The Pennsylvania State University

The Graduate School

TRANSFORMING MEDICAL EDUCATION THROUGH SIMULATION DESIGN: THE DEVELOPMENT AND VALIDATION OF A CENTRAL LINE TRAINING SYSTEM

A Dissertation in

Industrial Engineering

by

Haroula M. Tzamaras

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2024

The dissertation of Haroula Tzamaras was reviewed and approved by the following:

Scarlett R. Miller
Professor of Engineering Design and Industrial Engineering
Dissertation Advisor
Chair of Committee

Ling Rothrock
Professor of Industrial and Manufacturing Engineering

Yiqi Zhang
Associate Professor of Industrial and Manufacturing Engineering

Jason Z. Moore
Professor of Mechanical Engineering
Outside Field/ Unit Member of Committee

Elizabeth Sinz
Professor of Anesthesiology & Neurosurgery
Associate Dean of Clinical Simulation
Director, Clinical Simulation Center
Penn State University College of Medicine
Special Member

Steven Landry
Department Head of Industrial and Manufacturing Engineering

ABSTRACT

Central Venous Catheterization (CVC) is a commonly performed medical procedure used for medication delivery to the heart. While CVC is conducted over 5 million times annually it is plagued with high complication rates, resulting in adverse effects on patients, and in the worst cases, death. These complications are directly related to the experience level of the performing physician. A physician who has conducted less than 50 CVCs, is *twice as likely* to incur complications than a physician with more experience, reiterating the need for robust training of CVC for new medical residents.

To better train physicians in complex procedures like CVC, many residency programs utilize simulation-based training (SBT). SBT is an imitation of a procedure or environment that allows trainees to practice hands-on medical procedures risk-free to a predefined mastery level of performance before conducting the procedure on patients. When determining the effectiveness of SBT, instructors will sometimes employ self-assessment to gauge trainee knowledge gains. Self-assessment can be useful; however, when using self-assessment for measuring learning success, other factors like gender can potentially cause trainees to rate themselves lower even if their learning and performance is equivalent. This gender-gap is not widely researched in SBT, but is important to understand in the context of learning.

Additionally, most SBT methods require residents to already know how to conduct the steps of the procedure on their first usage, without checking for understanding. In this way, many simulators are designed for practicing procedures, but not for effective learning, indicating a need for innovative training methods that can do both. For CVC SBT, manikin trainers are commonly utilized and are useful because they provide hands-on practice but are limited in that they only provide practice on one anatomy and do not provide automated feedback to the trainee. The Dynamic Haptic Robotic Trainer (DHRT) addresses these deficits of manikin training for CVC by providing force tissue profiles to simulate multiple patient anatomies, along with providing automated, personalized feedback on performance to help the trainee learn and improve. While the DHRT has been shown to train residents as effectively as manikin trainers without the need of a trained preceptor, it only teaches *part of the mechanical portions* of CVC residents need to know to be proficient in the clinical environment. In addition, while DHRT has been validated for its educational effectiveness, it lacks clinical validation.

Considering these gaps in medical training for CVC SBT, the objective of this dissertation was to transform CVC education through assessment of current training methods, development of new training methods, and validation of new training methods. Specifically, this dissertation focused on: (1) evaluating the impact of sequential learning on initial skill gain and learning over time, (2) assessing DHRT training for differences in self-efficacy between men and women, (3) developing and analyzing the impact of a novel comprehensive simulator on resident self-efficacy and proficiency, and (4) validating the comprehensive simulator through eye gaze in the operating room and on the simulator with novice and expert physicians. The results of this dissertation indicate that sequential learning significantly increased initial skill gain, decreased the number of trials required to complete training, and reduced learning curves, women rate their self-efficacy significantly lower than men despite no performance differences for SBT and neither men nor women are able to accurately self-assess performance, comprehensive simulation is more effective than the original DHRT training for resident performance and self-efficacy, and the comprehensive simulator exhibits both predictive validity by aligning expert gaze between the simulator and the operating room and construct validity by distinguishing between expert and novices. This dissertation also provides novel methodology for conducting validity studies in the clinical environment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

When I decided to leave my job to come to graduate school, I had no idea what was waiting for me at Penn State, and I have been overwhelmed by the support I have received from the people in my life. Graduate school is challenging, rewarding, and set me on a path I never would have discovered without Penn State. I could never give enough thank yous to the people who have supported me throughout this degree.

First and foremost, to my committee members - I want to give a heartfelt thank you to my advisor, Dr. Scarlett Miller. I came to Penn State specifically to work with you, and am so grateful for all that has happened since then. Pursuing this PhD has given me so much direction into my desired career path and really helped me improve as a writer and researcher. Thank you for pushing me, for believing in me, and for giving me the funding and time to travel to conferences, present my research, and build a network. These experiences have been invaluable to my personal and professional growth. Your bright (BRITE?) energy and drive are inspiring, and I appreciate to have had the opportunity to learn from you. To Dr. Jason Moore, for your continuous positivity and project support, and for always believing that things will work out even when we don't have the solution yet. To Dr. Lisa Sinz, for basically adopting me while I was in Hershey for my data collection, for being such a joyful (and inspiring!) person to be around, and for being vital to the success of me being able to collect data in the operating room. And finally, to Dr. Ling Rothrock and Dr. Yiqi Zhang for their willingness to serve on my committee and provide feedback on my work.

To my lab mates – both current and past. Courtney, for always checking in and providing advice and guidance. For Aoran, for sharing snacks and being hilarious and fun to sit next to. To Abby, Sam, and others who have graduated before me, for being so kind and supportive. To Divyaa, for being a great support and always having a nice word to say when grad school got stressful. To Jessica, for being the most positive, sweetest person to ever exist. To Israa for being such a bright spot in my life, reminding me not to overthink, and being such a hype person. To Aseel, for being such an inspiration and example of hard work, and such a lovely human. To Marc, for being fellow Swiftie and trying to keep it positive even when life is rough. To Yi, even though we only overlapped for a short time, I have really enjoyed getting to know you and having you and your kind presence around the lab. And to all of the undergrads that I have had the pleasure of mentoring in some capacity (Aayod, Maggie, Meredith, Joe, Leilani, and Rachel) for teaching me how to teach, for teaching me how to be a better mentor, and for letting me feel like I am making a difference. Finally, to my not-quite-lab mates Dailen and Hang-Ling. Dailen for going through the ups and downs of product design, but being such a kind and positive spirit. And Hang-Ling for hyping me up and making me feel like I know what I am doing (plus the bike rides!).

To my biggest mentor outside of my project, Dr. Cindy Reed. I truly would not be finishing this degree without you, and I say that genuinely and sincerely. Thank you for not letting me give up on myself. And to the all of friends that I have made at Penn State, whether I have known you for a short time or a long time, thank you for providing a venting space, a hug, or a laugh whenever it was needed. To Gaby, you have been a colleague, a role model, one of the best people that I know, and an amazing friend, thank you. To Jessica Gonzalez-Vargas, for meeting me when I came to visit Penn State and being part of the reason, I came here – thank you for always being my friend, helping me practice my Spanish, and showing me your beautiful island. To Justin and Michelle Rist, for being some of my favorite people and the most wholesome people in State College (and to Michelle for gifting me the rainbow keyboard that I wrote this dissertation with). To Erin Johnson, you are a breath of fresh air in my life, thank you for spending time with me, reminding me I am not crazy, and always being a voice of reason. To Sandeep Krishnakumar, I missed you my last year! My grad school experience wouldn't have been the same without your friendship and I appreciate you immensely. To Stephanie Bunt, for climbing mountains with me metaphorically and literally, and always supporting me in my feelings; can't wait to visit you in New Mexico. To Rob and Robyn Newton, I forgive you for leaving, and appreciate how you always have and had my back. To Debrina Roy, for being one of my first few friends here, sharing your amazing artwork with me, and being such a sweet presence in my life.

To the friends I had before Penn State. Leah, for being a supportive role model who is one of the most thoughtful people in my orbit. To Jess, for bonding over the hard parts of grad school with me, being one of the loveliest humans, and always making time for a phone call. To Kellen, Amanda, and Val for providing consistent friendship and letting me vent to you when life was hard. And finally, to Josh, for checking in every couple of months to make sure we're both still surviving and thriving. To my mentors from UMD, Ken Kiger and Pam Armstrong, for supporting my decision to pursue a PhD, writing me letters of recommendation, and providing guidance whenever I need it.

To my family. Mom and Dad, thank you for supporting me in my endeavor to pursue a PhD, and aim to become the first Dr. Tzamaras. Without your consistent support, checking in, and undying belief in my ability to succeed, I would not have made it through this degree. You are two of my favorite people in the world, and I am so grateful for everything you have done for me. To my Yiayia, for calling me to check in every Sunday no matter what and being able to tell if something was wrong just by the sound of my voice. To my Papou, for being my biggest supporter who was always so proud of me. I am sorry you did not live to see me finish, but I thought of you often. To my big sister Dema, for being my first friend in this world, for being one of my biggest inspirations and cheerleaders, and for your general "you can do hard things" vibe. To my little brother Lexi, for letting me work on my leadership skills early on by bossing you around, for always supporting me, and for being one of the funniest people and meme providers in my life. To my in-laws' Susan, Andrew, Pat, Tom, Sue, Chuck, Irving, and Sam for truly treating me like another member of your family and supporting me in my endeavors, but especially to Susan for being a listening ear and constant motivator.

To my partner, Adam. I truly cannot thank you enough for your support from being the first person I call when good and bad things happen, for being forever an optimist to counterbalance my pessimistic anxious side, and for listening to me say the same things over and over again when I am stressed. I can't wait for more adventures with you, and also congrats again on defending your own PhD yesterday (at the time of writing this). Your ability to support me so well while going through your own PhD journey is admirable.

To end my acknowledgements, I would like to thank a few entities that made this research possible. First, the NIH National Heart, Lung, and Blood Institute for funding my PhD and this research under award number RO1HL127316. Without their generous support, this PhD would not have been possible. Second, the Diefenderfer endowment at Penn State for funding a year of this research and allowing me to explore entrepreneurship as part of my PhD. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH. And finally, to the public library system for existing, as it is where I did a large portion of my work during my degree.

Chapter 1

INTRODUCTION

Upwards of 50 million major surgeries are performed in the US each year with more than 14% of patients experiencing adverse events [1]. Importantly, researchers have shown that nearly half of these adverse events are *preventable* [1], attributing them to human performance deficiencies (HPD), or cognitive and mechanical errors in the execution of care from the clinician performing the procedure [2]. Despite nearly two decades of research on human error in healthcare [3], adverse events in surgery remain a significant cause of preventable injury, and death [4]. These errors have caused researchers to explore the critical training that occurs for these clinicians, the evaluation of their proficiency, and the translation of these skills to the clinical environment.

One procedure that has a high adverse event rate is Central Venous Catheter (CVC). CVC is a complex medical procedure performed in emergency in intensive care units or non-emergent in operating rooms for cardiac and other procedures, where a catheter gets inserted into a central vein site, most commonly the right internal jugular (IJ) vein, for quick and efficient administration of medication [5], with many physicians choosing to conduct IJCVC with the assistance of ultrasound [6] (US-IJCVC). While more than five million central lines are placed in the United States annually, the procedure is plagued with high infection and complication rates due to its complexity and proximity of the IJ to the carotid artery [7]–[9]. In addition, research has shown that surgeons who have inserted a central line less than fifty times are more than twice as likely to incur complications [7]. Because of the direct link between patient volume, years of experience, and patient outcomes, a clear need exists for evaluating and transforming US-IJCVC education in medicine to ensure safe and competent practice.

Historically, US-IJCVC training has followed the Halstedian method, as developed by surgeon Dr. William Stewart Halsted [10], the creator of the common phrase used in medical training, "see one, do one, teach one" [11]. This idea stems from the basic apprenticeship model that was the foundation of medical training, where doctors would observe a more experienced physician

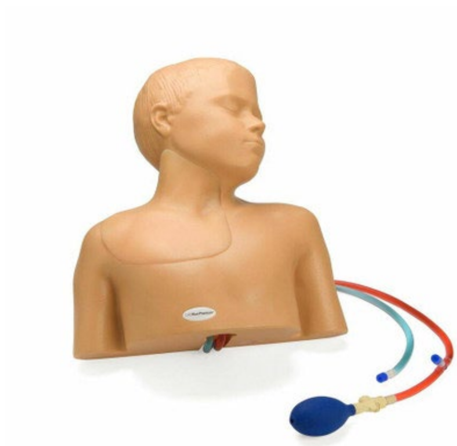

Figure 1: Blue Phantom Gen 11 Ultrasound Central Line Training Model [26]

conduct a procedure on a patient (see one), then learn by doing the procedure themselves (do one), and then walk another physician through the same procedure as they were walked through it before (teach one) [12], [13]. While "see one, do one, teach one" has been used in medical training for decades, critics argue that it should not remain the preferred method of training for new physicians because it puts patients at unnecessary risk [13]. Instead, simulation-based training (SBT) has facilitated a paradigm shift of medical training from "see one, do one, teach one" to "see one, *simulate many*, do one *competently*, teach *everyone*" [14]. SBT provides experiential, hands-on learning without putting patients at risk by exposing them to unskilled trainees [15]–[18]. SBT has been applied to a variety of procedures in medical education from laparoscopic surgery [19]–[21] to childbirth [22], [23] to CVC [24], [25]. A commonly used SBT method for US-IJCVC is a manikin trainer that includes a hand-pump to facilitate arterial pulse, false vein and arterial channels, a replaceable neck for the insertion site, and ultrasound guidance, **see Figure 1** [26]**.**

While these manikin simulators have been highly integrated into SBT for many procedures, including US-IJCVC, for their ability to provide risk-free, hands-on training to residents, they are static, represent only one patient anatomy and require an instructor to be present to provide performance feedback to the learner [25], [27]. Patients have a wide range of anatomical features [28] making practice on more than one patient anatomy important. On top of this, restricted clinical hours for resident trainees [29] emphasizes their need for practice with real-time feedback at any time, not just at the availability of trained instructors, demonstrating a clear need to both improve and validate other methods of CVC SBT.

In response to these US-IJCVC SBT deficits with manikin trainers, the Dynamic Haptic Robotic Trainer (DHRT) was developed by an interdisciplinary team of Penn State researchers [30], see **Figure 2**. The DHRT provides users with the ability to learn US-IJCVC on *diverse* patient anatomies and receive real-time, personalized performance feedback through a graphical user interface (GUI) [31]. Specifically, the DHRT generates a simulated ultrasound image that changes in response to a needle and mock US-probe movement, and can represent 17 different patient cases. The responsive haptic robot, Geomagic Touch X, with a custom designed and fabricated retractable syringe end effector, provides a realistic haptic feel of inserting a CVC into different patients [30]. Specifically, the GUI assigns the user diverse patient scenarios that vary the IJ and carotid artery depths, locations in relation to the US plan and each other, and diameters and wall thicknesses in accordance with realistic anatomical variations of patients that impact central line insertion [28]. In this way, the DHRT provides detailed training and feedback on key ***mechanical skills*** needed for US-IJCVC [30].

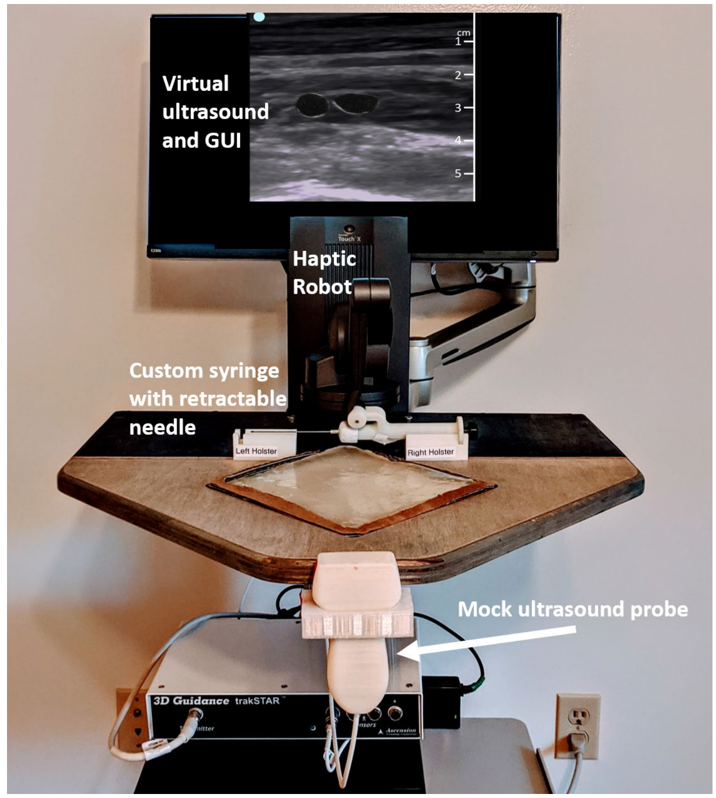

Figure 2: The original DHRT system designed by Penn State Researchers with labeled parts

While the DHRT provides critical training on some mechanical skills for US-IJCVC, it focuses solely on the initial mechanical steps of the procedure before and up to "achieving venous access", but none of the remaining steps past this. See table 1 for a breakdown of the steps required for the

Table 1: Breakdown of mechanical and procedural steps for CVC and which simulators cover them

	Main Steps of CVC	*Manikin*	*DHRT*	*DHRT+*
Procedural	Verbalize consent, universal precautions, and time out	–	–	✓
	Preparing catheter kit: flushing catheter and checking equipment	–	–	✓
	Maintaining sterile technique	–	–	–
	Selecting site for insertion	–	✓	–
	Injecting local anesthesia	–	–	✓
	Select correct ultrasound probe and use correct orientation	–	✓	–
Procedural + Mechanical	Obtaining clear image of target vessels using ultrasound	✓	✓	–
Procedural	Correctly distinguishing between the vein and the artery	✓	✓	–
Mechanical	Inserting introducer needle at 35-45°	–	✓	✓
	Locating the needle's position on the ultrasound	–	✓	–
	Advancing the introducer needle	✓	✓	✓
	Achieving venous access	✓	✓	✓
	Confirming vessel entry with needle aspiration	✓	✓	–
	Removing syringe while occluding hub	✓	–	✓
	Inserting guidewire into needle and advances without resistance	✓	–	✓
	Maintaining control of the guidewire	–	–	✓
	Removing introducer needle	✓	–	✓
	Using scalpel to make skin incision	✓	–	✓
	Inserting and removes dilator	✓	–	✓
	Passing catheter into vessel and removes wire	✓	–	✓
	Inserting catheter to correct distance (14-17cm)	–	–	✓
	Aspirating blood through the catheter	✓	–	✓
	Securing catheter into place with suture and dressing	–	–	✓
Procedural	Placing order for an X-ray and monitoring catheter	–	–	✓

procedure as defined by the New England Journal of Medicine [32], and which steps the manikin trainers and the DHRT cover. Manikin trainers also only focus on these main mechanical steps up to "achieving venous access" and also do not provide comprehensive training or automated feedback on the remainder of the steps [25], [27]. In CVC, trainees often struggle with both mechanical and procedural steps on all parts of the procedure, not just needle insertion [33], [34], indicating a need for a more robust, comprehensive education on CVC. Additionally, the gender diversity of physicians going to medical school and in resident training is continuously increasing [35], and SBT must be assessed to ensure that training is not biased, as previous research has indicated gender differences in simulator performance and learning [36].

While SBT is regarded as a useful and necessary training tool for medical residents, a common criticism of this method is the way that simulators are validated for training [37], [38]. To validate a simulator means to verify either that it looks like what it is trying to represent (face validity), is measuring what it is intended to be measuring (construct validity), is covering the entirety of information needed to properly convey what it is trying to train (content validity), or it is able to accurately project how a trainee will perform on a real-case based on how they perform on the simulator (predictive validity) [39], [40]. While much attention is put on the first three types of validity, less research has been done on predictive validity, or understanding and measuring the transfer of skills from the simulator to the actual clinical environment. When considering the DHRT from inception and development of the design through validation and use for training medical residents, it is useful to apply a systems-thinking approach. Systems thinking is approaching problems from all interconnected parts and the complex way that they fit together to create one whole part, or system [41] rather than focusing on the individual parts themselves[42]. This type of thinking has been applied to other aspects of healthcare, including for quality management [43], training [44], and medical education [45], though less research has been done on applying it to medical simulation.

Finally, while the DHRT has been shown to improve pre- to post-training US-IJCVC mechanical-skill performance and increase resident US-IJCVC self-efficacy compared to traditional manikin-training [46], [47], the system needs to be assessed for gender differences and modified to address the significant needs for training that have been identified by the Association for Surgical Education (ASE) Simulation Committee [48]. Two of these significant needs are: **1.** Need for validated training on all aspects of surgical performance, and **2.** Need for effective, evidence-based curricula that include objectives, assessment tools, and feedback mechanisms that are beneficial to all trainees. Fulfilling these needs are necessary to enhance the utility of SBT (including the DHRT) for improving patient outcomes and for the widespread adoption of SBT in medical education for US-IJCVC and other procedures.

1.1 Dissertation Goals

The current dissertation was developed to **transform CVC education through assessment of learning effectiveness and efficiency and validation of a novel medical simulator for Ultrasound Guided Central Venous Catheterization training.** Specifically, the goals of this dissertation were to (1) determine the efficiency and effectiveness of sequential learning on the DHRT through assessing initial skill gain, number of trials required, and learning curves, (2) assess if and how the gender-confidence gap and the Dunning-Kruger effect exist for CVC self-efficacy and performance on the DHRT, (3) evaluate the impact of comprehensive training with an advanced CVC simulator (DHRT+) on self-efficacy and checklist performance, and finally (4) evaluate the DHRT+ for construct and predictive validity utilizing eye tracking in the operating room and on the simulator with novices and experts. The remainder of this chapter provides a review on CVC and its complications, simulation-based training (SBT) and assessment methods used in CVC education, and simulator validation methods. The four manuscripts integrated into this dissertation are summarized below.

1.2 Central Venous Catheterization Procedure and Complications

Central Venous Catheterization (CVC), is a complex medical procedure where a catheter, also referred to as a central line, gets inserted into a central vein site for quick and efficient administration of medication [5]. This procedure is most commonly conducted with the use of ultrasound assistance [49]. The catheter can be inserted into one of three central vein sites, the femoral vein in the leg, the subclavian vein in the lower neck/collarbone, or the internal jugular vein (IJV) in the neck with the right IJV being the most common site of insertion due to ease of access and lower risk of complication [5],

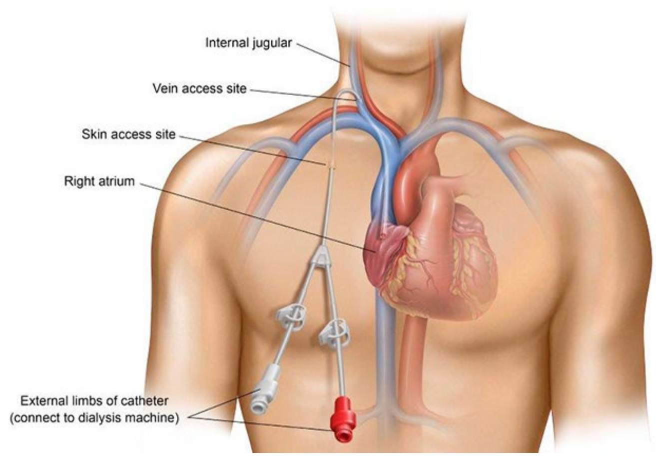

Figure 3: Catheter in the Internal Jugular Vein (IJV) (Central Venous Catheter Placement, n.d.). The IJV is the blue vessel and the red vessel next to it is the carotid artery, an important distinction in CVC.

[7], [50] (US-IJCVC). An example of a catheter inserted into the IJV can be seen in **Figure 3** [51]**.** When a central line is inserted into the IJV, the end of the catheter sits directly at the entrance to the heart. From the diagram, the proximity of the IJV (blue vessel) to the carotid artery (red vessel) is evident. Accidental needle insertion into the carotid artery is dangerous, but a common mistake made by new trainees learning CVC [34]. Therefore, it is important for trainees to receive an in-depth CVC training including how to distinguish these anatomical markers.

Additionally, US-IJCVC requires both mechanical and procedural competence [32], review table 1. These skills range from procedural skills such as the use of proper sterile technique and differentiating vessels using ultrasound, to mechanical skills such as appropriately accessing the IJV with an introducer needle directed by ultrasound. These skills also require a broad spectrum of instruments that are to be used sometimes with one hand at the same time as another tool, like using the ultrasound with the left hand to guide the needle with the right hand. CVC also requires medications that are to be applied at specific times in the procedure, like lidocaine used to numb the skin prior to any needle sticks. To understand the full complexity of the procedure and importance of understanding and tracking medical tool usage, an example of the medical tray used for CVC can be seen in **Figure 4.** The broad range of mechanical and procedural steps needed to

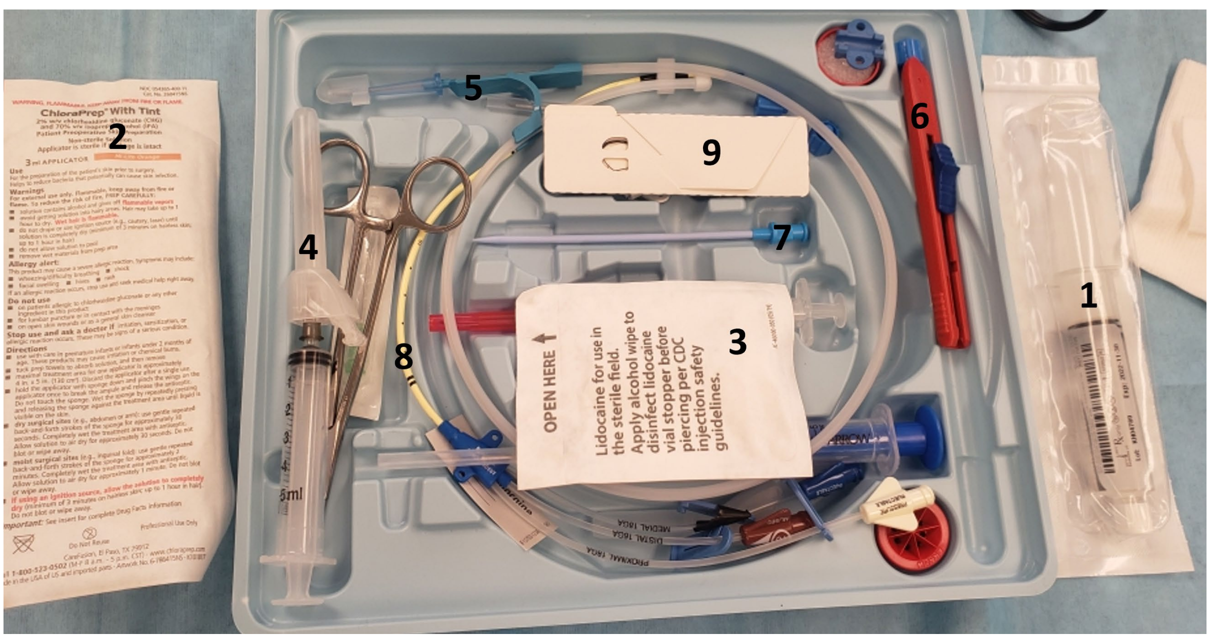

Figure 4: Central Line Tray including – 1) saline for flushing, 2) chlorohexidine for prepping sterile field, 3) lidocaine to numb the patient, 4) introducer needle, 5) guidewire, 6) scalpel, 7) dilator, 8) catheter, 9) thread for suturing

perform CVC along with the breadth of tools have led to challenges with training residents in this area when they have to apply multiple new skills at once.

While millions of central lines are placed in the United States annually, the procedure is plagued with high infection and complication rates [7]–[9], [52]. CVC complications can be broken down into three main categories: thrombotic, infectious, or mechanical [7], see **table 2**. Thrombotic complications are those related to thrombosis, or blood clots while infectious complications are those related to infection at the catheter site [7], [53]. Infectious complications are most often caused by clinical errors in the *procedural* steps of CVC, see Table 1. *Mechanical* complications are those related to operator error surrounding the use of CVC tools within the body that increase the complexity of the procedure, such as arterial puncture – inserting the needle into the carotid artery instead of the vein during initial venipuncture [7], or puncturing through the backwall of the vein [54]. One of the causes of this mechanical error is that the anatomical structure of the IJV – e.g. the location and size of the IJV and carotid artery changes from patient to patient. As such, the clinician needs to have a strong understanding of how to differentiate between the two vessels – e.g. the carotid is pulsatile and less compressible than the IJV, both of which can be observed on the ultrasound [6]. Another mechanical complication is caused by a lack of understanding of how to properly track medical tools during catheter insertion leading to the potential loss of the guidewire in the body [55]. Some of the most dangerous complications of CVC are those that lead to infection including not replacing a catheter after an arrythmia or

Table 2: Breakdown of the common types of complications associated with CVC and example causes [7]–[9]

Type of complication	Definition	Example cause
Thrombotic	Partial or complete blockage of a vessel via a blood clot	Multiple insertion attempts into the IJV [53]
Infectious	Infection or illness at the site of the catheter or in the bloodstream [7]	Breach of sterile field during procedure
Mechanical	Procedural complexity caused by issues with tool use during insertion	Puncturing the carotid artery instead of the jugular vein

perforating the walls of the heart; the consequences of these complications are longer hospitalization, higher treatment costs, and in the most severe cases, death [52].

One of the main drivers of these high complication rates is the experience of the physician performing the procedure [9], [56]. In fact, research has shown that surgeons who have inserted central lines less than fifty times are more than twice as likely to incur complications [7]. Because of the direct link to training and performance a need exists for transforming CVC education to better prepare residents to practice on patients.

1.3 Simulation based training (SBT)

One way to increase competence in medical procedures prior to clinical exposure is through simulation-based training (SBT). SBT has been shown to be valuable in medical education and is used in several fields including nursing [57], neurosurgery [58], and orthopedic surgery [59] to name a few. These patient simulators provide a low-stress, no-risk method for surgical training, and have the capability to transform the medical curriculum from a "see one, teach one, do one" model to a "see one, *simulate many*, do one *competently*, and teach *everyone*" model [14]. This is important because most procedures, including CVC, require a combination of both cognitive and motor skills and prior research has indicated that for effective acquisition of motor skills, repetitive training is needed [60].

1.3.1 Current State of SBT for CVC

While SBT is not formally required in most residency programs in the United States, it is recommended as a way to further didactic lectures and improve hands-on skills [61]. US-IJCVC SBT for medical residents and students is not standardized between institutions, so a wide range of simulators have been deployed in US-IJCVC education from chicken tissue [62] to vinyl phantom models [63] to partial body manikins [64]. These simulators are used to teach medical residents how to use ultrasound, identify anatomical structures, and perform venipuncture. Many of simulators have several instructional design features in common including repetitive practice, cognitive interactivity, and clinical variation [65]. While chicken tissue models have been found to have high ultrasound quality and haptic tissue feel [62], manikin trainers are preferable because they more closely resemble human anatomical structures and do not have to be specially stored. Currently, the most widely used simulators for CVC are partial body manikin trainers, see **Figure 5** for CAE Healthcare's Blue Phantom [64]. These manikin systems include an upper torso and neck model with vascular anatomy, a hand pump for arterial pulsing, and the ability to use ultrasound while going through CVC on the model. While these simulators are useful for learning the steps of CVC, they provide minimal automated feedback to the user. Specifically, the feedback provided from the manikin alone is limited to what color liquid is retrieved from the introducer needle *if* access to the vessel is attained: blue liquid represents access to the IJV while red access represents arterial access. Instead, these manikin models require a trained proctor to oversee the training, watch trainees, and provide feedback on performance along the way – a costly and inefficient method for providing feedback. This is problematic because numerous studies have pointed to the limited utility of SBT that fail to provide performance feedback. To put it in the researcher's own words "no feedback, no learning" [66]. In addition, these simulators are limited in that they only simulate one patient anatomy, they lack accurate force profiles for different types of skin and bodies, and they degrade

Figure 5: CAE Healthcare Blue Phantom [64]

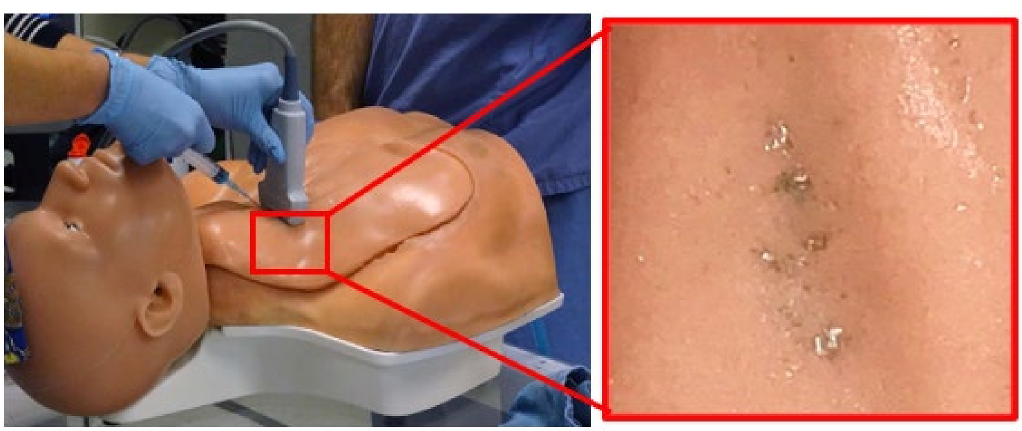

Figure 6: Simulab CentralLineMan [67] with red box indicating degradation over time that provides visual external evidence to residents on where to insert introducer needle

over time, providing visual evidence on where the introducer needle should be placed, see **Figure 6** for an example with Simulab's CentraLineMan [67]. As such, it is not surprising that prior research has shown that after undergoing manikin-based SBT and a few clinical rotations, residents are still not comfortable performing even the most common bedside procedures, including CVC, on their own [68].

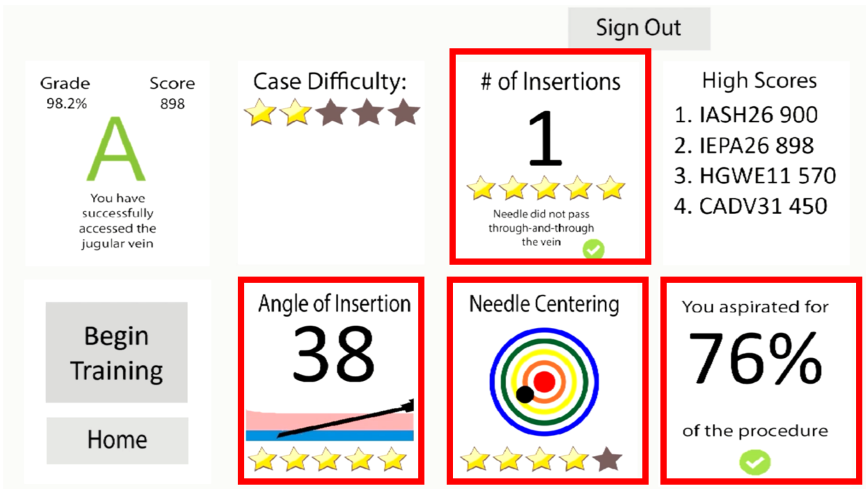

Figure 7: Summary Screen for DHRT with metrics of interest boxed in red

In order to overcome the deficits of existing training methods, the DHRT system, presented in Figure 2 was developed [69]. The DHRT includes a personalized learning interface with the goal of individualizing US-IJCVC training and removing the need for a trained proctor to give feedback [31], see **Figure 7**. Specifically, to provide automated feedback, the DHRT system collects performance data on the angle of the needle during insertion (degrees), the distance from the needle to the center of the vein (cm), number of attempts (whole number), percent of time spent aspirating and if the trainee punctures through the backwall of the vein [31], [46]. Higher performance in each of these factors contributes to higher chance of successful insertion and lower risk of complications [32], [46], [52]. In addition, the difficulty of the case is presented. Case difficulty is based on variables that impact how hard it would be for a doctor to insert a central line in a real patient case such as vein depth, skin thickness, vein diameter, and closeness between the IJV and the carotid artery [46]. This data is then aggregated into a final score.

The DHRT has been demonstrated to be an improvement over existing training methods due to its ability to provide training on diverse patient anatomies, its elimination of durability issues of manikin trainers, and its integration of individualized, non-proctor-based feedback [46], [70].

