Behavioral Assessment in Virtual Reality:
An Evaluation of Multi-User Simulations in Healthcare Education

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

by
Steven J. Anbro

Dr. Ramona A. Houmanfar / Dissertation Advisor

December, 2020
We recommend that the dissertation prepared under our supervision by

STEVEN JAMES ANBRO

entitled

Behavioral Assessment in Virtual Reality:
An Evaluation of Multi-User Simulations in Healthcare Education

be accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Ramona A. Houmanfar, Ph.D., Advisor

Mark P. Alavosius, Ph.D., Committee Member

Matthew P. Lewon, Ph.D., Committee Member

Amber M. Maraccini, Ph.D., Committee Member

Frederick C. Harris, Jr., Ph.D., Committee Member

Laura H. Crosswell, Ph.D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

December, 2020
Abstract

Human error in medicine – medical error – has been identified as the third leading cause of death within the United States. Analyses of deaths attributable to medical error conclude that, overwhelmingly, faulty communication plays a central role in medical error. Faulty communication can occur at any time. Patient handoffs, the transfer of patient care from one medical professional to another, are frequently occurring behavioral events in healthcare settings where communication accuracy is vital. Using lessons learned from other highly technical, risk-inherent industries such as aviation, the medical industry has put together a training package called TeamSTEPPS® to address this systemic problem. This initiative began in clinical settings and has worked its way into healthcare education. Among such initiatives, a fundamental challenge is the objective measurement of specific, critical skills. It is increasingly important to develop valid measurement and assessment protocols for the medical industry. Behavior science offers a robust history of objective behavioral measurement and assessment. Virtual reality (VR) provides a measurement-rich platform for assessing behavior in simulations, but its replacement of in-person (Direct) simulations lacks validation research. This study evaluated the validity of using VR simulations in healthcare education to measure and assess critical skills identified by the TeamSTEPPS® framework for healthcare professionals during simulated patient handoffs.
Dedication

This dissertation is dedicated to the memory of my grandmother, Myrna Tipps. Thank you for being a light in my life, and for bringing magic to my childhood. Though you are no longer with us, your “grandma-isms” live on – bringing joy and laughter to our hearts. Thank you for unapologetically being yourself; I strive to live up to the example you set. I love you.
Acknowledgements

First and foremost, I would like to acknowledge and thank my family for their unyielding support throughout my life – particularly throughout my time in graduate school. Mom and Dad: your love, support, and wisdom have meant so much to me over these years. This dissertation is as much a product of your behavior as it is of mine; it could have never happened without you. Words cannot adequately express my gratitude, so I will simply say this: Thank you. I love you.

To my advisor, Dr. Ramona Houmanfar. Thank you for your support of this research project, from its early stages of inception through its execution and dissemination. This has been an incredible research opportunity that could not have happened without your expertise and help. Thank you for giving me the freedom to work on such a unique project; your feedback and support throughout this project has been instrumental in its success.

To my dissertation committee: Drs. Mark Alavosius, Laura Crosswell, Fred Harris, Matthew Lewon, & Amber Maraccini. Thank you for bringing your wide-ranging insight and expertise to the table for this project. Your conceptual, methodological, and analytic contributions to this project helped shape it into a successful research accomplishment. A special thank you to Dr. Amber Maraccini for laying the empirical foundations of this research with your own, and to Dr. Fred Harris for your on-site support and supervision of simulation sessions.

To Julie Thomas and Kim Baxter. Thank you both for your significant contributions in the development and execution of this study. You are amazing collaborators to work with, and it was such a pleasure to organize this study with your help. From scheduling the study activities and handling many behind-the-scenes logistics, to co-running the training sessions’ experiential activities and even helping me analyze student data, you both have been a godsend. Thank you for being so responsive and accommodating, especially in light of the public health restrictions we’ve all had to work under while running this study.

To the VR team: Luka Starmer, Lucas Calabrese, Christopher Lewis, and Yifan Zhang. I cannot say thank you enough. Without your hard work, we would not even have a VR sim to use in this study! Thank you for the significant amount of time and effort you invested into creating and refining this simulation, and thank you for your on-site support running participants through the VR simulation. The hard work you all put into making this happen is greatly valued and appreciated, and the final simulation is just so cool to see and experience. You should all be very proud of the work you put into this, and of the final product you generated.
To the @One Digital Media team: Michelle Rebaleati, Kyle Weerheim, and Daniel Fergus. Thank you to Daniel for your help in the photogrammetry process used to begin creating the VR simulation for this study. Thank you to Michelle and Kyle for helping us share this research. The time you invested video recording simulation activities, interviewing student participants, and creating multimedia materials for dissemination will no doubt generate further interest in the ongoing challenge of medical error– and the research efforts aimed at mitigating these errors. Thank you for sharing this research in a captivating way.

To the healthcare professionals serving as training workshop instructors: Jill Packman, Jen Richards, and Paul Sierzenski. Thank you for volunteering your time and expertise to run fantastic training workshops for our participants. Your experiences and perspectives helped demonstrate the importance and significance of interdisciplinary work, and the positive impact it has on patients. The student participants were fortunate indeed to learn from you.

To the medical school’s simulation lab team: Kelly Farley, Devin Saxon, and Nicole Preston. Thank you for your help coordinating and running student participants through non-VR simulation sessions. The primary research question of this study focuses on the comparison between VR and non-VR simulations, and without your help we would have no comparison to make! A special thank you to Kelly Farley for your help in behind-the-scenes logistics: from coordinating photogrammetry sessions and simulation schedules, to setting up non-VR sim rooms and recording those simulation sessions with your team for our analysis.

Finally, a special acknowledgement to my outstanding team of research assistants: Kian Assemi, Daniel Benitez, Jonathan Corona Rubio, Anne George, Jose Ruiz, and Liz Stevenson. Thank you all so much for the time you dedicated to helping me run this study, and subsequently transcribing & coding all simulation sessions. This project was successful thanks in large part to your efforts. I wish you all the best in your future pursuits. Keep demonstrating the effort you showed in this project, and you’ll accomplish great things.

This dissertation would have been impossible without the contributions and support of these many individuals, across several academic and professional disciplines, whom I have had the privilege of working with. I would like to formally acknowledge my colleagues who have supported this research project throughout its many stages of development, implementation, and analysis. This project is truly a testament to interdisciplinary teamwork; I thank you all.

“...we are only as strong as we are united, as weak as we are divided.”

– Albus Dumbledore

(Rowling, 2000, p. 723)
# Table of Contents

List of Tables............................................................................................................... v

List of Figures............................................................................................................. vi

Introduction.................................................................................................................. 1

Method.......................................................................................................................... 29

Results......................................................................................................................... 46

Discussion..................................................................................................................... 64

References.................................................................................................................... 80

Tables......................................................................................................................... 88

Figures......................................................................................................................... 95

Appendices.................................................................................................................. 130
List of Tables

Table 1. Timeline of simulation events
Table 2. A behavior analytic interpretation of situational awareness
Table 3. Distribution of participants
Table 4. Simulation social validity responses
Table 5. Estimated cost analysis
Table 6. Statistically significant differences in performance across simulations
Table 7. Summary of statistical comparisons
List of Figures

Figure 1. TeamSTEPPS® model
Figure 2. Participant communication accuracy
Figure 3. Situational awareness pre-test vs. post-test measures
Figure 4. Cumulative simulation-based research in behavior science
Figure 5. History of virtual reality
Figure 6. I-PASS handoff tool
Figure 7. Virtual simulation AOIs
Figure 8. Study methodology
Figure 9. Simulation process
Figure 10. Handoff provider performance during simulated patient handoffs
Figure 11. Handoff durations
Figure 12. Handoff provider response rates by simulation type
Figure 13. Pretest to posttest response rate change
Figure 14. Handoff provider response rates by training workshop attended
Figure 15. Patient handoff scoring
Figure 16. Participant handoff change scores
Figure 17. Handoff receiver performance during simulated patient handoffs
Figure 18. “Filler words” during simulations
Figure 19. Communication accuracy in sample dyads
Figure 20. Situational awareness level 1 pattern classifications
Figure 21. *Inattentive* eye fixation pattern sample
Figure 22. *Attentive* eye fixation pattern sample
Figure 23. *Focused* eye fixation pattern sample
Figure 24. *Distracted* eye fixation pattern sample
Figure 25. *Divided* eye fixation pattern sample
Figure 26. *Observant* eye fixation pattern sample
Figure 27. Eye fixation patterns during VR simulations
Figure 28. Eye fixation patterns in a sample dyad
Figure 29. Situational awareness level 2 scores
Figure 30. Situational awareness level 3 response frequencies
Figure 31. Situational awareness in VR: Levels 1 and 2
Figure 32. Situational awareness levels 1 and 2, by stability
Figure 33. Situational awareness levels 1 and 2, by AOI fixation
Figure 34. Situational awareness in VR: Levels 2 and 3
Figure 35. T-TAQ survey results
Behavioral Assessment in Virtual Reality: Evaluating the Validity of Multi-User Simulations in Healthcare Education

The prevalence of medical error within the United States is an ongoing challenge requiring innovative solutions. In November 1999, the Institute of Medicine published a groundbreaking report titled, To Err is Human: Building a Safer Health System. This report led to the beginning of the patient safety movement and marked the “first time the impact and consequences of medical errors were quantified” (Patient Safety Movement Foundation, n.d.). The numbers were staggering: an estimated 44,000 – 98,000 annual American fatalities were attributed to medical error. These estimates exceeded the concurrent death toll from motor vehicle accidents, breast cancer, and AIDS; estimated monetary costs ranged from $17 – 29 billion annually (Institute of Medicine, 1999).

Following the Institute of Medicine’s initial report, newer estimates have suggested this problem’s escalation. James (2013) reviewed studies published between 2008 – 2011, concluding the number of preventable American deaths attributable to medical error had risen to 210,000 – 400,000+ annually. More recently, this increasing trend has led to medical error being ranked as the third leading cause of death in the United States (Makary & Daniel, 2016), following closely behind heart disease (n = 611,000) and cancer (n = 585,000). It is worth noting that human/system factors (e.g., “medical error”) are not included in the International Classification of Disease (ICD) codes, unlike other leading causes of death. ICD codes are used to assign causes of death on death certificates,
and summaries of these data are reported annually by the US Centers for Disease Control and Prevention (CDC). Without accurate ICD codes to track medical error as a cause of death, these reported numbers may underestimate the problem's significance and prevalence.

