Written Exam (B)

1. List 5 relative contra-indications for subclavian central venous catheter placement?
   i.____
   ii.____
   iii.____
   iv.____
   v.____

2. What barrier equipment must be worn by personnel directly assisting the physician inserting a central venous catheter?

3. Following placement of a right-sided subclavian catheter, where should the tip of the catheter be located on chest Xray?

4. List 2 reasons why the right-sided subclavian insertion site is preferred to the left-sided subclavian insertion site for central venous access?
   i.____
   ii.____

5. What sized central venous catheter should be inserted for initiation of inotrope/vasopressor therapy?

6. To minimize infectious complications, how long should the topical antiseptic (2% chlorhexidine gluconate) be allowed to remain on the site prior to central venous catheter insertion?

7. How frequently should central venous catheters be routinely replaced in the intensive care unit?

8. Identify 5 potential acute complications of an attempted femoral central venous catheter insertion.
   i.____

ii.____
iii.____
iv.____
v.____

9. A patient being ventilated with pressure control ventilation becomes agitated following an attempted subclavian catheter insertion. Identify 3 physical exam signs that would suggest an acute air embolism as a complication of the attempted line placement.
i.____
ii.____
iii.____

10. What are three immediate treatment priorities following suspicion of an air embolism?
i.____
ii.____
iii.____

11. When inserting a right internal jugular central line, how should the patient be positioned?

12. Approximately what percentage of patients with femoral central venous catheters are complicated by clinically significant DVT?

13. Identify the anatomical landmarks identifying the preferred site of insertion for a femoral central venous catheter?

14. Which site should be utilized for central venous catheter access in the cardiac arrest patient? List your reasons why.

15. Which anatomic site is preferred in the hypovolemic patient? Why?

APPENDIX C: CHECKLIST TOOL

Subject: **Reviewer:**

Competence Tool #1 (Checklist):
Right Internal Jugular Central Venous Catheter Insertion

Procedure Step	Successfully Completed? Yes	No
1. Appropriately applies sterile solution (At least two passes with sterile solution in circular fashion, moving from centre outwards, new swab per pass)	☐	☐
2. Correctly drapes patient in a sterile fashion (Appropriate use of sterile drapes, with appropriate field of view)	☐	☐
3. Appropriately utilizes anatomic landmarks for technique (Clearly utilizes landmarks)	☐	☐
4. Competently inserts angiocatheter or steel needle into IJ vein (Needle advanced at 30-45 degrees, towards ipsilateral nipple)	☐	☐
5. Safely inserts guide wire (Advances J-wire into vein, J-tip first, avoids excess pressure, maintains control of the wire, maintains control of needle)	☐	☐
6. Effectively uses scalpel to make skin incision (Single appropiate sized stab incision, dull side of blade along wire)	☐	☐
7. Smoothly passes dilator over wire and dilates IJ vein	☐	☐
8. Effectively inserts central venous catheter over wire (Maintains direct visualization of wire at all times)	☐	☐
9. Adequately flushes all catheter lumens (This may be done at start or end of procedure or combination)	☐	☐

APPENDIX D: CRITICAL ERROR TOOL

Competence Tool #2 (Critical Errors):
Please indicate whether any of the following critical errors were identified:

i.	Inadequate/inappropriate application of sterile solution	☐
ii.	Inadequate/inappropriate application of sterile drapes	☐
iii.	Critical mishandling of needle (e.g. no syringe, inappropriate depth, grossly incorrect direction)	☐
iv.	Critical mishandling or loss of guidewire	☐
v.	Not flushing all catheter lumens	☐
vi.	Other (please identify) ________________________________	☐

APPENDIX E: OSATS TOOL

Subject Number: **Reviewer:**

Competence Tool #1 (OSATS)

Specific components of technical ability:	**Not competent to perform independently**	**Borderline competent to perform independently**	**Competent to perform independently**	**Above average to perform independently**
Time & Motion	☐ Many unnecessary movements	☐	☐ Efficient but some unnecessary movements	☐ Economy of movement and maximum efficiency
Instrument handling	☐ Repeated awkward or tentative handling of instruments	☐	☐ Competent, occasionally appeared stiff or awkward	☐ Fluid handling of instruments; no stiffness or awkwardness
Knowledge of Instruments	☐ Frequently used inappropriate instrument	☐	☐ Used appropriate instrument for task	☐ Obviously familiar with required instruments
Flow of Procedure	☐ Frequently stopped and seemed unsure of next step	☐	☐ Demonstrated reasonable progression of procedure	☐ Obviously planned procedure; smooth flow
Knowledge of Specific Procedure	☐ Would have needed instruction to complete most steps	☐	☐ Seemed to know most steps of procedure	☐ Demonstrated familiarity with all aspects of procedure

APPENDIX F: O-SCORE EVALUATION TOOL

Competence Tool #2 (O-Score)

Overall Ability to Perform Procedure – Please circle candidate's overall score	
Score	**Descriptor:**
1	"I would have had to do it" (would have required complete hands on guidance) = NOT COMPETENT
2	"I would have had to talk them through it" (able to perform tasks but would require constant direction and supervision) = NOT COMPETENT
3	"I would have had to prompt them from time to time" (some independence but with intermittent direction) = NOT COMPETENT
4	"I would have needed to be in the room just in case" (independence but still required some supervision for safe practice) = BORDERLINE COMPETENT
5	"I would not have needed to be in the room" (complete independence; performs safely; practice ready) = COMPETENT