Additionally, training on the DHRT is shown to move trainee's performance toward expert performance over multiple training sessions; this is promising because it indicates notable procedural competence increases with the use of the system, see **Figure 8** [46]**.**

While there have been large advances in CVC SBT in recent years, there remain key areas for improvement. First, existing SBT methods focus purely on the steps involved in US-IJCVC for related to inserting the needle, see table 1 for comparison. While these skills are crucial for reducing mechanical complications such as arterial puncture [5] or puncturing the backwall of the vein [54] they do not work to reduce the high number of infectious and other mechanical complications associated with US-IJCVC that can occur at other points in the procedure such as losing control of the guidewire [55]. Moreover, focusing only on the needle insertion portions of US-IJCVC related to identifying and accessing the IJV does not allow residents to learn or practice usage with the other required tools in the CVC kit, and lacks exposure to the full breadth of knowledge required to safely conduct the procedure [32], [71]. As such, there is a need for a US-IJCVC simulator with more comprehensive learning, that includes a wider breadth of US-IJCVC skills.

In addition, US-IJCVC is a complex skill involving both cognitive and motor skill development; however, both the DHRT and manikin trainers require residents to apply all facets of their knowledge on their first insertion trial with limited feedback on how to improve. This

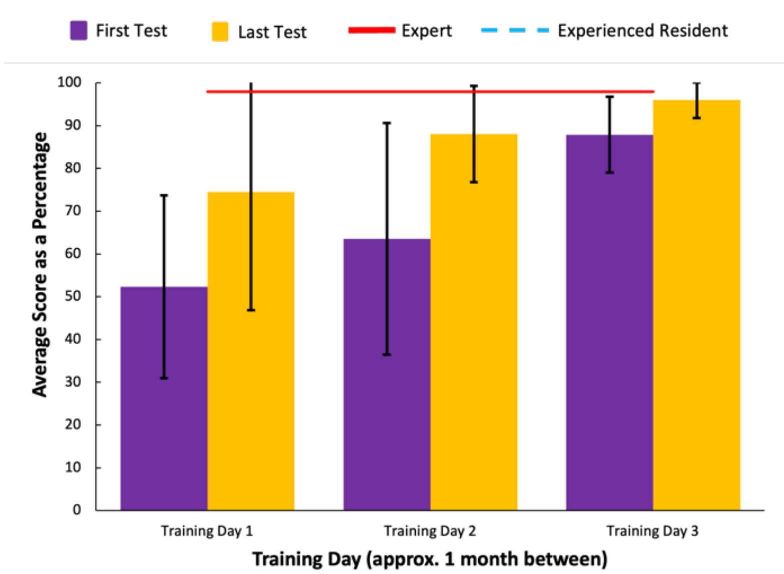

Figure 8: Graph taken from [46] showing novice performance over time compared to an expert standard

learning approach is in stark opposition to research on motor learning, which shows that motor skills are best learned in *stages*, and that complex tasks should be broken-down to increase effectiveness [60]. Sequential learning is the act of breaking down learning into small, incremental steps [72]. This concept was first introduced as a potential to include in SBT in 1989, when surgeons were exploring simulation and the “see one, do one, teach one” approach wondering if surgical motor skills would be best learned in “graduated sequences” [12]. In addition, in more recent years medical students have been shown to prefer learning broken down into sequences rather than presented at once [73], and research has shown that complex skills should be broken down for effective learning [60]. Along with sequences and sequential learning, incorporating learning through multimedia, the cognitive theory surrounding how perceive and learn information through words and images [74], may also be a useful tool for improving CVC SBT. This is because prior research has applied this method to simulation through incorporating videos or interactivity into their simulation rather than reading and observation [74] and found that training was more effective for retention and more engaging for learners. However, one of the disadvantages of medical simulation training is that participants generally approach simulators differently than they would real-life scenarios [75], a phenomena that is likely made worse when the simulator is poorly designed. As such, medical simulators must be catered to adult learners through ideas like controlling the sequence of tasks and offering guidance throughout the simulation [76].

1.3.2 Assessment of learning in SBT

To ensure that trainees are gaining skills and learning from SBT, there are several methods that have been employed to verify learning. Three of these methods that have been specifically applied in US-IJCVC SBT include self-assessment, checklist performance and learning curves. Self-assessment focuses on the trainee evaluating their own performance [77] or confidence with a procedure or skill [78]. Performance checklists utilize mastery-based learning and are graded by expert observers to determine when a trainee is competent and/or proficient enough in a procedure or skill to see patients [79]. Learning curves focus on plotting performance over time and verifying that there is a significant change, usually assessing for skill increase toward a pre-defined mastery level [80].

1.3.2.1 Proficiency Checklists

Performance checklists are usually graded by an expert observer and verify that the trainee is able to conduct all steps of a procedure proficiently before allowing them to practice on patients [81]. Checklists are used in many procedures [81], including US-IJCVC [82], and are sometimes referred to as Verification of Proficiency checklists. There are several variations of checklists used for US-IJCVC because of the lack of standardization of CVC curriculum. For example, one institution uses a 29-item checklist [83], while the DHRT and manikin training at another institution use a 24-item checklist [84]. This checklist used for the DHRT focuses on 24 actions that are required to successfully insert a central line from beginning to end, see table 1, and includes observer ratings for economy of time and motion, or how efficient hand and tool movements were during the procedure, and the number of insertion attempts, or how many times the person inserted the needle before achieving access to the vessel. For the DHRT, the US-IJCVC checklist has been employed to compare resident performance on a manikin trainer, finding that the DHRT is as effective as manikin training when it comes to mastery-based learning to a verification of proficiency [47]. While checklists can be useful to determine proficiency, there is some debate about subjectivity of observer ratings and how well they accurately measure competence [85], and as such they should be used in conjunction with other assessment methods.

1.3.2.2 Self-Assessment

Self-assessment is used in medical education to determine how well trainees are gaining skills from SBT [86]. Self-assessment in medical education generally means either self-assessment of performance [87], or confidence . Self-efficacy, or task-specific confidence [88], has been utilized in many procedures to determine effectiveness of and motivation for learning [89], including in SBT for CVC on the DHRT [70]. Self-efficacy in particular is a useful measure because of the confidence-competence relationship observed in learning medical procedures [90], [91]. This relationship posits that confidence and competence tend to grow together in new trainees. For example, as a trainee gains procedural proficiency, they should also be gaining confidence, though this relationship is not definite and there are some drawbacks to using it.

For example, there is an phenomena that has been observed with new trainees referred to as the Dunning-Kruger Effect [92]. In the late 1990's, Dunning and Kruger did a study on people's ability to recognize their own performance and found that people with little experience tend to grossly over estimate their ability to perform [92]. In medical education, this manifests as initial exposure to a skill giving trainees an unwarranted jump in confidence that can only be tempered with additional practice allowing them to realized how unskilled they truly are relative to experts [93]. In addition to the Dunning-Kruger effect, another potential problem with using self-assessment as the sole method for evaluating learning in medical education is the gender-confidence gap [94]. The gender-confidence gap in medical education posits that even when men and women perform equally, women have lower confidence in their own abilities then their men counterparts [95]. This gender-confidence gap is evident in self-efficacy [96], self-assessment of performance [97], and overall well-being of medical residents [96], [98]. There is limited research exploring both the Dunning-Kruger effect or gender differences and confidence in CVC, and the DHRT has yet to be assessed for confidence differences between genders before after training.

1.3.2.3 Learning Curves

Learning curves are graphed representations of the acquisition of knowledge over a specific interval [99]. More specifically in the application of medicine, learning curves are the tracing of performance and its improvement over time as trainees learn and build new skills [100], [101]. The learning curve is often plotted and analyzed for the existence of learning in simulation, where significant curves indicate significant learning [102], [103]. Insignificant learning curves can indicate two things [101]: either the lack of learning [66] or the plateau of learning because mastery level performance has already been reached [104]. Learning curves in medical skill acquisition have been estimated using a variety of models including linear, logarithmic, and power analyses with benefits to each method. Linear learning curves are useful for giving a basic understanding of knowledge progression [102]. Logarithmic learning curves are useful because they are able represent initial skill gain which tends to have a large slope that levels over time [101], [105]. Finally, power curves are useful because they are able to represent diminishing returns, or plateauing due to mastery level being reached [102], but are only mathematically viable if there are no negative values or zero values. Because they can identify when mastery level performance has been reached, learning curves can also differentiate novices and expert performers [106]. As such, plotting the learning curve for new trainees is useful to see how they are gaining skills toward a predefined level of mastery [101] and rates of the curve can vary depending on the difficulty level of the task being learned [80]. More complex procedures are noted to have more gradual learning curves because they are learned via small improvements over time, whereas simpler procedures will have steeper learning curves because the skills can be acquired faster [100]. SBT aims to minimize individual learning curves to improve patient safety and resident learning [107]. Finally, learning curves have also been assessed in US-IJCVC [108] and the DHRT [109]. More specifically, the DHRT has been shown to differentiate novices and experts based on learning curves, and significantly improved the performance of novices during training based on curve significance [109]. Using combinations of learning curves, proficiency checklists, self-assessment, and quantitative performance can give a strong indication of the presence and effectiveness of learning from SBT.

1.4 Simulator validation methods

SBT is only valid if the simulator elicits the same behavior as the real (clinical) environment [23], [110]–[112]. This is particularly important in medical education when the transfer of skills from simulator to patient has significant patient health implications. As such, for a simulator to be impactful in medical education, it must be able to develop residents' key skills that transfer to the clinic. However, researchers have identified that there is a lack of evidence for proper validation of simulators [113], [114], along with no standardized procedure or guidelines for how simulators should be validated for skill transfer [37]. As such, there are many types of validation that have been applied in SBT.

1.4.1 Types of Validity in SBT

Validity in SBT is commonly recognized through face validity, construct validity, content validity, and predictive validity. Face validity is how real a simulation looks, construct validity is how accurately a simulation measures the simulated task, content validity is how well a simulation covers the subject matter being simulated, and predictive validity is how well a simulator can project future performance. A breakdown of these types of validity with examples can be seen in **Table 3.**

Table 3: Breakdown of the types of validity in simulated learning [39]

Type of Validity	Definition	Example
Face	How real a simulation looks and feels	A full-body simulator with accurate visual fidelity, anatomy, and skin
Construct	If the simulator truly measures the concept of interest	A simulator that is able to differentiate between expertise levels
Content	The extent that a simulator covers the material of interest	A simulator that covers all aspects of a specific procedure
Predictive	If a simulator is able to reliably predict future performance	A simulator that is able to accurately reflect performance in the clinic or skill performance over time

While face validity is the easiest to assess because it can be done through survey [115], construct validity is a more objective measure than face validity. The main purpose of construct validity is to identify the extent to which the simulation accurately represents the real task and measures performance [116]. Construct validity has been shown to be a crucial piece for achieving transfer of learning, and a simulator with good construct validity is sensitive to variations in performance between individuals (e.g. novice and expert) [116], [117]. Content validity is often assessed with construct and face validity [118], also generally via survey or other measures [119]. Predictive validity is the hardest to measure and is used less often, however, it is closely related to construct validity. Generally, if construct validity is higher, a simulator is more likely to have predictive validity [39].

Many validation studies for surgical simulators focus solely on construct validity [120]. For example, one study compared the pass rates of CVC proficiency testing of residents who underwent a lecture-based didactic training to those who underwent both the didactic training and a simulation training. This study indicated that over 25% of the people without simulation training failed the proficiency testing on their first attempt vs the 3% failure rate seen for those who attended the simulator training [121]. These results indicate that residents more adequately learned and understood the steps of the procedure when exposed to it via simulation. Similarly, a study on the effectiveness on the DHRT system indicated that the use of the simulator increased resident performance relative to that of an experienced doctor, who elicited no learning curve due to expertise [46]. A motion analysis study on the DHRT system also identified that the simulator was able to distinguish expert and novice movement patterns, and that novices approached expert performance throughout training [109], examples of construct validity.

As defined in **Table 3**, predictive validity refers to how well a simulator is able to determine a person's performance in the true environment, and is considered the hardest type of validity to measure accurately [118]. Predictive validity is generally measured through skill-retention over time, proficiency checklists, or self-assess knowledge or confidence gains after simulation training [122], [123]. In this way, researchers have shown that CVC SBT with a manikin trainer can lead to long-term skill retention [124]. In addition, research on the DHRT identified simulation training as an effective means of improving trainee CVC skill and confidence [70]. This is relevant because prior work indicates that self-efficacy and performance are correlated [125], and that clinician procedural confidence could be used as a metric to determine the educational effectiveness of a simulator in mixed-fidelity simulation training [126]. While useful, these measures of predictive validity can be time consuming [124] and subjective [122], [123]. Another method of assessing predictive validity is the transfer of skills, yet few studies report the direct transfer of skills from simulators to a clinical environment [38] because it can be harder to measure. The validation methods here largely focus on the validation of simulators on resident mechanical skills and limited studies focus on actual clinical transfer or procedural skill gains, indicating a need for further validation of these items.

1.4.2 Eye Tracking as a Validation Method

One underutilized validation method for simulators is eye tracking. The idea behind tracing human eye movement goes back to the late 1800s when researchers began using mirrors to observe people's eyes during a task [127]. The idea was that eye movement and mental processes were correlated, and tracing eye movements could give a better understanding of cognitive function [127]. After hundreds of years of innovation, modern eye trackers are devices that are used to monitor and record the gaze patterns of a person [128]. This type of research is useful because it can give an understanding into how a person's cognitive process flows and also determine how to better help them learn [129], [130]. Eye tracking works by monitoring when the eye is moving and when it is still [128], [129]; this stillness is referred to as a fixation [129]. For fixations, the main metrics of interest are generally fixation duration, or how long a person was fixating, and fixation count, or how many times they fixated [131]. These fixations can be measured overall for a given task, referred to as scene-independent fixations [132], or on specific areas of interest (AOI), referred to as scene-dependent fixations [21], [132]. In addition to fixation, eye tracking research also often refers to saccades, the rapid movement of eyes from point to point, which gives an understanding on the relationship between fixations [127]. Another common measure in eye tracking is time to first fixation, or how long from the beginning of the provided stimulus to the first fixation on the pre-defined AOIs [130]. Researchers utilizing eye tracking are often also interested in the scan path of a person's eyes, also referred to as the fixation sequence, which gives an understanding of the flow of cognition during a given task [130], [133] by tracing fixations over time.

In medicine, eye tracking has been used in a variety of ways such as understanding how critical care nurses to divide their attention throughout a shift in the ICU [134]. In this study, researchers identified areas of interest and focused on fixation time on these areas. Another study on nursing utilized eye-tracking to better understand workload and mental stress on nurses throughout a regular shift [135], focusing on fixations and saccades. In addition to nursing, eye-tracking has been applied to medical diagnostics, vision quality in children, medical treatment, skills assessment, and expertise distinction [133], [136]. Most often in medicine, eye tracking has been employed to understand expertise differences in perception during visual processing [137].

For the application of *medical education*, eye tracking has been used to understand learning over time, feedback and assessment, perception during diagnostic interpretation, and expertise level [131], [138]. In fact, a plethora of prior literature in laparoscopic surgery [20], [139]–[144], urology [145], and microsurgery [146], [147], indicates that eye tracking can be used to distinguish expert performance, which can be used to determine construct validity. Eye tracking has also been found to be useful for monitoring the learning curve of novice surgeons, a useful tool to measure learning over time [145], [148]. Because of this, researchers have argued that eye-tracking technologies should be implemented into surgical education in order to improve surgeon performance [149]–[151] and provide an objective assessment of surgical skill.

There are several occurrences in the recent literature of eye tracking being applied specifically to training for CVC. A group of researchers implemented eye tracking to analyze the effectiveness of the learning interface for the DHRT and determine if performance improvements on the simulator could be predicted with gaze pattern data [152]. In this study, areas of interest were defined and the eye tracking metrics used were variations of fixation duration and scan path. Similarly, another study using eye tracking and the DHRT indicated that the system is able to

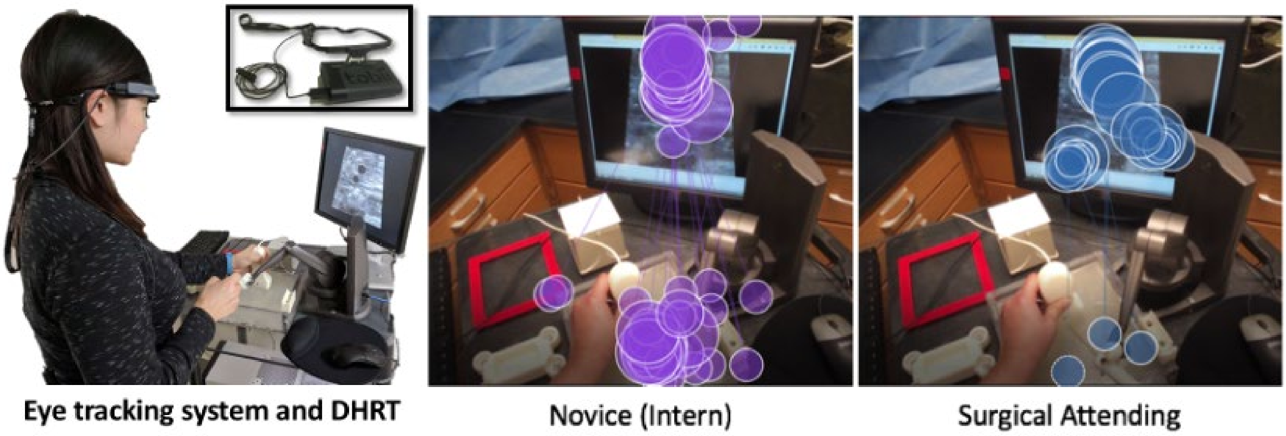

Figure 9: Eye tracking apparatus, areas of interest, and expert-distinguished results from previous study [153]

differentiate between novice and expert performance, as seen in **Figure 9** [153], demonstrating the construct validity of the system. Finally, a recent study exploring ultra-sound guided venipuncture indicated that experience level directly impacts gaze patterns during the procedure [154].

A common fallback with the application of eye-tracking in medical education is the complexity of analysis. When assessing scene-dependent metrics on specific areas of interest, there is no readily available software that can do so efficiently, timely, or automatically [155]. Each added area of interest adds complexity to the analysis [156] and each application of analysis software is specific, heavily dependent on the parameters of the defined task, and unable to be transferred between tasks and procedures [157]. However, scene-independent fixations do not require AOIs and are valid methods of understanding cognition and measuring gaze in the clinical and simulated environments [158].

Present day eye tracking is conducted non-invasively, and covertly through the examination of an individual's eye movements using either light-weight glasses, see **Figure 9**, or a fixed bar eye tracker that is attached to a computer or desktop monitor. Despite this, there is limited evidence into the predictive validity of these measures for monitoring expertise acquisition and utilizing eye-tracking to validate simulator performance to actual clinical performance in an operating room.

1.5 Summary of Areas for Investigation

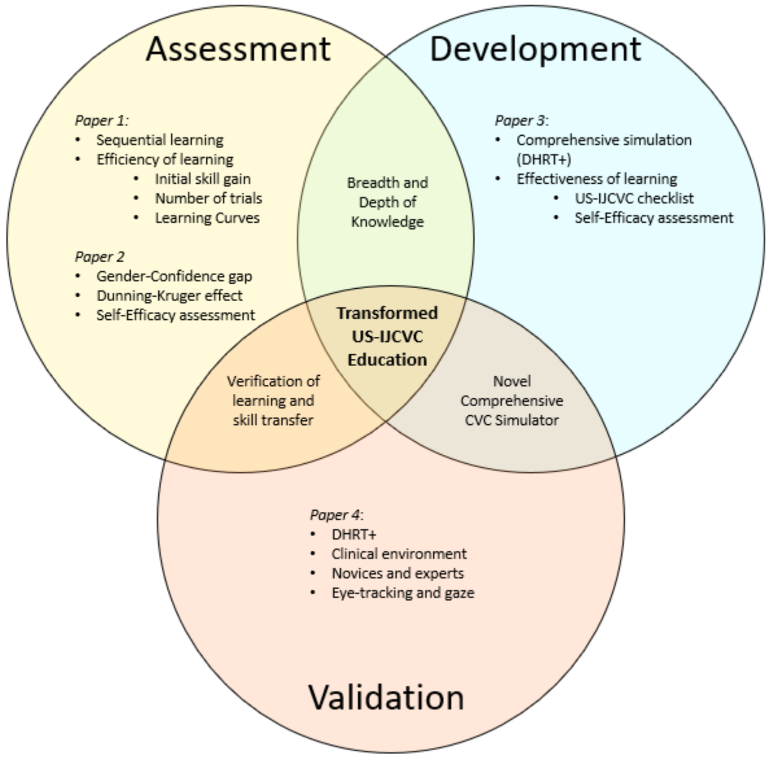

Figure 10: Summary of the areas of research to be investigated in this dissertation and how they contribute to the primary contribution of this dissertation

Previous work in SBT for US-IJCVC has focused on improving needle insertion skills for vessel access and determining the utility of new training methods compared to traditional methods including didactic lectures and manikin trainers. However, there is limited work in US-IJCVC on increasing simulator teaching efficiency through learning methods, evaluating gender and confidence as it relates to SBT, developing simulation to expand hands-on US-IJCVC education past vessel access, or validating the transferability of skills from the simulator to the clinical environment with eye-tracking. Therefore, the goal of this dissertation was to fill these gaps associated with US-IJCVC SBT through a systems-thinking focused approach, see **Figure 10** for a summary of investigation areas. More specifically, to contribute to the transformation of US-IJCVC education, this dissertation focuses on multiple parts of the simulation design process through assessing existing methods of US-IJCVC SBT for efficiency, developing and evaluating the impact of a more comprehensive US-IJCVC simulator on skills and self-efficacy during training, and validating the comprehensive simulator both to the operating room environment and for the distinction between expert and novice physicians.

1.6 Summary of Dissertation papers

This dissertation focuses on the improvement and validation of US-IJCVC education. An overview of the four articles presented in this dissertation can be found below.

Paper 1: Tapping into efficient learning: An exploration of the impact of sequential learning on skill gains and learning curves in central venous catheterization simulator training

The first objective of this dissertation was to explore efficient learning methods for CVC on the DHRT and determine if the implementation of sequential learning improves the efficiency and effectiveness of skill acquisition. Therefore, Chapter 2 of this dissertation presents a manuscript to be submitted to the Journal of Medical Education and Curricular Development. This study evaluates CVC DHRT performance metrics for cohorts trained on the DHRT with and without the implementation of sequential learning. Before 2022, to use the DHRT system, residents would participate in an online training and then watch one approximately eight-minute, non-interactive, instructional video outlining how to conduct CVC on the DHRT. For 2022 training, sequential learning was added to the DHRT system in the form of an interactive tutorial style walkthrough made up of eight videos and activity assessments highlighting seven key aspects of CVC as agreed on by three expert physicians. The DHRT system focuses on needle insertion to gain vessel access, so the walkthrough breaks down the mechanical components for this portion of the procedure. The walkthrough starts with basic skills including understanding how the patient is oriented with the table and their neck anatomy, and build on one another to the final step which is inserting the needle into the body and recognizing venous access. The performance of 59 residents who went through the sequential learning as part of their training was compared to that of 44 residents who did the DHRT training before sequential learning was implemented. All residents in the study participated in a pre-training online course. Results of this study showed that the introduction of sequential learning into the system gave residents a 3.58 times higher likelihood of successfully completing needle insertion on their first try without detrimental errors, such as puncturing the carotid artery by mistake or pushing the needle all the way through the vein. Results also indicated that sequential learning contributed to an overall reduction in the number of trials needed to reach mastery performance and learning curves compared to the nonsequential learning group, signifying an increase in learning efficiency and effectiveness.

Paper 2: Competence over confidence: gender-based differences in resident training for central venous catheterization

The second objective of this dissertation was to assess if gender-based differences exist in self-efficacy and performance for Central Venous Catheterization Training on the DHRT. As such, Chapter 3 of this dissertation presents a manuscript to be submitted to BMC Medical Education. The study in this paper evaluates the gender-confidence gap and the Dunning-Kruger effect in self-efficacy and performance on the DHRT system. Self-efficacy and self-assessment are widely integrated into medical training; however, new medical trainees struggle to accurately assess their performance, and women tend to rate themselves lower for clinical skills regardless if performance is equivalent to or better than their peers. In the central line training bootcamps where residents use the DHRT, they took a 14-item self-efficacy survey before and after training assessing their confidence on a 5-point Likert scale ranging from "not at all confident" to "extremely confident" in their ability for the skills and steps of the procedure, referred to as the Central Line Self-Efficacy (CLSE) survey. The skills asked about range from specific such as "using tactile feedback, identifying the correct vessel for puncture" to broad "conducting the entire procedure without any mistakes". The gender differences for 61 women and 112 men were assessed for self-efficacy both before and after training (gender-confidence gap), simulator performance based on the number of needle insertions, backwall punctures through and through the vein, and successful venipuncture without arterial puncture, and correlation between self-efficacy and performance (Dunning-Kruger effect). Results showed evidence of the gender-confidence gap, with women rating themselves significantly lower than men for nine of the 14 variables, despite no performance differences found between men and women. Results also indicate evidence of the Dunning-Kruger effect, indicating that neither men's nor women's self-efficacy was significantly correlated to their performance.

Paper 3: Evaluating the Effects of Comprehensive Simulation on Central Venous Catheterization Training: A Comparative Observational Study

The third objective of this dissertation was to evaluate the impact of the expansion of the DHRT into a comprehensive CVC simulator resident self-efficacy and performance. Therefore, Chapter 4 of this dissertation is of a manuscript to be submitted to the BMC Medical Education. The study in this paper evaluates efforts to fill the gaps in current CVC education by covering a more comprehensive and wider breadth of the procedure. Most training programs focus solely on vessel identification and access and neglect the remainders of the procedure which covers a plethora of other skills that can help to prevent mechanical complications that can arise after vessel access is achieved. An extension to the DHRT system was developed, the DHRT+, which when used in combination with the DHRT provides more comprehensive training on CVC by adding additional procedural skill steps and hands-on training for complex mechanical skills involving the full scope of medical tools required for the procedure. Fifty-nine residents trained with a combination of the DHRT and the DHRT+, referred to as comprehensive training, in 2022 and 72 residents trained on the DHRT alone in the previous year. All residents filled out a 19-item Central Line Self-Efficacy (CLSE) survey before and after training, and underwent a expert-observed skills assessment using a US-IJCVC checklist to test their CVC proficiency on procedural and mechanical skills. Results indicated that for two items on the US-IJCVC checklist including verbalizing consent and making an incision with the scalpel the comprehensive training group performed significantly better. These results were also found for two of the items for self-efficacy, using the proper equipment in the proper order and securing the catheter with suture. For all other items on both the CLSE survey and the US-IJCVC checklist, there were no significant differences between training groups.

Paper 4: Tracking success: Validating a novel central venous catheter trainer through eye gaze analysis in clinical and simulated environments

The final objective of this dissertation was to evaluate the integrated comprehensive system (DHRT+) for predictive and construct validity in CVC training through analyzing expert and novice gaze in the operating room and on the simulator. Chapter 5 of this dissertation presents a manuscript to be submitted to the Journal of Surgical Research. The dual part study in this paper explores predictive and construct validity for the DHRT+ system. To assess predictive validity, five expert physicians conducted CVC wearing a Tobii Pro Glasses 3 eye tracker in the operating room during nonemergent cardiac procedures. The same five experts then conducted CVC on the

DHRT+ simulator wearing a Tobii Pro Glasses 3 eye tracker. To assess construct validity, 12 novices also conducted CVC on the simulator while wearing a Tobii Pro Glasses 3 eye tracker. To account for variations in practice and align with previous research in the clinical environment, US-IJCVC was divided into the six most standard segments of the procedure and each segment was assessed separately for fixation metrics, specifically fixation count and fixation duration. Fixation metrics for the expert physicians were compared between the operating room and the simulator. Results indicated that there were only significant differences between the environments for one of the six segments, demonstrating predictive validity of the simulator. Similarly, fixation metrics between the novices and the experts on the simulator were compared. There were significant differences between novices and expert fixation counts and fixation durations with novices being significantly higher for three out of six segments of the procedure, indicating construct validity within those segments and partial construct validity overall.

1.7 Broader Impact and Contributions

This dissertation focuses on improving medical education through the assessment of current training methods and the development and validation of a new comprehensive training simulator for central venous catheterization. This dissertation provides evidence that implementing sequential learning into simulation training for complex procedures can increase the efficiency of learning, and decrease learning curves. This is beneficial to medical training as a whole because it indicates that there are ways to speed up the acquisition of skills gained from simulators to lessen the burden of training that residents have and to help them be prepared to work on patients sooner. Second, this dissertation indicates that there are gender differences in confidence gains from SBT training, and that both men and women residents struggle to self-assess their performance. This finding is important because it indicates that the way that feedback is presented during simulation should be re-evaluated to ensure that it is clear to learners where their strengths and deficits are. Third, findings in this dissertation indicate the utility of developing comprehensive US-IJCVC to include more than just the vessel access steps. Many training programs still rely solely on manikin trainings and checklist observation for residents to get exposure to CVC. Including the real tools that are used for the procedure, more of the mechanical and procedural steps, and automated real-time feedback without a proctor during the entire procedure can improve both retention of skills to a US-IJCVC checklist, and self-efficacy of trainees with their comfort on performing steps of the procedure. Finally, using eye tracking, this dissertation was able to prove both the partial construct validity and predictive validity of the new comprehensive training simulator, the DHRT+, for CVC. In addition, this portion of the dissertation also provides an example for how eye tracking studies can be conducted in the operating room, and how eye tracking can be utilized to objectively explore multiple types of validity and the transferability of skills. Overall, the content of this dissertation builds on what is known about SBT that can be expanded to other complex procedures normally trained by simulation, and provides evidence-backed methods for transforming SBT for CVC.

1.8 Document outline

A total of six chapters will be included in this document. CHAPTER 1 is the introduction outlining the scope of the research, the problem being solved, and an outline of the remainder of this document. CHAPTER 2 will outline the execution and completion of goal 1 of the dissertation, the impact of sequential learning on CVC skill acquisition and learning curves. CHAPTER 3 will outline the execution and completion of goal 2 of the dissertation, assessing gender differences in simulator training for CVC. CHAPTER 4 will outline the execution and completion of goal 3 of the dissertation, the impact of a CVC comprehensive simulator on resident self-efficacy and proficiency. CHAPTER 5 will outline the execution and completion of goal 4 of the dissertation, validating a CVC full procedural simulator through clinical and simulator-based eye gaze patterns**.** Finally, CHAPTER 6 will conclude this document and further explain the impact of this dissertation.

Chapter 2

TAPPING INTO EFFICIENT LEARNING: AN EXPLORATION OF THE IMPACT OF SEQUENTIAL LEARNING ON SKILL GAINS AND LEARNING CURVES IN CENTRAL VENOUS CATHETERIZATION SIMULATOR TRAINING

This paper was submitted to the journal of Medical Education and Curricular Development in February of 2024. This work is multiple authored by Haroula Tzamaras, Dailen Brown, Dr. Jason Moore, and Dr. Scarlett Miller. Haroula Tzamaras was the lead author on this paper. Dr. Jason Moore and Dr. Scarlett Miller helped advise this work. Dailen Brown helped with programming behind the development of the sequential learning modules.

2.1 Abstract

Medical residents are expected to learn how to perform many procedures in a short amount of time. Sequential learning, or learning in stages, is a method applied to complex motor skills to increase skill acquisition and retention, but has not been widely applied in simulation-based training (SBT). Central Venous Catheterization (CVC) is a complex medical procedure that could benefit from the implementation of sequential learning. CVC is typically taught with task trainers such as the Dynamic Haptic Robotic Trainer (DHRT) This study aims to determine the impact of sequential learning on skill gains and learning curves in CVC SBT by implementing a sequential learning walkthrough into the DHRT. 103 medical residents participated in CVC training in 2021 and 2022. One group (N=44) received training on the original DHRT system with one long video instruction while the other group (N=59) received training on the DHRT^{sequential} with interactive videos and assessment activities. All residents participated in online CVC training, pre- and post-training self-efficacy surveys, and received training and were quantitatively assessed on (e.g. first trial success rate, aspiration rate, distance to vein center) the DHRT or DHRT^{sequential} systems. Residents in the DHRT^{sequential} group exhibited a 3.58 times higher likelihood of successfully completing needle insertion on their first DHRT trial than those in the DHRT only group., and required significantly less trials to reach mastery level performance. Finally, the DHRT^{sequential} group has less learning curves compared to the DHRT only group. Implementing sequential learning into the DHRT

system had significant learning benefits in CVC training by increasing the efficiency of initial skill gain and reducing the learning curve by reaching higher performance in a shorter number of trials.

2.2 Introduction

Central Venous Catheterizations (CVC) is a commonly performed medical procedure that typically uses ultrasound guidance to insert a catheter into the internal jugular vein (US-IJCVC) [50], [159] to allow for quick medication delivery [9]. While this procedure is performed more than five million times annually in the US [5], [160], US-IJCVC is plagued with a high complication rate of 15% [7], including complications caused by mechanical errors like accidental puncture of the carotid artery [49] or puncturing through the backwall of the vein [54]. These errors are significant because they can cause complications such as bloodstream infection, stroke, or hemothorax, among others [9]. The number one driver of these error rates is the experience of the physician performing the procedure – a physician who has performed less than 50 lines is more than twice as likely to incur complications [7], [8]. Therefore, more practice by trainees before transitioning to patients could significantly decrease patient risk, reiterating the need to continuously improve CVC training methods [8], [56].

In order to be successful in US-IJCVC, a sequence of actions requiring both hands and specific motor skills must be followed including: (1) utilizing an ultrasound probe to identify the appropriate vessel, (2) distinguishing the carotid artery from the internal jugular vein in the ultrasound, (3) identifying where to insert the needle with respect to the ultrasound probe, (4) identifying where the needle tip is in the ultrasound image, (5) identifying when the needle has been appropriately centered in the vessel and (6) understanding how to aspirate the needle and verify when the vein has been accessed [32]. Current simulation-based training (SBT) [13], [14] in CVC education typically relies on residents to know and apply these critical steps in order without assessing individual skill mastery [161]. Skill mastery is an important concept in medical education [162], [163]. The idea of tailoring simulation training to achieving mastery performance has been seen in ventilator management [164], bronchoscopy [165], and pediatric cannulation [166], indicating that training to reach a predefined mastery performance level can improve skill acquisition and knowledge .

One of the most recent advancements in US-IJCVC training, the Dynamic Haptic Robotic Trainer (DHRT), see **Figure 11**, relies on comprehensive knowledge assessment rather than sequential learning. The DHRT system is advanced compared to traditional manikin simulators in that it provides risk-free practice with the ultrasound probe and the needle [167], but also allows diverse patient anatomies by changing the locations, depths, and sizes of the IJV and carotid artery [46]. The DHRT includes a mock ultrasound probe and simulated, reactive ultrasound image, a haptic robot, a specialized retractable needle that provides force to the trainee to simulate insertion, and a feedback screen that provides a personalized performance summary for learning [31], [168]. For each trial on the simulator, the DHRT tracks the aspiration rate, the number of needle insertions, needle centering in the vein, angle of insertion, and puncture through the backwall of the vein or of the carotid artery, and provides a total overall DHRT performance score and a post-trial performance summary screen [169]. The DHRT has been shown to be as effective as manikin training without the need of a trained preceptor [170], and has also been shown to significantly improve learning over time and to also identify learning curve changes as levels of mastery are

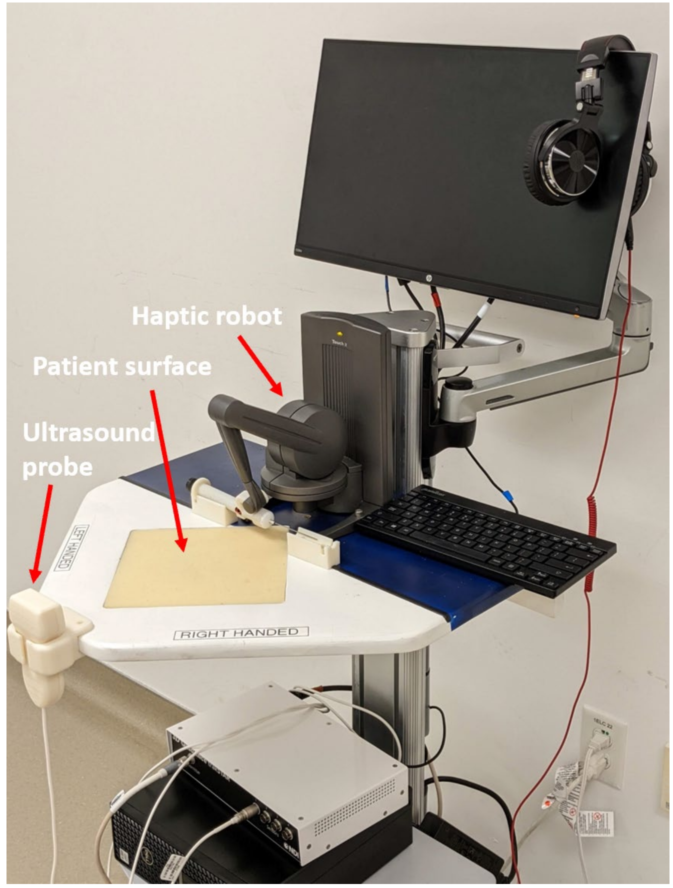

Figure 11: The DHRT system developed for US-IJCVC training

reached [109]. However, the DHRT requires residents to apply all facets of their knowledge on their first insertion trial without ensuring an understanding of steps. Additionally, the current training approach for the DHRT is for each resident to conduct six preset trials regardless of performance and level of proficiency reached [171]; however, literature indicates that medical education should be tailored to the individual's learning and performance needs [172] for optimal outcomes.