The Joint Commission conducts ongoing analyses of data pertaining to sentinel events, defined as “a patient safety event (not primarily related to the natural course of the patient’s illness or underlying condition) that reaches a patient and results in any of the following: death, permanent harm, or severe temporary harm and intervention required to sustain life” (2019, p. 2). Sentinel events are reported to the Joint Commission voluntarily; therefore, these data represent only a small sample of actual events. Given the available sample data, the Joint Commission lists the following sentinel events among the top ten most frequent: unintended retention of a foreign object, wrong-patient/wrong-site/wrong-procedure, patient fall, patient suicide, and delay in treatment (2019). Furthermore, similar potential causative factors for these sentinel events are identified; communication failures are almost universally included among these lists of causative factors, maintaining research findings from nearly two decades ago. Sutcliffe, Lewton, and Rosenthal (2004) analyzed reported estimates of death due to medical error; they attributed approximately 2/3 of sentinel events to communication errors among healthcare teams. Communication errors are hypothesized to increase in frequency given rigid organizational hierarchies, ambiguity, and lack of proper team communication training.
Given the prevalence of medical error leading to sentinel events (death, in particular) coupled with increased public awareness of the problem, the medical industry has looked to high-reliability organizations (HROs) across other industries for guidance. HROs are organizations whose employees conduct regularly occurring, highly technical operations in working conditions that range from moderate to high levels of potential risk (Anbro et al., 2020). A defining feature of HROs is the interlocking of behavior across managerial and operational levels of an organization; this fosters systemic learning from accidents and helps institutionalize performance/safety corrections based on lessons learned and process innovation (Alavosius, Houmanfar, Anbro, Burleigh, & Hebein, 2017; Dekker & Woods, 2009). The medical industry has looked to the aviation industry and their development of *Crew Resource Management (CRM)* as a guide to address persisting human/systems challenges.

**CRM Origins**

CRM originated in the aviation industry as a response to several high-casualty events, including the crash of Eastern Airlines Flight 401 into the Florida Everglades (101 fatalities in December 1972) and the collision of two planes on the tarmac in the Canary Islands: known colloquially as the Tenerife Airport Disaster (583 fatalities in March 1977). The National Aeronautics and Space Administration (NASA) conducted a workshop in 1979 to identify common factors leading to these events (Weiner, Kanki, & Kelmreich, 1993). Subsequent incident investigations into these (and other) catastrophic events revealed that an estimated 60 – 80% of incidents were directly attributable to human error
Human error in these incidents included ineffective interpersonal communication and poor decision-making. These errors were also an outcome of rigid hierarchical structures between the captain and other flight crew members. In these rigid hierarchies, the captain held ultimate authority and was never to be questioned; this is predicated on the faulty assumption that captains cannot and will not make errors themselves. Similar paradigms were present across other industries. In the medical industry, for instance, rigid hierarchies among patient care teams resulted in physicians playing a "sage on stage" role. Physicians’ decisions were not to be questioned by physician assistants, nurses, or any other member of the patient care team.

In response to the overwhelming contribution of human error to catastrophic incidents in aviation, CRM emerged as a method of training and behavior modification which shifted organizational cultures in aviation. CRM is defined as "the effective use of all resources, including hardware, software, and people, to achieve the highest possible level of safety" (Northwest Airlines, 2005, p.1). A concise summary of CRM principles is included in (now-defunct) Northwest Airlines' flight operations manual. They summarized CRM using the Four Words: "authority with participation, and assertiveness with respect" (Northwest Airlines, 2005, p.1). CRM also utilizes high-fidelity cockpit simulators to train and assess pilots' technical skills and human factors. Today, the utilization of CRM across the board in aviation has resulted in the industry emerging as the leader in transportation safety, with 0.07 passenger fatalities per billion passenger miles traveled (Savage, 2013). When an airplane crash occurs
in the present day, it is striking due to its rarity. A short case study illustrates flight crews' empowerment to make decisions through the systemic implementation of CRM. Note the outcome differences (notably, the death toll) between pre-CRM disasters discussed above and post-CRM incidents such as the following.

**CRM: A Sample Case Study**

In January 2009, US Airways Flight 1549 took off from LaGuardia airport in New York City. Shortly after takeoff, Captain Chesley “Sully” Sullenberger and First Officer (FO) Jeffrey Zaslow were faced with a perilous situation. A larger-than-average bird strike occurred when a flock of Canada Geese was caught in both engines of their Airbus A320, causing complete engine failure and an immediate halt in the plane's forward thrust and ascension. Because of the institutionalization of CRM within aviation, Sully and his FO were trained to communicate quickly and efficiently. Standard protocols found in the pilots’ flight operations manual were attempted unsuccessfully, leaving an opportunity for one of two outcomes: acceptable variability in pilot behavior or death for all 155 people aboard. The pilots decided to perform a controlled water landing in the Hudson River (known as “ditching” the aircraft; FAA, 2020). This course of action resulted in 95 minor injuries, five more serious injuries, and no casualties. Subsequent incident investigations and simulated re-creations of Flight 1549's path and engine failure confirmed the pilots’ correct action. Had they attempted to land at either LaGuardia or Teterboro (the two nearest airports), the plane would have crashed – causing substantial property destruction and death (Sullenberger & Zaslow, 2009).
A Behavioral Analysis of CRM

From a behavior analytic perspective, CRM is an initiative at the organization level that facilitates the selection and reinforcement of certain classes of behavior. At the task level (Malott, 2003), operations (e.g., a specific flight in aviation or a surgical procedure in medicine) can be viewed “as a cascading chain of behavioral events where the leader and crew members effectively utilize available resources to:

1. Plan a work process,
2. Brief everyone on roles/functions,
3. Monitor the process as it occurs,
4. Detect & report deviations from the plan,
5. Communicate corrections from the top down,
6. Adjust actions as needed,
7. Debrief at important moments (at significant change or conclusion of work),
8. Learn to refine the human-machine interface.

In complex and dynamic situations, CRM orchestrates cooperation among crew members that have different vantage points on the process” (Alavosius, Houmanfar, Anbro, Burleigh, & Hebein, 2017, p. 145-146). A central tenet of CRM, and all subsequent adaptations across industries, is a balance between two classes of behavior: procedural adherence and behavioral variability. Procedural adherence is requisite during normal operations. Behavioral variability is regarded as beneficial, even necessary, in situations where procedural adherence yields no useful result. By empowering crew members to make decisions in the moment, CRM helps mitigate the adverse effects of a bureaucratic organizational structure. Rather than figuratively paralyzing someone in the field with rules and regulations during a catastrophe, those same
operators are instead empowered to think and behave in novel or creative ways to solve the problem at hand, saving as many lives as possible. While empowerment via CRM is evident in unique incidents such as Flight 1549, other principles found in CRM are essential for standard operations.

Retired US Navy Captain and former NASA astronaut Jim Wetherbee has illustrated the necessary balance between adherence to standard procedures and behavioral variability in perhaps the most novel of environments: outer space (2016). Certain procedures, Wetherbee argues, should never allow for deviation; among these procedures is a standard method, vocabulary, and etiquette used in all communications – among the team itself, and when communicating with Mission Control. By standardizing both the topography and the function of communication (e.g., inter-team humor is appropriate until a particular time before launch, to ease pre-flight tension), the potential for communication error is reduced.

**CRM Translational Efforts in Medicine**

Every iteration of CRM across the aviation industry targets several critical skills, which vary depending upon the specific airline. Examples of these skills include communication, situational awareness, decision-making, leadership, teamwork, and management of limits of crew members' capabilities (Alavosius et al., 2017). Other industries adopting CRM include medicine, nuclear power, and oil & gas. These industries often identify similar critical skills in their CRM models. The medical industry was among the first outside of aviation to adopt CRM, through multiple initiatives.
Anesthesia Crisis Resource Management

The original adaptation of CRM into the medical field was known as Anesthesia Crisis Resource Management (ACRM; Howard, Gabba, Fish, Yang, & Sarnquist, 1992). ACRM applied principles from CRM in aviation to anesthesiologist training. The goal of ACRM was to better equip anesthesiologists to handle crisis scenarios while working within interdisciplinary teams. During this training, anesthesiologists acclimated to the simulation environment before participating in 2-hour simulation sessions. Videotapes of these simulation sessions were reviewed during debrief sessions. Written test scores showed significant improvement following the initial course using anesthesiology residents as subjects. More experienced anesthesiologists in a second course did not demonstrate similarly significant improvements in their written test scores. Social validity assessments of ACRM training were also conducted. Participants subjectively rated video examples of real anesthetic mistakes, the inclusion of simulation sessions, and post-simulation debriefings as highly enjoyable elements of the training that were helpful to their practice.

MedTeams

While ACRM solely targeted anesthesiologists, subsequent iterations of CRM in medicine expanded the population of interest. MedTeams is the second adaptation of CRM that targets entire emergency room teams as the population of interest. Morey et al. (2002) evaluated the effectiveness of MedTeams training across nine teaching and community hospital emergency departments, using a quasi-experimental, untreated control group design. Trained observers reported
a statistically significant improvement in the quality of team behaviors shown by the experimental group. Team behaviors were five dimensions assessed using "behaviorally anchored rating scales [that] were validated in previous military aviation research…” (Morey, et al., 2002, p. 1558). The clinical error rate was also measured, and participants receiving the MedTeams training reduced their rate of errors from 30.9% to 4.4% throughout the study. Social validity assessment of MedTeams training showed improved attitudes towards teamwork and improved staff assessment of institutional support.