Time to Cannulation with CVC (seconds):

Number of Passes with needle:

APPENDIX G: ROTATION EXIT SURVEY

Rotating Resident Exit Survey – Central Venous Catheter Study

1. Which Academic Training Program are you with?

Anaesthesia	○	CCFP:Emergency Medicine	○	ENT	○	Family Medicine (Rural)	○
Family Medicine (Urban)	○	FRCP:Emergency Medicine	○	General Surgery	○	Medicine	○
Neurology	○	Neurosurgery	○	OB/GYN	○	Ophthalmology	○
Orthopedic Surgery	○	Pathology	○	Plastic Surgery	○	Psychiatry	○
Radiology	○	Other	○				

2. What is your current year of training? 1 2 3 4 5

3. What is your Sex? Male ○ Female ○

4. What is your age? ____________________

5. Prior to this ICU rotation, have you inserted a central venous catheter? Yes No

6. If Yes, approximately how many have you inserted? 1-5 5-10 >10

Description Please indicate how well you agree with the following statements	**Strongly Disagree** **1**	**Disagree** **2**	**Neutral** **3**	**Agree** **4**	**Strongly Agree** **5**
2. *Prior to this rotation*, I felt comfortable inserting a central venous catheter *under direct* supervision	○	○	○	○	○
3. *Prior to this rotation*, I felt comfortable inserting a central venous catheter *without* direct supervision	○	○	○	○	○
4. *Following this ICU rotation*, I feel comfortable inserting a central venous catheter *under direct* supervision	○	○	○	○	○
5. *Following this ICU rotation*, I feel comfortable inserting a	○	○	○	○	○

central venous catheter *without* direct supervision					
6. I feel my training in central venous catheter insertion during this rotation was adequate	○	○	○	○	○
7. I understand the indications for central venous catheter insertion	○	○	○	○	○
8. I understand the contra-indications for central venous catheter insertion	○	○	○	○	○
9. I understand the potential complications of central venous catheter insertion	○	○	○	○	○
10. I achieved the clinical objectives pertaining to central venous catheter insertion during this rotation	○	○	○	○	○

APPENDIX H: VARIANCE COMPONENT RESULTS

H.1. Right IJ checklist variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	52	.795	0.020	11.36
Raters (*r*)	2	.954	0.002	1.14
Items *(i)*	8	2.080	0.011	6.25
p x *r*	104	.121	0.003	1.70
p x *i*	411	.229	0.046	26.14
r x *i*	16	.170	0.001	0.57
pri, e	820	.093	0.093	52.84
Total				100.00

H.2. Right SC checklist variance component results

Source of Variation	df	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	53	.650	0.017	12.59
Raters (*r*)	2	.656	0.001	0.74
Items *(i)*	8	.841	0.003	2.22
p x *r*	105	.112	0.005	3.71
p x *i*	413	.178	0.037	27.41
r x *i*	16	.254	0.004	2.96
pri, e	816	.068	0.068	50.37
Total				100.00

H.3. Right IJ critical error tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	52	.676	0.025	13.30
Raters (*r*)	2	1.215	0.003	1.60
Items *(i)*	5	3.729	0.022	11.70
p x *r*	104	.109	0.003	1.59
p x *i*	260	.214	0.041	21.81
r x *i*	10	.183	0.002	1.06
pri, e	520	.092	0.092	48.94
Total				100.00

H.4. Right SC critical error tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	53	.537	0.021	14.79
Raters (*r*)	2	.605	0.001	0.70
Items *(i)*	5	1.649	0.008	5.63
p x *r*	106	.076	0.000	0.00
p x *i*	265	.163	0.028	19.72
r x *i*	10	.306	0.004	2.82
pri, e	530	.080	0.080	56.34
Total				100.00

H.5. Right IJ OSATS tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	52	10.605	0.627	58.54
Raters (*r*)	2	6.022	0.016	1.49
Items *(i)*	4	.875	0.000	0.00
p x *r*	104	1.065	0.173	16.16
p x *i*	208	.330	0.043	4.02
r x *i*	8	.837	0.012	1.12
pri, e	415	.199	0.200	18.67
Total				100.00

H.6. Right SC OSATS tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	52	6.943	0.370	42.82
Raters (*r*)	2	5.777	0.015	1.74
Items *(i)*	4	.508	0.000	0.00
p x *r*	103	1.351	0.229	26.50
p x *i*	208	.315	0.036	4.17
r x *i*	8	.577	0.007	0.81
pri, e	412	.207	0.207	23.96
Total				100.00

H.7. Right IJ O-SCORE tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	52	4.115	1.182	65.49
Raters (*r*)	2	3.415	0.054	2.99
pr, e	104	.569	0.569	31.52
Total				100.00

H.8. Right SC O-SCORE tool variance component results

Source of Variation	*df*	Mean Squares	Variance Components	Variance Explained (%)
Participants (*p*)	53	3.340	0.897	54.20
Raters (*r*)	2	3.135	0.041	2.48
pr, e	103	.695	0.717	43.32
Total				100.00