Assessing the effectiveness of skill mastery is essential in complex motor skills like CVC because prior work has shown that these tasks are best learned in stages rather than all at once [60]. In addition, prior work has shown that medical students prefer learning in a sequential style, or breaking down learning into small steps [72], [73], [173] with an incremental progression of steps [73], [174]. Prior work has also shown that sequential learning is useful for multiple learning styles [175]; however, some literature on sequential learning in SBT is mixed, with one study indicating that the implementation of sequential learning did not significantly impact learning outcomes of medical students for emergency skills after SBT [176]. Sequential learning has not been explored extensively in SBT for CVC, and research focused on implementing SBT into medical education at the residency level is limited.

Sequential learning has also been shown to increase skill gain and reduce learning curves, or plots that show the number of repetitions required for a trainee to reach a desired level of performance [177]. The theory of learning curves posits that learning improves with experience, and can be plotted over time [101]. Learning curves have been studied in SBT. For example, learning curves have been plotted in CVC training [105] and in laparoscopic surgery [178], indicating that with SBT learning can improve significantly over a training period. In addition, learning curves have also been shown to differentiate expertise based on performance changes in laparoscopic simulation [19], simulated thoracentesis [179], and simulated US-IJCVC [108]. Learning curves also have been found to indicate the efficiency of learning by determining when mastery performance is reached based on curve plateaus [104], [180]–[182]. Minimizing and eliminating learning curves of new physicians by pushing them to reach mastery performance faster through robust SBT is important to optimizing patient safety [107], and applying sequential learning to CVC SBT has the potential to significantly improve learning curves in CVC.

To expand the body of knowledge on sequential learning and learning curves in CVC SBT, the purpose of this study was to measure the efficiency and effectiveness of sequential learning in

CVC SBT through assessing initial skill gain, the impact of the number of trials, and learning curves on the DHRT.

2.2.1 DHRT^{sequential} Learning Development

The design of the original DHRT system included one 7:30 video that trainees watched at the end of an online training required to be completed prior to attending an in-person training session on the DHRT, and a 27s refresher video that was played at the start of the in person DHRT training. The 7:30 video outlined how to log into the DHRT, the different parts of the system, and how to use and aspirate the needle. The 27s refresher briefly re-outlined how to use the DHRT system. Sequential learning was integrated into the DHRT system by breaking down US-IJCVC into seven key steps developed through guidance by three expert physicians and taken from the New England Journal of Medicine (NEJM) [32] and the National Library of Medicine (NLM) [71].

Specifically, seven videos were developed as part of the DHRT^{sequential} training ranging in time from 15s to 30s. In addition, an eighth video was developed to explain the post-trial summary screen and did not include a learning task. **Figure 12** breaks down the flow of the assessment

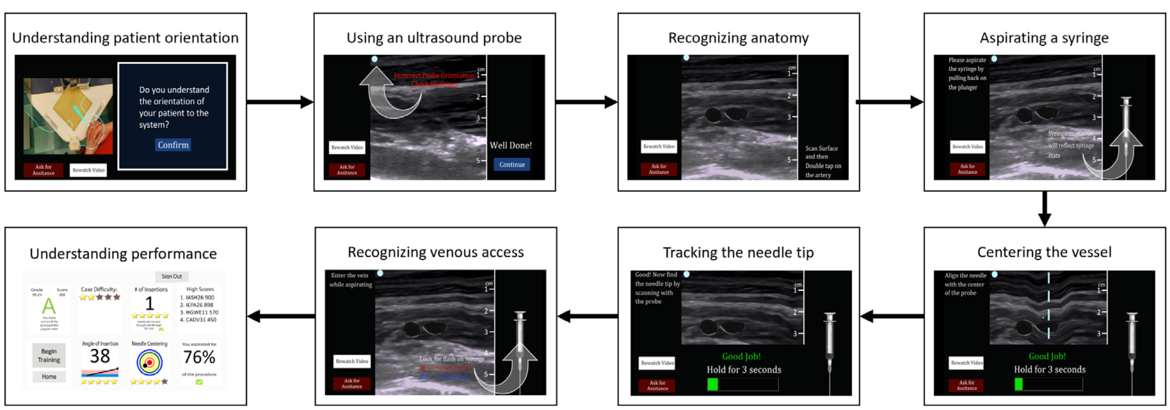

Figure 12: Flow of the learning assessment activities in the sequential learning walkthrough

tasks included in the sequential learning walkthrough training. For each step in the sequential training, a video explanation was provided and followed up with a hands-on activity and assessment that the user had to complete. The initial activities and assessments focus on learning individual skills such as distinguishing the vein from the artery and using the ultrasound probe. These skills are then re-emphasized through learning assessment activities to guide the user through how to conduct all of the steps of CVC at once. For example, to teach the skill of using the ultrasound probe, the user would watch a video explaining how to use the probe and then be prompted by the system to use the probe to scan the DHRT^{sequential} surface until they saw the internal jugular vein and carotid artery. The system would not prompt them to the next activity until they accomplished this task. Trainees were able to re-watch the explanation video for as many times as needed during the assessment activity. The final task and assessment combined all the steps learned throughout the training and had the trainee do an entire practice needle insertion.

In addition to the sequential learning, to better individualize learning to performance [172] the system was also modified for the sequential learning group to only require three successful insertions to finish the training depending on the user's score on the performance metrics. The required minimum of three trials was found by assessing a subset of previous trainees and finding by which trial residents were beginning to score within expert range as determined by Pepley in a previous study on the DHRT [46]. If the trainee is struggling, they can continue going through trials until they do up to six preset trials [171]. The last trial for each person is referred to as the verification of proficiency (VOP). For the sequential learning group, to finish the training in three trials, they may score anything on the first trial, must score at least 70% with successful insertion on the second trial, and at least 70% with successful insertion on the VOP. If these requirements were not met, trainees would continue going through the cases until they reached the correct score and could continue to the VOP trial, or until they reached trial five and were forced to do the VOP for trial six regardless of previous score. If participants reached the VOP in three, four, or five trials but then got less than a 70% on the VOP, they would repeat the VOP until they either reached 70% or reached six trials. The score required to move forward in training was not set to 100% because of normal variations in CVC technique that may differ from the optimal score programmed into the DHRT, and limitations of the system itself. Expert physicians are not likely to score 100% on the DHRT system [46] due to variations in angle of insertion, how often the needle is aspirated, overall procedural flow that is dependent on the physician, the patient anatomy, and CVC performance standards within a hospital. Additionally, students may have difficulties on the device that are not reflective of technical performance, but instead are reflective of limitations in the design and scoring of the system. Therefore, it would be difficult to expect new trainees to achieve a score of 100%.

2.3 Methods

The main objective of this study was to determine the efficiency and effectiveness of sequential learning in simulator training for CVC through assessing initial skill gain, the impact of the number of trials, and learning curves on the DHRT. For the remainder of this paper, the residents who received training on the original DHRT will be referred to as the DHRT cohort and those with the sequential learning will be referred to as the DHRT^{sequential} cohort. Specifically, this study aimed to answer the following research questions (RQs):

RQ1: Does the integration of sequential learning in the DHRT impact first-trial success and performance?

Our first research question was developed to examine the impact of sequential learning on resident first trial insertion performance on the DHRT systems. Specifically, we sought to understand the impact of the training on whether or not the participant was able to obtain venous access without puncturing the carotid artery during their first trial. We hypothesized that training type, distance to vein center, and aspiration rate would be significant predictors of successful performance on the first trial (*H1)* based on the inclusion of sequential and multimedia learning [73], [74] and prior literature on sequential learning in medical education in laparoscopic SBT [178].

RQ2: Does sequential learning in a CVC simulator impact number of trials required and performance at the end of training?

This research question was developed to determine if the inclusion of sequential learning impacted the number of trials conducted between the DHRT^{sequential} and the DHRT groups. We hypothesized that number of trials would be significantly different with the DHRT^{sequential} group finishing in less trials (*H2*) and that there would be no performance differences for both metrics between groups (*H3)* because both groups should reach mastery performance by their last trial. These hypotheses are based on prior literature indicating that complex skills are better learned when broken down [60] and individualized to the learner [172], and that the effectiveness of sequential learning can be seen in mastery level performance being reached faster [178].

RQ3: Does the integration of sequential learning into CVC SBT impact learning curves and time required for training?

The third research question was developed to explore learning curves on the DHRT system between groups for the DHRT performance score based on the number of trials conducted, and to determine if there was a difference in the amount of time required to complete training or reach the mastery level of performance. We hypothesized that the DHRT^{sequential} group would have significantly less learning curves present than the DHRT group (*H4) and that the overall training time would be reduced (H5)*. This is because prior research has shown that sequential learning is a better method for efficiently building competence in complex procedures and that competence levels can be differentiated via learning curves [19], [73], [103], [182]. Additionally, prior research has indicated that the steepness of learning curves can be reduced through incorporating more structure and feedback into learning, and previous studies on the DHRT have indicated that the presence of learning curves on the system signifies learning by the trainee [107], [109].

2.3.1 Participants

To answer our research questions, 103 participants were recruited from residency bootcamps at Hershey Medical Center in 2021 and 2022. Forty-four residents participated in the DHRT group, and 59 residents participated in the DHRT^{sequential} group. A summary of demographic data broken down by DHRT^{sequential} and DHRT groups can be seen in **Table 4.**

Table 4: Full demographic summary of residents included in the study

		DHRT^{sequential}	DHRT
Gender			
	Female	25	13
	Male	34	30
	Other	0	1
Race			
	Black or African American	1	1
	White	37	29
	Asian	17	8
	Other	0	3
	More than one race	3	1
	Prefer not to answer	1	2
Specialty			
	Anesthesiology	12	10
	Emergency Medicine	9	6
	Internal Medicine	14	12
	Other Non-surgery	5	7
	General Surgery	13	3
	Other Surgery	6	6

2.3.2 Procedures

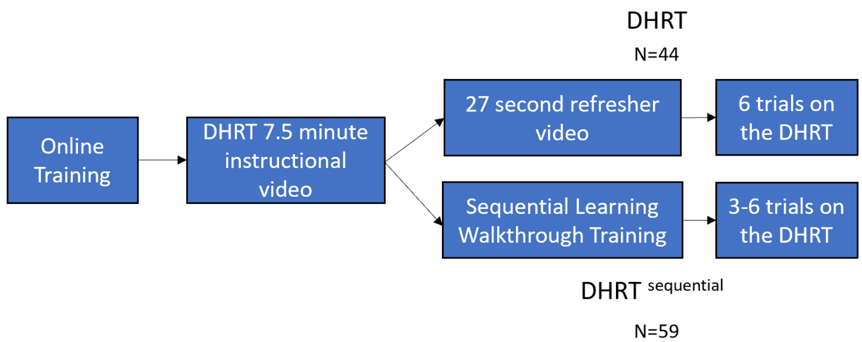

Figure 13: Complete procedural flow for the DHRT and DHRTsequential training groups

The data collected in this study is part of a larger investigation on residency training and CVC. As such, only the parts of the procedure that are relevant to the current study will be discussed. Before beginning training, residents consented to participate in this study by providing informed consent, as per the IRB protocol. Next, residents completed an online US-IJCVC training developed through prior work [84]. Specifically, the training includes eight video modules with embedded questions focused on teaching: (1) introduction to CVC, (2) an overview of CVC steps as defined by the New England Journal of Medicine [32], (3) an overview of the benefits and risks of each access site for CVC, (4) best practices to use CVC equipment, (5) rapid central vein assessment with ultrasound, (6) mechanical procedures for troubleshooting, (7) complication types and how to identify them, and (8) monitoring the patient and removing the catheter [84]. After completing the online training, participants completed a post-training US-IJCVC knowledge quiz which required a passing score of 80% or higher to attend in person DHRT training. Once the residents passed this assessment, they watched a 7:30 instructional video of how to use the DHRT system to conduct US-IJCVC. Then, DHRT residents watched a 27s recap video of how to use the DHRT system and then conducted 6 trials on the system. The DHRT^{sequential} residents underwent the full sequential walkthrough training with activity assessments and then conducted 3-6 trials on the DHRT depending on performance. See the flowchart in **figure 13** for a full breakdown of the procedures and how they differed between groups.

2.3.3 Metrics

The DHRT performance metrics used to answer our research questions included: needle distance to the center of the vein, aspiration rate, successful insertion, DHRT performance score, and last trial performance. These metrics were derived from previous work [171]. In addition, past training, DHRT performance score, predicted last trial, and patient case were also explored. Each of these metrics is detailed down below.

Distance to vein center: This distance is calculated as the radial distance from the tip of the needle at its final position to the center of the vein [183]. This metric is important because inserting the needle off-center decreases the chance of a successful insertion due to the tissue compression and requires more skin incisions to be made. For this variable, a lower number indicates better performance. The ideal score for distance to the center of the vein is 0.

Aspiration rate: Aspiration, or pulling back on the plunger of the syringe, is important because the influx of blood into the syringe, referred to as flash, is an indicator to the operator that the vessel has been accessed [32], [46], [52]. A higher percentage of aspiration time is beneficial for trainees so they understand when they have entered the vein; the ideal score for aspiration is 100%.

Successful Insertion*:* To have a successful insertion, a participant needed to end the trial in the vein *and not* have punctured the carotid artery *or* inserted the needle through the backwall of the vein. An arterial puncture, a backwall puncture, or ending the procedure when the needle is not in the vein resulted in an unsuccessful insertion.

DHRT Performance Score*:* For each trial on the DHRT, the participant was given a final score. This score is made up of their performance metrics and the difficulty of the trial [169]. The formula to determine the DHRT performance score is DHRT Performance Score= $Is^* (161.3 * (\theta s * Cs * as) - Bs - As)$ where Is refers to if the needle entered the artery or the vein, θs refers to the angle of the needle, Cs refers to the distance to the center of the vein, as refers to aspiration of the needle, Bs refers to puncturing the backwall of the vein, and As is the number of attempts. Refer to [183] for more details on how this was determined. If the artery is punctured, the score is automatically changed to zero (failing). A satisfactory score for progressing through the training in less trials is 70% or higher with successful insertion.

Number of Trials: To complete the training in the DHRT group, each participant was required to conduct between three and six trials on the DHRT depending on performance. This variable was automatically computed for the DHRT^{sequential} group; however, several participants did not see the GUI indicating they were finished training so their number of trials was corrected in post-processing based on score. Additionally, to compare the DHRT and DHRT^{sequential} last trial performance and normalize the trial in which the user ended, we post-processed the DHRT data by determining which trial the DHRT user would have ended their training based on their performance scores according to the process used on the DHRT^{sequential} group described above.

Past training: On the demographic survey conducted in the online training, residents were asked if they have had previous training in CVC. They could answer they had previous training through observation, manikin, robot, other, none, or more than one. Past training was treated as a binary metric for this study.

Patient Case: The DHRT system was programmed with multiple patient profiles based on hypothetical anatomical variations that could be seen in live patients, as determined by expert physicians [169]. Each trial conducted on the system is a different patient case, with the first trial and the VOP (last) patient case matching for comparison purposes.

Training Time: Training time (in seconds) was measured in the DHRT group as the time to watch the 27 second refresher video and undergo the 6 required trials and in the DHRT ^{sequential} group as the time required to undergo the walkthrough training and the required number of trials based on performance. The time on the system was recorded every time a case was actively running and during the walkthrough, and the instructional time was added after training for analysis.

2.4 Results

The main objective of this research was to determine if sequential learning impacts CVC skills gain and learning curves for CVC by answering the following RQs. All statistical analyses for this study were conducted in SPSS (v29).

RQ1: Does the integration of sequential learning in the DHRT impact first-trial success and performance?

The first research question was developed to determine if the likelihood of a successful first trial on the DHRT system, ending the procedure in the vein without puncturing the carotid artery or the backwall of the vein, could be predicted by the past experience and training type of the participant. Our hypothesis (*H1)* was that the sequential learning on the DHRT^{sequential} group would significantly impact first-trial success with the sequential learning group performing better [73]. To test this, a binomial logistic regression was performed with the dichotomous dependent variable, first trial success, being predicted based on the two dichotomous independent variables, past experience and training group (DHRT or DHRT^{sequential}), and two continuous performance variables, distance to vein center and aspiration rate. Before running the regression model, the percentage of people failing within each group was determined. To achieve a successful insertion, the trainee must have ended the procedure in the vein without puncturing the carotid artery or through the backwall of the vein. In the DHRT group, 28 people (63.6%) failed their first insertion. Of those who failed, 9 (20.4%) failed due to both puncturing the carotid artery and the backwall of the vein, 10 (22.7%) failed due to puncturing through the backwall of the vein, 4 (9.1%) people failed due to ending outside of the vein, and 5 (11.4%) failed due to puncturing the carotid artery. In the DHRT^{sequential} group, 15 (25.5%) people failed their first insertion. Of those who failed, 2 (3.4%) failed due to both puncturing the carotid artery and the backwall of the vein, 9 (15.3%) failed due to puncturing through the backwall of the vein, 1 (1.7%) failed due to ending outside of the vein, and 3 (5.1%) failed due to puncturing the carotid artery.

For the binomial regression, prior to computing the analysis, assumptions were verified. Past training was also included as a variable in this analysis to account for different exposure levels of trainees to CVC before attending the DHRT training. The binomial logistic regression model was statistically significant χ^2 (2) = 28.280, $p < .001$ with the model explaining 32.3% (Nagelkerke R^2, a medium effect size [184]), of the variance in successful first insertion and correctly classifying 71.8% of cases. Specifically, the results showed that the training group (*Wald* χ^2= 7.168, $p = .007$) and aspiration (*Wald* χ^2= 5.177, $p = .023$) were significant predictors of first trial success while past training (*Wald* χ^2 = .047, $p = .923$), and distance to vein center (*Wald* χ^2= 1.334, $p = .320$) were not. Specifically, DHRT^{sequential} group was associated with an increased likelihood of first-trial success (odds ratio = 3.58, *CI* 95% [1.41 – 9.17]) compared to the DHRT group. This result aligns with our hypothesis (*H1)* that DHRT^{sequential} would have a better first trial insertion performance.

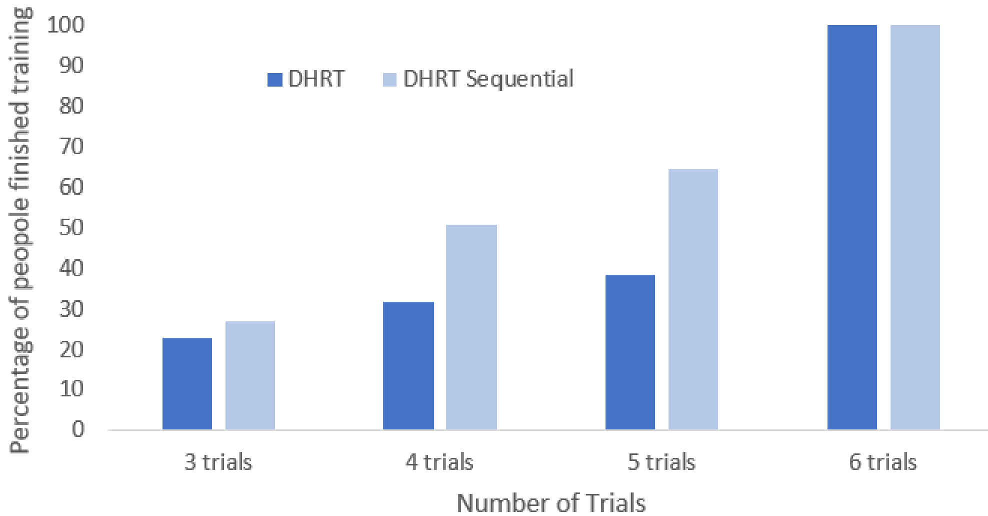

Figure 14: A cumulative percentage graph indicating what percent of each training group was finished at each number of trials

RQ2: Does sequential learning in a CVC simulator impact number of trials required and performance at the end of training?

The primary objective of RQ2 was to determine if the inclusion of sequential learning impacted the number of trials conducted between the DHRT^{sequential} and the DHRT groups. We hypothesized that number of trials would be significantly different with the DHRT^{sequential} group finishing in less trials (*H2*) and that there would be no performance differences for both metrics between groups (*H3).*This hypothesis is based on prior literature indicating that complex skills are better learned when broken down [60], and that skill mastery can be achieved through SBT [166]. **Figure 14** provides a breakdown of people in each training group that finished in each number of trials. For the DHRT^{sequential} group, 64.4% of learners finished in less than six trials compared to only 38.6% in the DHRT group. To test for differences in the number of trials, a Mann-Whitney U test was conducted. Distributions of performance metrics were similar, as assessed by visual inspection. Results of the Mann-Whitney U test indicated that there was a statistically significant difference in the number of trials performed (U=1012.500, z=-2.033, p=.042) the DHRT^{sequential} group (Md=4) and the DHRT group (Md=6). A follow-up analysis was done on the cumulative distributions of people finishing in each trial group through a Kaplan-Meier survival analysis with a log-rank test. All assumptions were met for this analysis. The results indicated that there was a significant difference (χ^2 = 5.558, p = .018) in the cumulative distributions of people finishing at different timepoints between the people who received sequential learning (M=4.57 trials) and those who did not (M=5.56 trials). Mann-Whitney U tests were also run within each number of trials group for aspiration and distance to vein center to determine performance differences between the DHRT^{sequential} group and the DHRT group. A Bonferroni adjustment was applied to account for testing for testing multiple performance variables [185], resulting in an family wise error rate (α) of .025. Distributions of performance metrics were similar, as assessed by visual inspection. For distance to vein center in the three-trial group, the median values were significantly different (*U*=23.5, z=-2.983, *p*=.002) for the DHRT^{sequential} group (Md=.3200) compared to the DHRT group (Md=.4650). No other significant differences were found between groups for either performance metrics. These results align with our hypothesis that the number of trials would be significantly different with the DHRT^{sequential} group finishing in less trials (*H2*). These results align with our hypothesis that there would not be significant performance differences at the end of training (*H3*), as a difference was only evident for a single performance metric in the three-trial group.

Table 5: Curve estimates for each trial in the DHRT and DHRT^{sequential} groups

	DHRT				*DHRT^{sequential}*			
Logarithmic	*B*	*R^2*	*F*	*p*	*B*	*R^2*	*F*	*p*
3	48.909	.603	42.598	**<.001***	9.502	.131	7.054	**.011***
4	34.854	.426	10.370	**.006***	9.108	.095	5.665	.021
5	32.794	.385	8.142	.014	13.108	.145	6.466	.015
6	14.403	.059	10.000	**.002***	2.360	.003	.338	.562
Linear								
3	25.250	.521	30.463	**<.001***	4.875	.118	6.268	.016
4	15.725	.399	9.312	**.009***	4.893	.126	7.812	**.007***
5	12.167	.328	6.349	.026	4.950	.128	5.596	.023
6	4.564	.073	12.691	**<.001***	.904	.003	.395	.531

RQ3: Does the integration of sequential learning into CVC SBT impact learning curves and time required for training?

The objective of this question was to explore learning curves and training time on the DHRT system between groups. Learning was assessed as a group, so logarithmic and linear learning curves were calculated to account for individual differences [102]. We hypothesized that the DHRT^{sequential} group would have significantly fewer learning curves and less training time than the DHRT group because prior research has shown that sequential learning is a more efficient method for building competence in complex procedures [73], [103], [182]. For time, a Mann-Whitney U test was conducted between the total time it took each person to go through the training in each group finding no significant differences (U = 1034.0, z = -1.760, p = .078) between the DHRT group (Md =494.3 seconds) and the DHRT^{sequential} group (Md = 382.8 seconds). To study learning curves, curve metric analyses were conducted. For each group, curves were analyzed separately for residents who finished the training in three, four, five, and six trials on the system to account for differences in learning speed. Linear and logarithmic curves were fit to each dataset and separately analyzed to determine significance. Specifically, a curve-fit estimation was conducted to identify statistically significant change in performance over time by modeling the DHRT performance score over the number of trials. A Bonferroni adjustment was applied to account for testing for multiple trial groups [185], resulting in an family wise error rate (α) of .0125. **Table 5** shows a breakdown of the results for this section. The DHRT group had 3 significant logarithmic learning curves and 3 significant linear learning curves. The DHRT^{sequential} group had 1 significant logarithmic learning curve and 1 significant linear learning curve. See **figure 15** for an example of the learning curves graphed for six trials in each group. These results align with our hypothesis that there would be less significant learning curves in the DHRT^{sequential} group, indicating that trainees started at a higher point and therefore saw less of a significant change from beginning to end of training.

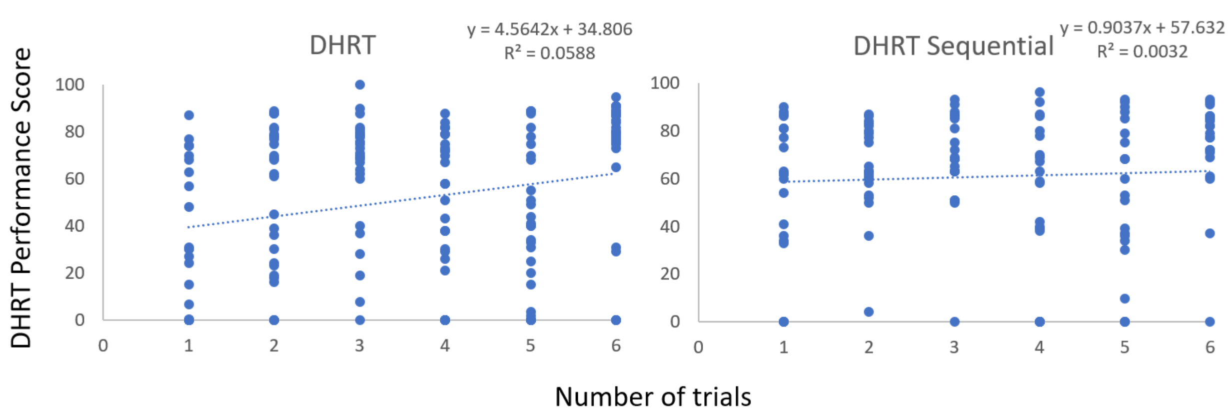

Figure 15: 6 trial learning curves for the DHRT group (significant) and the DHRTsequential group (nonsignificant)

2.5 Discussion

The goal of this study was to measure the efficiency and effectiveness of sequential learning in simulator training for CVC by assessing the initial skill gain, impact of the number of trials, and learning curves on the DHRT^{sequential} compared to on the DHRT alone. The main findings of this study were that

1) The DHRT^{sequential} group was 3.58 times more likely to have a successful insertion on their first trial, which was significantly impacted by aspiration rate
2) The number of trials required was significantly different between groups, but performance was not at the end of training
3) The DHRT^{sequential} group had fewer significant learning curves compared to the DHRT group.

We hypothesized that the implementation of sequential learning would improve CVC simulator training by increasing first trial performance (*H1*) on the DHRT, based on previous literature indicating that sequential learning is useful for learning complex skills [60], [178]. Sequential learning did impact initial success of the procedure by significantly increasing the success rate of learners making them 3.58 times more likely to achieve successful venous access in their first trial. This indicates that the challenge of having to apply multiple skills at the same time on the first use of the system was lessened by the exposure and breakdown of skills introduced in the sequential learning walkthrough and assessment activities. For specific performance metrics, aspiration rate was a significant predictor of successful first insertion. This was surprising because the DHRT^{sequential} group was explicitly taught how to align the needle to the center of the vein in the assessment activities and how to aspirate throughout the procedure and, so it was expected that both skills would have a significant impact on their ability to perform successfully.

We also hypothesized that the number of trials required to finish training would be significantly lower for the DHRT^{sequential} group than for the DHRT only group (*H2*) because indicating that complex skills are better learned when broken down [60] and individualized to the learner [172]. We also hypothesized that performance would not be significantly different between groups at the end of training (*H3*), because both groups should reach mastery performance by the end of SBT [166]. Number of trials was significantly different between groups (p=.042), as was the difference in the cumulative distributions finishing within each trial number (p = .018). The

DHRT^{sequential} group had a higher percentage of people finishing in less than six trials (64.4%) compared to the DHRT group (38.6%), indicating an increase in the efficiency of training with the addition of sequential learning. Additionally, performance was not significantly different between the groups at the end of training, except for distance to vein center in the three-trial group. Performance not differing at the end of training indicates the utility of individualized learning, and further emphasizes that allowing the required number of trials range from 3-6 depending on performance is a valid training method.

Finally, the we hypothesized that the number of significant learning curves would be lessened and that training time would be less in the DHRT^{sequential} group (*H4*) due to previous literature indicating the utility of deliberate practice and sequential learning on minimizing procedural learning curves [19], [107], [178], and that sequential learning is a more efficient method for building competence in complex procedures [73], [103], [182]. Our findings refute out hypothesis that training time would be less for the DHRT^{sequential} group; however, it is beneficial that there was no difference in training time between groups when considering the extra instructional time added with the walkthrough. Our findings align with our hypothesis that there would be less learning curves in the DHRT^{sequential} group. Learning curves were assessed in each training group based on the number of trials and the DHRT performance score. For the DHRT group, there were significant linear and logarithmic curves for all trial groups except for five trials. For the DHRT^{sequential} group, the only significant logarithmic curve was for three trials, and the only significant linear curve was for four trials. Overall, the learning curves indicated that the DHRT^{sequential} group had less room to improve than the DHRT group, helping to minimize the learning curves for CVC simulation and aligning with prior research on how practice impacts learning in simulation for other procedures [19], [179].

2.6 Conclusion

The goal of this research was to determine if the addition of sequential learning into a CVC simulator could improve the efficiency of learning, the impact of the number of trials, and learning curves on the DHRT. The first main finding of this study is that the implementation of sequential learning did significantly impact the rate of success of the first trial on a CVC simulator for the DHRT^{sequential} group, indicating increased initial acquisition of skills. Secondly, the number of trails required to complete training was significantly lower for those in the DHRT^{sequential} group.

Finally, the DHRT^{sequential} group had fewer significant learning curves. These results indicate that the implementation of sequential learning into medical education is beneficial to increasing the efficiency of initial skill gain, and minimizing the learning curve.

This study has several limitations that must be addressed. First, due to the nature of this study taking place during clinical training in medical centers, the sample sizes are limited and as such results from this paper should be considered exploratory. It is possible that with a larger sample size, differences between learning groups may have been more pronounced; future work will study this further. Second, this study took place at one U.S. medical center with one specific simulator and as such the generalizability of the findings is limited and further research would be needed to validate these findings for other simulators at other institutions. Third, the indication that training as completed for the DHRT^{sequential} group on the GUI was not clear to all participants, requiring post processing to correct. Fourth, this study assessed learning from a medical simulator for CVC, and as such it is impossible to tell from this study alone if the participants were truly learning the procedure or if they were learning only the device; future work should focus on validating this with a longitudinal study into the clinical environment. When using a simulator for training, it is hard to verify if the test is a perfect measure of real-world practice. Since we are unable to assume that the simulator is perfect, the score to pass cannot be set to 100%. Adding to this, this study assessed time differences between cohorts; however, the DHRT cohort spent all of their time training whereas the DHRT^{sequential} cohort spent part of their time learning with the walkthrough and part of it training. As such, future work should explore in more detail if it is the time spent learning or the time spent practicing that makes a bigger impact on learning efficiency. Finally, this study focused on group learning curves and did not explore individual learning curves; future work should explore individual learning curves in medical simulation for CVC to determine if findings would differ.

Chapter 3

COMPETENCE OVER CONFIDENCE: GENDER-BASED DIFFERENCES IN RESIDENT TRAINING FOR CENTRAL VENOUS CATHETERIZATION

This paper will be submitted to the BMC Medical Education in March of 2024. This work is multiple authored by Haroula Tzamaras, Dr. Lisa Sinz, Dr. Phillip Ng, Dr. Michael Yang, Dr. Jason Moore, and Dr. Scarlett Miller. Haroula Tzamaras was the lead author on this paper Dr. Jason Moore and Dr. Scarlett Miller helped advise this work. Drs. Sinz, Yang, and Ng were instrumental to the data collection for this study.

3.1 Abstract

While women make up over 50% of students enrolled in medical school, disparities in self-efficacy of medical skills between men and women have been observed throughout medical education. This difference is significant because low self-efficacy can impact learning, achievement, and performance, and thus create gendered skill gaps. Simulation-based training (SBT) is one area that relies heavily on assessments of self-efficacy, however, the Dunning-Kruger effect in self-assessment posits that trainees often struggle to recognize their skill level. Additionally, the impact of gender on self-efficacy during SBT has not been as widely studied. The objective of this study was to identify if the gender-confidence gap and the Dunning-Kruger effect exist in SBT for central venous catheterization (CVC) on the dynamic haptic robotic trainer (DHRT) by comparing self-efficacy and performance between men and women residents. Residents completed a 14-item Central Line Self-Efficacy survey (CLSE) before and after CVC training on the DHRT. DHRT CVC performance metrics of the number of insertion attempts, backwall puncture, and successful venipuncture were also collected. Gender differences of 173 surgical residents (N_{women}=61, N_{men}=112) pre- and post-CLSE, DHRT performance were compared. General estimating equation results indicated that women residents were significantly more likely to report lower self-efficacy for 9 of the 14 CLSE items ($p<.0035$). Mann-Whitney U and Fisher's exact tests showed there were no performance differences between men and women for successfully accessing the vein on the DHRT. These results indicate that despite receiving the

same SBT and performing at the same level, the gender-confidence gap exists in CVC SBT, and the Dunning-Kruger effect may be more evident for women than men.

3.2 Introduction

The percentage of women enrolled in medical school in the United States has been steadily increasing, from 27.9% in 2000 [186] to 54.6% in 2023 [187]. While medical school enrollment has been on the rise for women, so has their participation in clinical research. In 1986, the NIH enacted a policy encouraging researchers to use women participants in their clinical studies [188]; however, women were not required to be included in clinical trials until 1993 when it was written into law [188]. Since then, women account for around 40% of clinical trial participants for many diseases, often even if the disease is more prominent in women than men [189]. While science still has improvements to be made for gender parity in clinical research, despite reaching gender parity in medical education, researchers have identified a gender-confidence gap in medical training [190], [191]. This gender-confidence gap manifests as disparities in self-confidence where women underestimate and undervalue themselves compared to men [94]. This is worrisome because self-efficacy, or task specific confidence [192], [193], has been shown to be vital in challenging environments, like medicine, due to its relationship with an individuals' motivation to engage in tasks and to persevere when faced with training challenges [89]. In addition, self-assessment of confidence and performance are often used throughout medical education as a means to examine the effectiveness of trainee learning [47], [86]. When self-assessment of performance or confidence and actual knowledge are not aligned, it can give faculty perceptions that the trainee is lacking competence [194], [195] while simultaneously making the trainees more likely to embody this belief themselves [196].

It is important to note that higher self-efficacy does not always correlate with improved accuracy and performance [197]. For example, research has shown that people with lower skill levels often overestimate their abilities, or people with higher skill levels underestimate their abilities, a phenomenon referred to as the Dunning-Kruger Effect [92]. This effect has been found throughout medical training, from medical students [198], to medical residency [199], to attending physicians [93] where underperformers often rate themselves higher than their actual skill levels while high performers often rate themselves lower. Recent research has indicated that there may be gender effects in this phenomenon [200]. For example, women medical students often report lower confidence than men in bedside procedures [201], and in self-rated performance on surgical clerkship [95] regardless of actual performance. Similarly, women in surgical residency often rate their knowledge on patient care [202] and general competency [97] lower than men despite no gender-based performance differences [203]. While it has been shown that women have disproportionately lower confidence than men, it is less known how different types of training may impact this.