**TeamSTEPPS®**

The most recent iteration of CRM within the medical industry is Team Strategies and Tools to Enhance Performance and Patient Safety, or TeamSTEPPS® (Salas, Burke, & Stagl, 2004; TeamSTEPPS®, n.d.). TeamSTEPPS® emerged as a collaborative effort between the Agency for Healthcare Research and Quality (AHRQ) and the Department of Defense (DOD). Like most iterations of CRM, TeamSTEPPS® places emphasis on specific critical skills or competencies – in this case: communication, leadership, situation monitoring, and mutual support. The TeamSTEPPS® model is shown in Figure 1. The model depicts four core competencies encircled by the patient care team, consisting of the patient and all relevant healthcare providers (e.g., primary physician, specialized consultants, attending nurses, home care providers, etc.). The model also depicts three desirable outcomes (improved performance, knowledge, and attitudes) that can be achieved by patient care teams who successfully demonstrate mastery of the four core competencies.
Since its creation, TeamSTEPPS® training has been used in many research studies to evaluate its impact. The full TeamSTEPPS® training curriculum (currently, version 2.0) consists of an introductory “Essentials Course” followed by 7 “Fundamentals” modules and 5 “Supplemental” modules (AHRQ, 2019). The effects of this didactic training have been promising. Similar to MedTeams, TeamSTEPPS® has resulted in reduced rates of medical errors (Cima, et al., 2009; Haig & Sutton, 2006; Mann, Marcus, & Sachs, 2006), improved communication techniques (Turner, 2012; Ward, Zhu, Lampman, & Stewart, 2015), and improvements among the core competencies of the TeamSTEPPS® model (Sawyer, Laubach, Hudak, Yamamura, & Pocrnich, 2013; Sheppard, Williams, & Klein, 2013). The range of improvements shown in the medical literature has resulted in the ongoing use of TeamSTEPPS® training as a gold standard in the medical industry.

Following the successes of TeamSTEPPS®, a portion of its curriculum was further developed to create a tool used in patient handoffs called the I-PASS (Starmer et al., 2014). I-PASS is an acronym that stands for Illness severity, Patient summary, Action list, Situation awareness and contingency planning, and Synthesis by receiver (emphasis added). The I-PASS mnemonic structures the order and content of patient handoffs. This order standardizes interprofessional communication during critical patient transfer events. Initial research evaluating the I-PASS training bundle demonstrated a reduction in medical errors and reduced omissions of patient information during handoffs (Starmer et al., 2013). More recently, interdisciplinary research has also demonstrated the effects of a
condensed TeamSTEPPS® + I-PASS training on the performance of medical and nursing students.

**Behavioral Measurement in TeamSTEPPS®**

Maraccini, Houmanfar, Kemmelmeier, Piasecki, & Slonim (2018) observed medical and nursing student performance in simulated patient handoffs, before and after a 90-minute TeamSTEPPS® + I-PASS training. *Patient handoffs* in the TeamSTEPPS® protocol refer to the transfer of patient care from any medical professional to another. Handoffs are critical moments of inter-professional communication where opportunities for error abound. While previous research relied on subjective ratings or Likert scales to assess communication (along with the other three core competencies), Maraccini and colleagues utilized an objective method of measurement (2018). All simulated patient handoff sessions were video recorded, transcribed, and coded for varying levels of accuracy. The results indicated an increase in communication accuracy among medical and nursing students, following the shortened TeamSTEPPS® training. Notably, these results constitute the first published data using objective behavioral measures to measure the impact of TeamSTEPPS® on performance; this reveals perhaps the most significant limitation of CRM research across industries: the lack of objective behavioral measurement.

**Measurement of Critical Skills in CRM**

The current standard of behavioral measurement in the medical research literature reveals gaps that need filling. Self-report or subjective ratings of perceptions and attitudes (e.g., Likert scales) are used almost exclusively to
communicate behavior change in this literature. While these measures may be coupled with more objective outcome measures (e.g., reducing medical errors made by a team), behavior at the individual level is seldom accounted for objectively. From a behavior analytic perspective, these measures are non-dimensional. That is, they are determined on the basis of something other than the measurable dimensions of whatever behaviors are being emitted by research subjects participating in these studies. It is worth noting that subjective measures are not worse or less important than objective measures; rather, they give us information of a different sort. It has been argued that the "pragmatic utilization of non-behavioral research designs and methods (e.g., randomized control trials and survey research) alongside traditional behavioral ones (i.e., time-series designs, dimensional measures of behavior) by [contextual behavior science] researchers should be encouraged to the extent that it can bolster our claims and invite broader social acceptance" of research conducted within behavior science (Newsome, Newsome, Fuller, & Meyer, 2019, p. 348).

Conversely, behavior science relies primarily on objective measures of behavior. “Data may be considered behavioral when the numbers reported are direct reflections of the values obtained through measurement of dimensional qualities of behavior” (Newsome et al., 2019, p. 348). There are three primary dimensional qualities of behavior from which behavioral measures derive: repeatability (e.g., count or rate), temporal locus (e.g., latency), and temporal extent (e.g., duration). Derivative measures such as percent correct combine or aggregate two forms of behavioral data (Cooper, Heron, & Heward, 2007).
Derivative measures are typically less preferable to direct behavioral measures. From both a conceptual and practical perspective, some behaviors (or behavior classes) are more challenging to measure than others. The core competencies identified by most iterations of CRM, for instance, present measurement challenges. Across the majority of CRM implementations in aviation, oil & gas, nuclear power, and medicine, the two most commonly listed critical skills are communication and situational awareness. A review of measurement in previous research for these two skills is presented below.

**Communication**

The dynamic nature of HROs requires the behaviors of many individuals to be seamlessly integrated to execute complex tasks, often requiring concurrent and interlocked chains of behaviors. This complexity cannot be maintained consistently without effective communication (i.e., verbal behavior) among team members. Within the parameters of CRM, it is essential to note that communication is not a top-down, one-way street. Communication is most effective when one's rank in the organization is irrelevant to the equation. For instance, a nurse should be allowed and encouraged to speak up if they disagree with a physician's treatment decision. Following the principles outline by the "Four Words" (discussed above), the physician should exert their authority as a team leader by clearly inviting participation (i.e., feedback) – be it assenting or dissenting. Simultaneously, the nurse should assert their position while respecting the decision-making burden and authority of the physician. The
entirety of verbal exchanges between members of a team can be analyzed from a behavior analytic perspective.

**A Behavior Analytic Account of Verbal Behavior.** From a behavioral perspective, Skinner’s analysis of verbal behavior (1957) sets the stage for our understanding of communication in organizational teams. Skinner defines verbal behavior as "behavior reinforced through the mediation of other persons" (1957, p. 2), later refining his definition to include a requirement that the other person (listener) has previously learned a repertoire to reinforce speaker behavior. A defining characteristic of verbal behavior, Skinner argues, is that it is indirectly reinforced. An individual can pour themselves a glass of water, take a drink, and come into contact with the reinforcing properties of the water: cool temperature, alleviation of thirst, etc. Conversely, that same individual can act as a speaker, and ask someone else to please get them a glass of water. Those same reinforcing properties are then only available to the speaker as a result of the other individual's behavior – that is, the listener. In Skinner’s account of verbal behavior, this example is called a *mand* – “a verbal operant in which the response is reinforced by a characteristic consequence and is therefore under the functional control of relevant conditions of deprivation or aversive stimulation” (1957, p. 35-36). This simple example of someone *manding* for a glass of water can be extrapolated to interactions in everyday life, where the roles of speaker and listener often fluctuate rapidly; this is particularly true in organizational settings.
Skinner's analysis of verbal behavior provides a foundation for our understanding of communication in the workplace; on its own, it is insufficient. His analysis focused primarily on the behavior of the speaker. An additional account of language and cognition, known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), introduced a different conceptual class of verbal behavior, from the listener's perspective: relating. RFT posits that we learn not only by directly contacting stimulation in our environment, but also by relating stimulation in the present to stimulation in our learning history.

For example, hearing Christmas music at home while cooking can evoke a range of verbal behaviors. For one individual, this song was previously paired with family, good food, and opening presents. Hearing the song transfers the stimulus functions of family fun, good food, and opening presents to that individual's current environment. Another individual may have an aversive history of experiences around Christmastime. That same song was previously paired with family fights, mounting debt from shopping, and overcooked food. For this second individual, hearing this same Christmas song transfers the stimulus functions of those aversive stimuli from the past into the present. The past stimulus functions, good or bad, in the present environment can then evoke all sorts of different behaviors.

An RFT approach is generally a more useful lens through which organizational verbal behavior (communication) can be analyzed. Communication in organizations often serves to alter the function of workplace stimuli in an attempt to alter employee behavior (Houmanfar, Rodrigues, & Smith,
An RFT approach can be used to generate *augmentals* – rules which alter the reinforcing or punishing effectiveness of consequences for specific behaviors. For instance, the rule "if we have zero reported safety incidents this month, everyone gets a monetary bonus," will likely lead to suppression of reporting any safety incidents. The rule increases the reinforcing value of the monetary bonus – the bonus is only available contingent on certain behavior. Simultaneously, the punishing value of reporting safety incidents (e.g., a "near-miss" or close call, or even minor injuries) is increased by the rule – any report of safety incidents leads to a removal of the bonus. While this example demonstrates the value of an RFT approach to generating rules to evoke behavior we want (e.g., rule-following), more rapid exchanges of information between multiple team members (where neither is merely following a specific rule) warrant in-depth analyses.

**Referential Coding of Verbal Behavior.** Bijou, Chao, and Ghezzi (1988) introduced a referential verbal coding system based on Kantor’s (1977) analysis of referential verbal interactions. This methodology was later used by Johnson, Houmanfar, and Smith (2010) to assess productivity and customer service. This procedure is as follows: audio/video recording of relevant verbal interactions, transcription of audio/video recordings, development of operationally defined coding categories, and scoring of the transcribed segments with respect to the correctly identified coding category. Smith, Houmanfar, and Denny (2012) refined this method, allowing researchers to quantify communicative statements exchanged between pairs for enhanced measurement: frequencies of correct
responding and the duration of verbal interactions. Maraccini, Houmanfar, Kemmelmeier, Piasecki, and Slonim (2018) used this refined method to assess communication accuracy among medical/nursing student pairs during simulated patient handoffs. Simulated handoffs between medical and nursing students were videotaped and subsequently transcribed. These transcripts allowed for analysis of each statement made ("verbal units"), and subsequent coding of comments as "correct," "missing," "erred," or "omitted." Maraccini et al. (2018) defined these categories as follows:

- **Correct**: verbal units included patient information that completely and accurately matched the information provided in the original patient case. In other words, no information was missing or inaccurately stated.

- **Missing**: verbal units contained a portion of accurate patient information stated in the original patient case but did not address all components of information necessary to complete that segment correctly.

- **Erred**: verbal units referred to instances when patient information was provided but contained at least one characteristic that was inaccurate as compared to the original patient case.

- **Omitted**: verbal units referred to cases when the entire segment of patient information was skipped entirely.