One area that frequently uses self-assessments of confidence is simulation-based training (SBT). SBT using imitations of real procedures and environments [14], [76], or simulators, to allow physicians to practice before working with live patients [204], [205]. SBT has been widely integrated throughout medical education due to its low-risk, hands-on practice [206], [207]. SBT often relies on self-assessments of confidence because it provides an indication of trainee learning [86], [89]. The primary means of measuring self-confidence in SBT is through self-efficacy [208] and/or self-rated assessments of performance [209]. Self-efficacy is an important construct in SBT due to its relationship with learning [86], [192] and achievement [125], [210], and many studies focus on how SBT can be used to increase trainee confidence. For example, in medical training, participation in SBT has been shown to significantly increase trainee self-efficacy for acute skills [211], emergency room preparedness [212], intercostal drain insertion [213], and central venous catheterization (CVC) [214]. However, the gender-confidence gap has been shown to manifest in SBT [36] with one study showing that women had lower self-efficacy for obstetric emergency after SBT training despite the fact that there were no gendered performance differences [215]. However, few studies exist that explore the relationship between gender, confidence, and simulation training. Exploring these effects is important due to of the relationship between confidence and competence [90] in medicine.

One procedure that is useful for exploring the impact of the gender-confidence gap in SBT in medical residency is central venous catheterization (CVC). CVC is a complex medical procedure where a catheter is inserted into a central vein for critical medication delivery to the bloodstream [7]. CVC is performed over 5 million times annually [7], however, over 15% of these end procedures end with a complication [9]. The complication rate with CVC is largely associated with experience level of the physician - a physician who was performed less than 50 catheterizations is twice as likely to cause complications [7]. CVC is most commonly conducted with ultrasound guidance into the right internal jugular vein (US-IJCVC) [50], [111], and requires a series of bi-manual steps to complete the procedure. These steps include manipulating an ultrasound probe in one hand while inserting a needle into the internal jugular vein, while avoiding other anatomy like the carotid artery [52]. Once the vein is accessed, a catheter can be inserted; however, most training for CVC focuses just on the initial needle insertion.

CVC is typically taught using SBT [27] including one of the newest SBT systems, the Dynamic Haptic Robotic Trainer (DHRT) [24], see **figure 16**. The DHRT uses haptic robotic simulation and mock ultrasound to train residents on CVC needle insertion for multiple patient anatomies, because in live patients the location and depth of the internal jugular vein can vary [28]. The DHRT is as effective as manikin training for CVC in skill and self-efficacy gains [70], [216], and is more beneficial than manikin training due to its objective scoring and real-time feedback. Each time the trainee uses the system, they are presented with a graphical user interface (GUI) with personalized feedback on their performance, including number of insertion attempts and where to improve if insertion was not successful [31]. The focus of training on the DHRT is achieving successful venipuncture by inserting the needle into the vein in one attempt without puncturing through the backwall of the vein. The DHRT aims to reduce the likelihood of mechanical complications that are often caused by human error and training deficits [7] such as puncturing the vein backwall [54] or puncturing the carotid artery [7], [217].[7]

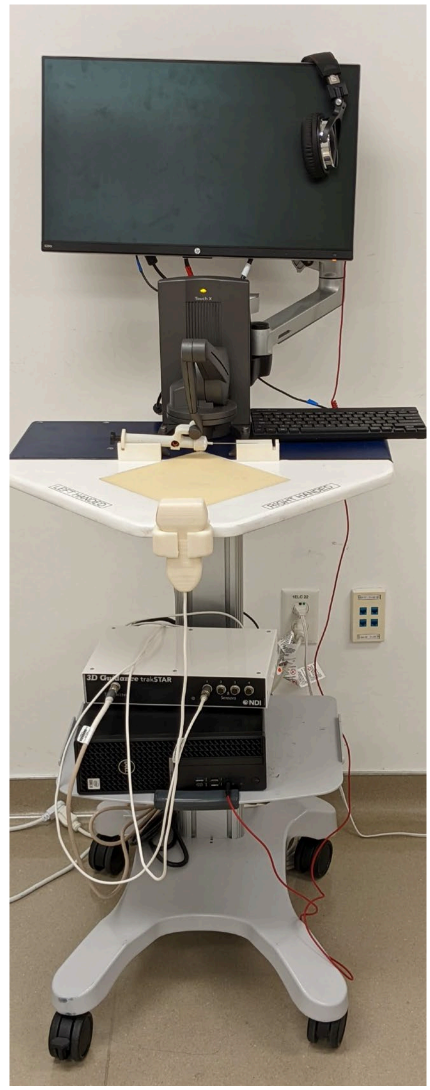

Figure 16: The DHRT Trainer used in residency training

In light of this previous work, the objective of this study was to assess if the gender-confidence gap or the Dunning-Kruger effect exists in CVC SBT by comparing self-efficacy and DHRT performance between men and women residents. The DHRT is the focus of the current study because it objectively measures performance of an individual, allowing for comparison between people and to other metrics like self-efficacy.

3.3 Methods

The goal of this study was to identify if the gender-confidence gap or Dunning-Kruger effect exists in CVC SBT. This was achieved by comparing self-efficacy and DHRT performance between men and women residents. Specifically, we sought to explore the following research questions (RQs):

RQ1: Is there a gender-confidence gap in CVC SBT pre- or post- training?

The first research question was developed to examine if the gender-confidence gap exists for residents learning CVC before or after SBT. We hypothesized that women residents would have lower CVC self-efficacy than men residents both pre- and post-training (*H1*) and that there would be no difference between men and women resident's self-efficacy gains pre- to post-training (*H2*). These hypotheses are based on previous work that indicated that women in graduate and post-graduate medical training rate themselves lower in perceived clinical skills [97], performance [95], and confidence [215] than men. To add to this, most SBT, including the DHRT, have focused on studying how SBT improves self-efficacy overall [70], [216], and not explored gender effects.

RQ2: Are there gender-based performance differences in CVC SBT at the end of training?

The second research question was developed to examine if there were gender-based performance differences in CVC SBT. We hypothesized that there would be no significant differences in CVC SBT performance between genders (*H3*). This hypothesis was based on prior work that found that men and women do not differ in performance for clinical knowledge or technical skills at the residency level in programs such as general and plastic surgery [97], and obstetric training [215]. Understanding if performance differences exist in SBT for CVC may allow further investigation into the equity of training methods.

RQ3: Does the Dunning-Kruger effect exist in CVC SBT post-training?

The final research question was developed to examine if the Dunning-Kruger effect exists post-CVC SBT training by exploring the relationship between self-efficacy and performance. We hypothesized that there would be no significant relationship between these variables for either gender, thus supporting the existence of the Dunning-Kruger effect in CVC SBT (*H4*). This hypothesis was based on prior literature that found that medical residents' ability to accurately self-assess skills was weak, with close to zero correlation between self-ratings and observed performance [199], and that medical trainees are often unaware of their actual skill level [92]. Lack of correlation between performance and self-efficacy for CVC SBT would provide insight into how to better provide feedback to trainees to increase understandings of performance strengths and deficits.

3.3.1 Participants

In order to answer these questions, 173 residents (N_{women}=61, N_{men}=112) from two residency cohorts (N_{2021}=72 and N_{2022}=101) and two medical centers (N_{M1}=103, N_{M2}=70) were recruited from the new resident bootcamp over a span of two summers with trainings running from June through September, see **table 6** for participant breakdown. While the bootcamp was mandatory for all residents, participation in this research was voluntary and only residents who consented to participate were included in this study.

Table 6: Demographic Information for residents in this study

		Women (N=61_)			***Men (N=112)***		
	Demographic Information	***2021***	***2022***	***Total***	***2021***	***2022***	***Total***
Specialty		3	11	14	3	7	10
	Urology	0	0	0	2	3	5
	Cardiothoracic surgery	0	0	0	1	1	2
	Neurosurgery	1	0	1	0	1	1
	Orthopedic surgery	1	1	2	3	2	5
	Otolaryngology	0	2	2	2	1	3
	Ophthalmology	1	0	1	1	0	1
	OBGYN	0	4	4	0	0	0
	Plastic surgery	0	0	0	1	2	3
		11	14	25	16	15	31
	Anesthesiology	3	2	5	10	16	26
	Emergency medicine	2	2	4	4	7	11
	Medical ICU	0	1	1	0	0	0
	Neurology	0	0	0	4	3	7
	Nephrology	0	1	1	1	0	1
	Pathology	0	0	0	0	2	2
	Radiology	0	1	1	0	0	0
	Podiatry	0	0	0	0	1	1
	Physical medicine and rehabilitation	0	0	0	1	0	1
	Pulmonary and critical care	0	0	0	1	0	1
	Pediatric critical care	0	0	0	0	1	1
Level	PGY1	22	35	57	47	54	101
	PGY2	0	1	1	0	2	2
	PGY3 or PGY4	0	1	1	3	1	4
	Did not report	0	2	2	0	5	5
Race/Ethnicity	White	7	18	25	34	34	68
	Asian	7	15	22	11	18	29
	Hispanic	1	0	1	1	1	2
	Black or African American	1	2	3	0	4	4
	Other	2	0	2	1	1	2
	More than one race	2	4	6	2	3	5
	Prefer not to answer	2	0	2	1	1	2

3.3.2 Procedures

At the start of the study, and prior to training, residents consented to participate in this research by providing informed consent, through an online platform. Next, participants completed an online central line training that consisted of a demographic survey, a pre-online training knowledge assessment, eight interactive video modules that focused on CVC content and steps of the procedure, and a post-online training knowledge assessment, see [84] for detailed description of training protocol. Specifically, the eight video modules trained residents on: (1) introduction to CVC, (2) an overview of CVC steps as defined by the New England Journal of Medicine [32], (3) an overview of the benefits and risks of each access site for CVC, (4) best practices to use CVC equipment, (5) rapid central vein assessment with ultrasound, (6) mechanical procedures for troubleshooting, (7) complication types and how to identify them, and (8) monitoring the patient and removing the catheter. To pass the online training, residents needed to receive a post-training assessment score of 80% or higher, multiple attempts were allowed.

After completing the online training, residents attended an in-person training session. At the start of the training, residents completed a pre-training central line self-efficacy (CLSE) survey. Next, each resident completed a set of trials on the DHRT. In the 2021 cohort, residents at both medical centers performed a total of six trials on the DHRT regardless of performance. In 2022 the system was updated so that the number of trials each resident completed was based on previous performance. To complete training, residents in 2022 had to complete two successful venipunctures on the DHRT after a mandatory training trial, defined as vessel access with minimal insertion attempts and no serious error (e.g. arterial puncture). Thus, the minimum number of insertions per 2022 resident was 3 and the maximum was 6. Additionally, surgical residents at medical center 1 in 2021, and trainees at medical center 1 in 2022 received additional hands-on procedural practice covering the steps of CVC in greater depth than provided on the DHRT system. For this extra procedural training, the DHRT was extended so residents had a full CVC kit and interactive feedback on steps of the procedure past the needle insertion that is covered on the

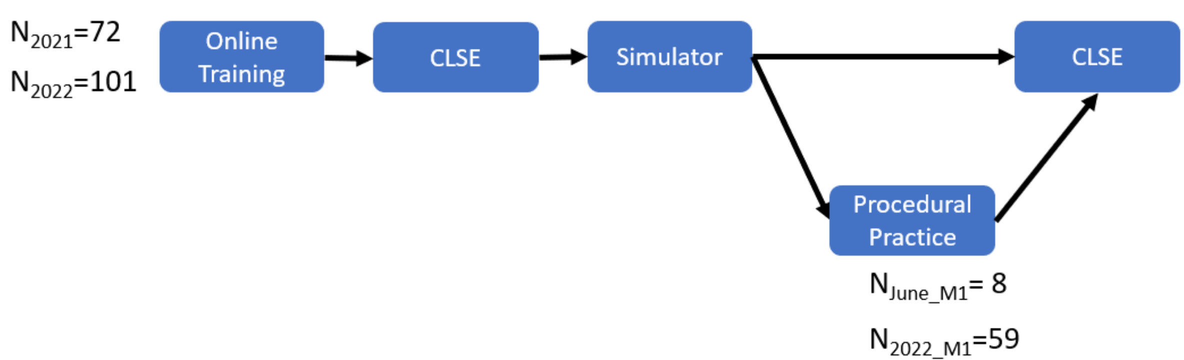

Figure 17: Procedural flow for medical residents in CVC training

DHRT. Finally, residents completed a post-training CLSE survey. See **Figure 17** for the complete procedural flow.

3.3.3 Metrics

In order to answer our research questions, the following metrics were computed.

Performance Metrics

The DHRT measures performance on each trial based on previous research [24], [183]. For the current study, the performance on the last trial was used as this was the Verification of Proficiency test. The performance variables of interest for the current study were number of insertion attempts, backwall puncture, and successful venipuncture without arterial puncture. These metrics are defined below.

Insertion attempts. Insertion attempts was computed by the system as the number of insertions it took to achieve access to the vein. For example, if the trainee pierced the needle into the DHRT and then removed the needle fully and re-insert it to readjust, two insertion attempts were computed. Limiting insertion attempts is important to reduce the likelihood of infectious complications associated with multiple needle sticks [218].

Backwall puncture. A backwall puncture was computed every time a resident inserted the needle into the vein but also punctured the back side of the vessel. Avoiding backwall puncture is necessary to limit the risk of accidental arterial puncture and decrease the risk of treatment complexity caused by mechanical complications [7].

Successful Venipuncture. A successful trial was computed when a resident accessed the vein without puncturing the carotid artery or through the backwall of the vein. Puncturing the carotid artery can lead to serious complications like stroke and death [7] and potentially the insertion of the catheter into the wrong vessel and as such needs to be avoided [217].

Central-Line Self-Efficacy (CLSE)

A five-point, 14-item Likert-scale CVC self-efficacy, referred to as the Central-Line Self-Efficacy (CLSE) developed in prior work [39] was used to assess resident confidence on the procedure. On the CLSE, residents rated themselves in their belief in their ability where a one represented not at all confident and five represented extremely confident. The first ten items on the CLSE survey focused on the specific steps of the procedure such as “modifying the needle trajectory” while the last four questions related to broader aspects of the procedure such as “conducting the entire procedure on a simulator”. The full CLSE survey can be found here, however, please note that in later trainings the CLSE was updated to 19-items, so only the first 14 items are relevant to the current study.

3.3.4 Statistical Analysis

To assess gender-confidence gaps in CVC SBT (RQ1), a general estimating equation (GEE) was computed with gender and self-efficacy type (pre- or post-training) and their interaction as the independent variables and the 14 CLSE questions as the dependent variables. To account for any potential effect of the additional procedural training in 2022 on self-efficacy, training year was also included as a variable. A GEE was used to extend the standard generalized linear regression model and account for the repeated measures of the pre and posttest. All assumptions were met for the GEE. To assess gender-based performance gaps (RQ2), a Mann-Whitney U test was conducted for the continuous variable, insertion attempts. Fisher’s exact test was conducted for the dichotomous performance variables, backwall puncture and successful venipuncture. All assumptions were met for both of these analyses. Finally, to assess the Dunning-Kruger effect (RQ3), regression analyses were conducted to determine if there was a correlation between self-efficacy and performance. Prior to this, the internal reliability of the CLSE was verified (Cronbach’s alpha = .952) justifying the aggregation of the 14 items on the CLSE into one average score. For each regression analysis, the performance metric was the response variable and post self-efficacy, gender, and their interaction were the predictor variables. Linear regression was conducted for the continuous variable, insertion attempts, and binary logistic regression was run for the two dichotomous variables, backwall puncture and successful venipuncture. The analysis was conducted with the entire dataset to determine the significance of the interaction term and then the dataset was split and a follow-up analysis was run within each gender to determine if one had a stronger significant relationship than the other. Assumptions were checked and outliers were found for all three variables, determined true outliers, and kept in for the analysis. All other assumptions were met for all regression models.

3.4 Results

The main the objective of this study was to identify if the gender-confidence gap or the Dunning-Kruger effect exist in CVC SBT on the DHRT. The following results are broken down by research question.

RQ1: Is there a gender-confidence gap in CVC SBT pre- or post- training?

The first research question was developed to determine if the gender-confidence gap existed in CVC self-efficacy for men and women in SBT on the DHRT. We hypothesized that women residents would have lower CVC self-efficacy than men residents both pre- and post-training (*H1*) and that there would be no difference between men and women resident's self-efficacy gains pre- to post-training (*H2*). A Bonferroni adjustment was applied to account for repeated measures on the CLSE survey [185], resulting in an family wise error rate (α) of .0035. GEE results indicated that gender was a significant predictor with women ranking lower for 9 of the 14 variables including using tactile feedback during placement (*Wald* χ^2 =18.814, p<.001), using tactile feedback to identify the vessel (*Wald* χ^2 =20.045, p <.001), advancing the introducer needle (*Wald* χ^2 =11.053, p<.001), modifying the needle trajectory (*Wald* χ^2 =12.492, p<.001), identifying the needle in location (*Wald* χ^2 =8.733, p =.003), using tactile feedback to guide the needle (*Wald* χ^2 =14.216, p<.001), placing the needle in one attempt (*Wald* χ^2 =17.888, p<.001), placing the needle in multiple attempts (*Wald* χ^2 =9.314, p=.002), and conducting the entire procedure without mistakes (*Wald* χ^2 =9.975, p=.002), aligning with our hypothesis (*H1*). Parameter estimates for the nine CLSE items where gender was a significant predictor indicated that a resident who identified as a woman was more likely to rate themselves lower than their men counterparts, see table II. Additionally, the interaction between self-efficacy type (pre or post) and gender was significant for conducting the entire procedure without mistakes (*Wald* χ^2 =12.350, p<.001) meaning that the impact of gender on this variable varied based on the test condition, though gender was not a significant predictor for this variable (p=.004). Positive parameter values [.987(.2822), <.001] for women indicate that identifying as a woman impacted pre-CLSE more than post-CLSE for this variable, though both were lower than for men. Training year did not have a significant impact on any of the 14 variables. See **Table 7** for full significant results. For all 14 CLSE questions, there were significant increases pre- to post-test for both genders (p<.001), aligning with our hypothesis (*H2)*. These results indicate that the gender-confidence gap is evident in CVC training both before and after exposure to SBT.

Table 7: Significant model effects of the GEE * indicates a significant value $p < 0.0035$

Self-Efficacy Item	***Predictor***	***Wald χ^2***	***df***	***p***	***Women Parameter Estimates [B(SE), p-value]***
Using tactile feedback during placement	Gender	18.814	1	**<.001***	[-.704(.1619), **<.001***]
	Pre or Post	115.894	1	**<.001***	
	Year	1.955	1	.162	
	Interaction	2.717	1	.099	
Using tactile feedback to identify the vessel	Gender	20.045	1	**<.001***	[-.766(.1710), **<.001***]
	Pre or Post	98.891	1	**<.001***	
	Year	4.622	1	.032	
	Interaction	4.314	1	.038	
Advancing the introducer needle	Gender	11.053	1	**<.001***	[-.416(.1431),**.004***]
	Pre or Post	84.239	1	**<.001***	
	Year	4.166	1	.041	
	Interaction	.008	1	.927	
Modifying the needle trajectory	Gender	12.492	1	**<.001***	[-.494(.1352),**<.001***]
	Pre or Post	134.188	1	**<.001***	
	Year	1.656	1	.198	
	Interaction	1.178	1	.278	
Identifying the needle in location	Gender	8.733	1	**.003***	[-.405(.1438),**.006***]
	Pre or Post	158.631	1	**<.001***	
	Year	2.761	1	.097	
	Interaction	.125	1	.723	
Using tactile feedback to guide the needle	Gender	14.216	1	**<.001***	[-.673(.1572), **<.001***]
	Pre or Post	79.977	1	**<.001***	
	Year	4.226	1	.040	
	Interaction	4.439	1	.035	
Placing the needle in one attempt	Gender	17.888	1	**<.001***	[-.710(.1612), **<.001***]
	Pre or Post	149.848	1	**<.001***	
	Year	1.214	1	.271	
	Interaction	3.803	1	.051	
Placing the needle in multiple attempts	Gender	9.314	1	**.002***	[-.858(.2990), **.004***]
	Pre or Post	115.584	1	**<.001***	
	Year	.523	1	.469	
	Interaction	.009	1	.924	
Conducting the entire procedure without mistakes	Gender	9.975	1	**.002***	[-.701(.1719), **<.001***]
	Pre or Post	80.437	1	**<.001***	
	Year	2.171	1	**.141**	
	Interaction	12.350	1	**<.001***	[.522(.1326), **<.001**]

RQ2: Are there gender-based performance differences in CVC SBT at the end of training?

The second research question was developed to determine if performance differences existed between men and women for CVC SBT. We hypothesized that there would be no significant differences in CVC SBT performance (*H3*) [97]. Statistical tests were run based on the variable type of the performance metric (backwall puncture, successful venipuncture, and number of insertion attempts). A Bonferroni adjustment was applied to account for repeat testing for the three performance variables [185], resulting in an family wise error rate (α) of .017. For backwall puncture, 95.5% of men avoided backwall puncture and 95.1% of women avoided backwall puncture, with Fisher's exact test finding no statistically significant difference (p = 1.00). For successful venipuncture, 84.8% of men successfully accessed the vein and 90.1% of women successfully accessed the vein, with Fisher's exact test finding no significant difference (p = .360). Finally, for the number of insertion attempts, a Mann-Whitney U-test found no significant differences between men and women (U = 271.94, z = .507, p = .612). These results support our hypothesis (*H3*) that no gender differences in performance in CVC SBT exist post-training.

RQ3: Does the Dunning-Kruger effect exist in CVC SBT post-training

The final research question was developed to determine if the Dunning-Kruger effect existed for men and women in CVC SBT. We hypothesized that there would be no significant relationship between self-efficacy and performance for either gender, indicating the existence of the Dunning-Kruger effect (*H4*) [199]. To test this, regression models were used based on the variable type (continuous or dichotomous) to determine if performance could be predicted based on the aggregated post CLSE. A Bonferroni adjustment was applied to account for repeat testing for the three performance variables [185], resulting in an family wise error rate (α) of .017. For insertion attempts, the linear regression model was unable to significantly predict performance for the whole population based on gender, self-efficacy, and their interaction $F(3,169) = 2.719$, p = .046. When divided by gender, the linear regression models for insertion attempts were unable to significantly predict performance for men, $F(1,109) = 4.214$, p = .042, or for women, $F(1,58) = 3.308$, p = .074 based on self-efficacy. For backwall puncture, the binary logistic regression model was not significant for the whole population $\chi^2(3) = .720$, p = .869. When divided by gender, the binary logistic regression models for backwall puncture were not significant for men, $\chi^2(1) = .570$,

$p = .450$ or for women $\chi^2(1) = .604$, $p = .437$. For successful insertion, the binary logistic regression model was not significant for the whole population $\chi^2(3) = 4.306$, $p = .230$. When divided by gender, the binary logistic regression models for successful insertion were not significant for men, $\chi^2(1) = 1.349$, $p = .245$ or for women $\chi^2(1) = .2.473$, $p = .116$. The results indicating that no models were able to significantly predict performance based on the aggregated post CLSE support our hypothesis (*H4*) that neither men nor women would be able to accurately assess their performance based on confidence. These results indicate that the Dunning-Kruger effect may exist for both genders in CVC SBT.

3.5 Discussion

The objective of this study was to identify if the gender-confidence gap or the Dunning-Kruger effect existed in CVC SBT by comparing self-efficacy and DHRT performance between men and women residents.

The main findings of this study were:

(1) The gender-confidence gap was evident for seven self-efficacy variables on the CLSE with women rating themselves lower both before and after training,

(2) there were no significant differences in CVC SBT performance differences between men and women,

(3) the Dunning-Kruger effect was evident for both men and women

These results support previous literature in obstetrics [215] and general and plastic surgery [97] that found that women had lower self-efficacy than men in training despite there being no performance differences [16-17], leading us to hypothesize that women would rate themselves lower than men [190], [191]. While this was not true for all 14 variables, over half of the variables on the CLSE survey aligned with our expected findings. For all seven significant variables, there were significant negative relationships for women indicating that identifying as a woman leads to a decrease in self-efficacy of varying rates per question. Of the variables that women had lower self-efficacy for, three were related to using tactile feedback, five were related to using the needle, and one was related to overall procedural confidence. Importantly, despite having lower self-efficacy for items related to the use of the needle, there were no actual differences in the ability to achieve successful venipuncture, avoid backwall puncture, or reduce insertion attempts on the DHRT, making these lower self-efficacy ratings unfounded and aligning with previous research

[97], [215], [219]. Additionally, the variable “conducting the entire procedure without mistakes” was significant (p=.002), with women rating themselves lower for this variable, indicating that women may be less confident in their overall ability regardless of SBT exposure and performance. Future work should explore how performance feedback is presented in the DHRT system to combat this finding.

Finally, previous literature has indicated the existence of the Dunning-Kruger effect in medical training for decades [92] positing that new medical trainees are unable to accurately assess their performance [198] regardless of gender [14], [34]. Our results align with previous literature in this area, finding that neither men nor women were able to accurately assess their performance. Regardless of inaccuracies, women still rated themselves lower on the majority of items on the post training CLSE compared to men, suggesting a potential variance in self-rating between genders. To fully explain this finding, a follow-up study should be conducted with a larger, more balanced sample size.

This study focused specifically on self-efficacy and performance for CVC SBT based on training with the DHRT, but the results are reflective of a greater problem with the gender-confidence gap in residency training. Programs have started to highlight resources for physicians to utilize to help fix gender disparities in medicine, including The American Medical Association (AMA) [220]. The field of medical education would benefit greatly by lessening the gender-confidence gap for trainees due to the relationships between confidence and competence in the medical profession [90].

3.6 Conclusion

While medical education has reached gender parity, the gender-confidence gap and the Dunning-Kruger effect are still found to impact self-efficacy at the residency level for SBT. SBT relies on self-efficacy and self-assessed performance to determine if trainees are learning effectively. Self-efficacy has also been shown to impact performance and achievement. As such, this study aimed to identify if the gender-confidence gap or the Dunning-Kruger effect existed in CVC SBT on the DHRT by comparing self-efficacy and performance between men and women. We found that women were significantly more likely to have lower self-efficacy for half of the CLSE survey items, there were no performance differences between men and women on the DHRT, and neither gender was able to predict performance based on self-efficacy for all three performance metrics. Men’s self-efficacy did predict insertion attempts, indicating that the

Dunning-Kruger effect was not found for that variable. Overall, the results of this study provide evidence of both the gender-confidence gap and the Dunning-Kruger effect in CVC SBT on the DHRT. These results indicate an increased need to evaluate gender-differences in resident SBT. Future work should be conducted to further evaluate these findings.

There are several limitations in this study that must be addressed. One limitation of this study was that we did not evaluate gender and race/ethnicity interactions due to the limited sample size in race/ethnicity. Another limitation is that the dataset lacked adequate representation for genders other than men and women and therefore we were only able to study gender as binary. Future work should explore larger sample sizes with more demographic representation to analyze self-efficacy on a larger scale. In addition, this study contained data from only two U.S. medical centers that integrated the DHRT training. As such, the generalizability of the findings is needed across training systems and across institutions. Another limitation of this work is the duality of the Dunning-Kruger effect meaning that it is impossible to know from this study if women were rating themselves lower than men for self-efficacy because they were truly less confident, or if it could be because they were learning more and more aware of where their skills lacked. To validate this, future work should include a longitudinal study to follow their progression of learning throughout training. Finally, the system flow of the DHRT changed between training years modifying how many trials each person needed to complete which may have contributed to changes in self-efficacy between years. As such, this should be explored in future work.

Chapter 4

EVALUATING THE EFFECTS OF COMPREHENSIVE SIMULATION ON CENTRAL VENOUS CATHETERIZATION TRAINING: A COMPARATIVE OBSERVATIONAL STUDY

This paper will be submitted to the journal of BMC Medical Education in March of 2024. This work is multiple authored by Haroula Tzamaras, Dailen Brown, Dr. Jessica Gonzalez-Vargas, Dr. Jason Moore, and Dr. Scarlett Miller. Haroula Tzamaras was the lead author on this paper. Dr. Jason Moore and Dr. Scarlett Miller helped advise this work. Dailen Brown assisted in data collection and Dr. Gonzalez-Vargas helped with data collection and editing.

4.1 Abstract

Simulation-based training (SBT) is vital to complex medical procedures such as Ultrasound Guided Central Venous Catheterization (US-IJCVC), where the experience level of the physician impacts the likelihood of incurring complications. The Dynamic Haptic Robotic Trainer (DHRT) was developed to train residents in CVC as an improvement over manikin trainers, however, the DHRT and manikin trainer both only provide training on one specific portion of CVC, needle insertion. As such, CVC SBT would benefit from more comprehensive training. An extended version of the DHRT was created, the DHRT+, to provide hands-on training and automated feedback on additional steps of CVC. The DHRT+ includes a full CVC medical kit, a false vein channel, and a personalized, reactive interface. When used together, the DHRT and DHRT+ systems provide comprehensive training on needle insertion and catheter placement for CVC. This study evaluates the impact of the DHRT+ on resident self-efficacy and CVC skill gains as compared to training on the DHRT alone. 47 medical residents completed training on the DHRT and 59 residents received comprehensive training on the DHRT and the DHRT+. Each resident filled out a central line self-efficacy (CLSE) survey before and after undergoing training on the simulators. After simulation training, each resident did one full CVC on a manikin while being observed by an expert rater and graded on a US-IJCVC checklist. For two items on the US-IJCVC

checklist, verbalizing consent and aspirating blood through the catheter, the comprehensive training group performed significantly better than the DHRT only group. While both groups showed significant improvements in self-efficacy from before to after training, training group was a significant predictor for using the proper equipment in the proper order, and securing the catheter with suture and applying dressing with the comprehensive training group showing higher post training self-efficacy. Learning gains and self-efficacy were significantly higher for the group with comprehensive training on the DHRT+ for several skills.

4.2 Introduction

For over a decade, simulation-based training (SBT) has been integrated into medical education because it is an imitation of real-life events and procedures that can provide hands-on practice [14] without putting patients at risk [221] and prevent harmful situations caused by unskilled trainees [222]. SBT allows educators to control the situations and tasks that are presented to learners [76]. This control exposes learners to typical and atypical patient cases, increasing their level of preparedness for the variations of patients encountered in the clinical environment [76], [223]. However, SBT is recommended by most residency programs, but it is not required [61]. As a result, there is currently no standardized approach to SBT in medical training [14], [224]. Although not standardized, one procedure that has seen an increase in the use of SBT is Central Venous Catheterization (CVC). With over five million CVCs conducted annually in the United States [5], this procedure involves a catheter insertion for quick and efficient medication delivery through the heart [5], [7]. Typically, CVC is inserted into the right internal jugular vein (IJV) [50], often assisted with ultrasound guidance [6] (US-IJCVC). To conduct US-IJCVC, a series of steps must be followed in order. **Figure 18** illustrates the four main steps and their sub steps, as outlined by the New England Journal of Medicine [32]. Mechanical skills require the manipulation of CVC tools, while procedural knowledge involves the ordering of CVC steps and management of complications [55], [56]. Due to its complexity, CVC is associated with high rates of complication [5], including mechanical complications related to catheter insertion, infectious complications

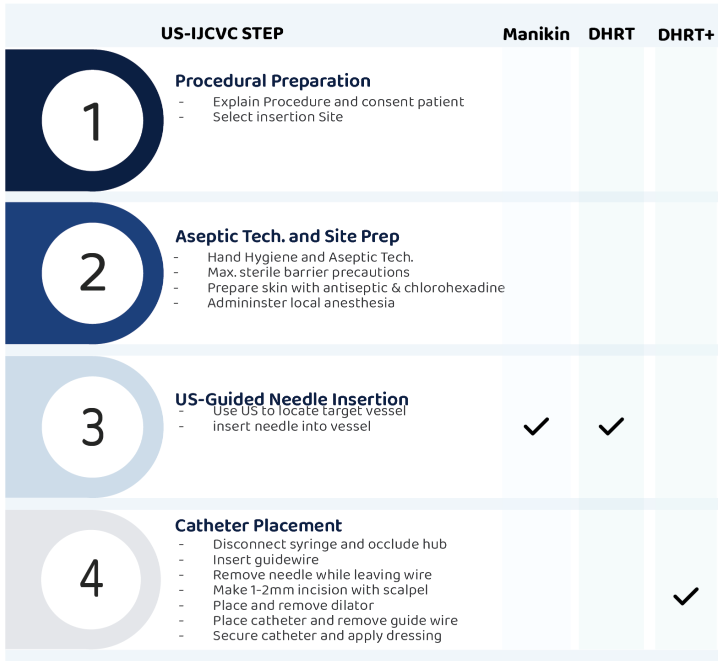

Figure 18: The main steps of CVC and which simulators cover which steps

related to infection, and thrombotic complications arising from clotting in the bloodstream caused by the catheter [9]. One factor contributing to these prevalent complication rates is the experience level of the person conducting the procedure [5], [7], [49]. A physician who has performed less than 50 catheterizations is two times more likely to incur complications than someone with more experience [7]. To reduce these complications, SBT is critical for providing more hands-on practice before performing CVC on patients [34].

Commonly, SBT for CVC includes a manikin trainer with a hand-pump arterial pulse and ultrasound guidance [5], [7] that allow trainees to practice real procedural skills [225] and are physically realistic [205]. However, manikin simulators have several limitations. First, manikin simulators are manufactured to represent only a single patient anatomy, limiting the diverse patient cases presented in the clinical environment. Second, they rely on an instructor to provide performance feedback to the learner [25], [27]. This is challenging as there is currently no standardized approach to training or assessing CVC trainee performance [27], [226], [227]. Instead CVC training and proficiency are determined individually by program [228], leading to potential discrepancies in competence between and within institutions [121]. Lastly, manikin simulators only train residents on vessel identification and access, see Figure 1. Consequently, manikin simulators do not provide practice in all the mechanical skills required for conducting US-IJCVC, including use of the guidewire, scalpel, dilator, and catheter (Figure 1, steps 3 and 4). This training gap is crucial, since a lack of practice on mechanical steps may impact increase the likelihood of mistakes among novice physicians', such as arterial puncture [7] or guidewire mismanagement [55].

To improve upon manikin trainers for US-IJCVC training, researchers developed the Dynamic Haptic Robotic Trainer (DHRT), see **Figure 19** [30]. Specifically, the DHRT provides users with a step-by-step realistic training experience of performing US-guided needle insertion (Figure 1, step 3) [46]. The DHRT includes 17 patient cases that differ based on the IJV size, depth, and location [167], [183]. These variations and force changes during US-guided needle insertion are achieved by using force tissue profiles, a haptic robotic arm, simulated ultrasound screen, and mock ultrasound probe [46]. In addition, the DHRT has a personalized learning graphical user interface (GUI) [31] that provides automated performance feedback on angle, number of insertion attempts, rate of aspiration, and needle centering [229], see figure 2. The DHRT was also validated for training, indicating that is was as effective as manikin simulators based on a US-IJCVC checklist [230]. Moreover, the DHRT distinguishes expert and novice users based on gaze patterns [153], and lessens the learning curve for CVC skills [109]. The DHRT has also indicated that self-efficacy, defined as confidence in oneself for specific tasks and outcomes [125], increases pre to post training [47]. Self-efficacy is a critical measure because evidence shows that performance and self-efficacy can be directly related, and can gauge the effectiveness of learning by the trainee [88] In CVC SBT, a correlation has been observed between confidence, performance, and simulation, indicating that SBT leads to

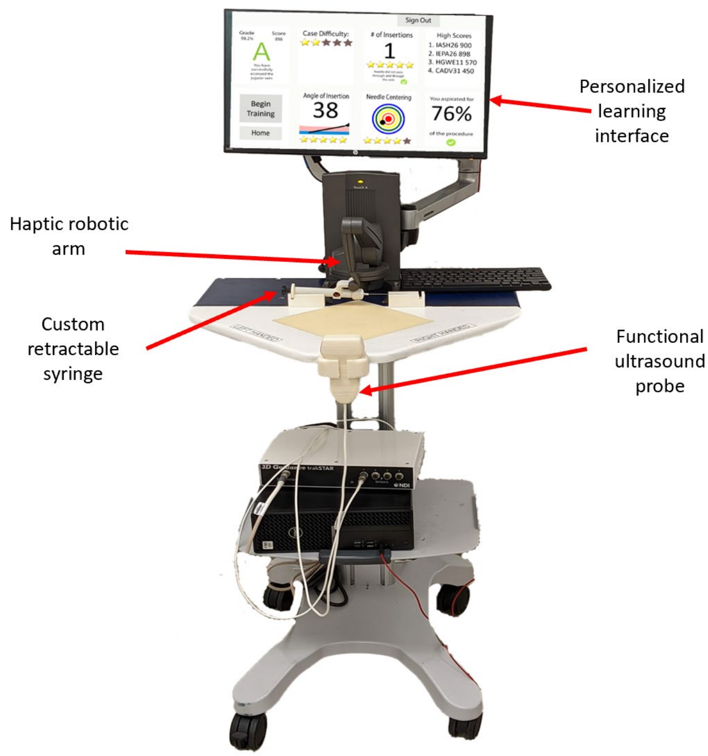

Figure 19: The DHRT system used for CVC training

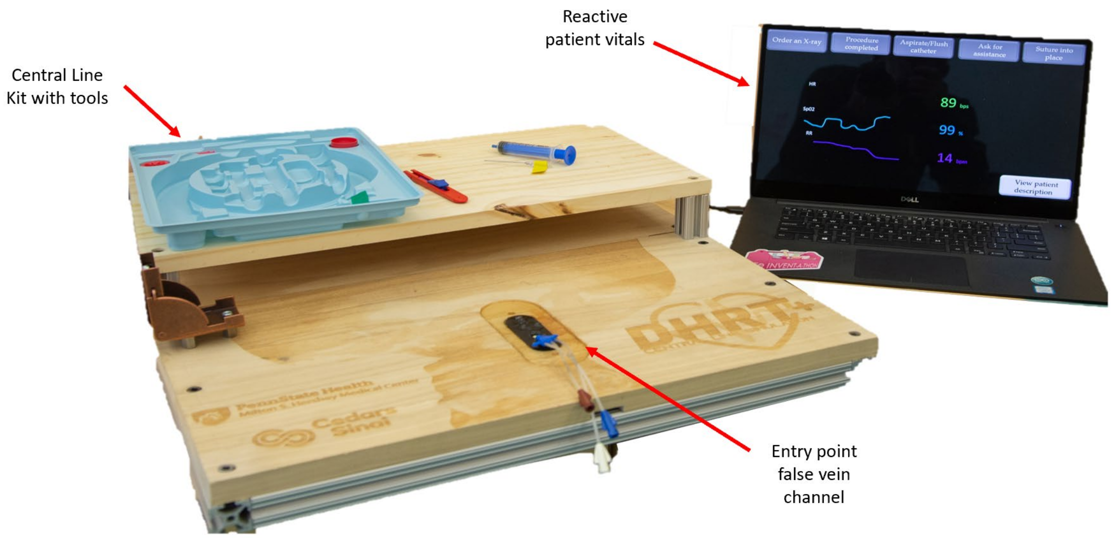

Figure 20: The DHRT+ system (not pictured: overhead camera for computer vision tracking)

confidence and performance increases [126], [210]. However, this correlation has limitations, and trainees may be overconfident or underconfident compared to their actual performance [90], [199], [231].

While the DHRT has been proven to be effective [230], it also only focuses on the mechanical skills for US-guided needle insertion (Figure 1, step 3). To improve upon this and since prior work demonstrated that comprehensive training can accelerate residents experience than those not trained with simulation [232], we sought to develop the DHRT+. The DHRT+ provides in-depth procedural knowledge and mechanical skills training on catheter placement (step 4), see Figure 1. Specifically, the DHRT+ allows users to interact with a real CVC kit (e.g. guidewire, dilator, catheter, scalpel, and suture) and includes an interactive screen that provides patient vitals that react based on performance, see **Figure 20**. Since learning in CVC is not effective without feedback [25], the DHRT+ provides automated performance feedback. Feedback is provided by tracking the order of tools that are used with computer vision, the accuracy of the tool usage (e.g., depth of catheter or guidewire insertion in the vessel) with a false vein channel, and a GUI that provides post-insertion feedback.

To provide a comprehensive US-IJCVC SBT, *when used together*, the DHRT and DHRT+ create a comprehensive training that allows trainees to practice with automated and reactive feedback, covering the critical mechanical skills for US-IJCVC (figure 1, steps 3 and 4). Building on the foundations of this previous work, the main objective of this study was to evaluate if the comprehensive training impacted resident performance and self-efficacy compared to DHRT only training.

4.3 Methods

Data for this study was collected at Hershey Medical Center (HMC) in the summers of 2021 and 2022 at resident training bootcamps during July through September of each year. This study sought to evaluate if the comprehensive training impacted resident performance and self-efficacy compared to the DHRT system alone. The remainder of this section highlights the methodology used to accomplish this objective.

4.3.1 Research Questions

Specifically, the study was designed to answer the following research questions (RQ)s:

RQ1: Is there a difference in performance on a US-IJCVC checklist between residents with comprehensive training on the DHRT+ and residents trained only on the DHRT?

The first research question was developed to determine if adding the DHRT+ to DHRT training led to differences in performance between the two training groups according to expert-observed performance scores on a US-IJCVC. The checklist includes economy of time and motion, the number of insertion attempts taken by the resident, and 24 pass/fail items on the steps of the procedure. We hypothesized that residents in the comprehensive training group would have more efficient movements, lower number of insertions, and higher pass rates on the US-IJCVC checklist than those who had training on the DHRT alone because prior work in other fields of medical education has shown that focusing SBT on the whole procedure positively impacts learning gains and improves trainee performance [233], [234].

RQ2: Is there a difference in self-efficacy between residents with comprehensive training on the DHRT+ and residents trained only on the DHRT?

The second research question was developed to determine if adding the DHRT+ to DHRT training led to differences in self-efficacy between the two training groups, as measured by a central line self-efficacy (CLSE) survey. We hypothesized that self-efficacy would increase pre to

Table 8: Summary of participant medical specialties

	DHRT only	Comprehensive training group	TOTALS
MEDICAL SPECIALTY			
ACUTE CARE	0	1	1
ANESTHESIOLOGY	12	13	25
EMERGENCY MEDICINE	7	8	15
GENERAL SURGERY	0	14	14
INTERNAL MEDICINE	18	17	35
ICU	0	1	1
NEPHROLOGY	1	2	3
NEUROLOGY	3	1	4
OPHTHALMOLOGY	2	0	2
PEDIATRIC CRITICAL CARE	0	1	1
PHYSICAL MEDICINE AND REHABILITATION	1	0	1
PRELIMINARY MEDICINE	1	0	1
PULMONARY	2	0	2
RADIOLOGY	0	1	1
TOTAL	47	59	106

post training for both groups, but would be higher for the comprehensive training group post training. This is because more steps required for CVC [32] were covered with the comprehensive training, and prior research in other medical fields has indicated that more comprehensive exposure to simulation training can increase procedural confidence [126], [235].

4.3.2 Participants

A total of 106 medical residents at Hershey Medical Center participated in the study. In total, there were 42 female participants and 63 male participants. One resident reported their gender as other. Seventy-one residents identified as White, 23 identified as Asian, 3 identified as Black, 2 identified as Hispanic, 1 identified as North African, 3 declined to answer, and 3 identified as more than one race. Of all residents in this study, 14 were general surgery, 25 were anesthesia, 35 were internal medicine, 15 were emergency medicine, and the remainder were various other specialties, see **Table 8**.

4.3.3 Procedures

For all participants, informed consent was obtained according to an Institutional Review Board (IRB) approved protocol. Before coming to the in-person simulation training, all participants completed a pre-simulator online training including a demographic survey, a pre-test on CVC knowledge, eight interactive video modules covering CVC content, and a posttest on CVC knowledge, see [84] for more details on this training protocol. The eight interactive modules included: (1) an introduction to CVC, (2) an overview of CVC steps as defined by the New England Journal of Medicine [32], (3) the benefits and risks of each access site, (4) best practices for CVC equipment, (5) rapid vein assessment with ultrasound, (6) mechanical procedures for troubleshooting, (7) complication types and identification, and (8) patient monitoring and catheter removal. After completing the online training with a posttest score of 80% or higher, residents were able to attend the in-person simulation training. Upon entering the simulation training, residents completed a 19-item 5-point Likert scale central line self-efficacy (CLSE) survey to assess their pre-training confidence on specific skills needed to successfully perform CVC. Next, all residents conducted trials on the DHRT system. In the DHRT only group in 2021, all residents conducted six trials on the DHRT and then went on to fill out the post training CLSE. In the comprehensive training group in 2022, the system was modified to include an interactive walkthrough of the procedure on the DHRT that showed residents how to use the simulator, and to modify the number of trials each resident completed on the DHRT based on performance. Residents who received a 70% or higher and avoided puncturing the carotid artery or backwall of the vein each trial were able to finish the training in as little as three trials, but could do up to six trials if more practice was needed. After the DHRT training, the comprehensive training group did one full US-IJCVC through to the final step of inserting the catheter using the DHRT+ system. After the full US-IJCVC on the DHRT+ system, the comprehensive training group filled out the post training CLSE form. Finally, each resident did one full US-IJCVC on a manikin trainer and

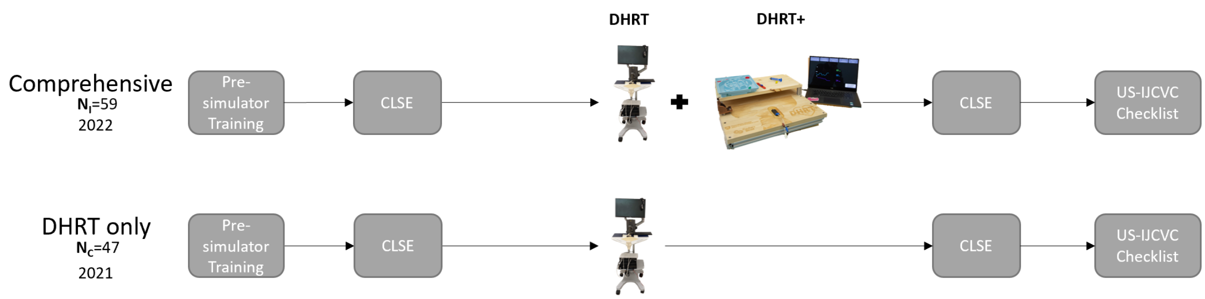

Figure 21: The methodological process of the training flow between the DHRT only and the comprehensive training groups

were scored by an expert observer using a US-IJCVC checklist. The procedure and how it differed between training cohorts can be seen in **Figure 21.**

4.3.4 Metrics

The following metrics were used to assess differences between the comprehensive training and DHRT groups.

US-IJCVC Checklist

The US-IJCVC checklist is a verification of proficiency checklist evaluated by an expert observer, someone who is trained on how to conduct the steps of the procedure. The US-IJCVC checklist is an assessment metric used to determine when a resident is proficient and can move from CVC simulation training to supervised CVC on patients [84]. The checklist includes 2 continuous variables. The first is insertion attempts, or how many tries it took with the needle before the resident was able to successfully insert it into the vein. The second is economy of time and motion, or the efficiency of the hand motion of the resident. The US-IJCVC checklist also includes 24 dichotomous items outlining all the mechanical and procedural steps and sub steps (refer to figure 1). For each dichotomous item, the observer would mark 1 for pass if the resident remembered to do the step *and* conducted it correctly or mark 0 for fail. The full checklist can be found here. If a resident passed every item on the checklist, they were considered competent in the procedure, otherwise they were recommended for further practice.

Central Line Self-Efficacy (CLSE) Survey

The 19-item 5-point Likert scale central line self-efficacy (CLSE) survey is used to assess confidence on skills needed to successfully perform CVC. These items include specific skills such as "locating vessels in an ultrasound image" or "securing the catheter with suture", to more general skills such as "conducting the procedure without mistakes" or "conducting the entire procedure on a simulator". The full CLSE survey can be found here.

4.3.5 Data Analysis

All analysis was conducted in SPSS (v. 29.0). To analyze differences on the US-IJCVC checklist, different statistical tests were run for each variable type. For the continuous variable, number of attempts, and the ordinal variable, economy of time and motion, Mann-Whitney U-tests were run to determine if there were differences between the DHRT and comprehensive training groups. To analyze differences in the 24 dichotomous pass/fail variables, a Pearson Chi-Square was used to test for significant differences in proportions. Fisher's Exact Test was used in place of chi-square for any variable that did not have at least 5 residents fail in both the DHRT only and comprehensive training groups. All assumptions were met for both of these analyses.

To analyze differences in self-efficacy, a General Estimating Equation (GEE) was computed. Training group, CLSE type (pre or post-training), and their interaction were the independent variables and the CLSE questions were the dependent variables. All assumptions were met for GEE. For each variable with a significant interaction term, post hoc pairwise comparisons were conducted via an analysis of estimated marginal means.

4.4 Results

The main objective of this research was to evaluate if comprehensive impacted resident performance and self-efficacy compared to the DHRT system alone. The following results are presented by research question.

RQ1: Is there a difference in performance on a US-IJCVC checklist between residents with comprehensive training on the DHRT+ and residents trained only on the DHRT?

The first research question was developed to determine if comprehensive training with the DHRT+ led to differences in performance for residents. We hypothesized that the additional training would lead to better performance on the US-IJCVC checklist. For insertion attempts, a

Mann-Whitney U test found no significant differences (*U*=1101.5*, z*=-.401*, p*=.688) between the DHRT (Md=1) and the comprehensive training group (Md=1) groups. For economy of time and motion, a Mann Whitney U test found no significant differences (*U*=1466.5, *z*=.696, *p*=.486) between the DHRT (Md=3) and the comprehensive training group (Md=3) groups. For the 24 dichotomous items on the US-IJCVC a Bonferroni correction was applied to account for repeated measures [185], resulting in a family-wise error rate adjusted alpha value of .002. With the Bonferroni correction, Pearson's chi-square indicated that two items on the US-IJCVC checklist differed significantly between the DHRT and comprehensive training groups. For "*verbalizing consent"*, there was a statistically significant difference (χ^2 =14.252, p<.001) between the proportion of residents who passed in the comprehensive training group (86.4%) compared to the DHRT group (53.2%). For "*aspirating blood through the catheter*", there was a statistically significant difference (χ^2 =11.229, p<.001) between the proportion of residents who passed in the comprehensive training group (81%) compared to the DHRT group (50%). Differences for all other dichotomous variables were nonsignificant. These results confirm our hypothesis for two items on the US-IJCVC checklist that the comprehensive training group would have a significantly higher pass rate, and indicate that comprehensive training with the DHRT+ may positively impact performance. Full results from the Pearson Chi-Square and Fisher's Exact test can be found in the appendix.

RQ2: Is there a difference in self-efficacy between residents with comprehensive training on the DHRT+ and residents trained only on the DHRT?

The second research question was developed to determine if comprehensive training with the DHRT+ led to differences in self-efficacy. We hypothesized that self-efficacy would be higher for the comprehensive training group due to comprehensive training in more steps of the procedure covering more mechanical skills and procedural knowledge. To account for the repeated measures of the 19-item CLSE, a Bonferroni correction was applied [185], resulting in a family-wise error rate of .0026. GEE analysis revealed that for two of items on the CLSE survey, treatment was a significant predictor with the comprehensive training group rating higher than the DHRT only group for each one. These variables included "*using the proper equipment in the proper order"* (*Wald* χ^2 =12.258, p<.001), and "*securing the catheter with suture"* (*Wald* χ^2 =16.343, p<.001). While the change from pre to post test was significant for all variables (p<.001), there were significant interactions between the self-efficacy type (pre or post) and training group (DHRT or comprehensive) for one of the items on the CLSE survey, "*placing the needle in multiple attempts*" (*Wald* χ^2 =10.173, *p*=.001). Post hoc analysis via estimated marginal means (SE= .3403, p=.001, 95% CI [-1.752, -.418]), revealed that while the pre-CLSE for this variable was significantly higher for the DHRT group than for the comprehensive training group (Mean difference = .51, *p*=.011), there were no significant differences after training. Overall, these results confirm our hypothesis that the comprehensive training group has higher self-efficacy for two items on the CLSE survey, and indicate that the DHRT+ positively impacts self-efficacy.

4.5 Discussion

The DHRT+ system was developed because existing training methods used in US-IJCVC SBT focus only the US-guided needle insertion portion of CVC (refer to figure 1) [25], [27], indicating a dire need to continuously create more comprehensive US-IJCVC education by covering more steps of the procedure. The main objective of this study was to evaluate if comprehensive training impacted resident self-efficacy and performance rated compared to training on the DHRT system alone. The main findings of this study indicated that

- The comprehensive training group had better US-IJCVC checklist performance for verbalizing consent and aspirating blood through the catheter
- The comprehensive training group had higher self-efficacy for using the proper equipment in the proper order and securing the catheter with suture
- For all other items on the CLSE survey and the US-IJCVC checklist, comprehensive training was as effective as DHRT group since there were no significant differences between the training groups

For the US-IJCVC checklist, we hypothesized that the comprehensive training group would perform better on the checklist with more efficient movements, a lower number of insertions, and a higher pass rate for dichotomous items. This hypothesis was based on prior literature, which indicated that exposure to a more comprehensive training with more steps of US-IJCVC covered would lead to more successful performance [126], [234]. Specifically, since the DHRT+ included training in tool usage and equipment required for US-IJCVC for catheter placement, it was expected that the comprehensive training group would have more efficient hand motions. On the US-IJCVC, there no differences in economy of time and motion or number of insertion attempts between groups, refuting this part of our hypothesis. For two pass/fail items on the US-IJCVC checklist, "*Verbalizing consent*" and "*Aspirating blood through the catheter*", the comprehensive training group had a significantly higher pass rate than the DHRT group, aligning with our hypothesis that there would be a higher pass rate on the US-IJCVC. For the other 22 dichotomous pass/fail items on the US-IJCVC checklist, the DHRT only and the comprehensive training groups performed similarly. While aspirating blood through the catheter was explicitly taught in the comprehensive training during step 4, catheter placement, verbalizing consent was not. This may have been due to residents thinking about the procedure as a whole since more steps were covered; however, further experimentation should be done to verify this effect of comprehensive training. Overall, these results provide evidence that a comprehensive CVC training with more steps of the procedure and automated performance feedback [25], [66] may be more effective for learning than trainers that focus only on needle insertion [233].

For the CLSE survey, we hypothesized that both groups would improve on self-efficacy pre to post training, but that the comprehensive training group would have higher self-efficacy post training because of exposure to comprehensive training. This hypothesis was based on prior literature indicating that SBT leads to confidence increases post training [47], and that more extensive procedural training can better increase procedural confidence [126], [235]. Our results indicated significant improvement from pre to post training for both groups, aligning with previous literature indicating the utility of SBT for US-IJCVC [47]. For two items on the CLSE survey, *"Preparing the proper equipment in the proper order"* and *"Securing the catheter with suture*", the comprehensive training group had significantly higher self-efficacy after training than the DHRT only group. For all other items on the CLSE, the DHRT and comprehensive training group groups both improved significantly from pre to post training with no significant differences in effectiveness. These findings align with prior work that has showed that medical trainee confidence increases with more exposure to simulation and hands-on training [125], [235], [236].

Although, prior work has also indicated that confidence and proficiency in surgical skills increase together [91], this was not the case for the comprehensive training. Specifically, self-efficacy items with higher ratings were not related to higher performance on the US-IJCVC checklist. For example, the residents in the comprehensive training group had a higher pass rate for *"verbalizing consent*", but did not have higher self-efficacy for this item on the CLSE. These findings require further experimentation to determine if comprehensive training may overinflate resident confidence in their ability to perform parts of the procedure [92], as observed in prior work on medical residents and training [93]. Overall, the integration of a comprehensive training by adding the DHRT+ training on the DHRT, shows potential for improving US-IJCVC education.

4.6 Conclusion

The main objective of this paper was to evaluate if the integration of a comprehensive training impacted resident performance and self-efficacy. We found that the inclusion of comprehensive DHRT+ training increased self-efficacy for "*preparing the proper equipment in the proper order*", and "*securing the catheter with suture*". We also found that performance was improved for this group for "*verbalizing consent"* and "*aspirating blood through the catheter*". Future work should focus on validating these findings with a larger sample size, and the integration of the DHRT and the DHRT+ into one comprehensive training tool instead of two separate devices used together.

There were some limitations of the study that must be addressed. For the US-IJCVC checklist and the self-efficacy survey, the data is filled on paper and is prone to human error and sections being skipped or missed. This led to small sample size changing slightly between variables if an observer missed a checkmark on the US-IJCVC; this can be observed in the appendix. Additionally, since they were multiple expert observers for the US-IJCVC checklist, there can be subjectivity in grading. Another limitation is the modification in required trials between training groups, which may have impacted self-efficacy in ways that were not evident from this study. Finally, data was collected at only one medical center in the United States which may limit the generalizability of these results.

Chapter 5

TRACKING SUCCESS: VALIDATING A CENTRAL VENOUS CATHETER TRAINER THROUGH EYE GAZE ANALYSIS IN CLINICAL AND SIMULATED ENVIRONMENTS

This paper will be submitted to the journal of Simulation in Healthcare in March of 2024. This work is multiple authored by Haroula Tzamaras, Joseph Mast, Dr. Lisa Sinz, Dr. Jason Moore, and Dr. Scarlett Miller. Haroula Tzamaras was the lead author on this paper. Dr. Jason Moore and Dr. Scarlett Miller helped advise this work. Drs. Sinz was vital to the data collection in this study. Joseph Mast assisted in video segmentation.

5.1 Abstract

Simulation-based training (SBT) is commonly used in medical education to teach residents before they conduct complex procedures on patients. While SBT is a valuable tool, a common criticism is the lack of robust validation of existing simulators. Two common types of validity are predictive validity and construct validity. Predictive validity is how well future performance can be projected, and can be measured as skill transfer from the simulator to the clinic. Construct validity is how accurately a simulation measures a task, and can be measured as distinction of expertise levels. Validity can be difficult to measure in simulation because there is no standard or expectation for how it should be conducted. Eye-tracking can be used to determine differences in gaze patterns between the clinical environment and a simulator (predictive validity), as well as between novices and experts (construct validity). This study aims to utilize eye tracking to assess predictive and construct validity of a simulator for central venous catheterization, the DHRT+. Experts physicians (N=5) conducted CVC wearing an eye tracker in the operating room for nonemergent cardiac procedures, and on the DHRT+. Novice residents (N=12) conducted CVC wearing an eye tracker on the DHRT+. For analysis, CVC was divided into six standard segments and gaze metrics, fixation count and fixation duration, were assessed. The gaze metrics were compared for experts between the two environments, operating room and simulator, and between experts and novices on the simulator. Paired sample t-tests between the operating room and the DHRT+ indicated predictive validity ($p<.05$) for five out of six segments of CVC. Mann-Whitney

U tests between novices and experts on the DHRT+ indicated construct validity (p<.05) for three out of six segments of the procedure. Based on gaze metrics, the results of this study indicate evidence of both predictive and construct validity on the DHRT+, and help to lay the groundwork for conducting clinical research for simulator validation.

5.2 Introduction

Simulation based training (SBT) has been heavily integrated throughout medical education and residency training [237]–[239], because it allows residents to practice procedures risk-free and build skills and confidence before working with patients [14], [240]. However, a common criticism of SBT is the lack of robust validation from the simulator into the clinical environment [38], [241]. Validation in medical education is important because it ensures that simulators being used to teach new physicians complex medical procedures are teaching them the correct procedure the correct way every time [120], [242]. Two types of validity used for SBT are construct validity, or how accurately a simulation represents the simulated task, and predictive validity, or how well a simulator can project future performance [39], [40], [243]. These two types of validity measures are vital because they are closely related to how well a trainee's skills will transfer from the simulator to the clinical environment [39].

Construct validity is generally measured by variations in performance between individuals (e.g. novice and expert [117]), and the majority of validation studies for simulators focus on construct validity [116], [117], [120], [121], [244], [245]. One way to measure construct validity is through eye tracking as it can be used to determine expertise distinction[137], [138], [143]. Mobile eye trackers are glasses that are worn by a participant and measure when the eye is moving and when it is still; the stillness of the eye is referred to as a fixation [129].This type of eye tracking application has been seen in simulated nerve block insertion [246], laparoscopic simulation [145], [247], anesthesia [248], and healthcare in general[249]. On the other hand, while there has been less focus on predictive validity of medical simulators, it is generally measured through skill gain toward expert level, [250] skill retention, [251] or resident confidence increases [252], [253]. For simulators in other fields, such as driving, predictive validity has been assessed based on performance and observation over time [254], but also to predict workload [255]. Finally, while eye tracking is not a common method of measuring predictive validity in medical simulation, this method has been applied to predict user performance in information technology [256], failure detection [257], and human-robot interaction [258].

Few studies in medical simulation directly compare performance in the operating room (OR) and on a simulator to assess predictive validity through direct skill transfer [38]. One study on intubation compared the performance for intubating actual patients of residents who received simulation training to those who did not, finding that residents trained with simulation were scored higher by expert observers [259]. Another study utilizing eye tracking compared fixations between experienced physicians in the OR and on a simulator for transurethral resection of the prostate, finding that physician gaze metrics did not align between environments and indicating a lack of simulator predictive validity [158]. SBT is used for many medical procedures in training, yet most simulators have not been validated for predictive validity or clinical skill transfer [37]. Understanding clinical transferability of skills is important for knowing when residents are ready to transition from simulators to patients [260].

One such procedure that has lacked validation of SBT is central venous catheterization (CVC). CVC is a complex procedure used to get direct venous access to the heart through the insertion of a catheter [5], [7]; most commonly into the right internal jugular vein using ultrasound guidance (US-IJCVC) [50], [159]. In US-IJCVC, the physician goes through a series of steps to insert a catheter into the IJ. [32]. These steps can vary depending on the physician performing the procedure, but the steps that are most standardized between physicians [71] include (1) using an ultrasound probe to identify the vein, (2) accessing the vein through inserting the needle into the neck, (3) inserting a guidewire through the needle to trace the vessel to the heart, (4) using a scalpel to make a larger incision in the neck where the needle is inserted, (5) using a dilator to widen the IJV, and finally (6) inserting a triple-lumen catheter [261] or a Swan-Ganz [262], [263] in more complex patients [32]. CVC is typically trained with manikin trainers which are beneficial in that they allow hands-on practice with vessel identification and access, indicate if the IJV was properly accessed with a colored liquid, are physically realistic [205], and have been evaluated for construct validity [264]. However, these trainers lack automated performance feedback to allow learning without supervision [27] and variation of patient anatomy [27], [265] reflective of what will be seen in the clinical environment [28]. While construct validity has been proven for manikin trainers, the predictive validity of skills from manikin trainers into the clinical environment is less studied; less than 25% of commercially available simulators specifically provide evidence of predictive validity [38]. To combat the gaps associated with manikin trainers, the Dynamic Haptic Robotic Trainer (DHRT) was developed. The DHRT trains residents in the vessel access and identification steps of CVC using haptic robotic simulation and a mock ultrasound probe [28]. The DHRT has been shown to be as effective of a method for learning CVC as manikin trainers [266], but improves upon manikins by providing objective scoring and real-time feedback [31] without the need for a trained proctor to observe. The DHRT has also been evaluated for construct validity with findings indicating that the system could distinguish expertise based on performance [24] and gaze patterns [109], but has not been evaluated for predictive validity and clinical skill transfer.

The DHRT was updated in 2022 to a comprehensive training simulator, the advanced DHRT (DHRT+) which extends the original DHRT to include a full CVC medical tray, computer vision to track the usage of tools and the order of steps, patient vitals, and a false vein channel to track insertion depth, see **Figure 22**. The DHRT+ is a mixed reality simulator, meaning that while the ultrasound screen and patient vitals are simulated and reactive to trainee performance, the tools used to carry out the procedure are the same as those used in the OR. The following study aims to evaluate the predictive and construct validity of the DHRT+ through gaze comparisons of experts on the DHRT+ and in the clinical environment, and gaze patterns of novices and experts on the DHRT+.

Figure 22: The updated DHRT system used for training CVC

5.3 Methods

To answer these research questions, an empirical study was conducted to compare eye gaze patterns during US-IJCVC across the following conditions: expert in a clinical environment, expert in a simulated environment, and novice in a simulated environment.

5.3.1 Research Questions

The objective of this study was to evaluate the predictive and construct validity of the DHRT+. Specifically, the study was designed to answer the following research questions (RQs):

RQ1: Does the DHRT+ exhibit predictive validity between the OR and the simulated environment?

The first research question was developed to determine if the DHRT+ exhibits predictive validity based on the transferability of gaze between the OR and the DHRT+. A simulator has predictive validity when it is able to accurately predict clinical performance for a specific procedure [39]. Previous studies comparing expert gaze patterns in simulated versus the clinical environment for other procedures found that there were more fixations in the clinical environment but they were shorter than in the simulated environment, and that simulators tend to lack predictive validity to the OR due to the level of distraction and movement in the surgical environment [158].Therefore, we hypothesized that there would be significantly more fixations in the OR with significantly shorter durations for all of the segments of the procedure, indicating a lack of predictive validity and aligning with previous studies.

RQ2: Does the DHRT+ exhibit construct validity as defined by differences in gaze patterns between novices and experts?

The second research question was developed to determine if the DHRT+ exhibits construct validity. Prior research has indicated that construct validity of simulators can be demonstrated through distinguishing novices and experts [116], [117] including research on the original DHRT indicating that eye tracking on the system could distinguish between expert and novice gaze patterns [109], [153]. The purpose of this research question was to examine if the new DHRT+ can distinguish expert and novice performance using eye gaze broken within specific procedural segments. Therefore, we hypothesized that there would be significant differences in gaze between novices and experts for all segments of the procedure, aligning with results of previous studies.

5.3.2 Case selection in the OR

Only non-emergent cardiac cases with pre-planned central lines were considered for inclusion in the study per IRB protocol, subject to both patient and physician consent. Only expert physicians,

Table 9: Participant demographics for experts in the operating room, experts on the simulator, and novices on the simulator

	Experts	Novices
Gender		
Male	4	6
Female	1	6
Race		
Asian	2	2
Hispanic	0	1
White	3	9
More than one race	0	1
Specialty		
Anesthesia	5	0
Internal Medicine	0	3
Neurology	0	1
Pathology	0	1
General Surgery	0	4
Urology	0	1
Plastic Surgery	0	1
Otolaryngology	0	1

defined as attendings or fellows having conducted at least 50 central line insertions into the internal jugular vein were recruited for the OR portion of this study.

5.3.3 Participants

The expert participants were five anesthesiologists from one east coast medical center. One participant identified as female and the rest identified as male. The participants previous experiences with central lines ranged from 75 to 1000+ insertions. The novice participants were 12 PGY1 residents at two medical centers on the east and west coasts of the United States. Of the novice participants, six identified as female and six identified as male. The novice physicians had a range of previous CVC experience with two never having received any training on CVC, two having observed CVC on a person, four having practiced CVC previously on a manikin, three having both observed and practiced on a manikin, and one not reported. A demographic summary of participants can be found in **Table 9.**

5.3.4 Procedure

Three separate procedural flows were followed depending on the level of expertise of the participant and the environment of the study. **Figure 23** outlines the experimental flows followed in this study.

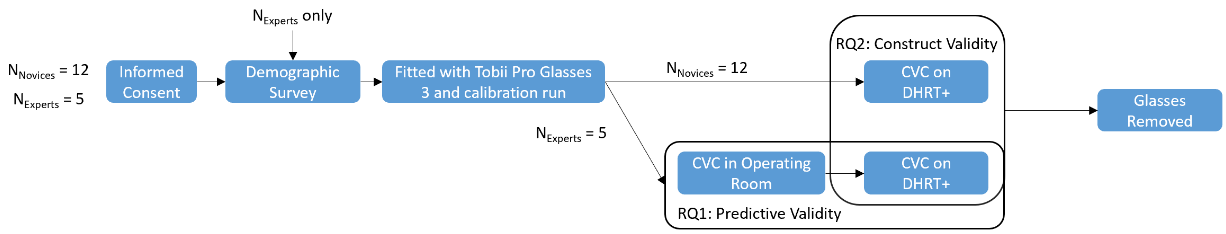

Figure 23: Experimental flow for novices and experts between environments

5.3.4.1 Expert Physicians – Clinic and Simulator

For the clinical portion, at the start of the study, the appropriate Summary Explanations of Research form was disseminated to both the patient and the physician. Physicians were pre-screened to be experts as per the definition of having inserted 50 central lines. After both parties consented, the participating expert filled out a prior experience and demographics survey. After entering the OR, the physician was fitted with Tobii Pro Glasses 3. Proper fit was checked and calibration was run. The physician then underwent the full CVC procedure on the patient. Once the central line insertion was finished and the doctor de-gowned, and the eye tracker was removed. After conducting CVC in the OR, physician then went to the simulation center at a predetermined time. Physicians provided verbal consent for participation in the simulator portion of this study, and were compensated with $50 gift cards, as per IRB protocol. In the simulation center, they were refitted with the Tobii Pro Glasses 3 which was subsequently calibrated to the new environment with the DHRT+. Next, the physician did up to three trials on the DHRT+. Because physicians have a tight clinical schedule, some participants were paged away after one or two trials an unable to complete the three trials. Once the physician completed the trials, the were tracker was removed.

5.3.4.2 Novice Physicians – Simulator Only

The novices in this study participated during their new resident central line bootcamp. The data collected during this study was part of a larger investigation on residency training and CVC. Only the parts of the procedure that are relevant to the current study will be discussed. Upon arriving for training, residents who were randomly selected to participate in novice eye tracking data collection verbally consented to have their eye tracking video and gaze recorded, as per IRB protocol. Then, residents were fitted with the Tobii Pro Glasses 3, fit was checked, and calibration was run. Following the calibration of the eye tracker, residents underwent a training module on the DHRT+ to teach them how to do the procedure and use the simulator. After the training modules, the novice did one full CVC trial on the DHRT+. Finally, the eye tracker was removed.

5.3.5 Metrics

Prior to analysis, the eye tracking recordings, saved as videos, were cleaned so that only fixations actively in the surgical field, ultrasound screen, patient vitals, or medical tray were used. Next, the video was segmented and analyzed as outlined in the remainder of this section. All statistical analysis for this study was conducted in SPSS (v. 29.0).

5.3.5.1 Eye Tracking Video Segmenting

Each eye tracking video was segmented by two independent raters by identifying the start and end point for the US-IJCVC procedure as outlined by the New England Journal of Medicine [32] and the National Library of Medicine [71]. Specifically, the segment breakdown was as follows: (1) "vessel identification" - locating the correct anatomy on the ultrasound screen, (2) "vessel access" - inserting the needle into the IJV, (3) "guidewire insertion"- inserting the guidewire through the needle into the IJV and removing the needle , (4) "scalpel incision" - making an incision in the skin, (5) "dilation" - inserting the dilator into the IJV and removing it, (6) "catheter insertion" - inserting the catheter into the IJV and removing the guidewire. For cases where a Swan-Ganz catheter was used, the sixth segment was removed due to the differences in how this is inserted and where it rests in the body compared the standard triple-lumen catheter [262], [263]. In order to segment the video, two independent raters coded three eye tracking videos into segments in Microsoft Excel by watching videos in Windows Media Player and identifying beginning and ending points of each segment for the clinic (kappa=.718), and for the simulator (kappa=.708). A single rater coded the remainder of the videos to facilitate segmental procedure analysis, or analysis focusing on specific segments based on what is the most standard for the procedure regardless of the patient scenario [267], [268].

5.3.5.2 Eye Tracking Metrics

Eye tracking metrics for both the clinic and were analyzed within each video segment using Tobii Pro Lab (Version 1.171). In Tobii Pro, a fixation was defined by a pause in the movement of the eyes of at least 60 milliseconds with eye movement less than 70 degrees per second (°/s), as recommended for increased effectiveness of eye tracking recording in dynamic situations [269]–[272]. Times of interest based on the defined beginning and ending points of each segment were created in Tobii to assess the fixation metrics within each segment. The fixation metrics of interest for this study were *fixation duration* and *fixation count*.

> *Fixation duration* refers to the total amount of time that a person was fixating in seconds [128], [129], [158], in this case, the times of all the individual fixations within a segment summed.
>
> *Fixation count* refers to the total number of fixations [128], [129], [158], in this case, within each segment.

Each recording was separately analyzed for fixation duration and count within each segment, and then averages were aggregated for each participant within each segment across recordings to account for differences in the total number of trials done by each person according to the clinical schedule and time.

5.4 Results

The main objective of this research was to evaluate the predictive and construct validity of the DHRT+ system through analyzing expert and novice gaze patterns in the OR and on the simulator. The following results are broken down by RQ.

RQ1: Does the DHRT+ exhibit predictive validity between the OR and the simulated environment?