This same verbal coding scheme was later utilized in a technical feasibility study by Anbro and colleagues (2020). This study involved creating a 360° video simulation in which a simulated, pregnant patient suffers from a pre-eclamptic seizure. Medical and nursing students viewed this simulation using virtual reality headsets, allowing them to see the simulated environment by moving their head and eyes around as in a typical environment. In this simulation, students were
asked by the simulated nurse to serve as a recorder and provide verbal
checkbacks. **Checkbacks** are a technique used in TeamSTEPPS®. The recipient
of patient information repeats back (or "checks back") the information they have
just received; this allows the initial speaker to confirm whether the patient
information was communicated correctly. **Figure 2** shows the change in
communication accuracy among medical and nursing students before and after a
3-hour TeamSTEPPS® training.

**Table 1** shows the simulation timeline in more detail. The rightmost
column ("Sample Verbal Checkback") lists the explicit criteria for each checkback
opportunity to be counted as correct. Given the variable nature of
communication, the number of required elements per checkback is not
standardized across every opportunity. Simpler statements (e.g., "bed rails up")
have only one criterion to score as "Correct." More complex statements (e.g., "6
grams magnesium sulfate IV over 15 minutes") have more "Correct" criteria – in
this example, there are four. Anbro et al. (2020) replicated findings shown by
Maraccini et al. (2018): after training, more participants made correct statements
during verbal checkbacks, while simultaneously decreasing the number of
omitted checkback opportunities. Erred and missing statements during
checkbacks demonstrated little to no change following TeamSTEPPS® training.
It is worth noting that the frequencies of erred and missing statements were low
before training, therefore there were few opportunities for improvement in
posttest. Continued use of this assessment method is warranted, as it provides
an objective assessment of communication accuracy in team settings.
**Situational Awareness**

Situational awareness remains one of the most challenging components of CRM to train and accurately measure. Historically, situational awareness research has its origins in cognitive psychology. Endsley (1995a; 1995b) conceptualized situational awareness as having three levels: 1) perception, 2) comprehension, and 3) projection. The majority of research that has followed has either been conceptual or has relied on more subjective measures. A behavioral interpretation of situational awareness (Killingsworth, Miller, & Alavosius, 2016) analyzed Endsley's model and re-defined each of the three levels of situational awareness. This interpretation is shown in Table 2. In summary, perception features a combination of stimulus control, conditional discrimination, and observing responses; comprehension is shown through verbal responses that tact the features and functions of stimuli, and their relation to other stimuli and events; projection involves behaving with respect to one's learning history and that history's interaction with current contingencies – in other words, "predicting" responses (Skinner, 1974).

Situational awareness training and measurement are best conducted in high-fidelity simulators, as is common practice in the aviation industry. This environment provides a realistic depiction of the highly technical working environment, without exposing anyone to actual risk. Environmental manipulations can be implemented in a simulator and in real work settings to promote situational awareness. The construction of data readouts should always involve the collaboration of those who will be using the readouts (e.g., pilots,
doctors). The simpler a data display can be, the more accessible it becomes (Jenkins & Gallimore, 2008). The use of objective measures, such as physiological readouts, can lead to increased situational awareness (Vidulich, Stratton, Crabtree, & Wilson, 1994). Some research has also shown that there are benefits to using subjective questionnaires as measures of perceived situational awareness. This applies to questionnaires completed by both the performer and another evaluator (e.g., their supervisor, Waag & Houck, 1994).

**Measuring Situational Awareness.** Several tools have emerged which intend to measure situational awareness, with varying levels of objectivity. The Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1987) is a technique where a simulation is frozen, data screens blackout, and the performer is asked a series of objective questions. These questions pertain to the status of relevant simulated measures (e.g., current airspeed), what these measures mean, and what is likely to happen if the simulation continues. Recent studies (Ikuma, Harvey, Taylor, & Handal, 2014; Kass, Cole, & Stanny, 2007; Morgan et al., 2015; Salmon et al., 2009) have verified a positive correlation between performance in the SAGAT and performance in complex tasks in dynamic environments.

The Situation Present Assessment Method (SPAM; Durso & Dattel, 2004) is similar to the SAGAT in terms of having a performer interact with a high-fidelity simulator and respond to queries about the simulation. The key difference with the SPAM is that a measure of *workload* is attempted. A warning signal denotes when a query is ready, and the performer gets to press a button when they are
prepared to respond. The latency of button pressing is said to measure the performer's workload. High accuracy and rapid speed of responding to these queries are supposed to indicate situational awareness. Initial research has validated SPAM as a useful tool to measure situational awareness (Durso et al., 1998; Loft & Morrell, 2013).

The objective measurement of perception in simulators, as a component of situational awareness, has only been possible when an overt response occurs following the presentation of some stimulus (Schlinger, 1995). However, advances in eye-tracking technology may help solve this problem. For instance, Tobii eye-tracking sensors provide a new method for objectively tracking where a person looks and for how long. Used with their software, an analysis across time can be conducted, revealing exactly where a person is looking, at what point in time, and for how long. These data can be quantitatively analyzed, and patterns can be identified for measures such as latency (i.e., orienting to some change within a simulation).

**Behavior Analytic Research into Situational Awareness.** Killingsworth, Miller, and Alavosius (2016) call for a research line that “could focus on an assessment methodology for determining the minimal simulation components that allow for maximum transfer to on-the-job performance and novel problem solving” (p. 316). Following this recommendation, Anbro et al. (2020) conducted a technical feasibility study using virtual reality simulations with medical and nursing students. Of their 105 participants, a sub-set of 22 participants used a virtual reality headset with embedded eye-tracking sensors. Figure 3 shows a
scatterplot of those 22 participants who used the HTC Vive headset with embedded eye-tracking. The x-axis shows the change in eye fixation duration on the speaker as they speak, from pretest to posttest. Data points to the right side of the scatterplot indicate those participants who fixated on the speaker as they spoke for a longer duration in posttest than pretest. Data points to the left side indicate a decrease in eye fixation duration on the speaker as they spoke (i.e., participants looking at the environment more). The y-axis shows the change in each participants' number of correct verbal responses from pretest to posttest. Data points in the top half of the scatterplot indicate those participants who emitted correct responses more frequently in posttest than in pretest. Data points in the bottom half indicate those participants who emitted fewer correct responses in posttest than in pretest.

These results show that 17 participants increased their frequency of correct verbal responses in posttest (5 looked at the speaker longer, 11 looked at the environment longer; one did not change). Four participants did not increase their frequency of correct verbal responses in posttest (2 looked at the speaker longer; two looked at the environment longer). One participant decreased their frequency of correct verbal responses in posttest (while looking at the speaker longer). These patterns provide an initial insight into the measurement and assessment of situational awareness (Level 1: perception) in virtual simulations: there is no one "right" way to be situationally aware. Some individuals raised their awareness of events (as measured by communication accuracy) by looking at a speaker longer than the environment. Others raised their awareness by looking
at the environment to which the speaker is referring, longer than at the speakers themselves. Individual histories likely bear substantial influence on which method is most useful for which individuals.

Researchers in the field of behavior science agree that eye gazing is an operant, susceptible to the processes of conditioning (Madelain, Paeye, & Darcheville, 2011; Schroeder & Holland, 1969). As VR and eye-tracking technology advance, algorithms can develop awareness profiles that allow individuals to correlate their observing responses with response accuracy in simulated scenarios. In this manner, individuals would familiarize themselves with the pattern of observing responses that work best for them and, therefore, attend to clinical environments using such a pattern. From a didactic training perspective, it has been suggested that Acceptance and Commitment Training (ACT) may have a beneficial impact on situational awareness. "Practicing the skill of being in the here and now has been shown to have remarkable effects on health and well-being. The more often you are focused on what you are doing, the more likely your actions will lead to optimal performance" (Moran, 2013, p. 110).

**Using Simulations**

The science of human behavior relies on empirical research to further our understanding, prediction, and influence of behavioral phenomena. The use of analog environments for research often results in more experimental control than natural environments. The role of simulations is, therefore, increasingly relevant. Ward and Houmanfar (2011) analyzed studies published between the years
1987-2010 in *The Journal of Applied Behavior Analysis (JABA)*, *The Journal of Experimental Behavior Analysis (JEAB)*, *The Journal of Organizational Behavior Management (JOBM)*, and *The Psychological Record (PR)*. Empirical studies in these journals were sorted into the following categories, with the identified criteria. Note that some categories include two requirements, while some only include one.

1. **Analog experiment**: Includes simulations and non-simulations; it is one in which the researcher explicitly attempts to extrapolate their findings as an explanation for events found in naturalistic settings.

2. **Non-simulation-based analog experiment**:
   - a. An analog study, meaning the process found as a result of the study is explicitly extrapolated as an explanation for events found in naturalistic settings; and
   - b. This type of study is not a simulation, meaning the physical features of the experimental apparatus were not designed to recreate any features of a naturalistic setting.

3. **Simulation-based analog experiment**:
   - a. An analog study in that the process found as a result of the study is explicitly extrapolated as an explanation for events found in naturalistic settings; and
   - b. The process that is extrapolated from a simulation-based analog was produced with an apparatus whose physical features were designed to recreate particular features of the naturalistic setting to which the process is extrapolated.

4. **Non-analog experiment**: Any experiment in which the researcher does not explicitly extrapolate their findings as an explanation of events in naturalistic settings is not an analog, even if the apparatus used was designed to recreate features of a naturalistic setting.

This review identified 46 simulation-based studies (avg. 2 per year), which were conducted and published within these journals. These simulation-based studies utilized various levels of simulation, including laboratory space constructed to
resemble a natural environment, computer tasks simulating real-world job tasks, etc.

Anbro et al. (2020) extended this initial review to include studies published between 2011-2018 in these same behavioral journals, with the addition of studies published in *Behavior and Social Issues (BSI)* and *The Journal of Contextual Behavior Science (JCBS)*. Our extended review (30 years total) indicates a rapid increase of simulation-based studies within behavior science: an additional 49 simulation-based studies were published in the journals named above between 2011-2018 (avg. 7 per year). The results of our extended review are shown in Figure 4. Despite the steadily increasing rate of simulation-based studies within behavior science, none of these studies have utilized virtual reality (VR) simulation technology.

**Development of Virtual Simulation Technologies**

The history of VR technology can be traced back for many decades. Figure 5 (Poetker, 2019) shows the development of this technology from early conceptualization in sci-fi writings to modern-day immersive hardware and software. A summarized history of VR’s development is presented below. It is worth noting the summation of factors leading to the creation and innovation of VR technology. These factors range from innovation by leading researchers, to the invigoration of public interest through film and gaming applications.