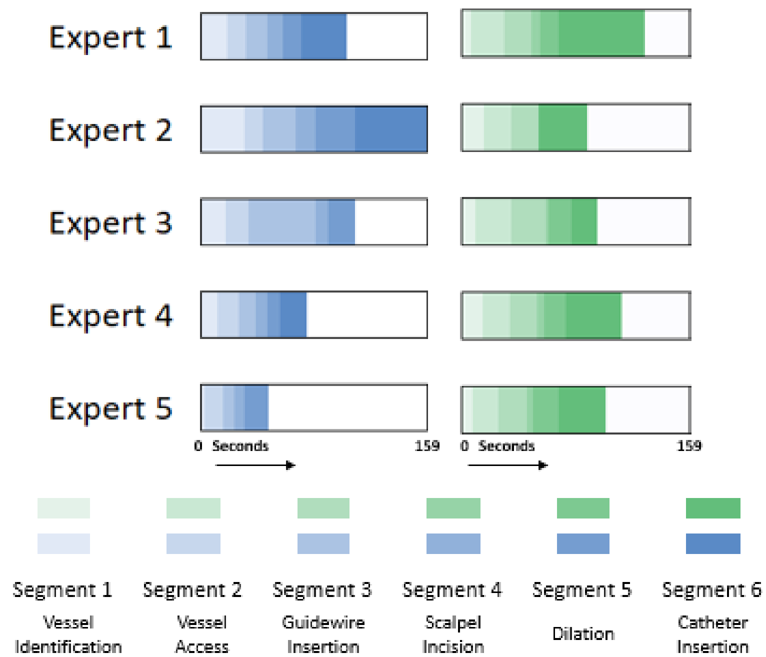

Figure 24: Graph of time measured in seconds depicting the length of time spent fixating by expert physicians in the operating room (left) and on the simulator (right) on each segment of the procedure

The first research question was developed to determine if the DHRT+ exhibits predictive validity based on the transferability of gaze between the clinical environment of the OR and the simulated environment of the DHRT+. We hypothesized that there would be differences in the count of fixations and the overall duration of fixations for each segment of the procedure based on similar findings in prior literature [158]. In order to answer this question, a paired samples t-test was used to examine if there was a statistically significant mean difference between experts in the OR and in the clinic for fixation duration and fixation count, see **Figure 24** for visual summary. Prior to the analysis, assumptions were checked - no outliers were detected and the assumption of normality was not violated, as assessed by Shapiro-Wilk's test ($p > .05$) for each segment. The results showed that for ***segment 2, vessel access***, expert participants fixated significantly longer with the simulator (M = 20.7, SD = 2.9) than in the OR (M = 13.9, SD = 1.4), a statistically significant mean difference of 6.79 seconds, 95% CI[3.95, 9.63], $t(4) = 6.649$, $p < .003$, $d = 2.973$. For ***segment 2, vessel access***, participants also had a higher fixation count with the simulator (M = 41.4, SD = 10.9) than in the OR (M=16.9, SD=4.3), a statistically significant mean difference of 24.43, 95% CI[8.92, 39.93], $t(4) = 4.375$, $p = .012$, $d = 1.956$. No other significant differences were found between environments in any other segments ($p > .05$). Summary statistics for each

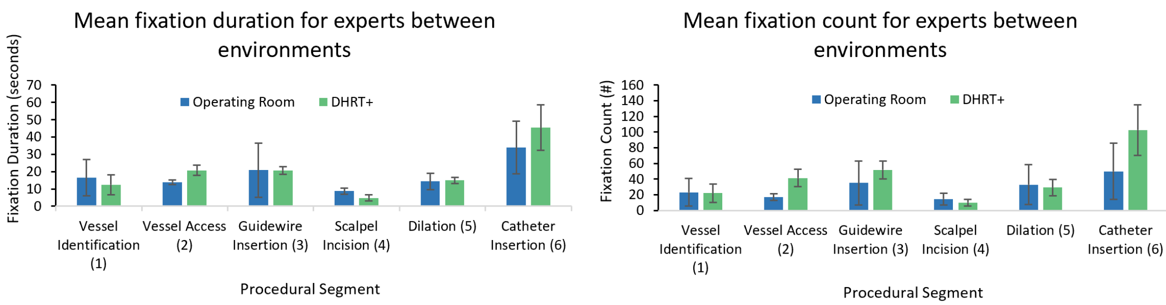

Figure 25: Mean values for fixation duration and fixation count with standard deviation bars for experts in the operating room and on the DHRT+

environment broken down by segment can be seen in **Figure 25**. For five of the six segments, these results indicate predictive validity of the DHRT+, and refute our hypothesis that fixation duration and count would be different in the OR than on the DHRT+.

RQ2: Does the DHRT+ exhibit construct validity as defined by differences in gaze patterns between novices and experts?

The second research question was developed to determine if the DHRT+ exhibits construct validity through distinguishing novices and experts based on gaze. We hypothesized that there would be significant differences in gaze between novices and experts for all segments of the procedure, aligning with results of previous studies indicating that simulators can distinguish expert performance based on eye tracking metrics [153]. Summary statistics for novices and experts for each segment can be found in **Figure 26**.

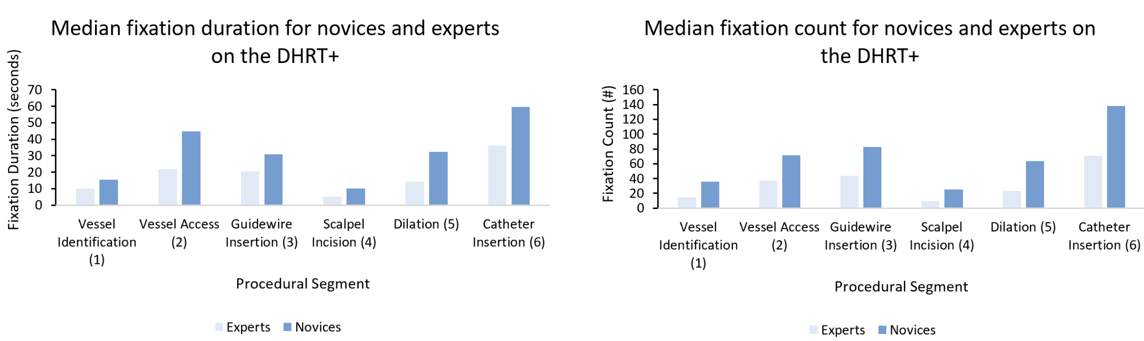

Figure 26: Median values fixation duration and fixation count for experts and novices on the DHRT+

To answer this question, a Mann-Whitney U test was run to determine if there were differences between novices and experts for fixation duration and fixation gaze. Distributions of the gaze patterns were similar, as assessed by visual inspection. The results showed that median fixation duration for ***segment 2, vessel access***, was significantly higher (U = 1.00, z = -3.057, p = .001) for the novices (44.7s) than the experts (21.9s). Median fixation count for vessel access was also significantly higher (U = 4.0, z = -2.744, p = .004) for novices (71.5 fixations) than for experts (37 fixations). For ***segment 4, scalpel incision***, median fixation count was significantly higher (U = 2, z = -2.672, p = .004) for novices (25 fixations) than for experts (9.5 fixations). For ***segment 5, dilation,*** median fixation duration was significantly higher (U = 0, z = -1.9792, p = .082) for novices (32.2s) than for experts (14.3s); Similarly, median fixation count was significantly higher (U = 1, z = -2.793, p = .002) for novices (63.5 fixations) than for experts (23 fixations). There were no significant differences between novices and experts found for segment 1, 3, or 6 (p > .05). These results align with our hypothesis that the DHRT+ exhibits construct validity, but refute our hypothesis that construct validity would be evident for all segments of the procedure.

5.5 Discussion

The main objective of this study was to evaluate the predictive and construct validity of the DHRT+ system through analyzing expert and novice gaze patterns in the OR and the DHRT+. The main findings of this study were

(1) the DHRT+ exhibits predictive validity for 5 of 6 segments of the procedure

(2) the DHRT+ exhibits construct validity in 3 of 6 segments of the procedure

The first hypothesis of this study was evaluating the predictive validity of the DHRT+. Based on limited prior literature directly comparing physician gaze between the OR and a simulator [158], we hypothesized that the simulator would not exhibit predictive validity due to the distractions of live surgery significantly changing the physician's gaze during the procedure. Unexpectedly, the results of this study indicate that physician gaze was not significantly different between the DHRT+ and the OR. ***Segment 2, vessel access*** was the only segment that had a significant difference between environments for both count and duration showing both a significantly higher average count (41.4 vs 16.9) and duration (20.7 vs 13.9) on the simulator. Theneedle insertion portion of CVC on the simulator involves the use of a modified syringe and requires the participant to hold the needle steady after venous access for five seconds while a bar on the screen fills up to indicate success. It is possible that this requirement contributed to the added duration and count on the simulator for this step. Additionally, it is likely that more time was needed to identify and track the needle on the ultrasound screen due to the differences between a real ultrasound and the simulated ultrasound; further experimentation is needed to verify this finding. However, the lack of difference between gaze in the OR and the simulator for all other segments indicates the utility of including real tools with simulated training to increase the transferability to the OR and predictive validity of the simulator. These findings indicate that the DHRT+ demonstrates predictive validity based on gaze patterns [25], [39], and provides information for future studies about methods for assessing predictive validity through clinical skill transfer between the OR and a simulator.

There were significant gaze differences between novices and experts for ***segment 2, vessel access, segment 4, scalpel incision, and segment 5, dilation***, of the procedure as identified by both fixation duration and fixation count. These results indicate construct validity for these segments, aligning with previous literature indicating that gaze distinguishes expertise in simulation training [137], [246], [267], [268]. For these differences, the novices had higher fixation durations and counts. For all other segments, there were no significant differences between novices and experts, refuting our hypothesis of construct validity for ***segment 1, vessel identification, segment 3, guidewire insertion, or segment 4, catheter insertion***. These results also refute previous research on gaze with the original DHRT system [153], which indicated gaze distinction between novices and experts for vessel identification and vessel access. The lack of significant differences between novices and experts for all segments in the current study is likely due to the uneven sample size leading to the study being too underpowered to see significant results across segments. Future experimentation with a larger sample size should be conducted to further validate this finding. Overall, these findings align with our hypothesis that the DHRT+ demonstrates construct validity several of segments of the procedure [117]. These findings also add to previous literature on the use of eye tracking for prediction of future performance, and the application of this method to medical simulation.

5.6 Conclusion

The goal of this research was to evaluate the predictive and construct validity of the DHRT+ system through analyzing expert and novice gaze patterns in the OR and the DHRT+. The first main takeaway from this study is that the DHRT+ exhibits predictive validity between the OR and the simulator based on expert gaze patterns for five of six segments of CVC. Secondly, construct validity is evident in the DHRT+ for half of the segments of the procedure based on comparisons between novices and experts. Overall, these results indicate that the DHRT+ is a valid teaching method for CVC based on both construct and predictive validity, and that including hands-on components in virtual simulator training may increase the transferability of clinical skills from the simulator to the OR. This study also helps lay the groundwork for conducted research in the OR to contribute to the current state of simulator validity in medical education.

This study has several limitations that must be addressed. Due to the nature of part of this study taking part in the clinical environment, one limitation was that the data collection was dependent on the availability of clinicians for nonemergent procedures, which in this case biased all expert participants into one specialty. Another limitation of the study was reading errors with the Tobii Pro Glasses. The glasses can only capture gaze when the participant was looking directly through the lenses; if they were looking above or below the rim of the glasses, the glasses did not register their gaze caused missing datapoints, or if the participant had dry eyes or other eye problems their gaze was not tracked accurately and data had to be voided. Additionally, sample size of the expert population was largely limited by the number of non-emergent CVC procedures, the availability of the onsite researcher, and the availability of the expert physicians. The initial recruited sample size of physicians was larger, but some ultimately had to be cut due to issues with the eye tracking. A larger sample size should be explored in future work. Adding to this, due to the focus of this paper being the tool usage in the six segments of the procedure, and the complexity of analysis when wanting to include multiple areas of interest, this paper focused on independent gaze metrics and did not breakdown fixations into areas of interest within each segment. This should be visited in future work. Finally, this paper focused on just fixation duration and segments, and did not include other useful gaze metrics such as pupil dilation or saccades.

Chapter 6

Conclusions

Based on rates of complication associated with US-IJCVC [52], the focus of current training methods on vessel identification and access[25], [30], [32], and the lack of simulator validation in medical education [39], [120], there exists a clear need for innovative training methods to be explored to transform US-IJCVC. Improved training methods can ensure that physicians are learning efficiently and effectively, and that the skills being taught through SBT are adequately transferable from the simulation center to the operating room when trainees begin to work with patients. Previous work in these areas has shown the benefits of increasing the usage of SBT in medical education [181], [273], focusing on mastery-based learning [80], [213], and validating simulators for learning to ensure that they are teaching the correct constructs [116], [158]. The lack of validation of new simulation technology for prediction of skill transfer into the operating room is a big criticism of SBT in medical education [38]; despite this, there is still limited research on this US-IJCVC. It is necessary to assess current training methods, and develop a more comprehensive, validated US-IJCVC SBT to ensure skill transferability, efficiency of learning, and effectiveness of SBT. To improve the state-of-the-art of SBT for CVC, this dissertation focused on the transformation of US-IJCVC education with a systems-thinking approach, through the assessment of existing US-IJCVC SBT methods, and the development and validation of an improved, comprehensive SBT US-IJCVC simulator. **Figure 27** summarizes the main findings of this dissertation as they relate to assessment, development, and validation.

Specifically, paper 1 (chapter 2) explored methods of efficient learning by assessing the impact of a sequential learning-based interactive walkthrough for US-IJCVC on initial skill gain and performance learning curves of residents trained with and without sequential learning. Paper 2 (chapter 3) explored the gender-confidence gap and the Dunning-Kruger Effect in SBT for US-IJCVC and how they relate to resident performance. Paper 3 (chapter 4) introduced a new, comprehensive version of the DHRT system, the DHRT+, that contains more steps of US-IJCVC than previous simulators, and compared the proficiency and self-efficacy of residents trained with the comprehensive training to those without. Finally, paper 4 (chapter 5) utilized eye tracking and

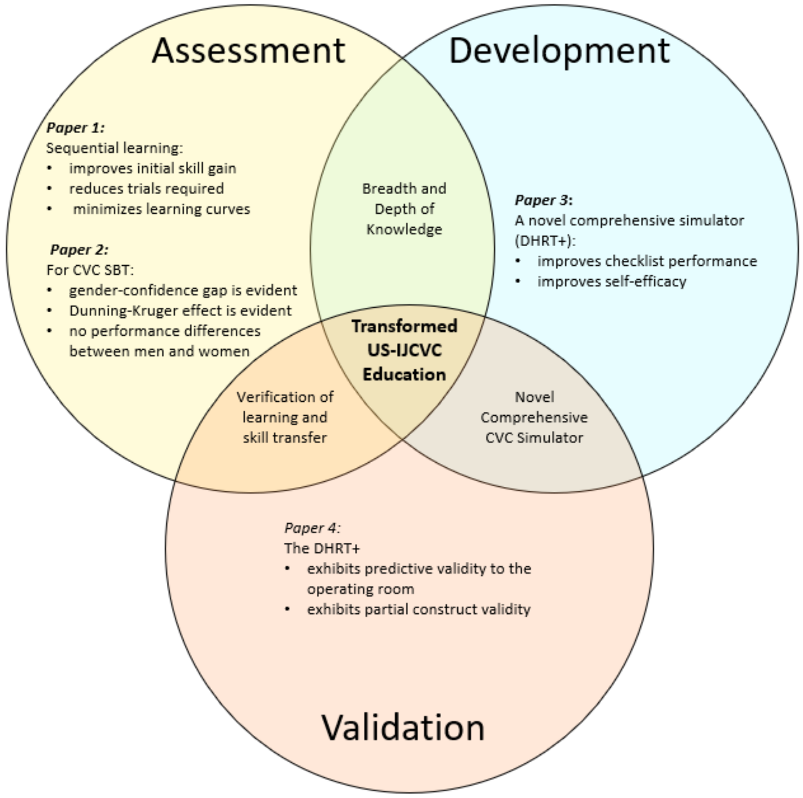

Figure 27*:* Summary of the areas of research investigated in this dissertation and the main findings of each paper

gaze metrics to validate the comprehensive DHRT+ by assessing both how well the simulator measured the construct being trained through comparing expert and novice gaze (construct validity), and if the simulator could predict how someone would perform in the operating room based on simulator performance (predictive validity).

This dissertation provides a roadmap for how SBT for US-IJCVC can be structured to be efficient, comprehensive, and validated. It lays groundwork for conducting validation studies in the operating room using eye tracking devices, which there is limited existing research on. The results found in this dissertation have been disseminated in conferences and will be submitted as journal articles to add to the existing body of knowledge surrounding SBT and US-IJCVC. Conferences where this research has already been presented include the Association for Surgical Education, American College of surgeons, the International Meeting of Simulation in Healthcare, and the Annual Meeting for the Human Factors and Ergonomics Society. Hopefully, the dissemination of these results will inspire other researchers to explore learning methods for SBT, clinical validation methods, eye tracking, and more comprehensive simulation training for various procedures.

6.1 Contributions

The following sections of this chapter summarize the contributions of the four papers presented in this dissertation and provide insight into future areas of research.

6.1.1 Adding sequential learning to the DHRT significantly increased initial skill gain, decreased the number of trials required to complete training, and reduced learning curves

The first contribution of this dissertation was a demonstration that sequential learning is a more efficient method for teaching US-IJCVC on the DHRT. Paper 1 (chapter 2) exhibits that residents trained with sequential learning had a significant higher likelihood of successfully inserting the needle on their first trial on the DHRT system. This result provides evidence that utilizing sequential learning is more efficient for training, because residents are able to recognize patient anatomy and adequately use the needle and ultrasound faster than they were with traditional DHRT methods. Results from this paper also showed that the number of trials required to reach proficient performance on the DHRT was significantly reduced with the implementation of sequential learning. This result provides evidence that implanting sequential learning can allow residents to progress through training faster, which also has the potential to lessen the strain on resident training hours. Lastly, results also showed that sequential learning improved the efficiency of skill gains by demonstrating fewer significant learning curves than training on the original DHRT. This provides evidence that sequential learning can minimize the learning curve and allow residents to reach an expert level of performance at a faster rate. Overall, these results provide evidence that sequential learning is ***a more efficient*** training method for SBT than traditional teaching methods. These results could also be applied to simulators to help minimize the learning curves of other procedures.

6.1.2 The Gender-Confidence Gap and the Dunning-Kruger Effect exist in SBT for US-IJCVC despite no significant performance differences

The second contribution of this dissertation provides evidence of the gender-confidence gap and the Dunning-Kruger effect in SBT for US-IJCVC. Paper 2 (chapter 3) demonstrates that despite men and women both having significant increases in self-efficacy on the 14-item CLSE survey from pre- to post-SBT, women residents rated their self-efficacy significantly lower than their peers both before and after training for 9 out of 14 items. This result indicates that SBT is useful for increasing self-efficacy in US-IJCVC; however, there is a gender-confidence gap in this increase and women have lower self-efficacy for US-IJCVC skills. This provides evidence that even when receiving the same training, women will still rate themselves lower than men. The second result of this paper showed that there were no performance differences on the DHRT between men and women for any metric. This result provides evidence that men and women can both learn effectively from SBT for US-IJCVC, despite women having lower self-efficacy. Finally, the last finding from this study indicated that neither men or women's self-efficacy is significantly correlated with their performance, demonstrating the existence of the Dunning-Kruger Effect. This provides evidence that at the residency level, men and women are unable to accurately assess their own performance, therefore, more objective performance measures other than self-efficacy should be used for evaluation. Overall, the results of this paper provide evidence that the gender-confidence gap exists in SBT for US-IJCVC ***despite no differences in performance*** and that ***medical residents cannot accurately self-assess*** for US-IJCVC.

6.1.3 Comprehensive simulation with the DHRT and DHRT+ system is more effective than the original DHRT training focusing solely on vessel access and identification

The third goal of this dissertation demonstrates the effectiveness of comprehensive simulation for US-IJCVC over the original DHRT that focused only on two main parts of the procedure. Paper 3 (chapter 4) indicates that on a 24-item US-IJCVC proficiency checklist, residents trained on the combined DHRT and DHRT+ performed significantly better than residents trained on the DHRT system alone for two items on the checklist, and no significant differences were found for other items. This result provides evidence that more comprehensive simulation has the potential to improve overall US-IJCVC proficiency and make residents more prepared for the clinical environment. This study also showed that for self-efficacy assessments on the 19-item CLSE survey, residents trained with the DHRT and DHRT+ had significantly higher self-efficacy for two of the items on the CLSE, and no significant differences were found for other items. This result indicates that the more comprehensive training with the DHRT+ exposing residents to more parts of the procedure can improve US-IJCVC self-efficacy. Overall, these results indicate the utility of expanding and developing US-IJCVC to be more comprehensive with the DHRT+ by demonstrating that comprehensive training is ***more effective*** for improving US-IJCVC checklist performance and self-efficacy than the original DHRT system.

6.1.4 The DHRT+ system exhibits construct validity by distinguishing expert and novice gaze and predictive validity by aligning expert gaze between the DHRT+ and the operating room

The fourth and final contribution of this dissertation demonstrates the predictive validity between the DHRT+ and the operating room and construct validity of the DHRT+ system between novices and experts. Paper 4 (chapter 5) indicates that the DHRT+ exhibits predictive validity for 5 out of 6 procedural segments of US-IJCVC in the operating room by finding no significant differences between environments for expert physicians. This result indicates that skills gained through training on the DHRT+ should transfer from the simulator to the clinical environment. Results also indicated construct validity of the DHRT+ by showing that for 3 of 6 procedural segments of US-IJCVC, expert and novice gaze patterns significantly differed. This result provides evidence that the DHRT+ is able to accurately differentiate expert and novice performance based on gaze. Overall, these results provide methodology for conducting predictive validity studies in the operating room, and provide evidence that the novel, comprehensive ***DHRT+ system exhibits predictive and construct validity.***

6.2 Limitations and Future Directions

While this dissertation provides valuable evidence for sequential learning, the gender-confidence gap, comprehensive training, and validation of SBT in US-IJCVC, there are several limitations to this dissertation that should be noted. First, data collected in this dissertation was from two medical centers on the east and west coast of the United States that have implemented the DHRT and DHRT+ simulators into their residency trainings. As such, the generalizability of these findings to other medical centers and training devices may be limited. Second, the sample size for many of these studies is small. This is due to the data being collected as part of existing residency trainings at these medical centers limiting the availability of medical residents to participate. Third, for research conducted in the clinic, the hours and availability of physicians, as well as the surgical schedule, largely dictated when data could be collected and which physicians would be performing US-IJCVC, biasing the participants to one specialty. Fourth, the US-IJCVC checklist and the CLSE survey are both filled out on paper, and as such are prone to human error and people skipping questions. This both limited the sample size, and caused changes in population sizes for some parts of this study. For the eye tracking portion of this dissertation, eye tracking analysis focused on independent fixations, and did not breakdown fixations into areas of interest. Future work should add this to further understand differences between expertise levels and environments. Finally, this dissertation employed Tobii Pro Glasses, which have limitations in function including those dependent on the dryness of the participant's eyes, which caused some data to be unusable.

There are several future directions that this work could take. First, the results on sequential learning could be used within the DHRT to modify training. When considering the result that the training takes the same amount of time on average when including the walkthrough as it does when residents just immediately start on trials, it is possible that the walkthrough could be expanded without impacting training time significantly. One potential avenue could be to add another module that walks the trainee through their own performance and ensure that they understand the portions that they need to improve on. Another avenue could be finding a way to better help the people who are still struggling after maxing out the training with 6 trials. For example, introducing another walkthrough type of module or help for residents who are scoring below a certain amount each trial before they reach trial 6. Future work on the DHRT+ should also include analysis of the parameters set for passing the training, including both the 70% score required to move forward and the 3-trial minimum. Based on the increase in efficiency from the sequential learning, it is possible that the 3-trial minimum could be decreased. A large-scale assessment of resident learning after the sequential learning has been used in further training sessions would be useful for re-evaluating these parameters. A future study, once more people have been trained on the sequential learning system, comparing if the number of trials conducted during sequential learning the score on the US-IJCVC checklist could be useful to further validate the adaptive learning method. This could also be useful for further verifying that the DHRT is a fair measure of real-world practice and clinical skill. Better understanding the accuracy of the DHRT as compared to actual clinical skill transfer would give insight into whether the passing score could be set to 100% for future training. Additionally, the sequential learning method could be applied to other procedures and simulators. While this dissertation focused on US-IJCVC, there are other needle-based and non-needle-based procedures that could benefit from increasing the efficiency of learning and decreasing learning curves. It would also be interesting to see sequential learning studies with larger sample sizes to evaluate how sample size may impact learning gains and learning curves. This dissertation also focused on group learning curves rather than individual learning curves, which could be another direction for future research. Second, continued exploration of the gender-confidence gap in US-IJCVC and in SBT in general would likely benefit women residents by further evaluating how the gender-confidence gap could be mediated. Self-efficacy and confidence are sometimes used as evidence of learning, and women are at a detriment in this case because even though they perform the same as their men counterparts, their self-efficacy is lower, likely due to external factors. A longitudinal study to determine if there is a true gender-confidence gap or if women may actually be learning better would be beneficial for better understanding why there were gender differences in self-efficacy before and after CVC simulation training. Third, for comprehensive training on the DHRT+, a study with a larger sample size may provide even more evidence of the utility of this learning method. Additionally, future work could focus on the expansion from part task training to comprehensive training in other procedures. Additionally, this dissertation explores taking a systems-thinking approach to simulation by considering the design process as a whole from development, to validation and actual use in training. Further exploration of this approach as well as including the lifecycle approach and more stages of medical education would be interesting and potentially beneficial to medical education as a whole. Finally, this research lays groundwork for conducting validation studies for SBT in the operating room. There is little research on this in the literature. For this type of research, it would be interesting to incorporate more metrics of gaze, including pupil dilation. A follow-up study focused on pupil dilation and cognitive load would be interesting to understand if the mental load on physicians is the same in the operating room and on the simulator. This study would need to be controlled for environmental factors such as lighting of the room, and baseline measures would need to be taken for each participant. Utilizing these metrics could help get an even fuller understanding of simulator validity. Future work could focus on increasing the amount of studies utilizing eye tracking to measure predictive validity and clinical skill transfer to ensure that we are teaching medical residents as accurately as possible. Overall, this research provides several avenues for future work that could be explore in SBT for US-IJCVC and beyond. This work was supported by the national Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) under Award Number RO1HL127316. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Coauthors Dr. Moore and Miller owns equity in Medulate, which may have a future interest in this project. Company ownership has been reviewed by the University's Individual Conflict of Interest Committee.

REFERENCES

[1] O. Anderson, R. Davis, G. B. Hanna, and C. A. Vincent, “Surgical adverse events: A systematic review,” *Am. J. Surg.*, vol. 206, no. 2, pp. 253–262, 2013, doi: 10.1016/j.amjsurg.2012.11.009.

[2] J. W. Suliburk *et al.*, “Analysis of Human Performance Deficiencies Associated with Surgical Adverse Events,” *JAMA Netw. Open*, vol. 2, no. 7, pp. 1–12, 2019, doi: 10.1001/jamanetworkopen.2019.8067.

[3] J. Reason, “The contribution of latent human failures to the breakdown of complex systems,” *Hum. Error Aviat.*, vol. 484, pp. 5–14, 2017, doi: 10.4324/9781315092898-2.

[4] C. Skelly, M. Cassagnol, and S. Munakomi, “Adverse Events,” *StatPearls*, 2022.

[5] R. W. Taylor and A. V. Palagiri, “Central venous catheterization,” *Crit. Care Med.*, vol. 35, no. 5, pp. 1390–1396, 2007, doi: 10.1097/01.CCM.0000260241.80346.1B.

[6] P. Brass, M. Hellmich, L. Kolodziej, G. Schick, and A. F. Smith, “Ultrasound guidance versus anatomical landmarks for internal jugular vein catheterization,” *Cochrane Database Syst. Rev.*, vol. 2017, no. 6, 2015, doi: 10.1002/14651858.CD006962.pub2.

[7] D. C. McGee and M. K. Gould, “Preventing complications of Central Venous Catheterization,” *N. Engl. J. Med.*, vol. 348, no. 12, pp. 1123–1133, 2003, doi: 10.1016/j.jamcollsurg.2007.01.039.

[8] J. I. Sznajder, F. R. Zveibil, H. Bitterman, P. Weiner, and S. Bursztein, “Central Vein Catheterization: Failure and Complication Rates by Three Percutaneous Approaches,” *Arch. Intern. Med.*, vol. 146, no. 2, pp. 259–261, 1986, doi: 10.1001/archinte.1986.00360140065007.

[9] R. E. Kusminsky, “Complications of Central Venous Catheterization,” *J. Am. Coll. Surg.*, vol. 204, no. 4, pp. 681–696, 2007, doi: 10.1016/j.jamcollsurg.2007.01.039.

[10] S. Y. Tan and P. Uyehara, “William stewart Halsted (1852-1922): Father of American surgery,” *Singapore Med. J.*, vol. 51, no. 7, pp. 530–531, 2010.

[11] W. T. M. Mason and P. W. Strike, “See one, do one, teach one - Is this still how it works? A comparison of the medical and nursing professions in the teaching of practical procedures,” *Med. Teach.*, vol. 25, no. 6, pp. 664–666, 2003, doi: 10.1080/01421590310001605705.

[12] R. W. Barnes, N. P. Lang, and M. F. Whiteside, “Halstedian technique revisited. Innovations in teaching surgical skills,” *Ann. Surg.*, vol. 210, no. 1, pp. 118–121, 1989, doi: 10.1097/00000658-198907000-00018.

[13] S. V. Kotsis and K. C. Chung, “Application of the ‘see one, do one, teach one’ concept in surgical training,” *Plast. Reconstr. Surg.*, vol. 131, no. 5, pp. 1194–1201, 2013, doi: 10.1097/PRS.0b013e318287a0b3.

[14] J. Vozenilek, J. S. Huff, M. Reznek, and J. A. Gordon, “See one, do one, teach one: Advanced technology in medical education,” *Acad. Emerg. Med.*, vol. 11, no. 11, pp. 1149–1154, 2004, doi: 10.1197/j.aem.2004.08.003.

[15] R. Datta, K. K. Upadhyay, and C. N. Jaideep, “Simulation and its role in medical education,” *Med. J. Armed Forces India*, vol. 68, no. 2, pp. 167–172, 2012, doi: 10.1016/S0377-1237(12)60040-9.

[16] T. M. Maus, “Simulation: The importance of ‘hands-on’ learning,” *J. Cardiothorac. Vasc. Anesth.*, vol. 25, no. 2, pp. 209–211, 2011, doi: 10.1053/j.jvca.2011.01.007.

[17] A. Khaled, J. Gulikers, H. Biemans, M. van der Wel, and M. Mulder, “Characteristics of hands-on simulations with added value for innovative secondary and higher vocational education,” *J. Vocat. Educ. Train.*, vol. 66, no. 4, pp. 462–490, 2014, doi: 10.1080/13636820.2014.917696.

[18] D. J. Grant and S. C. Marriage, “Training using medical simulation,” *Arch. Dis. Child.*, vol. 97, no. 3, pp. 255–259, 2012, doi: 10.1136/archdischild-2011-300592.

[19] T. P. Grantcharov, L. Bardram, P. Funch-Jensen, and J. Rosenberg, “Learning curves and impact of previous

operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills," *Am. J. Surg.*, vol. 185, no. 2, pp. 146–149, 2003, doi: 10.1016/S0002-9610(02)01213-8.

[20] B. Law, A. J. Lomax, M. S. Atkins, C. L. Mackenzie, and A. E. Kirkpatrick, "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment," *Eye Track. Res. Appl. Symp.*, vol. 1, no. 212, pp. 41–47, 2004, doi: 10.1145/968363.968370.

[21] C. S. Kulkarni *et al.*, "Scene - dependent , feedforward eye gaze metrics can differentiate technical skill levels of trainees in laparoscopic surgery," *Surg. Endosc.*, 2022, doi: 10.1007/s00464-022-09582-3.

[22] S. Cooper *et al.*, "Simulation based learning in midwifery education: A systematic review," *Women and Birth*, vol. 25, no. 2, pp. 64–78, 2012, doi: 10.1016/j.wombi.2011.03.004.

[23] R. Ramlogan, A. U. Niazi, R. Jin, J. Johnson, V. W. Chan, and A. Perlas, "A virtual reality simulation model of spinal ultrasound: Role in teaching spinal sonoanatomy," *Reg. Anesth. Pain Med.*, vol. 42, no. 2, pp. 217–222, 2017, doi: 10.1097/AAP.0000000000000537.

[24] D. F. Pepley *et al.*, "Training Surgical Residents With a Haptic Robotic Central Venous Catheterization Simulator," *J. Surg. Educ.*, vol. 74, no. 6, pp. 1066–1073, 2017, doi: 10.1016/j.jsurg.2017.06.003.

[25] J. H. Barsuk, W. C. McGaghie, E. R. Cohen, J. S. Balachandran, and D. B. Wayne, "Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit," *J. Hosp. Med.*, vol. 4, no. 7, pp. 397–403, 2009, doi: 10.1002/jhm.468.

[26] "Gen 11 ultrasound central line training model.pdf." https://anatomywarehouse.com/gen-ii-ultrasound-central-line-training-model-a-108369.

[27] M. I. Soffler, M. M. Hayes, and C. Smith, "Central venous catheterization training: current perspectives on the role of simulation," *Adv. Med. Educ. Pract.*, pp. 9–395, 2018, [Online]. Available: http://dx.doi.org/10.2147/AMEP.S142605.

[28] S. Parmar, S. Parikh, and H. Mehta, "Anatomical variations of the internal jugular vein in relation to carotid artery : an ultrasound study," *Int. J. Med. Sci. Public Heal.*, vol. 2, no. 2, p. 223, 2013, doi: 10.5455/ijmsph.2013.2.215-220.

[29] K. Y. Bilimoria *et al.*, "National Cluster-Randomized Trial of Duty-Hour Flexibility in Surgical Training," *N. Engl. J. Med.*, vol. 374, no. 8, pp. 713–727, 2016, doi: 10.1056/nejmoa1515724.

[30] D. Pepley, M. Yovanoff, S. Miller, and D. Han, "A Virtual Reality Haptic Robotic Simulator for Central Venous Catheterization Training," *J. Med. Device.*, vol. 10, no. September, pp. 1–2, 2016, doi: 10.1115/1.4033867.

[31] M. Yovanoff, D. Pepley, K. Mirkin, J. Moore, D. Han, and S. Miller, "Personalized learning in medical education: Designing a user interface for a dynamic haptic robotic trainer for central venous catheterization," *Proc. Hum. Factors Ergon. Soc.*, vol. 2017-Octob, no. 2014, pp. 615–619, 2017, doi: 10.1177/1541931213601639.

[32] A. Graham, C. Ozment, K. Tegtmeyer, S. Lai, and D. Braner, *Central Venous Catheterization*. 2007.

[33] J. N. Nathwani *et al.*, "The Relationship Between Technical Errors and Decision Making Skills in the Junior Resident ACGME Competencies-Medical Knowledge, Practice Based Learning and Improvement, Systems Based Practice HHS Public Access," *J Surg Educ*, vol. 73, no. 6, pp. 84–90, 2016, doi: 10.1016/j.jsurg.2016.08.004.The.

[34] A. Chao *et al.*, "Performance of central venous catheterization by medical students: A retrospective study of students' logbooks," *BMC Med. Educ.*, vol. 14, no. 1, pp. 1–7, 2014, doi: 10.1186/1472-6920-14-168.

[35] AAMC, "Total U.S. Medical School Enrollment by Race/Ethnicity (Alone) and Gender 2018-2023," 2022. [Online]. Available: https://www.aamc.org/media/6116/download?attachment.

[36] S. A. Amacher *et al.*, "Influence of Gender on the Performance of Cardiopulmonary Rescue Teams: A

Randomized, Prospective Simulator Study," *Crit. Care Med.*, vol. 45, no. 7, pp. 1184–1191, 2017, doi: 10.1097/CCM.0000000000002375.

[37] B. M. A. Schout, A. J. M. Hendrikx, F. Scheele, B. L. H. Bemelmans, and A. J. J. A. Scherpbier, "Validation and implementation of surgical simulators: A critical review of present, past, and future," *Surg. Endosc.*, vol. 24, no. 3, pp. 536–546, 2010, doi: 10.1007/s00464-009-0634-9.

[38] J. J. Stunt, P. H. Wulms, G. M. Kerkhoffs, J. Dankelman, C. N. van Dijk, and G. J. M. Tuijthof, "How valid are commercially available medical simulators?," *Adv. Med. Educ. Pract.*, vol. 5, pp. 385–395, 2014, doi: 10.2147/AMEP.S63435.

[39] D. J. Harris, J. M. Bird, P. A. Smart, M. R. Wilson, and S. J. Vine, "A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training," *Front. Psychol.*, vol. 11, no. March, pp. 1–10, 2020, doi: 10.3389/fpsyg.2020.00605.