During the late 1950s, cinematographer Morton Heilig created a *Sensorama* – a large, cabinet-like device that included sources of stimulation for sight, sound, touch, and smell. The 1960s saw the advent of the first head-
mounted display with two video screens (one for each eye), the first flight simulator, and a refined head-mounted display that displayed computer-generated graphics as opposed to video. The movie *TRON* introduced the concept of VR to the public at large in the early 1980s. The actual term "virtual reality" was introduced in 1987 by researcher and computer scientist John Lanier.

Throughout the 1990s and early 2000s, VR technology was refined and brought to the market primarily by video game companies (e.g., Sega, Nintendo). *The Matrix* premiered in 1999, reinvigorating public interest in the potential for fully immersive virtual experiences. In the past decade, VR hardware and software has exponentially improved. VR headsets such as Valve Index, HTC Vive, Oculus Quest, and PlayStation VR are common on the market and are often sold for private, at-home use by gamers. VR innovation is also closely tied to innovation in augmented reality (AR). AR combines real-world perception with computer-generated graphics. AR is used in televised football games (e.g., the "yellow yard line" overlaid on video feed), commercial hardware (e.g., Google Glass), and even mobile gaming applications (e.g., Pokémon Go and Harry Potter: Wizards Unite). While deeply rooted in modern gaming applications, both AR and VR have the potential to enhance training and assessment within HROs.

VR simulation technology provides an immersive audio/visual experience within a simulated environment. As described above, this technology has steadily gained popularity for gaming applications in which the user neither creates nor modifies the simulated environment. In potential research applications,
simulations are entirely under the control of the programmer/experimenter. Promising advances in recent VR simulation technology include the addition of eye-tracking sensors within VR headsets. These sensors track users' pupil movements as they look throughout a simulated environment, allowing data collection that heretofore has been impossible. While previous studies have examined observational responses, these are often at the level of simple orienting responses (e.g., the subject orients towards some stimulus); the precision of measurement offered by eye-tracking is lacking in such research (e.g., Buzsáki, 1982). By utilizing VR simulation technology, behavior scientists can capture objective measures conceptually related to previously discussed behavioral phenomena – some of which lack empirical measurement to date. Skills identified by CRM (e.g., situational awareness) are prime examples.

**Simulations in Medical Education**

Medical education often relies on in-person simulations (referred to as Direct simulations from here on) of events to assess performance targets among medical and nursing students (e.g., Sectish et al., 2010). A challenge to conducting Direct simulations with entire cohorts of students lies in the resources required – specifically, the use of simulated patient rooms and personnel to act in the role of patients or other medical staff. Virtual reality (VR) technology offers a potential solution to this challenge. Once completed, VR simulations do not require as many ongoing resources to use, other than a predetermined amount of physical space that the user occupies while interacting with the simulation. While VR applications predominantly lie in gaming, they have potential
applications outside of this domain. The potential advantages of VR simulations vs. direct simulations lie in the robust degree of experimental control that VR offers. The experimenter controls every visual and auditory element within a virtual simulation. This control ensures the internal validity of the actual simulation(s) used across participants within any research preparation. The issue that is raised, however, is whether or not VR simulations are a valid substitute for direct simulations.

The future of high-precision measurement and assessment of behavior in simulations is conceivably tied to VR simulation technology. The nature of VR simulations, with complete control of programmed content, may allow for a high degree of experimental control heretofore unachievable in other simulated scenarios (e.g., computer tasks in a research lab), greater measurement precision of difficult-to-measure responses (e.g., eye fixations), greater variety and customization potential for scenarios used, and the emerging ability of simulations to support multiple users at once (e.g., for team-based simulation assessment). While common sense may indicate the face validity of using VR rather than direct simulations, empirical criterion validation has yet to occur. Criterion validity “refers to comparing the procedure under examination to an existing instrument that is already accepted as an adequate indicator of the characteristics of interest” (Johnston & Pennypacker, 2009, p. 143). The procedure under examination in the current study is the use of VR simulations for performance assessment, while the currently accepted instrument is the use of Direct simulations to assess performance – standard practice in medicine.
Purpose

Anbro et al. (2020) provided a preliminary basis for developing a procedure to assess communication and situational awareness during simulated critical events in medical and nursing education. While the simulation used in this study was a 360° pre-recorded video, the current study will further utilize VR simulation technology by creating and programming fully interactive, multi-user virtual environments (e.g., a virtual patient room as opposed to a filmed one). A fully immersive VR environment will allow participants to pick up and manipulate objects, communicate directly with a partner in the same simulation (e.g., by speaking in real-time with another participant's avatar), and interact directly with a simulated virtual patient (e.g., administering medications). A VR environment also allows for easier addition of virtual stimuli, which eye-tracking sensors can measure fixations on, than when using a 360° video simulation.

Considering the potential benefits of VR simulations, the purpose of the current study is twofold: 1) to assess the criterion validity of VR simulations as an assessment tool to measure performance change in healthcare education through performance comparisons to established, direct simulation methodology; 2) to assess changes in healthcare students’ communication accuracy and situational awareness across both simulation modalities.

Method

Participants

This study was conducted as a curriculum-based, inter-professional education (IPE) intervention designed for 45 participants: 23 first-year physician
assistant (PA) students attending the University of Nevada, Reno School of Medicine (UNR Med), and 22 first-year nursing students attending the Orvis School of Nursing (OSON). Thirty-six participants were female, and 9 male.

**Setting**

Four standardized patient rooms (SP rooms) and six medium-sized classrooms at the UNR Med campus were used in the current study. SP rooms were set up and equipped as fully functioning clinician offices, used for conducting direct simulations as part of medical students' practical training. Each SP room was equipped with a patient exam table, medical supply cabinets, a computer, and two video cameras mounted at opposing angles near the ceiling. A control room was located at the end of the SP rooms' corridor, which included multiple computer monitors with video feeds from each SP room. From the control room, simulations in SP rooms were monitored for participant safety and recorded for post-hoc data analysis. Two large-sized lecture halls were used to run the small group experiential exercises for each training.

An additional medium-sized classroom was briefly used before the onset of simulation sessions. This room was used to gather and brief participants regarding the content and process of the simulations they would complete. This room was then used for pre-simulation case reviews. Students were allotted 10 minutes in this room to familiarize themselves with the fictional patient case they were responsible for handing off during their simulations. They were also allowed to take any notes they wanted to bring into the simulation with them. Direct simulation participants took notes on a physical notecard, which they could use
during their simulation. VR simulation participants took notes in a text-entry program on laptops provided for their use. VR participants saved these notes onto USB flash drives, which were later used to load their notes onto a blank virtual notecard for use in their VR simulation.

Design

A pre/post comparative group design was used to evaluate changes in communication accuracy during simulated patient handoffs, as a function of two different training workshops. Physician Assistant (PA) students and nursing students were randomly assigned as partners. Each set of PA student/nursing student partners (henceforth referred to as "dyads") was randomly assigned to experimental or control conditions. As this study was conducted during the COVID-19 pandemic, all students completed online university safety trainings prior to any participation in this study.

An interprofessional TeamSTEPPS® training was given to dyads in the experimental condition, and Adverse Childhood Experiences (ACE) training was given to dyads in the control condition. Dyads in each condition were also randomly assigned to participate in either direct or VR simulations before and after their respective training. Dyads did not change simulation modalities across simulation sessions. Dyads assigned to direct simulations completed both pretest and posttest patient handoff simulations in the same physical SP room. Dyads assigned to VR simulations completed both pretest and posttest patient handoff simulations in separate physical SP rooms, wearing VR headsets and occupying the same virtual SP room. Virtual simulations are described in greater procedural
detail below. **Table 3** shows the participants’ distribution into four groups: Group
1) VR-TeamSTEPPS®; Group 2) VR-ACE; Group 3) Direct-TeamSTEPPS®;
Group 4) Direct-ACE.

**Independent Variables: Training Workshops and Simulation Materials**

A three-hour TeamSTEPPS® training workshop, previously used by Anbro et al. (2020), was delivered by the Chief Medical Officer and Chief Nursing Officer of a large, local hospital system. This workshop was created using existing materials and research from the TeamSTEPPS® curriculum, publicly available online (TeamSTEPPS®, n.d.). The content expanded on a 90-minute training workshop used by Maraccini et al. (2018), which itself was created using a previous medical student curriculum (O'Toole et al., 2014) available from the MedEdPORTAL online library. This training was used for dyads in the experimental condition. The training was delivered in a hybrid format; this was due to governmental restrictions on large gatherings, imposed in response to COVID-19. A two-hour lecture was delivered online via Zoom, followed by an additional hour of in-person experiential exercises detailed below.

A three-hour Adverse Childhood Experiences (ACE) training workshop was delivered by the Director of Interprofessional Education and Collaboration for UNR Med. The workshop was previously created and used by Pletcher, O’Connor, Swift-Taylor, and DallaPiazza (2019). Dyads in the control condition attended this training workshop, as there was no identified overlap in content with that of the TeamSTEPPS® workshop. This training was also delivered in a hybrid
format. A two-hour lecture was delivered online via Zoom, followed by an additional hour of in-person experiential exercises detailed below.

Simulation materials varied, given the simulation modality. Two SP rooms used for direct simulations were equipped with a mannequin on the patient bed, an IV stand, a computer monitor, and a cart with various medical instruments and equipment. Two SP rooms used for VR simulation each contained one desktop computer hooked up to a VR headset. The VR simulations themselves were virtual recreations (to scale) of the SP room.

To virtually recreate the SP rooms, VR specialists used a 3D scanning process called photogrammetry. Photogrammetry is a method of extracting 3D information from objects or scenes from photographs. The process involves taking overlapping photographs that render digitally to recreate the digital space. To create this particular scene, VR specialists began by taking 174 pictures of an SP room. The images were taken at varying elevations (eye-level, "bird's eye" view from a ladder, and floor-level view) at multiple locations throughout the room. These photographs were then digitally stitched together using specialized photogrammetry software called Agisoft Metashape. The software creates a 3D mesh and applies the photographic data as a texture. Once the textured model itself was completed and rendered in 3D, computer science experts populated the scene with various stimuli described in detail below. They also prepared the simulation for use by two simultaneous users. Groups 1 and 2 completed patient handoffs in this simulation using HTC Vive Pro Eye VR headsets, which come complete with embedded eye trackers.
TeamSTEPPS® Workshop

A three-hour TeamSTEPPS® workshop, previously used by Anbro et al. (2020), was used as the experimental condition training workshop. The workshop was conducted in a hybrid manner: a two-hour interactive lecture was followed by one hour of relevant experiential exercises. The lecture included historical discussions of the patient safety movement, an introduction to the TeamSTEPPS® model, and instruction on using the I-PASS tool to structure patient handoffs.