[40] M. Racy, A. Barrow, J. Tomlinson, and F. Bello, "Development and Validation of a Virtual Reality Haptic Femoral Nailing Simulator," *J. Surg. Educ.*, vol. 78, no. 3, pp. 1013–1023, 2021, doi: 10.1016/j.jsurg.2020.10.004.

[41] J. P. Monat and T. F. Gannon, "What is Systems Thinking? A Review of Selected Literature Plus Recommendations," *Am. J. Syst. Sci.*, vol. 4, no. 1, pp. 11–26, 2015, doi: 10.5923/j.ajss.20150401.02.

[42] R. D. Arnold and J. P. Wade, "A definition of systems thinking: A systems approach," *Procedia Comput. Sci.*, vol. 44, no. C, pp. 669–678, 2015, doi: 10.1016/j.procs.2015.03.050.

[43] D. McNab, J. McKay, S. Shorrock, S. Luty, and P. Bowie, "Development and application of systems thinking' principles for quality improvement," *BMJ Open Qual.*, vol. 9, no. 1, pp. 1–10, 2020, doi: 10.1136/bmjoq-2019-000714.

[44] M. M. Plack *et al.*, "Systems Thinking and Systems-Based Practice Across the Health Professions: An Inquiry Into Definitions, Teaching Practices, and Assessment," *Teach. Learn. Med.*, vol. 30, no. 3, pp. 242–254, 2018, doi: 10.1080/10401334.2017.1398654.

[45] C. Y. Colbert, P. E. Ogden, A. R. Ownby, and C. Bowe, "Systems-based practice in graduate medical education: Systems thinking as the missing foundational construct," *Teach. Learn. Med.*, vol. 23, no. 2, pp. 179–185, 2011, doi: 10.1080/10401334.2011.561758.

[46] D. F. Pepley *et al.*, "Training Surgical Residents With a Haptic Robotic Central Venous Catheterization Simulator," *J. Surg. Educ.*, vol. 74, no. 6, pp. 1066–1073, 2017, doi: 10.1016/j.jsurg.2017.06.003.

[47] M. A. Yovanoff *et al.*, "Investigating the Effect of Simulator Functional Fidelity and Personalized Feedback on Central Venous Catheterization Training," *J. Surg. Educ.*, vol. 75, no. 5, pp. 1410–1421, 2018, doi: 10.1016/j.jsurg.2018.02.018.

[48] D. Stefanidis *et al.*, "Simulation in Surgery: What's Needed Next?," *Ann. Surg.*, vol. 261, no. 5, pp. 846–853, 2015, doi: 10.1097/SLA.0000000000000826.

[49] B. Saugel, T. W. L. Scheeren, and J. Teboul, "Ultrasound-guided central venous catheter placement : a structured review and recommendations for clinical practice," *Crit. Care*, pp. 1–11, 2017, doi: 10.1186/s13054-017-1814-y.

[50] M. Ishizuka, H. Nagata, K. Takagi, and K. Kubota, "Right internal jugular vein is recommended for central venous catheterization," *J. Investig. Surg.*, vol. 23, no. 2, pp. 110–114, 2010, doi: 10.3109/08941930903469342.

[51] "Central Venous Catheter Placement," *Azura Vascular Care*. https://www.azuravascularcare.com/medical-services/dialysis-access-management/central-venous-catheter-placement/ (accessed Jun. 10, 2022).

[52] S. Hodzic, D. Golic, J. Smajic, S. Sijercic, S. Umihanic, and S. Umihanic, "Complications Related to Insertion and Use of Central Venous Catheters (CVC)," *Med. Arch. (Sarajevo, Bosnia Herzegovina)*, vol. 68, no. 5, pp. 300–303, 2014, doi: 10.5455/medarh.2014.68.300-303.

[53] C. Wall, J. Moore, and J. Thachil, “Catheter-related thrombosis: A practical approach,” *J. Intensive Care Soc.*, vol. 17, no. 2, pp. 160–167, 2016, doi: 10.1177/1751143715618683.

[54] M. Blaivas and S. Adhikari, “An unseen danger: Frequency of posterior vessel wall penetration by needles during attempts to place internal jugular vein central catheters using ultrasound guidance,” *Crit. Care Med.*, vol. 37, no. 8, pp. 2345–2349, 2009, doi: 10.1097/CCM.0b013e3181a067d4.

[55] S. Zhao, Z. Wang, and Y. Zhao, “Loss of guidewire and its sequelae after central venous catheterization A case report,” *Med. (United States)*, vol. 98, no. 29, 2019, doi: 10.1097/MD.0000000000016513.

[56] W. T. McGee, “Central venous catheterization: Better and worse,” *J. Intensive Care Med.*, vol. 21, no. 1, pp. 51–53, 2006, doi: 10.1177/0885066605281705.

[57] B. S. Clark and B. Hammond, “Patient-centered simulation,” *Nurs. Crit. Care*, pp. 45–47, 2017.

[58] S. Konakondla, R. Fong, and C. M. Schirmer, “Simulation training in neurosurgery: Advances in education and practice,” *Adv. Med. Educ. Pract.*, vol. 8, pp. 465–473, 2017, doi: 10.2147/AMEP.S113565.

[59] J. Cecil, A. Gupta, and M. Pirela-Cruz, “An advanced simulator for orthopedic surgical training,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 2, pp. 305–319, 2018, doi: 10.1007/s11548-017-1688-0.

[60] A. R. Luft and M. M. Buitrago, “Stages of motor skill learning,” *Mol. Neurobiol.*, vol. 32, no. 3, pp. 205–216, 2005, doi: 10.1385/MN:32:3:205.

[61] Accreditation Council for Graduate Medical Education, “ACGME Common Program Requirements (Residency),” 2023. [Online]. Available: https://www.acgme.org/globalassets/pfassets/programrequirements/cprresidency_2023.pdf.

[62] J. D. Mitchell, A. Mueller, V. M. Banner-goodspeed, and J. I. Mcsparron, “Expert Evaluation of a Chicken Tissue-based Model for Teaching Ultrasound-guided Central Venous Catheter Insertion,” *J. Educ. Perioper. Med.*, vol. 19, no. 1, pp. 1–13, 2017, doi: 10.46374/volxix-issue1-nachshon.

[63] D. F. Pepley *et al.*, “Building ultrasound phantoms with modified polyvinyl chloride: a comparison of needle insertion forces and sonographic appearance with commercial and traditional simulation materials,” *Simul. Healthc.*, vol. 13, no. 3, pp. 149–153, 2019, doi: 10.1097/SIH.0000000000000302.Building.

[64] CAE, “CAE Healthcare Blue Phantom,” *CAE Website*. https://medicalskillstrainers.cae.com/central-line-and-regional-anesthesia-ultrasound-training-models/p?skuId=99 (accessed Apr. 10, 2022).

[65] D. A. Cook *et al.*, *Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis*, vol. 35, no. 1. 2013.

[66] T. Mahmood and A. Darzi, “The learning curve for a colonoscopy simulator in the absence of any feedback: No feedback, no learning,” *Surg. Endosc. Other Interv. Tech.*, vol. 18, no. 8, pp. 1224–1230, 2004, doi: 10.1007/s00464-003-9143-4.

[67] Simulab, “CentraLineMan,” *Simulab Website*. https://simulab.com/products/centralineman-system (accessed Apr. 10, 2022).

[68] G. C. Huang *et al.*, “Beyond the comfort zone: Residents assess their comfort performing inpatient medical procedures,” *Am. J. Med.*, vol. 119, no. 1, pp. 71.e17-71.e24, 2006, doi: 10.1016/j.amjmed.2005.08.007.

[69] N. Luedicke *et al.*, “Novel central venous catheterization simulation for medical training,” *Proc. - 29th South. Biomed. Eng. Conf. SBEC 2013*, pp. 93–94, 2013, doi: 10.1109/SBEC.2013.55.

[70] M. Yovanoff, D. Pepley, K. Mirkin, J. Moore, D. Han, and S. Miller, “Improving medical education: Simulating changes in patient anatomy using dynamic haptic feedback,” *Proc. Hum. Factors Ergon. Soc.*, pp. 603–607, 2016, doi: 10.1177/1541931213601138.

[71] J. Kolikof, K. Peterson, and A. Baker, “Central Venous Catheter,” *National Library oF medicine*, 2023. https://www.ncbi.nlm.nih.gov/books/NBK557798/?report=printable.

[72] K. Mangold, K. Kunze, M. Quinonez, L. Taylor, and A. Tenison, "Learning Style Preferences of Practicing Nurses Kara," *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2017, doi: 10.1097/NND.0000000000000462.Learning.

[73] D. Hernández-Torrano, S. Ali, and C. K. Chan, "First year medical students' learning style preferences and their correlation with performance in different subjects within the medical course," *BMC Med. Educ.*, vol. 17, no. 1, pp. 1–7, 2017, doi: 10.1186/s12909-017-0965-5.

[74] R. E. Mayer, "Applying the science of learning to medical education," *Med. Educ.*, vol. 44, no. 6, pp. 543–549, 2010, doi: 10.1111/j.1365-2923.2010.03624.x.

[75] D. Krishnan, A. Keloth, S. U.- Education, and U. 2017, "Pros and cons of simulation in medical education: A review," *Int. J. Med. Heal. Res.*, vol. 3, no. 6, pp. 84–87, 2017, [Online]. Available: https://www.researchgate.net/profile/Mohamed-Mourad-Lafifi/post/What_are_the_pros_and_cons_of_using_simulation_as_a_research_method/attachment/5a53e1df4cde266d58812066/AS%3A580616928612353%401515441271505/download/Pros+and+cons+of+simulation+in+medical+ed.

[76] H. Y. So, P. P. Chen, G. K. C. Wong, and T. T. N. Chan, "Simulation in medical education," *J. R. Coll. Physicians Edinb.*, vol. 49, no. 1, pp. 52–57, 2019, doi: 10.4997/JRCPE.2019.112.

[77] D. a Davis, P. E. Mazmanian, M. Fordis, R. Van Harrison, K. E. Thorpe, and L. Perrier, "CLINICIAN ' S CORNER Accuracy of Physician Self-assessment Compared A Systematic Review," *JAMA*, vol. 296, no. 9, pp. 1094–1102, 2006.

[78] C. M. Hsieh and N. J. Nolan, "Confidence, knowledge, and skills at the beginning of residency : A survey of pathology residents," *Am. J. Clin. Pathol.*, vol. 143, no. 1, pp. 11–17, 2015, doi: 10.1309/AJCP3TAFU7NYRIQF.

[79] A. M. Whalen, D. L. Boyer, and A. Nishisaki, "Checklist-Based Assessment of Procedural Skills: A Missing Piece in the Link between Medical Education Interventions and Patient Outcomes," *J. Pediatr.*, vol. 188, no. September, pp. 11–13, 2017, doi: 10.1016/j.jpeds.2017.05.040.

[80] N. M. Howard, D. A. Cook, R. Hatala, and M. V. Pusic, "Learning Curves in Health Professions Education Simulation Research: A Systematic Review," *Simul. Healthc.*, vol. 16, no. 2, pp. 128–135, 2021, doi: 10.1097/SIH.0000000000000477.

[81] A. M. Beason, C. E. Hitt, J. Ketchum, H. Rogers, and H. Sanfey, "Verification of Proficiency in Basic Skills for PGY-1 Surgical Residents: 10-Year Update," *J. Surg. Educ.*, vol. 76, no. 6, pp. e217–e224, 2019, doi: 10.1016/j.jsurg.2019.08.019.

[82] W. C. McGaghie, W. H. Adams, E. R. Cohen, D. B. Wayne, and J. H. Barsuk, "Psychometric Validation of Central Venous Catheter Insertion Mastery Learning Checklist Data and Decisions," *Simul. Healthc.*, vol. 16, no. 6, pp. 378–385, 2021, doi: 10.1097/SIH.0000000000000516.

[83] J. H. Barsuk *et al.*, "Attending Physician Adherence to a 29-Component Central Venous Catheter Bundle Checklist during Simulated Procedures," *Crit. Care Med.*, vol. 44, no. 10, pp. 1871–1881, 2016, doi: 10.1097/CCM.0000000000001831.

[84] J. M. Gonzalez-Vargas *et al.*, "Going the (social) distance: Comparing the effectiveness of online versus in-person Internal Jugular Central Venous Catheterization procedural training," *Am. J. Surg.*, vol. 224, no. 3, pp. 903–907, 2022, doi: 10.1016/j.amjsurg.2021.12.006.

[85] A. M. Hettinga, E. Denessen, and C. T. Postma, "Checking the checklist: A content analysis of expert- and evidence-based case-specific checklist items," *Med. Educ.*, vol. 44, no. 9, pp. 874–883, 2010, doi: 10.1111/j.1365-2923.2010.03721.x.

[86] S. K. Nayar, L. Musto, G. Baruah, R. Fernandes, and R. Bharathan, "Self-Assessment of Surgical Skills: A Systematic Review," *J. Surg. Educ.*, vol. 77, no. 2, pp. 348–361, 2020, doi: 10.1016/j.jsurg.2019.09.016.

[87] D. Blanch-Hartigan, “Medical students’ self-assessment of performance: Results from three meta-analyses,” *Patient Educ. Couns.*, vol. 84, no. 1, pp. 3–9, 2011, doi: 10.1016/j.pec.2010.06.037.

[88] L. Stankov, J. Lee, W. Luo, and D. J. Hogan, “Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety?,” *Learn. Individ. Differ.*, vol. 22, no. 6, pp. 747–758, 2012, doi: 10.1016/j.lindif.2012.05.013.

[89] R. M. Klassen and J. R. L. Klassen, “Self-efficacy beliefs of medical students: a critical review,” *Perspect. Med. Educ.*, vol. 7, no. 2, pp. 76–82, 2018, doi: 10.1007/s40037-018-0411-3.

[90] M. Gottlieb, T. M. Chan, F. Zaver, and R. Ellaway, “Confidence-competence alignment and the role of self-confidence in medical education: A conceptual review,” *Med. Educ.*, vol. 56, no. 1, pp. 37–47, 2022, doi: 10.1111/medu.14592.

[91] J. Clanton, A. Gardner, M. Cheung, L. Mellert, M. Evancho-Chapman, and R. L. George, “The relationship between confidence and competence in the development of surgical skills,” *J. Surg. Educ.*, vol. 71, no. 3, pp. 405–412, 2014, doi: 10.1016/j.jsurg.2013.08.009.

[92] J. Kruger and D. Dunning, “Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments,” *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1121–1134, 1999, doi: 10.1037/0022-3514.77.6.1121.

[93] M. Rahmani, “Medical Trainees and the Dunning–Kruger Effect: When They Don’t Know What They Don’t Know,” *J. Grad. Med. Educ.*, vol. 34, pp. 532–534, 2020, doi: 10.1023/B.

[94] K. Kay and C. Shipman, “The confidence gap,” *The Atlantic*, pp. 56–66, 2014.

[95] D. S. Lind, S. Rekkas, V. Bui, T. Lam, E. Beierle, and E. M. Copeland, “Competency-based student self-assessment on a surgery rotation,” *J. Surg. Res.*, vol. 105, no. 1, pp. 31–34, 2002, doi: 10.1006/jsre.2002.6442.

[96] L. A. Milam, G. L. Cohen, C. Mueller, and A. Salles, “The Relationship Between Self-Efficacy and Well-Being Among Surgical Residents,” *J. Surg. Educ.*, vol. 76, no. 2, pp. 321–328, 2019, doi: 10.1016/j.jsurg.2018.07.028.

[97] R. M. Minter, L. D. Gruppen, K. S. Napolitano, and P. G. Gauger, “Gender differences in the self-assessment of surgical residents,” *Am. J. Surg.*, vol. 189, no. 6, pp. 647–650, 2005, doi: 10.1016/j.amjsurg.2004.11.035.

[98] A. Brown, G. Bonneville, and S. Glaze, “Nevertheless, They Persisted: How Women Experience Gender-Based Discrimination During Postgraduate Surgical Training,” *J. Surg. Educ.*, vol. 78, no. 1, pp. 17–34, 2021, doi: 10.1016/j.jsurg.2020.06.027.

[99] L. E. Yelle, “the Learning Curve: Historical Review and Comprehensive Survey,” *Decis. Sci.*, vol. 10, no. 2, pp. 302–328, 1979, doi: 10.1111/j.1540-5915.1979.tb00026.x.

[100] A. N. Hopper, M. H. Jamison, and W. G. Lewis, “Learning curves in surgical practice,” *Postgrad. Med. J.*, vol. 83, no. 986, pp. 777–779, 2007, doi: 10.1136/pgmj.2007.057190.

[101] N. Khan, H. Abboudi, M. S. Khan, P. Dasgupta, and K. Ahmed, “Measuring the surgical ‘learning curve’: Methods, variables and competency,” *BJU Int.*, vol. 113, no. 3, pp. 504–508, 2014, doi: 10.1111/bju.12197.

[102] M. V. Pusic, K. Boutis, R. Hatala, and D. A. Cook, “Learning Curves in Health Professions Education,” *Acad. Med.*, vol. 90, no. 8, pp. 1034–1042, 2015, doi: 10.1097/ACM.0000000000000681.

[103] D. I. Watson, R. J. Baigrie, and G. G. Jamieson, “A learning curve for laparoscopic fundoplication: Definable, avoidable, or a waste of time?,” *Ann. Surg.*, vol. 224, no. 2, pp. 198–203, 1996, doi: 10.1097/00000658-199608000-00013.

[104] M. Pusic, M. Pecaric, and K. Boutis, “How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation,” *Acad. Med.*, vol. 86, no. 6, pp. 731–736, 2011, doi:

10.1097/ACM.0b013e3182178c3c.

[105] N. Tsuboi, M. Abe, S. Matsumoto, N. Nishimura, and S. Nakagawa, “The Effect of Clinical Experience on the Learning Curve of Pediatric Intensive Care Unit Residents for the Central Venous Catheter Placement Procedure,” *J. Pediatr. Intensive Care*, vol. 07, no. 01, pp. 039–042, 2018, doi: 10.1055/s-0037-1604153.

[106] B. Mema, M. Mylopoulos, A. Tekian, and Y. S. Park, “Using Learning Curves to Identify and Explain Growth Patterns of Learners in Bronchoscopy Simulation: A Mixed-Methods Study,” *Acad. Med.*, vol. 95, no. 12, pp. 1921–1928, 2020, doi: 10.1097/ACM.0000000000003595.

[107] A. Hasan, M. Pozzi, and J. R. L. Hamilton, “New surgical procedures: Can we minimise the learning curve?,” *Bmj*, vol. 320, no. 7228, pp. 171–173, 2000, doi: 10.1136/bmj.320.7228.171.

[108] I. Kim, S. R. Miller, and A. Freivalds, “Using learning curves to assess resident surgical skill acquisition: A case study of ultrasound-guided catheter insertion,” *Proc. Hum. Factors Ergon. Soc.*, vol. 2014-Janua, pp. 2350–2354, 2014, doi: 10.1177/1541931214581489.

[109] H. E. Chen *et al.*, “Can Haptic Simulators Distinguish Expert Performance? A Case Study in Central Venous Catheterization in Surgical Education,” *Simul. Healthc.*, vol. 14, no. 1, pp. 35–42, 2019, doi: 10.1097/SIH.0000000000000352.

[110] T. Grundgeiger, C. Klöffel, S. Mohme, T. S. Wurmb, and O. Happel, “Visual attention in anesthesiology: The effect of the environment (simulated vs. real) and expertise (junior vs. senior),” *Proc. Hum. Factors Ergon. Soc.*, p. 531, 2016, doi: 10.1177/1541931213601121.

[111] M. Gann and A. Sardi, “Improved results using ultrasound guidance for central venous access,” *Am. Surg.*, vol. 69, no. 12, pp. 1104–1107, 2003, doi: 10.1177/000313480306901217.

[112] N. Vaughan, V. N. Dubey, M. Y. K. Wee, and R. Isaacs, “Synthetic methods for ultrasound- guided epidural insertion simulation,” *J. Med. Devices, Trans. ASME*, vol. 8, no. 2, pp. 3–4, 2014, doi: 10.1115/1.4027053.

[113] F. J. Carter *et al.*, “Consensus guidelines for validation of virtual reality surgical simulators,” *Surg. Endosc. Other Interv. Tech.*, vol. 19, no. 12, pp. 1523–1532, 2005, doi: 10.1007/s00464-005-0384-2.

[114] T. Rutherford-Hemming, “Determining content validity and reporting a content validity index for simulation scenarios,” *Nurs. Educ. Perspect.*, vol. 36, no. 6, pp. 389–393, 2015, doi: 10.5480/15-1640.

[115] B. Wickens, J. Lewis, D. P. Morris, M. Husein, H. M. Ladak, and S. K. Agrawal, “Face and content validity of a novel, web-based otoscopy simulator for medical education,” *J. Otolaryngol. - Head Neck Surg.*, vol. 44, no. 1, pp. 1–8, 2015, doi: 10.1186/s40463-015-0060-z.

[116] H. B. Colaco, K. Hughes, E. Pearse, M. Arnander, and D. Tennent, “Construct Validity, Assessment of the Learning Curve, and Experience of Using a Low-Cost Arthroscopic Surgical Simulator,” *J. Surg. Educ.*, vol. 74, no. 1, pp. 47–54, 2017, doi: 10.1016/j.jsurg.2016.07.006.

[117] P. A. A. Milcent, A. L. Kulcheski, F. M. Rosa, L. Dau, and E. Stieven Filho, “Construct Validity and Experience of Using a Low-cost Arthroscopic Knee Surgery Simulator,” *J. Surg. Educ.*, vol. 78, no. 1, pp. 292–301, 2021, doi: 10.1016/j.jsurg.2020.06.007.

[118] S. Chawla, S. Devi, P. Calvachi, W. B. Gormley, and R. Rueda-Esteban, “Evaluation of simulation models in neurosurgical training according to face, content, and construct validity: a systematic review,” *Acta Neurochir. (Wien).*, vol. 164, no. 4, pp. 947–966, 2022, doi: 10.1007/s00701-021-05003-x.

[119] A. H. Sadeghi *et al.*, “Virtual Reality Simulation Training for Cardiopulmonary Resuscitation After Cardiac Surgery: Face and Content Validity Study,” *JMIR Serious Games*, vol. 10, no. 1, 2022, doi: 10.2196/30456.

[120] S. S. Van Nortwick, T. S. Lendvay, A. R. Jensen, A. S. Wright, K. D. Horvath, and S. Kim, “Methodologies for establishing validity in surgical simulation studies,” *Surgery*, vol. 147, no. 5, pp. 622–630, 2010, doi: 10.1016/j.surg.2009.10.068.

[121] A. A. Alsaad, V. Y. Bhide, J. L. Moss, S. M. Silvers, M. M. Johnson, and M. J. Maniaci, “Central line

proficiency test outcomes after simulation training versus traditional training to competence," *Ann. Am. Thorac. Soc.*, vol. 14, no. 4, pp. 550–554, 2017, doi: 10.1513/AnnalsATS.201612-987OC.

[122] R. Kleinert, N. Heiermann, R. Wahba, D. H. Chang, A. H. Hölscher, and D. L. Stippel, "Design, Realization, and First Validation of an Immersive Web-Based Virtual Patient Simulator for Training Clinical Decisions in Surgery," *J. Surg. Educ.*, vol. 72, no. 6, pp. 1131–1138, 2015, doi: 10.1016/j.jsurg.2015.05.009.

[123] E. A. Patel, A. Aydın, A. Desai, P. Dasgupta, and K. Ahmed, "Current status of simulation-based training in pediatric surgery: A systematic review," *J. Pediatr. Surg.*, vol. 54, no. 9, pp. 1884–1893, 2019, doi: 10.1016/j.jpedsurg.2018.11.019.

[124] J. H. Barsuk, E. R. Cohen, W. C. Mcgaghie, and D. B. Wayne, "Long-Term Retention of Central Venous Simulation-Based Mastery Learning," *Acad. Med.*, vol. 85, no. 10, pp. 10–13, 2010, doi: 10.1097/ACM.0b013e3181ed436c.

[125] K. Talsma, B. Schüz, R. Schwarzer, and K. Norris, "I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance," *Learn. Individ. Differ.*, vol. 61, no. April 2017, pp. 136–150, 2018, doi: 10.1016/j.lindif.2017.11.015.

[126] T. R. Spencer and A. J. Bardin-Spencer, "Pre- and post-review of a standardized ultrasound-guided central venous catheterization curriculum evaluating procedural skills acquisition and clinician confidence," *J. Vasc. Access*, vol. 21, no. 4, pp. 440–448, 2020, doi: 10.1177/1129729819882602.

[127] M. Płużyczka, "The First Hundred Years: a History of Eye Tracking as a Research Method," *Appl. Linguist. Pap.*, vol. 4/2018, no. 25, pp. 101–116, 2018, doi: 10.32612/uw.25449354.2018.4.pp.101-116.

[128] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005, doi: 10.1016/j.cviu.2004.07.010.

[129] K. Holmqvist and R. Andersson, *Eyetracking: A comprehensive guide to methods, paradigms and measures*, no. March. New York: Oxford University Press, 2011.

[130] M. L. Lai *et al.*, "A review of using eye-tracking technology in exploring learning from 2000 to 2012," *Educ. Res. Rev.*, vol. 10, no. 88, pp. 90–115, 2013, doi: 10.1016/j.edurev.2013.10.001.

[131] T. T. Brunyé, T. Drew, D. L. Weaver, and J. G. Elmore, "A review of eye tracking for understanding and improving diagnostic interpretation," *Cogn. Res. Princ. Implic.*, vol. 4, no. 1, 2019, doi: 10.1186/s41235-019-0159-2.

[132] S. Deng, J. Oh, T. Wang, S. H. Parker, and N. K. Lau, "Assessing Laparoscopic Surgical Skills of Trainees with Scene Independent and Dependent Eye Gaze Metrics," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, pp. 1–6, 2023, doi: 10.1177/21695067231192642.

[133] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Comput. Med. Imaging Graph.*, vol. 65, no. 2018, pp. 176–190, 2018, doi: 10.1016/j.compmedimag.2017.04.006.

[134] D. A. Hofmaenner *et al.*, "Use of eye tracking in analyzing distribution of visual attention among critical care nurses in daily professional life: an observational study," *J. Clin. Monit. Comput.*, vol. 35, no. 6, pp. 1511–1518, 2021, doi: 10.1007/s10877-020-00628-2.

[135] N. Ahmadi *et al.*, "Quantifying Workload and Stress in Intensive Care Unit Nurses: Preliminary Evaluation Using Continuous Eye-Tracking," *Hum. Factors*, 2022, doi: 10.1177/00187208221085335.

[136] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Z. Yang, and A. Darzi, "Eye tracking for skills assessment and training: A systematic review," *J. Surg. Res.*, vol. 191, no. 1, pp. 169–178, 2014, doi: 10.1016/j.jss.2014.04.032.

[137] S. E. Fox and B. E. Faulkner-Jones, "Eye-tracking in the study of visual expertise: Methodology and approaches in medicine," *Front. Learn. Res.*, vol. 5, no. 3 Special Issue, pp. 43–54, 2017, doi:

10.14786/flr.v5i3.258.

[138] H. Ashraf, M. H. Sodergren, N. Merali, G. Mylonas, H. Singh, and A. Darzi, "Eye-tracking technology in medical education: A systematic review," *Med. Teach.*, vol. 40, no. 1, pp. 62–69, 2018, doi: 10.1080/0142159X.2017.1391373.

[139] D. R. Leff, J. J. H. Leong, R. Aggarwal, G. Z. Yang, and A. Darzi, "Could variations in technical skills acquisition in surgery be explained by differences in cortical plasticity?," *Ann. Surg.*, vol. 247, no. 3, pp. 540–543, 2008, doi: 10.1097/SLA.0b013e31815fa42e.

[140] T. Krings *et al.*, "Cortical activation patterns during complex motor tasks in piano players and control subjects. A functional magnetic resonance imaging study," *Neurosci. Lett.*, vol. 278, no. 3, pp. 189–193, 2000, doi: 10.1016/S0304-3940(99)00930-1.

[141] D. Stefanidis, W. W. Hope, J. R. Korndorffer, S. Markley, and D. J. Scott, "Initial Laparoscopic Basic Skills Training Shortens the Learning Curve of Laparoscopic Suturing and Is Cost-Effective," *J. Am. Coll. Surg.*, vol. 210, no. 4, pp. 436–440, 2010, doi: 10.1016/j.jamcollsurg.2009.12.015.

[142] L. Tetzlaff and D. R. Schwartz, "The use of guidelines in interface design," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 5, pp. 329–333, 1991, doi: 10.1145/108844.108936.

[143] F. Hermens, R. Flin, and I. Ahmed, "Eye movements in surgery: A literature review," *J. Eye Mov. Res.*, vol. 6, no. 4, pp. 1–11, 2013, doi: 10.16910/jemr.6.4.4.

[144] R. Naik, A. Kogkas, H. Ashrafian, G. Mylonas, and A. Darzi, "The Measurement of Cognitive Workload in Surgery Using Pupil Metrics: A Systematic Review and Narrative Analysis," *J. Surg. Res.*, vol. 280, pp. 258–272, 2022, doi: 10.1016/j.jss.2022.07.010.

[145] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters, "Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts," *Surg. Endosc.*, vol. 24, no. 10, pp. 2458–2464, 2010, doi: 10.1007/s00464-010-0986-1.

[146] Y. Chen, L. Dong, A. G. Gale, B. Rees, and C. Maxwell-Armstrong, "Laparoscopic surgical skills training: an investigation of the potential of using surgeons' visual search behaviour as a performance indicator," *Med. Imaging 2014 Image Perception, Obs. Performance, Technol. Assess.*, vol. 9037, no. March 2014, p. 903704, 2014, doi: 10.1117/12.2044387.

[147] J. Hanley, D. Warren, N. Glass, D. Tranel, M. Karam, and J. Buckwalter, "Visual Interpretation of Plain Radiographs in Orthopaedics Using Eye-Tracking Technology," *Iowa Orthop. J.*, vol. 37, pp. 225–231, 2017.

[148] C. Diaz-Piedra, J. M. Sanchez-Carrion, H. Rieiro, and L. L. Di Stasi, "Gaze-based Technology as a Tool for Surgical Skills Assessment and Training in Urology," *Urology*, vol. 107, pp. 26–30, 2017, doi: 10.1016/j.urology.2017.06.030.

[149] S. Eivazi *et al.*, "Optimal eye movement strategies: a comparison of neurosurgeons gaze patterns when using a surgical microscope," *Acta Neurochir. (Wien).*, vol. 159, no. 6, pp. 959–966, 2017, doi: 10.1007/s00701-017-3185-1.

[150] S. Eivazi, M. Slupina, W. Fuhl, H. Afkari, A. Hafez, and E. Kasneci, "Towards automatic skill evaluation in microsurgery," *Int. Conf. Intell. User Interfaces, Proc. IUI*, pp. 73–76, 2017, doi: 10.1145/3030024.3040985.

[151] M. L. Sánchez-Ferrer, M. D. Grima-Murcia, F. Sánchez-Ferrer, A. I. Hernández-Peñalver, E. Fernández-Jover, and F. Sánchez del Campo, "Use of Eye Tracking as an Innovative Instructional Method in Surgical Human Anatomy," *J. Surg. Educ.*, vol. 74, no. 4, pp. 668–673, 2017, doi: 10.1016/j.jsurg.2016.12.012.

[152] H.-E. Chen *et al.*, "Can Eye Tracking be Used to Predict Performance Improvements in Simulated Medical Training? A Case Study in Central Venous Catheterization," *Proc. Int. Symp. Hum. Factors Ergon. Heal. Care*, vol. 8, no. 1, pp. 110–114, 2019, doi: 10.1177/2327857919081025.

[153] H. E. Chen *et al.*, “Looks can be deceiving: Gaze pattern differences between novices and experts during placement of central lines,” *Am. J. Surg.*, vol. 217, no. 2, pp. 362–367, 2019, doi: 10.1016/j.amjsurg.2018.11.007.

[154] K. Tatsuru *et al.*, “The evaluation of eye gaze using an eye tracking system in simulation training of real-time ultrasound-guided venipuncture,” *J. Vasc. Access*, pp. 1–5, 2021, doi: 10.1177/1129729820987362.

[155] H. M. Tzamaras, H.-L. Wu, J. Z. Moore, and S. R. Miller, “Shifting Perspectives: A proposed framework for analyzing head-mounted eye-tracking data with dynamic areas of interest and dynamic scenes,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, 2023, doi: 10.1177/21695067231192929.

[156] C. Jongerius *et al.*, “Eye-tracking glasses in face-to-face interactions: Manual versus automated assessment of areas-of-interest,” *Behav. Res. Methods*, vol. 53, no. 5, pp. 2037–2048, 2021, doi: 10.3758/s13428-021-01544-2.

[157] N. Kumari, V. Ruf, S. Mukhametov, A. Schmidt, J. Kuhn, and S. Küchemann, “Mobile eye-tracking data analysis using object detection via YOLO v4,” *Sensors*, vol. 21, no. 22, pp. 1–17, 2021, doi: 10.3390/s21227668.

[158] S. J. Vine, J. S. McGrath, E. Bright, T. Dutton, J. Clark, and M. R. Wilson, “Assessing visual control during simulated and live operations: Gathering evidence for the content validity of simulation using eye movement metrics,” *Surg. Endosc.*, vol. 28, no. 6, pp. 1788–1793, 2014, doi: 10.1007/s00464-013-3387-4.

[159] A. C. Gordon, J. C. Saliken, D. Johns, R. Owen, and R. R. Gray, “US-guided puncture of the internal jugular vein: Complications and anatomic considerations,” *J. Vasc. Interv. Radiol.*, vol. 9, no. 2, pp. 333–338, 1998, doi: 10.1016/S1051-0443(98)70277-5.

[160] C. Kornbau, K. Lee, G. Hughes, and M. Firstenberg, “Central line complications,” *Int. J. Crit. Illn. Inj. Sci.*, vol. 5, no. 3, pp. 170–178, 2015.

[161] A. N. Musits *et al.*, “Physician Versus Nonphysician Instruction: Evaluating an Expert Curriculum-Competent Facilitator Model for Simulation-Based Central Venous Catheter Training,” *Simul. Healthc.*, vol. 14, no. 4, pp. 228–234, 2019, doi: 10.1097/SIH.0000000000000374.

[162] D. Richardson *et al.*, “Growth mindset in competency-based medical education,” *Med. Teach.*, vol. 43, no. 7, pp. 751–757, 2021, doi: 10.1080/0142159X.2021.1928036.

[163] E. L. Scahill, N. G. Oliver, V. R. Tallentire, S. Edgar, and J. F. Tiernan, “An enhanced approach to simulation-based mastery learning: optimising the educational impact of a novel, National Postgraduate Medical Boot Camp,” *Adv. Simul.*, vol. 6, no. 1, pp. 1–10, 2021, doi: 10.1186/s41077-021-00157-1.

[164] C. J. Schroedl, A. Frogameni, J. H. Barsuk, E. R. Cohen, L. Sivarajan, and D. B. Wayne, “Impact of Simulation-based Mastery Learning on Resident Skill Managing Mechanical Ventilators,” *ATS Sch.*, vol. 2, no. 1, pp. 34–48, 2021, doi: 10.34197/ats-scholar.2020-0023OC.

[165] E. Kattan *et al.*, “Simulation-based mastery learning of bronchoscopy-guided percutaneous dilatational tracheostomy: Competency acquisition and skills transfer to a cadaveric model,” *Simul. Healthc.*, vol. 16, no. 3, pp. 157–162, 2021, doi: 10.1097/SIH.0000000000000491.

[166] H. A. Ballard *et al.*, “Use of a simulation-based mastery learning curriculum to improve ultrasound-guided vascular access skills of pediatric anesthesiologists,” *Paediatr. Anaesth.*, vol. 30, no. 11, pp. 1204–1210, 2020, doi: 10.1111/pan.13953.

[167] D. Pepley, M. Yovanoff, K. Mirkin, D. Han, S. Miller, and J. Moore, “Design of a virtual reality haptic robotic central venous catheterization training simulator,” *Proc. ASME Des. Eng. Tech. Conf.*, vol. 5A-2016, pp. 1–8, 2016, doi: 10.1115/DETC2016-59560.

[168] D. Pepley, M. Yovanoff, K. Mirkin, D. Han, S. Miller, and J. Moore, “Design of a virtual reality haptic robotic central venous catheterization training simulator,” *Proc. ASME Des. Eng. Tech. Conf.*, vol. 5A-2016, pp. 1–8, 2016, doi: 10.1115/DETC2016-59560.