The final hour of the workshop required students to participate in experiential exercises in small groups. The exercises provided students an opportunity to practice effective communication in teams by restricting communication and generally easing the restrictions. Two activities were selected for use: House of Cards and Code: Mr. Potato Head. Both activities required students to complete tasks in groups – build the largest house of cards possible and complete a specific Mr. Potato Head configuration. Each activity was completed in three rounds. During round 1, no verbal communication was permitted. During rounds 2 and 3, all team members were allowed to talk. During round 2, only a designated team leader was allowed to see instructions in the Code: Mr. Potato Head activity; 1 minute of briefing time was allotted to students before round 2 in both activities. During round 3, all team members were permitted to see the instructions in the Code: Mr. Potato Head activity. Briefing time was again allotted to teams before this round. Additionally, a large-group
Debrief was conducted between round 2 and 3 of the House of Cards activity to share successful strategies across teams.

**Adverse Childhood Experiences (ACE) Workshop**

A three-hour ACE workshop, previously developed and used by Pletcher et al. (2019), was used as the control condition training workshop. The workshop was conducted in a hybrid manner: a two-hour interactive lecture was followed by one hour of relevant experiential exercises. The lecture included a discussion regarding the consequences of adverse childhood experiences on patients throughout their lives, review of trauma-informed care and how it can benefit patients, and instruction on using the ACE survey tool.

The final hour of the workshop required students to participate in experiential exercises in small groups. Interactive historical case discussions and practice using the ACE survey tool were selected as experiential exercises for this workshop. Note that the exercises chosen did not focus on team communication, nor was any feedback or performance appraisal given to students in this workshop regarding their skills communicating among their activity groups.

**Primary Dependent Variables**

Simulated patient handoffs were used as the primary events for behavioral assessment in the current study. All participants' verbal behavior across all simulation sessions was audio/video recorded. Participants in Direct simulation groups had their verbal behavior recorded using pre-installed cameras in SP rooms. Participants in VR simulation groups had their verbal behavior recorded...
using microphones embedded in the VR headsets; these audio recordings were merged with a "teacher view" video of the virtual simulation room. Transcriptions for all simulation sessions were generated manually. Research assistants viewed each simulation's video recording and produced a transcript of the simulation as they watched it. The total duration of each simulation session was also recorded.

Handoff transcriptions were used to analyze the verbal behavior of both the handoff provider and the handoff receiver. The handoff provider's verbal statements regarding their patient was compared to the written case example and coded for accuracy during the first four stages of the handoff identified by the I-PASS mnemonic tool. The handoff receiver's verbal statements in the final "synthesis by receiver" step were also coded for accuracy, as a measure of situational awareness level 2 (comprehension). Communication accuracy was scored using the same four categories as previous research: Correct, Missing, Erred, or Omitted (Anbro, et al., 2020; Maraccini et al., 2018). The specific scoring criteria for each fictional patient case are shown in Appendix A. These criteria were created using the I-PASS handoff tool, shown in Figure 6. Additionally, the post-simulation assessments asked participants multiple questions about the patient they received care of during their simulated handoff. These questions included recall and predictive questions.

First, recall questions asked specifics regarding their patient: name, age, presenting problem/concern, action list, etc. These questions constituted a secondary measure of situational awareness level 2 (comprehension). The questions were presented to participants after the handoff event in question; from
a behavioral perspective, their responses constitute measures of *understanding* or comprehension. Next, predictive questions asked participants to respond regarding this patient’s next steps of care. Participants were asked to provide action steps, given particular patient case developments. These responses were evaluated by professionals from the PA and nursing programs for correctness.

**Inter-Observer Agreement (IOA)**

Trial-by-trial IOA was calculated for 100% of simulation sessions in both pretest and posttest. Following the transcription for each simulation session, one primary observer and one of six secondary observers independently coded transcribed statements. Statements were coded into the coding categories described above: Correct, Missing, Erred, or Omitted. IOA was calculated for verbal statements made by both participants in both handoff roles: handoff provider and handoff receiver.

**Secondary Dependent Measures**

Secondary measures were collected via self-report Likert scales, which were administered using Qualtrics. These assessments are described in more detail below. Secondary dependent measures were collected only for groups 1 & 2 – the VR groups. These measures were collected using eye-tracking sensors embedded in the VR headsets, and are described in more detail below.

**TeamSTEPPS® Teamwork Attitudes Questionnaire (T-TAQ)**

The T-TAQ is a 30-item assessment tool that asks individuals to rate each item on a 5-point Likert scale, ranging from "strongly agree" to "strongly disagree." The T-TAQ is used increasingly often with TeamSTEPPS® research,
to assess participants’ attitudes towards the core components of TeamSTEPPS®. These components are used as categories that the 30 items are sorted into (6 items each): team structure, leadership, situation monitoring, mutual support, and communication (Baker, Amodeo, Krokos, Slonim, & Herrera, 2010). The T-TAQ questions are shown in Appendix B.

**Eye-Tracking Measures**

Eye trackers were embedded into the virtual reality headsets used by Groups 1 & 2. Using the dark pupil tracking technique described below, these eye trackers automatically track participant eye gazes relative to certain stimuli in the virtual environment, called *areas of interest* (AOIs). The stimuli, described below, include the simulated patient, partner’s avatar, notecard (with patient handoff notes), IV stand, cart (with various medical instruments and supplies), and computer monitor. *Figure 7* (top image) shows these AOIs in an overhead screenshot of the virtual simulation used. *Figure 7* (bottom image) also shows a first-person POV of the simulated standard patient room. Two types of measures were taken by embedded eye trackers: fixation count and fixation duration. Fixation count is a simple frequency count of the total number of eye fixations made by each participant for each AOI. Fixation duration measures were derived from the start time, duration, and stop time of each eye fixation. Fixation duration was a previously used metric to assess situational awareness (level 1) in virtual simulations (Anbro et al., 2020). Together these measures allow for the generation of time series graphs that show participant eye fixations throughout their simulated patient handoffs, described in more detail below.
Procedure

Before the onset of this study, students were randomly assigned a dyad partner, and dyads were randomly sorted into one of four groups as described above. Students also completed a brief online questionnaire before the onset of this study; the questionnaire is shown in Appendix C. The questionnaire asked students to respond on a Likert scale indicating their previous use of VR and their susceptibility to motion sickness. Students were also asked if they consented to be randomly assigned to participate in a VR simulation. Students who did not consent or who indicated susceptibility to motion sickness were excluded from Groups 1 and 2 (VR groups). Groups were reorganized as needed to ensure all students participating in VR simulations a) gave consent, and b) did not suffer from any history of motion sickness. The online questionnaire separately asked students for general consent to participate in this study and to have their data anonymously aggregated. A separate photo consent was included; two multimedia production specialists recorded sample video clips of participants in both simulation types for future dissemination.

This study was comprised of three phases: Phase I) Pre-intervention assessment (“Pre-test”); Phase II) Intervention; Phase III) Post-intervention assessment (“Post-test”). An optional Phase IV was provided for all participants, following the completion of Phases I-III. Phases I-III were completed during the Summer 2020 semester of PA and nursing students’ education. Figure 8 shows the outlined procedures for the current study, and each Phase is described in detail below.
Phase I: Pretest

A process map for the patient handoff procedure used during pretest and posttest is shown in Figure 9. Student dyads were first brought into a small meeting room adjacent to the SP rooms used in the current study. A member of the research team conducted an initial 10-minute briefing for students. The purpose of the pre-simulation briefing was to provide instruction regarding the simulation content and process. Dyads were then split up. PA students and nursing students were brought to separate areas, and each received a fictional patient case summary. Students in the VR simulation groups (1 & 2) took notes on a computer; these notes were used to populate virtual notecards in the VR simulations. Virtual notecards were held at all times by participants' avatars, and participants could view their notes by raising their left-hand controller while in the simulation. Students in the direct simulation groups (3 & 4) were given notecards and pens to take notes on; these participants held onto their notecards for use during their simulated patient handoffs.

All PA students received the same fictional patient case (Case 1), while all nursing students received another fictional patient case (Case 2). Cases 1 & 2 were previously created for, and used by, Maraccini et al. (2018) and are shown in Appendix D. Members of the research team were stationed in each room to supervise, while students were given 10 minutes to review their fictional patient case. Each fictional patient case was written in paragraph format and included general patient information: illness severity, medical history, and action items for the medical team. Students were instructed not to assume any information that
wasn't explicitly stated in the case summary, nor to make any additional diagnoses. After the 10-minute review period, students were led to randomly assigned SP rooms. Beginning at this stage in the process, procedures for VR and direct simulations differ. Each will be described in detail below.

**Phase I: Pretest VR Simulations (Groups 1 & 2).** Dyads assigned to complete VR simulations did not have any direct, in-person interaction from this point onward in Phase I. Students in these dyads were led to separate SP rooms across a hallway from one another. Upon students' arrival in these SP rooms, a research assistant helped the students calibrate their VR headset to optimize the audio and video calibration. A brief calibration was also completed to ensure the embedded eye trackers functioned properly. The calibration process uses the binocular dark pupil tracking technique (Tobii Pro, n.d.) to detect absolute pupil size, pupil position, gaze direction, gaze origin, and time for each eye fixation. The process is an automated 5-point calibration procedure conducted to ensure accuracy before each participant's use of the eye-tracking hardware, which takes approximately 1–2 min on average. Once their equipment was fitted correctly and calibrated, participants were virtually placed in an avatar customization room.

Avatars in the current simulation were made of standardized models for all participants to enhance experimental control. Each avatar was dressed in either a set of scrubs complete with white lab coat (PA students) or in scrubs alone (nursing students). VR specialists helped the participant choose the correct base avatar model (PA/nurse), and participants customized the gender, hair color, and skin color of their avatars to resemble themselves. Once their avatars were
customized, participants used their VR remotes to click a "Begin" button displayed in the avatar customization room; this virtually teleported the user into the virtually simulated SP room.