[169] J. M. Gonzalez-Vargas *et al.*, “Objective Assessment Metrics for Central Line Simulators: An Exploration of Causal Factors,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 64, no. 1, pp. 2008–2012, 2020, doi: 10.1177/1071181320641487.

[170] H. E. Chen *et al.*, “Evaluating surgical resident needle insertion skill gains in central venous catheterization training,” *J. Surg. Res.*, vol. 233, pp. 351–359, 2019, doi: 10.1016/j.jss.2018.07.040.

[171] D. C. Brown *et al.*, “Evaluating the Impact of Assessment Metrics for Simulated Central Venous Catheterization Training,” *Simul. Healthc. J. Soc. Simul. Healthc.*, vol. Publish Ah, no. 00, pp. 1–8, 2022, doi: 10.1097/sih.0000000000000704.

[172] N. Sharma, I. Doherty, and C. Dong, “Adaptive learning in medical education: The final piece of technology enhanced learning?,” *Ulster Med. J.*, vol. 86, no. 3, pp. 198–200, 2017.

[173] C. Lwande, L. Muchemi, and R. Oboko, “Identifying learning styles and cognitive traits in a learning management system,” *Heliyon*, vol. 7, no. 8, p. e07701, 2021, doi: 10.1016/j.heliyon.2021.e07701.

[174] W. Jiraporncharoen, C. Angkurawaranon, M. Chockjamsai, A. Deesomchok, and J. Euathrongchit, “Learning styles and academic achievement among undergraduate medical students in Thailand,” *J. Educ. Eval. Health Prof.*, vol. 12, p. 38, 2015, doi: 10.3352/jeehp.2015.12.38.

[175] A. Dünser and M. Jirasko, “Interaction of hypertext forms and global versus sequential learning styles,” *J. Educ. Comput. Res.*, vol. 32, no. 1, pp. 79–91, 2005, doi: 10.2190/1J25-LWQF-PQ3W-LABM.

[176] R. Greif, L. Egger, R. M. Basciani, A. Lockey, and A. Vogt, “Emergency skill training-A randomized controlled study on the effectiveness of the 4-stage approach compared to traditional clinical teaching,” *Resuscitation*, vol. 81, no. 12, pp. 1692–1697, 2010, doi: 10.1016/j.resuscitation.2010.09.478.

[177] T. Viering and M. Loog, “The Shape of Learning Curves: A Review,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7799–7819, 2023, doi: 10.1109/TPAMI.2022.3220744.

[178] F. Nickel *et al.*, “Sequential learning of psychomotor and visuospatial skills for laparoscopic suturing and knot tying—a randomized controlled trial ‘The Shoebox Study’ DRKS00008668,” *Langenbeck’s Arch. Surg.*, vol. 401, no. 6, pp. 893–901, 2016, doi: 10.1007/s00423-016-1421-4.

[179] G. Jiang *et al.*, “Learning curves and long-term outcome of simulation-based thoracentesis training for medical students,” *BMC Med. Educ.*, vol. 11, no. 1, 2011, doi: 10.1186/1472-6920-11-39.

[180] K. A. Ericsson, K. Nandagopal, and R. W. Roring, “Toward a Science of Exceptional Achievement Deliberate Practice,” *Longevity, Regen. Optim. Heal.*, vol. 217, pp. 199–217, 2009, doi: 10.1196/annals.1393.001.

[181] N. O. Kolozsvari, L. S. Feldman, M. C. Vassiliou, S. Demyttenaere, and M. L. Hoover, “Sim one, do one, teach one: Considerations in designing training curricula for surgical simulation,” *J. Surg. Educ.*, vol. 68, no. 5, pp. 421–427, 2011, doi: 10.1016/j.jsurg.2011.03.010.

[182] E. D. Grober, M. Roberts, E. J. Shin, M. Mahdi, and V. Bacal, “Intraoperative assessment of technical skills on live patients using economy of hand motion: establishing learning curves of surgical competence,” *Am. J. Surg.*, vol. 199, no. 1, pp. 81–85, 2010, doi: 10.1016/j.amjsurg.2009.07.033.

[183] J. M. Gonzalez-vargas *et al.*, “OBJECTIVE ASSESSMENT METRICS FOR CENTRAL LINE SIMULATORS : AN EXPLORATION OF CAUSAL FACTORS,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 64, no. 1, pp. 2008–2012, 2020, doi: 10.1177/1071181320641487.

[184] D. Lakens, “Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs,” *Front. Psychol.*, vol. 4, no. NOV, pp. 1–12, 2013, doi: 10.3389/fpsyg.2013.00863.

[185] P. Schober and V. Thomas, “Adjustments for Multiple Testing in Medical Research,” *Anesth. Anaglesia*, vol. 130, no. 1, p. 2020, 2020.

[186] K. E. Campbell and H. J. Mccammon, “Elizabeth blackwell’s heirs: Women as physicians in the united

states, 1880-1920," *Work Occup.*, vol. 32, no. 3, pp. 290–318, 2005, doi: 10.1177/0730888405277634.

[187] AAMC, "Total U.S. Medical School Enrollment by Race/Ethnicity (Alone) and Gender 2019-2024," 2023.

[188] NIH, "NIH Inclusion Outreach Toolkit: How to Engage, Recruit, and Retain Women in Clinical Research," *National Institutes of Health Office of Research on Women's Health*. .

[189] A. Z. Sosinsky, J. W. Rich-Edwards, A. Wiley, K. Wright, P. A. Spagnolo, and H. Joffe, "Enrollment of female participants in United States drug and device phase 1–3 clinical trials between 2016 and 2019," *Contemp. Clin. Trials*, vol. 115, no. March, p. 106718, 2022, doi: 10.1016/j.cct.2022.106718.

[190] N. Zern, S. Shalhub, D. Wood, and K. Calhoun, "Association of Sex With Perceived Career Barriers Among Surgeons," *J. Bone Jt. Surg. - Am. Vol.*, vol. 154, no. 12, pp. 1155–1157, 2019, doi: 10.2106/JBJS.18.00657.

[191] J. M. Colbert-Getz, C. Fleishman, J. Jung, and N. Shilkofski, "How do gender and anxiety affect students' self-assessment and actual performance on a high-stakes clinical skills examination?," *Acad. Med.*, vol. 88, no. 1, pp. 44–48, 2013, doi: 10.1097/ACM.0b013e318276bcc4.

[192] A. R. Artino, "Academic self-efficacy: from educational theory to instructional practice," *Perspect. Med. Educ.*, vol. 1, no. 2, pp. 76–85, 2012, doi: 10.1007/s40037-012-0012-5.

[193] S. Morony, S. Kleitman, Y. P. Lee, and L. Stankov, "Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries," *Int. J. Educ. Res.*, vol. 58, pp. 79–96, 2013, doi: 10.1016/j.ijer.2012.11.002.

[194] J. K. Heath, M. E. Alvarado, C. B. Clancy, T. D. Barton, J. R. Kogan, and C. J. Dine, "The Context of 'Confidence': Analyzing the Term Confidence in Resident Evaluations," *J. Gen. Intern. Med.*, vol. 37, no. 9, pp. 2187–2193, 2022, doi: 10.1007/s11606-022-07535-z.

[195] A. Hoffman, R. Ghoubrial, M. McCormick, P. Matemavi, and R. Cusick, "Exploring the gender gap: Letters of recommendation to pediatric surgery fellowship," *Am. J. Surg.*, vol. 219, no. 6, pp. 932–936, 2020, doi: 10.1016/j.amjsurg.2019.08.005.

[196] K. A. Ladonna, S. Ginsburg, and C. Watling, "'Rising to the Level of Your Incompetence': What Physicians' Self-Assessment of Their Performance Reveals about the Imposter Syndrome in Medicine," *Acad. Med.*, vol. 93, no. 5, pp. 763–768, 2018, doi: 10.1097/ACM.0000000000002046.

[197] L. Zwaan and W. E. Hautz, "Bridging the gap between uncertainty, confidence and diagnostic accuracy: Calibration is key," *BMJ Qual. Saf.*, vol. 28, no. 5, pp. 352–355, 2019, doi: 10.1136/bmjqs-2018-009078.

[198] M. Sawdon and G. Finn, "The 'unskilled and unaware' effect is linear in a real-world setting," *J. Anat.*, vol. 224, no. 3, pp. 279–285, 2014, doi: 10.1111/joa.12072.

[199] T. Gude *et al.*, "Do medical students and young physicians assess reliably their self-efficacy regarding communication skills? A prospective study from end of medical school until end of internship," *BMC Med. Educ.*, vol. 17, no. 1, pp. 1–7, 2017, doi: 10.1186/s12909-017-0943-y.

[200] S. P. Vajapey, K. L. Weber, and J. B. Samora, "Confidence Gap Between Men and Women in Medicine: A systematic review," *Educ. Pract. Manag.*, vol. 31, no. 5, pp. 494–502, 2020.

[201] J. Barr and C. S. Graffeo, "Procedural Experience and Confidence Among Graduating Medical Students," *J. Surg. Educ.*, vol. 73, no. 3, pp. 466–473, 2016, doi: 10.1016/j.jsurg.2015.11.014.

[202] L. Kwasny, S. Shebrain, G. Munene, and R. Sawyer, "Is there a gender bias in milestones evaluations in general surgery residency training?," *Am. J. Surg.*, vol. 221, no. 3, pp. 505–508, 2021, doi: 10.1016/j.amjsurg.2020.12.020.

[203] R. S. Watson *et al.*, "A Multicenter Prospective Comparison of the Accreditation Council for Graduate Medical Education Milestones : Clinical Competency Committee vs . Resident," *J. Surg. Educ.*, vol. 74, no. 6, pp. e8–e14, 2017, doi: 10.1016/j.jsurg.2017.06.009.

[204] K. Kunkler, “The role of medical simulation: an overview,” *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 2, no. April, pp. 203–210, 2006, doi: 10.1002/rcs.

[205] J. B. Cooper and V. R. Taqueti, “A brief history of the development of mannequin simulators for clinical education and training.,” *Postgrad. Med. J.*, vol. 84, no. 997, pp. 563–570, 2004, doi: 10.1136/qshc.2004.009886.

[206] G. Pinar, “An Educational Revolution and Innovative Technologies: The Role of Simulation,” *Creat. Educ.*, vol. 11, no. 11, pp. 2218–2232, 2020, doi: 10.4236/ce.2020.1111162.

[207] Y. Okuda *et al.*, “National growth in simulation training within emergency medicine residency programs, 2003-2008,” *Acad. Emerg. Med.*, vol. 15, no. 11, pp. 1113–1116, 2008, doi: 10.1111/j.1553-2712.2008.00195.x.

[208] H. Wu, S. Li, J. Zheng, and J. Guo, “Medical students’ motivation and academic performance: the mediating roles of self-efficacy and learning engagement,” *Med. Educ. Online*, vol. 25, no. 1, 2020, doi: 10.1080/10872981.2020.1742964.

[209] D. A. Cook, W. G. Thompson, K. G. Thomas, M. R. Thomas, and V. S. Pankratz, “Impact of self-assessment questions and learning styles in web-based learning: A randomized, controlled, crossover trial,” *Acad. Med.*, vol. 81, no. 3, pp. 231–238, 2006, doi: 10.1097/00001888-200603000-00005.

[210] A. A. Hayat, K. Shateri, M. Amini, and N. Shokrpour, “Relationships between academic self-efficacy, learning-related emotions, and metacognitive learning strategies with academic performance in medical students: A structural equation model,” *BMC Med. Educ.*, vol. 20, no. 1, pp. 1–11, 2020, doi: 10.1186/s12909-020-01995-9.

[211] Z. Paskins and E. Peile, “Final year medical students’ views on simulation-based teaching: A comparison with the Best Evidence Medical Education Systematic Review,” *Med. Teach.*, vol. 32, no. 7, pp. 569–577, 2010, doi: 10.3109/01421590903544710.

[212] F. Stroben, T. Schröder, K. A. Dannenberg, A. Thomas, A. Exadaktylos, and W. E. Hautz, “A simulated night shift in the emergency room increases students ’ self-efficacy independent of role taking over during simulation,” *BMC Med. Educ.*, pp. 1–7, 2016, doi: 10.1186/s12909-016-0699-9.

[213] J. Kerins, E. McCully, S. A. Stirling, S. E. Smith, J. Tiernan, and V. R. Tallentire, “The impact of simulation-based mastery learning, booster session timing and clinical exposure on confidence in intercostal drain insertion: a survey of internal medicine trainees in Scotland,” *BMC Med. Educ.*, vol. 22, no. 1, pp. 1–9, 2022, doi: 10.1186/s12909-022-03654-7.

[214] K. Schwab, J. Friedman, M. Lazarus, and J. Williams, “Preparing Residents for Emergent Vascular Access: The Comparative Effectiveness of Central Venous and Intraosseous Catheter Simulation-Based Training,” *Int. J. Crit. Care Emerg. Med.*, vol. 5, no. 1, pp. 1–8, 2019, doi: 10.23937/2474-3674/1510069.

[215] J. Fritz *et al.*, “Training in obstetric and neonatal emergencies in Mexico: effect on knowledge and self-efficacy by gender , age , shift , and profession,” *BMC Med. Educ.*, vol. 20, no. 97, pp. 1–10, 2020.

[216] M. A. Yovanoff *et al.*, “Investigating the Effect of Simulator Functional Fidelity and Personalized Feedback on Central Venous Catheterization Training,” *J. Surg. Educ.*, vol. 75, no. 5, pp. 1410–1421, 2018, doi: 10.1016/j.jsurg.2018.02.018.

[217] A. De Cassai *et al.*, “Safety in training for ultrasound guided internal jugular vein CVC placement: a propensity score analysis,” *BMC Anesthesiol.*, vol. 21, no. 1, pp. 1–6, 2021, doi: 10.1186/s12871-021-01460-0.

[218] C. B. IJsselmuiden and R. R. Faden, “Complications and Failures of Subclavian-vein catheterization,” *N. Engl. J. Med.*, vol. 331, no. 26, pp. 1735–1738, 1992.

[219] S. R. Reder *et al.*, “Gender differences in self-assessed performance and stress level during training of basic interventional radiology maneuvers,” *Eur. Radiol.*, no. 0123456789, 2023, doi: 10.1007/s00330-023-09993-

3.

[220] "Advancing Gender Equity in Medicine: Resources for Physicians," *American Medical Association*, 2022. https://www.ama-assn.org/delivering-care/health-equity/advancing-gender-equity-medicine-resources-physicians.

[221] S. Abrahamson, J. Denson, and R. Wolf, "Effectiveness of a Simulator in Training Anesthesiology residents," *J. Med. Educ.*, vol. 44, no. 6, pp. 515–519, 1969.

[222] K. Kunkler, "The role of medical simulation : an overview †," *Int. jounral Med. Robot. Comput. Assist. Surg.*, no. July, pp. 203–210, 2006, doi: 10.1002/rcs.

[223] A. Żytkowski, R. S. Tubbs, J. Iwanaga, E. Clarke, M. Polguj, and G. Wysiadecki, "Anatomical normality and variability: Historical perspective and methodological considerations," *Transl. Res. Anat.*, vol. 23, no. December 2020, 2021, doi: 10.1016/j.tria.2020.100105.

[224] C. Vaitsis1 *et al.*, "Standardization in Medical Education: Review, Collection and Selection of Standards To Address Technical and Educational Aspects in Outcome-Based Medical Education," *Mefanet J*, vol. 5, no. 1, pp. 28–39, 2017, [Online]. Available: http://www.scottishdoctor.org.

[225] A. Cheng, A. Lockey, F. Bhanji, Y. Lin, E. A. Hunt, and E. Lang, "The use of high-fidelity manikins for advanced life support training-A systematic review and meta-analysis," *Resuscitation*, vol. 93, pp. 142–149, 2015, doi: 10.1016/j.resuscitation.2015.04.004.

[226] N. Moureau *et al.*, "Evidence-based consensus on the insertion of central venous access devices: Definition of minimal requirements for training," *Br. J. Anaesth.*, vol. 110, no. 3, pp. 347–356, 2013, doi: 10.1093/bja/aes499.

[227] M. Reppucci, C. Prendergast, K. Flynn, S. Scarbro, S. C. Derderian, and J. Diaz-Miron, "Trainee educational curriculum to standardize central venous catheter repair," *BMC Med. Educ.*, vol. 23, no. 1, pp. 1–7, 2023, doi: 10.1186/s12909-023-04977-9.

[228] J. D. Davis, M. M. Treggiari, E. A. Dickson, and P. M. Schulman, "A Training Program for Real-Time Ultrasound-Guided Catheterization of the Subclavian Vein," *J. Med. Educ. Curric. Dev.*, vol. 8, p. 238212052110258, 2021, doi: 10.1177/23821205211025849.

[229] D. C. Brown *et al.*, "Evaluating the Impact of Assessment Metrics for Simulated Central Venous Catheterization Training," *Simul. Healthc.*, no. 00, pp. 1–8, 2022, doi: 10.1097/sih.0000000000000704.

[230] H. E. Chen *et al.*, "From the simulation center to the bedside: Validating the efficacy of a dynamic haptic robotic trainer in internal jugular central venous catheter placement," *Am. J. Surg.*, vol. 219, no. 2, pp. 379–384, 2020, doi: 10.1016/j.amjsurg.2019.10.026.

[231] D. O'Donoghue, G. Davison, L. J. Hanna, B. McNaughten, M. Stevenson, and A. Thompson, "Calibration of confidence and assessed clinical skills competence in undergraduate paediatric OSCE scenarios: A mixed methods study 13 Education 1303 Specialist Studies in Education," *BMC Med. Educ.*, vol. 18, no. 1, pp. 1–8, 2018, doi: 10.1186/s12909-018-1318-8.

[232] J. Day *et al.*, "The Impact of a Comprehensive Simulation-Based Training and Certification Program on Resident Central Venous Catheter Complication Rates," *Simul. Healthc.*, vol. 16, no. 2, pp. 92–97, 2021, doi: 10.1097/SIH.0000000000000500.

[233] D. Nestel *et al.*, "Evaluating training for a simulated team in complex whole procedure simulations in the endovascular suite," *Med. Teach.*, vol. 31, no. 1, 2009, doi: 10.1080/01421590802337104.

[234] M. Vandewaetere, D. Manhaeve, B. Aertgeerts, G. Clarebout, J. J. G. Van Merriënboer, and A. Roex, "4C/ID in medical education: How to design an educational program based on whole-task learning: AMEE Guide No. 93," *Med. Teach.*, vol. 37, no. 1, pp. 4–20, 2015, doi: 10.3109/0142159X.2014.928407.

[235] N. E. Anton, T. Mizota, L. R. Timsina, J. A. Whiteside, E. M. Myers, and D. Stefanidis, "Attentional selectivity, automaticity, and self-efficacy predict simulator-acquired skill transfer to the clinical

environment," *Am. J. Surg.*, vol. 217, no. 2, pp. 266–271, 2019, doi: 10.1016/j.amjsurg.2018.11.028.

[236] C. Nitschmann, D. Bartz, and N. R. Johnson, "Gynecologic Simulation Training Increases Medical Student Confidence and Interest in Women's Health," *Teach. Learn. Med.*, vol. 26, no. 2, pp. 160–163, 2014, doi: 10.1080/10401334.2014.883984.

[237] ACGME, "ACGME Program Requirements for Graduate Medical Education in General Surgery," *ACGME Website*, pp. 1–31, 2017, [Online]. Available: http://www.acgme.org/portals/0/pfassets/programrequirements/440_general_surgery_2016.pdf.

[238] S. Hamstra and I. Philibert, "Simulation in Graduate Medical Education: Understanding Uses and Maximizing Benefits," *J. Grad. Med. Educ.*, vol. 4, no. 4, pp. 539–540, 2012, doi: 10.4300/jgme-d-12-00260.1.

[239] L. Bokken, J. J. Rethans, A. J. J. A. Scherpbier, and C. P. M. Van Der Vleuten, "Strengths and weaknesses of simulated and real patients in the teaching of skills to medical students: A review," *Simul. Healthc.*, vol. 3, no. 3, pp. 161–169, 2008, doi: 10.1097/SIH.0b013e318182fc56.

[240] J. M. Rodriguez-Paz *et al.*, "Beyond 'see one, do one, teach one': Toward a different training paradigm," *Postgrad. Med. J.*, vol. 85, no. 1003, pp. 244–249, 2009, doi: 10.1136/qshc.2007.023903.

[241] P. B. Andreatta, D. T. Woodrum, P. G. Gauger, and R. M. Minter, "LapMentor metrics possess limited construct validity," *Simul. Healthc.*, vol. 3, no. 1, pp. 16–25, 2008, doi: 10.1097/SIH.0b013e31816366b9.

[242] E. Johnson, "Simulating medical patients and practices: Bodies and the construction of valid medical simulators," *Body Soc.*, vol. 14, no. 3, pp. 105–128, 2008, doi: 10.1177/1357034X08093574.

[243] A. W. Calhoun and M. W. Scerbo, "Preparing and Presenting Validation Studies: A Guide for the Perplexed," *Simul. Healthc.*, vol. 17, no. 6, pp. 357–365, 2022, doi: 10.1097/SIH.0000000000000667.

[244] J. E. Benson, A. M. Hillebrand, and T. Auguste, "Design and evaluation of a low-cost speculum examination training model," *Simul. Healthc.*, vol. 12, no. 6, pp. 414–421, 2017, doi: 10.1097/SIH.0000000000000252.

[245] M. Chuang, H. Purswani, M. J. Fazzari, J. Kaplan, S. Pardanani, and E. H. Banks, "A Low-Cost Trainer for the Surgical Management of Postpartum Hemorrhage," *Simul. Healthc.*, vol. 15, no. 4, pp. 289–294, 2020, doi: 10.1097/SIH.0000000000000434.

[246] E. Capogna, F. Salvi, L. Delvino, A. Di Giacinto, and M. Velardo, "Novice and expert anesthesiologists' eyetracking metrics during simulated epidural block: A preliminary, brief observational report," *Local Reg. Anesth.*, vol. 13, pp. 105–109, 2020, doi: 10.2147/LRA.S267879.

[247] E. Fichtel *et al.*, "Eye tracking in surgical education: gaze-based dynamic area of interest can discriminate adverse events and expertise," *Surg. Endosc.*, vol. 33, no. 7, pp. 2249–2256, 2019, doi: 10.1007/s00464-018-6513-5.

[248] G. A. Wright *et al.*, "Eye-Tracking Technology to Determine Procedural Proficiency in Ultrasound-Guided Regional Anesthesia," *J. Educ. Perioper. Med.*, vol. 24, no. 1, pp. 1–10, 2022, doi: 10.46374/volxxiv_issue1_zurca.

[249] E. A. Henneman, J. L. Marquard, D. L. Fisher, and A. Gawlinski, "Eye Tracking: A Novel Approach for Evaluating and Improving the Safety of Healthcare Processes in the Simulated Setting," *Simul. Healthc.*, vol. 12, no. 1, pp. 51–56, 2017, doi: 10.1097/SIH.0000000000000192.

[250] P. Culligan, E. Gurshumov, C. Lewis, J. Priestley, J. Komar, and C. Salamon, "Predictive Validity of a Training Protocol Using a Robotic Surgery Simulator," *Female Pelvic Med. Reconstr. Surg.*, vol. 26, no. 5, pp. 287–298, 2020, doi: 10.1097/SPV.

[251] J. H. Barsuk, W. C. McGaghie, E. R. Cohen, K. J. O'Leary, and D. B. Wayne, "Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit," *Crit. Care Med.*, vol. 37, no. 10, pp. 2697–2701, 2009, doi: 10.1097/CCM.0b013e3181a57bc1.

[252] M. B. Antonoff, R. C. Shelstad, C. Schmitz, J. Chipman, and J. D'Cunha, "A Novel Critical Skills Curriculum for Surgical Interns Incorporating Simulation Training Improves Readiness for Acute Inpatient Care," *J. Surg. Educ.*, vol. 66, no. 5, pp. 248–254, 2009, doi: 10.1016/j.jsurg.2009.09.002.

[253] L. Hazell, H. Lawrence, and H. Friedrich-Nel, "Simulation based learning to facilitate clinical readiness in diagnostic radiography. A meta-synthesis," *Radiography*, vol. 26, no. 4, pp. e238–e245, 2020, doi: 10.1016/j.radi.2020.03.006.

[254] H. L. Lew, J. H. Poole, E. H. Lee, D. L. Jaffe, H. C. Huang, and E. Brodd, "Predictive validity of driving-simulator assessments following traumatic brain injury: A preliminary study," *Brain Inj.*, vol. 19, no. 3, pp. 177–188, 2005, doi: 10.1080/02699050400017171.

[255] O. V. Bitkina, J. Park, and H. K. Kim, "The ability of eye-tracking metrics to classify and predict the perceived driving workload," *Int. J. Ind. Ergon.*, vol. 86, no. September, p. 103193, 2021, doi: 10.1016/j.ergon.2021.103193.

[256] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, "Real-time Prediction of User Performance based on Pupillary Assessment via Eye-Tracking," *AIS Trans. Human-Computer Interact.*, vol. 10, no. 1, pp. 26–60, 2018, doi: 10.17705/1thci.00103.

[257] C. Hasse, D. Grasshoff, and C. Bruder, "Eye-tracking parameters as a predictor of human performance in the detection of automation failures," *Proc. HFES Eur. Chapter Conf. Toulouse*, pp. 133–144, 2012.

[258] S. Upasani, D. Srinivasan, Q. Zhu, J. Du, and A. Leonessa, "Eye-Tracking in Physical Human–Robot Interaction: Mental Workload and Performance Prediction," *Hum. Factors*, 2023, doi: 10.1177/00187208231204704.

[259] A. Ryason *et al.*, "Validation of a Virtual Simulator With Haptic Feedback for Endotracheal Intubation Training," *Simul. Healthc. J. Soc. Simul. Healthc.*, vol. Publish Ah, no. 00, pp. 1–9, 2023, doi: 10.1097/sih.0000000000000708.

[260] G. Wulf, C. Shea, and R. Lewthwaite, "Motor skill learning and performance: A review of influential factors," *Med. Educ.*, vol. 44, no. 1, pp. 75–84, 2010, doi: 10.1111/j.1365-2923.2009.03421.x.

[261] L. E. Kirson and J. Friedman, *Central Venous Catheterization and Pressure Monitoring*, 4th ed. Elsevier Inc, 2010.

[262] K. Chatterjee, "The Swan-Ganz catheters: Past, present, and future: A viewpoint," *Circulation*, vol. 119, no. 1, pp. 147–152, 2009, doi: 10.1161/CIRCULATIONAHA.108.811141.

[263] C. Wilde, "Understanding the Difference: Swan-Ganz Catheterization Vs. Right Heart Catheterization," *Medical Record Associations, LLC*, 2018. https://www.mrahis.com/understanding-the-difference-swan-ganz-catheterization-vs-right-heart-catheterization/.

[264] G. C. Huang *et al.*, "Procedural Competence in Internal Medicine Residents: Validity of a Central Venous Catheter Insertion Assessment Instrument," *Acad. Med.*, vol. 84, no. 8, pp. 1127–1134, 2009, doi: 10.1097/ACM.0b013e3181acf491.

[265] J. H. Barsuk, E. R. Cohen, D. B. Wayne, W. C. McGaghie, and R. Yudkowsky, "A comparison of approaches for mastery learning standard setting," *Acad. Med.*, vol. 93, no. 7, pp. 1079–1084, 2018, doi: 10.1097/ACM.0000000000002182.

[266] M. Yovanoff, D. F. Pepley, K. Mirkin, J. Z. Moore, D. C. Han, and S. R. Miller, "IMPROVING MEDICAL EDUCATION: SIMULATING CHANGES IN PATIENT ANATOMY USING DYNAMIC HAPTIC FEEDBACK," *Proc Hum Factors Erg. Soc Annu. Meet.*, vol. 60, no. 1, pp. 603–607, 2016, doi: 10.1177/1541931213601138.Abstract.

[267] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Z. Yang, and A. Darzi, "Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair," *Surg. Endosc.*, vol. 29, no. 2, pp. 405–413, 2015, doi: 10.1007/s00464-014-3683-7.

[268] S. Erridge, H. Ashraf, S. Purkayastha, A. Darzi, and M. H. Sodergren, "Comparison of gaze behaviour of trainee and experienced surgeons during laparoscopic gastric bypass," *Br. J. Surg.*, vol. 105, no. 3, pp. 287–294, 2018, doi: 10.1002/bjs.10672.

[269] A. T. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements," *Math. Biosci.*, vol. 24, no. 3–4, pp. 191–204, 1975, doi: 10.1016/0025-5564(75)90075-9.

[270] H. Collewijn, C. Erkelens, and R. M. Steinmant, "BINOCULAR CO-ORDINATION OF HUMAN VERTICAL SACCADIC EYE MOVEMENTS," *J. Physiol.*, vol. 404, pp. 183–197, 1988.

[271] O. V. Komogortsev, S. Jayarathna, D. H. Koh, and S. M. Gowda, "Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms," *Eye Track. Res. Appl. Symp.*, pp. 65–68, 2010, doi: 10.1145/1743666.1743682.

[272] T. Pro, "Tobii Pro Lab Gaze Filter," *Tobii Connect*. https://connect.tobii.com/s/article/Gaze-Filter-functions-and-effects?language=en_US.

[273] J. B. Cooper and V. R. Taqueti, "A brief history of the development of mannequin simulators for clinical education and training.," *Postgrad. Med. J.*, vol. 84, no. 997, pp. 563–570, 2008, doi: 10.1136/qshc.2004.009886.

APPENDIX

Complete results of the chi-square analysis to compare pass rates on the IJ-CVC Checklist from chapter 4

Checklist Item	Treatment	Fail	Pass	Chi-square	p-value
Verbalizes Consent	Comprehensive	8 (13.6%)	51 (86.4%)	14.252	**<.001***
	DHRT	22(46.8%)	25 (53.2%)		
Prepares Kit	Comprehensive	15 (27.3%)	40(72.7%)	5.970	.015
	DHRT	23 (51.1%)	22(48.9%)		
Sterile Technique	Comprehensive	1 (1.7%)	57 (98.3%)	NA	.170
	DHRT	4 (8.5%)	43 (91.5%)		
Appropriate Site	Comprehensive	1 (1.7%)	57 (98.3%)	NA	1.00
	DHRT	1 (2.1%)	46 (97.6%)		
Applied Anesthesia	Comprehensive	7(12.1%)	51(87.9%)	2.514	.113
	DHRT	11(23.9%)	35(76.1%)		
Ultrasound Orientation	Comprehensive	3 (5.2%)	55 (94.8%)	NA	.462
	DHRT	5 (10.6%)	42 (89.4%)		
Ultrasound Clear Image	Comprehensive	0 (0%)	59 (100%)	NA	.194
	DHRT	2(4.3%)	45 (95.7%)		
Distinguish Anatomy	Comprehensive	2 (3.4%)	57 (96.6%)	NA	.502
	DHRT	0 (0%)	47 (100%)		
Needle Insertion Angle	Comprehensive	7(11.9%)	52(88.1%)	NA	.507
	DHRT	3 (6.4%)	44 (93.6%)		
Locating Needle on Ultrasound	Comprehensive	7 (12.1%)	51 (87.9%)	NA	.751
	DHRT	4 (8.5%)	43 (91.5%)		
Advancing Needle	Comprehensive	5 (8.6%)	53 (91.4%)	NA	.750
	DHRT	5 (10.6%)	42 (89.4%)		
Successful Venipuncture	Comprehensive	2 (3.6%)	53 (96.4%)	NA	.402
	DHRT	4 (9.1%)	40 (90.9%)		
Confirm Entry with Aspiration	Comprehensive	5 (8.5%)	54 (91.5%)	NA	1.000
	DHRT	3 (6.5%)	43 (93.5%)		
Remove Syringe	Comprehensive	22 (37.9%)	36 (62.1%)	1.284	.257
	DHRT	23 (48.9%)	24 (51.1%)		
Guidewire Insertion	Comprehensive	10 (17.5%)	47 (82.5%)	.044	.833
	DHRT	9 (19.1%)	38 (80.9%)		
Guidewire control	Comprehensive	8 (14%)	49 (86%)	.036	.850
	DHRT	6 (12.8%)	41 (87.2%)		
Needle Removal	Comprehensive	5 (8.8%)	52 (91.2%)	1.715	.190
	DHRT	8 (17.4%)	38 (82.6%)		
Verbalizes Incision	Comprehensive	1 (1.7%)	57 (98.3%)	7.752	.005
	DHRT	8 (17.0%)	39 (83.0%)		
Verbalizes Dilation	Comprehensive	1 (1.7%)	57 (98.3%)	NA	.170
	DHRT	4 (8.5%)	43 (91.5%)		
Catheter Insertion and Wire Removal	Comprehensive	4 (6.9%)	54 (93.1%)	NA	.723
	DHRT	4 (9.1%)	40 (90.9%)		
Verbalizes Catheter Distance	Comprehensive	14 (24.6%)	43 (75.4%)	6.517	.011
	DHRT	22 (48.9%)	23 (51.1%)		
Aspirates Blood through Catheter	Comprehensive	11 (19%)	47 (81.0%)	11.229	**<.001***
	DHRT	23 (50%)	423 (50%)		
Verbalizes Suture	Comprehensive	4 (6.9%)	54 (93.1%)	NA	.126
	DHRT	0 (0%)	47 (100%)		
Verbalizes X-ray	Comprehensive	2 (3.5%)	55 (96.5%)	NA	.135
	DHRT	6 (13.0%)	40 (87.0%)		

Note: Fisher's exact test was used for all chi-square columns of NA; ***** indicates significant p values for p <.05

VITA

EDUCATION

The Pennsylvania State University, University Park, PA
PhD, Industrial Engineering (Focus: Human Factors) *June 2020 to May 2024*
MS, Industrial Engineering *June 2020 to December 2021*

University of Maryland, College Park, MD *August 2014 to May 2018*
BS Mechanical Engineering, Minor International Engineering, Departmental Honors
Study Abroad: Nanyang Technological University, Singapore, 2017

FELLOWSHIP

Diefenderfer Entrepreneurial Fellowship (August 2022 – August 2023) and two follow up scholarships

JOURNAL ARTICLES

Tzamaras, H.M., Gonzalez-Vargas, J.M., Brown D., Moore, J.Z., and Miller, S.R., "Evaluating the effects of comprehensive simulation on central venous catheterization training: A comparative observational study". BMC medical Education. In Preparation to be submitted March 2024

Tzamaras, H.M., Brown D., Moore, J.Z., and Miller, S.R., "Tapping into efficient learning: An exploration of the impact of sequential learning on skill gains and learning curves in central venous catheterization" Medical Education and Curricular Development. Under Review as of February 2024

Tzamaras, H.M., Sinz, L., Mast, J, Moore, J.Z., and Miller, S.R., "Tracking success: validating a central venous catheterization trainer through eye gaze analysis in clinical and simulated environments" Journal of Simulation in Healthcare. In Preparation to be submitted March 2024

Tzamaras, H.M., Sinz, L., Ng, P., Yang, M., Moore, J.Z., and Miller, S.R., "Competence over confidence: gender based-differences in resident training for central venous catheterization" BMC Medical Education. In Preparation to be submitted March 2024

PEER REVIEWED CONFERENCE ARTICLES

Tzamaras, H. M., Wu, H., Moore, J. Z., & Miller, S. R. (2023). "Shifting Perspectives: A proposed framework for analyzing head-mounted eye-tracking data with dynamic areas of interest and dynamic scenes". In Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Tzamaras, H.M., Hicks, S., Sallai, G., Cummings, C., Dennis, L., Nolte, H., Restrepo, A., Reed, C., (2023). "Fostering community at the graduate level: one university's student-led approach." In 2023 ASEE Annual Conference & Exposition.

Tzamaras, H. M., Brown, D. C., Gonzalez-Vargas, J. M., Moore, J. Z., & Miller, S. R. (2022). "Slow and Steady: Examining the impact of hands-on instructions on learnability of a training simulator to enhance development of core skills in Central Venous Catheterization". In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 66, No. 1, pp. 1867-1871). Sage CA: Los Angeles, CA: SAGE Publications.

Tzamaras, H. M., Brown, D., Gonzalez-Vargas, J. M., Moore, J., & Miller, S. R. (2021). "Sensorization of Medical Trays for Tray-Based Surgical Procedure Simulation." In Frontiers in Biomedical Devices (Vol. 84812, p. V001T13A004). American Society of Mechanical Engineers.

Tzamaras, H. M., Martinez, J., Brown, D. C., Gonzalez-Vargas, J. M., Moore, J. Z., & Miller, S. R. (2021). "Fun and Games: Designing a Gamified Central Venous Catheterization Training Simulator." In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 65, No. 1, pp. 267-271). Sage CA: Los Angeles, CA: SAGE Publications.