The virtual SP room was an exact reproduction of the physical SP room each student was physically located in, as described above. Students could look in all directions, as in a physical SP room. Students in VR simulations also saw virtual representations of one another - their avatars. The virtual SP room was populated with several stimulus objects: a patient exam table with a simulated patient lying on top, the partner's avatar, a virtual notecard in the participant's avatar hand, an IV stand, and a table with various medical instruments and drugs. As described above, each of these stimulus objects was an area of interest (AOI). Designating AOIs triggered the eye-tracking sensors embedded in the headset to automatically capture and record participants' eye fixations (start time, duration, and stop time) on each AOI. Metrics automatically captured included fixation count and fixation duration, used to create time series graphs. Each VR headset was also equipped with a microphone that served two functions: 1) they allowed users to talk with one another in real-time, as in real life; 2) they allowed recording of all verbal behavior from each users, which was overlaid on a video recording of the simulation from within the virtual room.

Once both students entered the VR SP room, their simulation session began. Virtual notecards with participants’ case notes were available at all times during the simulation. The participant need only raise the VR controller in their left hand, and their avatar would raise its left hand, which held the virtual
notecard. Each student then handed off their fictional patient case to their partner, using the virtual mannequin as a substitute for an actual patient. Of note, these handoffs were conducted bedside rather than isolating the patient care team away from the patient for a handoff. Following the completion of both simulated patient handoffs, the simulation ended and prompted both participants to remove their VR headset. VR specialists sterilized all VR hardware prior to the next participants’ simulations, and replaced the disposable face lining of the VR headsets. Participants were then prompted to complete a post-simulation questionnaire. The questionnaire, described in detail above, included 1) a social validity assessment of the simulation modality used; 2) an assessment of students’ retention regarding the fictional patient case they were briefed on during the simulation; 3) an assessment of students’ predictions for that patient.

Phase I: Pretest Direct Simulations (Groups 3 & 4). Dyads assigned to complete direct simulations were led to opposing doors on either side of the same SP room. Participants were instructed to enter the room and handoff their fictional patient case to one another in a predetermined order. PA students handed off their patient to nursing students first, and then nursing students handed off their patient case to PA students. These simulated patient handoffs also occurred bedside. Cameras built into the SP rooms were used to audio and video record all direct handoff simulations. Following the completion of both simulated patient handoffs, participants handed their notecards to a member of the research team before completing a post-simulation questionnaire. The questionnaire, described in detail above, included questions regarding students’
retention of the fictional patient case they received care of during their simulation (Appendix E); an assessment of students' predictions for that patient case (Appendix E); and the T-TAQ survey tool described above (Appendix B). Students were also asked to provide feedback regarding their simulation modality (Appendix F).

**Phase II: Intervention**

The intervention phase was completed the week after students completed the pretest phase. As described above, two concurrent training workshops were given: a TeamSTEPPS® training workshop (experimental training) and an Adverse Childhood Experiences (ACE) workshop (control training). As described above, student dyads were randomly assigned to receive either the experimental or control training. Both workshops lasted three hours and were delivered in a hybrid format. The first two hours of both training workshops were interactive lectures delivered live over Zoom. During the final hour, students met in small groups led by a member of the research team to complete the respective experiential exercises for their assigned training. Experiential activities were completed on the same day as training. Summaries of the lecture content and experiential activities for both training workshops are described above.

**Phase III: Post-Test**

Two days after the Phase II training workshops were conducted, students returned for the second round of simulated patient handoffs. As previously indicated, students did not change simulation modalities from Phase I. Students who completed VR simulations in Phase I also completed VR simulations in
Phase III. Students who completed direct simulations in Phase I also completed direct simulations in Phase III. Procedures for Phase III (posttest) were identical to those in Phase I (pretest), although new fictional patient cases were used. Case 3 was handed off by PA students and Case 4 by nursing students. Like Cases 1 & 2, Cases 3 & 4 were previously used by Maraccini et al. (2018) and are shown in Appendix D. The order of patient handoffs was also reversed from Phase I. While PA students began the handoffs in Phase I, nursing students began the handoffs in Phase III.

**Data Exclusion Criteria**

Due to uneven participant numbers (23 PA students, 22 nursing students), two different instructors served as a partner for one of the PA participants – one during pretest, the other during posttest. The instructors’ data are not included in any analyses presented or discussed below. Furthermore, two participants’ performance data are excluded from analysis due to technical errors. One of these participants misunderstood the initial instructions and did not hand off their patient during their pretest simulation. This participant, therefore, has no pretest handoff data to report. The second participant did hand off their patient in posttest, but the video recording system cut off the beginning of the simulation recording. Rather than include an incomplete handoff’s data, these data were also excluded. Both of these participants’ responses were included in social validity summary data. It is worth noting that both of these technical errors occurred in direct simulation groups, not VR groups. These exclusions are shown as part of Table 3.
Results

The following sections provide an overview of participants' performance data during their simulated patient handoffs. First, summaries of changes in communication accuracy for both the handoff provider and handoff recipient are provided. Next, a summary of participants' situational awareness across all three previously described levels is presented. Finally, social validity data and a cost analysis conclude the comparisons between VR and direct simulations.

Communication Accuracy: Handoff Provider

Figure 10 shows the average number of comments made by handoff providers in the VR and Direct simulation groups. The top graph shows pretest results, and the bottom graph shows posttest results. Both groups demonstrated comparable performance in pretest. VR simulation participants emitted an average of 12.8 “correct,” 1.2 “missing,” and 0.3 “erred” comments, while omitting an average of 2.7 comments. Direct simulation participants emitted an average of 13.4 “correct,” 1.1 “missing,” and 0.3 “erred” comments, while omitting an average of 2.2 comments. The groups’ overall performance did not change much from pretest to posttest; change was comparable across groups.

In posttest, VR simulation participants emitted an average of 12.8 “correct,” 1.3 “missing,” and 0.2 “erred” comments; on average, 2.3 comments were scored as “omitted.” Comparatively, Direct simulation participants emitted an average of 13 "correct," 1.1 “missing,” and 0.3 “erred” comments; on average, 2.1 comments were scored as “omitted.” The variability of responding is shown in Figure 10 as error bars, denoting the ranges of participant responses. These
ranges are comparable across participants in both simulation types. One statistically significant difference was found in posttest: VR simulation participants made fewer errors (M = 0.22, SD = 0.18) than Direct simulation participants (M = 0.48, SD = 0.36); t = -1.66408, p < 0.10.

Figure 11 shows the average simulation duration in minutes across groups. The difference in pretest simulation duration between VR simulations (M = 1.77, SD = 0.62) and Direct simulations (M = 2.05, SD = 0.73) was statistically significant; t = -1.36643, p < 0.10. The difference in posttest simulation duration between VR simulations (M = 1.60, SD = 0.50) and Direct simulations (M = 2.18, SD = 0.60) was even more statistically significant; t = -3.47783, p < 0.001. Note that while VR simulations decreased in average duration from pretest to posttest, Direct simulations increased in average duration. Due to these significantly different trends, the comparable responding of handoff providers shown in Figure 10 was further analyzed using rate of response as a metric. Rate of response was chosen as it is “the metric of choice for the experimental analysis of behavior, and was critical to the discovery of much of what is known about the principles of behavior (Skinner, 1966)” (Newsome, Newsome, Fuller, & Meyer, 2019, p. 348).

Figure 12 shows the response rate per minute of handoff providers in VR and Direct simulations, in both pretest and posttest. “Correct” responses are shown in the top graph, “missing” responses in the middle graph, and “erred” responses in the bottom graph. “Omitted responses” are not shown, as this coding designation refers to a lack of verbal responses - which cannot be
converted to rate. During both pretest and posttest, VR simulation participants emitted "correct" responses at a higher average and median rate than Direct simulation responses, albeit with greater variability. The posttest difference between VR simulations (M = 8.71, SD = 2.63) and Direct simulations (M = 6.35, SD = 1.72) was statistically significant; t = 3.3223, p < 0.001. Direct simulation participants emitted "missing" responses at a lower average and median rate than VR participants during pretest and posttest. The posttest difference between VR simulations (M = 0.72, SD = 0.64) and Direct simulations (M = 0.44, SD = 0.37) was statistically significant; t = 1.70342, p < 0.10. VR participants demonstrated a reduction in the variability of "missed" rates from pretest to posttest, while Direct participants demonstrated increased variability of "missed" rates from pretest to posttest. Comparable "erred" response rates are seen across groups in pretest, but VR participants alone greatly reduced their variability of "erred" responses during posttest.

**Figure 13** shows the pretest to posttest change in participant response rate. Changes in response rate for "correct," "missing," and "erred" comments are categorized as either *increased* response rate or *decreased* response rate. An increased rate for "correct" responses is indicative of a shift towards greater communication accuracy. 65% of VR simulation participants increased their rate of "correct" responses from pretest to posttest, while only 35% of Direct simulation participants increased their "correct" response rate. A decreased rate for "missing" and "erred" responses also indicates a shift towards greater communication accuracy. These results are mixed. 35% of VR participants
decreased their rate of "missing" comments, while 55% of Direct participants decreased their rate of "missing" comments. 83% of VR participants decreased their "erred" response rate, while 75% of Direct participants decreased their "erred" response rate.

Figure 14 shows data similar to Figure 12. However, these data are compared across training groups rather than simulation modality. Participants assigned to the ACE training ($M = 8.36$, $SD = 2.46$) demonstrated a higher rate of "correct" responses before training, compared to TeamSTEPPS® participants ($M = 6.96$, $SD = 2.73$); $t = -1.56928$, $p < 0.10$. This difference is nullified in the posttest; the average response rates of both groups are comparable after training. The TeamSTEPPS® group also shows less variability than the ACE group in their rate of "correct" responses during posttest. Rates of "missing" responses show a change in trend from pretest to posttest. In pretest both groups had comparable average "missing" response rates, while the ACE group showed less variability and a lower median. In posttest, both groups had comparable average "missing" response rates, while the TeamSTEPPS® group showed less variability and a lower median. ACE participants’ rate of "erred" responses ($M = 0.11$, $SD = 0.24$) was, on average, lower and less variable than TeamSTEPPS® participants ($M = 0.23$, $SD = 0.29$) during pretest; $t = 1.38761$, $p < 0.10$. In posttest, the ACE group slightly increased their average "erred" response rate, while the TeamSTEPPS® group slightly decreased their average "erred" response rate. The ACE group’s "erred" response rate became more variable in posttest, while the TeamSTEPPS® group’s became less variable. No
significant difference was seen in erred response rates across training groups in posttest. To summarize, TeamSTEPPS® training led to improvements in performance such that there were no statistically significant differences between groups in posttest, despite differences observed in pretest performance. This replicates performance improvements seen in previous research utilizing the same training workshop (Anbro et al., 2020; Maraccini et al., 2018).

Due to the lack of clear delineation in performance across the training groups, the remainder of the analyses presented will instead focus only on differences between simulation groups (VR and Direct). It is worth noting that increased safety precautions (use of face masks, completion of online training that promoted self-monitoring, awareness of social interactions, etc.) mandated by the university’s COVID-19 response may have affected the degree to which TeamSTEPPS® training improved performance. In the absence of these additional requirements and accompanying stress on student participants, the TeamSTEPPS® training likely would have shown an even greater impact on performance as in previous literature. Specifically, we could have expected to see higher rates of “Correct” responding and lower rates of “Missing” and “Erred” responding among the TeamSTEPPS® groups compared to the ACE groups. Future research without such requirements in place can address this hypothesis.

**Communication Accuracy: Handoff Provider Weighted Change Scores**

In addition to the handoff provider analyses above, a weighted scorecard was created to calculate one overall simulation session score for each handoff provider. Scores were calculated for each participant's performance during both
pretest and posttest. The weighted scorecard is shown with one participant’s data as a sample in Figure 15. In this scorecard, "correct," "missing," "omitted," and "erred" responses all have different values ("multipliers"). "Correct" statements are assigned a multiplier of 1.0, such that each correct statement counts as “+1 point” towards an overall handoff score. "Missing," “Omitted,” and “Erred” responses are all assigned negative multipliers (-0.25, -0.50, and -1.0, respectively), as these represent undesirable performance. Abernathy (1996; 2000) discusses weighted scoring, including negative multipliers, in the context of developing scorecards. Furthermore, the scoring system used to calculate overall handoff scores was reviewed and approved by the three external experts serving as training workshop instructors in the current study.

The frequency of coded responses was populated into the scorecard for each handoff provider's pretest and posttest simulation. The frequencies of "correct" responses were multiplied by a value of +1.0. Their opposite, "erred" response frequencies, were multiplied by a value of -1.0. The frequencies of "omitted" (i.e., entirely omitted) responses were multiplied by -0.50, and the frequencies of "missing" (i.e., partially omitted) responses were multiplied by -0.25. The four weighted scores are then added together and divided by the total possible score, or the total number of comments made during the simulated patient handoff.

Cases 1-4 varied in their total possible scores, with a range of 16-18 possible points maximum. Change scores are calculated by subtracting each participant's pretest score from their posttest score. Individual participant pretest
and posttest scores, along with their change scores, are shown in Figure 16. Among the VR simulation participants, 14 (61%) showed a positive change score, while 9 (39%) showed a negative change score from pretest to posttest. Among the Direct simulation participants, 7 (35%) showed a positive change score, and 13 (65%) showed a negative change score from pretest to posttest.

**Communication Accuracy: Handoff Receiver**

Figure 17 shows handoff receiver performance during simulated patient handoffs across simulation groups. The top graph shows pretest results, and the bottom graph shows posttest results. "Correct," "missing," and "erred" responses are shown. There is no requirement for handoff receivers to repeat all patient information during the "synthesis by receiver" portion of the I-PASS handoff structure. For this reason, "omitted" responses were not counted for handoff receivers, and are not included in this analysis. During pretest, handoff receivers in Direct simulations made over twice as many "correct" comments on average as those in VR simulations. The pretest difference between "correct" synthesized comments in VR simulations (M = 1.70, SD = 0.82) and Direct simulations (M = 3.62, SD = 2.14) was statistically significant; t = -2.66796, p <0.01. This difference was seen in posttest as well. The posttest difference between "correct" synthesized comments in VR simulations (M = 2.62, SD = 1.85) and Direct simulations (M = 5.05, SD = 3.04) was also statistically significant; t = -2.59322, p <0.01.

The VR group made fewer average "missing" and "erred" comments than the Direct group, both in pretest and posttest. In pretest, handoff receivers in VR
simulations made zero comments scored as “missing” across the board and an average of 0.1 erred comments. Concurrently, handoff receivers in Direct simulations made an average of 0.2 “missing” comments and 0.07 “erred” comments. In posttest, VR handoff receivers made an average of 0.2 “missing” comments and zero “erred” comments across the board. Concurrently, handoff receivers in Direct simulations made an average of 0.5 “missing” comments and 0.4 “erred” comments. The posttest difference between “erred” comments for VR simulation participants (M = 0, SD = 0) and Direct simulation participants (M = 0.43, SD = 0.75) was statistically significant; t = -2.05798, p < 0.10.

It is also worth noting the number of participants who are represented by these data. In pretest, 10 VR simulation participants and 12 Direct simulation participants made synthesis by receiver comments. In posttest, the number of participants synthesizing information increased: a slight increase for VR participants (13 of 23 participants) and a substantial increase for Direct simulation participants (20 of 20 participants).

“Filler” Words

The rate of filler words (“um” and “uh”) was calculated for each handoff during pretest and posttest as a secondary assessment. Figure 18 shows the distribution of pretest and posttest rate of filler words per minute. In VR simulations (top graph), 14 participants increased their rate of filler words from pretest to posttest, and 9 participants decreased their rate of filler words. In Direct simulations (bottom graph), 7 participants increased their rate of filler words from pretest to posttest, and 13 decreased their rate of filler words.
Communication Accuracy IOA

Trial-by-trial inter-observer agreement was calculated for 100% of pretest and posttest simulation sessions. Each sub-category of the I-PASS handoff structure (shown in Figure 6) was scored independently by one primary scorer and one of six secondary scorers. Pretest session scoring resulted in 94.9% agreement (range, 90 – 100%), and posttest session resulted in 95.4% agreement (range, 90 – 100%).

Dyad Samples: Communication Accuracy

Figure 19 shows two time-series graphs depicting communication accuracy across time in simulated patient handoffs. The top graph shows communication accuracy for an exemplary dyad. Both of these participants made frequent “correct” comments, few “missing” comments, and no “erred” comments as handoff providers. As handoff receivers, both participants synthesized information correctly, with only one “missing” synthesis comment for participant 211’s handoff receiver segment. The combined duration of both these handoffs was 230s. It is also worth noting that the exemplary dyad was one who completed their simulations in VR.

The bottom graph shows communication accuracy for a typical dyad. Both of these participants made fewer “correct” comments and more “missing” comments than the exemplary dyad. These participants did synthesize more information, but the frequencies of synthesized information included higher “missing” and “erred” comments than the exemplary dyad. The combined duration of both these handoffs was 313s.
Situational Awareness: Level 1 (Perception)

As described above, eye-tracking sensors embedded in the VR headsets were used to capture and record the frequency of participant eye fixations – the number of times participants looked at each of the predefined AOIs shown in Figure 7. The start time, duration, and stop time for each eye fixation was automatically recorded. These data were plotted onto individual time-series graphs for each participant. A visual analysis of the resulting patterns yielded six distinct categories of eye gaze patterns. These patterns were classified based on two criteria: 1) the rate of eye fixations for the AOI with the highest fixation count, and 2) that AOI’s pattern variability with other, simultaneous AOI patterns. Patterns were characterized as high rate (slope ≥ 1.0), medium rate (slope = 0.5 – 0.99), or low rate (slope < 0.5), and as either stable or variable. The resulting six patterns identified were labeled Inattentive (low rate, stable), Distracted (low rate, variable), Attentive (medium rate, stable), Divided (medium rate, variable), Focused (high rate, stable), and Observant (high rate, variable). Figure 20 shows the classification criteria, along with these resulting six patterns.

Examples of these six patterns are shown in Figures 21 – 26. Figure 27 shows the number of participants demonstrating each pattern during pretest and posttest. During pretest, the most common patterns were Attentive and Distracted. During posttest, the most common patterns were Attentive and Divided. The highest-fixation AOIs for participants in both pretest and posttest were predominantly the partner avatar or the simulated patient, as these sources of information would be richest in non-simulated scenarios. While other stimulus
objects (e.g. computer monitor displaying vitals) in the environment would certainly provide information for healthcare providers the most information likely comes from the patient and/or other healthcare providers.

The relationship between eye fixation patterns and handoff comprehension measures is discussed below, following data summaries of situational awareness levels 2 & 3. Additionally, a sample of one VR dyad’s eye fixations across successive patient handoffs are shown in Figure 28.

**Situational Awareness: Level 2 (Comprehension)**

Following each simulation, participants were given a brief patient assessment questionnaire. This questionnaire asked participants about the patient they received care of (i.e., handed off by their partner) during their simulation. The questions were identical for all fictional patient cases, but correct responses varied from case to case. Participants were asked to report the patient’s name, condition, reason for hospital admittance, action steps already taken, medications (if any) ordered but not yet administered, and future action steps to take. All four sets of questions totaled the same possible score: 11 points. These question sets and acceptable answers are shown in Appendix G.

Whenever a participant scored less than a perfect score, the transcripts for the simulation were revisited. If the handoff provider left out any patient information, the corresponding number of points was deducted from the overall possible total. For example, if a handoff provider gave all the relevant patient information for Case 1 (John Kim) except that patient’s first name (i.e., “Mr.
Kim*), the possible total score for their handoff receiver was reduced from 11 to 10. These adjusted scores are shown in Figure 29.

The difference between pretest scores for VR simulation participants (M = 40.70, SD = 17.74) and Direct simulation participants (M = 50.43, SD = 17.45) was statistically significant; t = -1.83179, p <0.10. The difference between posttest scores for VR simulation participants (M = 68.22, SD = 13.16) and Direct simulation participants (M = 79.23, SD = 12.44) was also statistically significant; t = -2.88091, p <0.01. The increase in average scores for VR and Direct simulations are comparable (+27% & +29%, respectively). Both groups reduced their variability of scores from pretest to posttest.

**Situational Awareness: Level 3 (Projection)**

Following the brief patient assessment questionnaire, participants were asked to freely respond to two hypothetical case developments for the same patient. Two likely scenarios were presented for each patient (see Appendix E). Participants provided information regarding their proposed plan of care for that patient, with the level of detail left to their discretion. The action steps identified in their responses were coded for appropriateness by experts from healthcare education. Responses were scored as appropriate actions, team situational awareness (SA) actions, or inappropriate actions. Appropriate actions were those that described specific tasks to functionally aid the patient with regard to the case development (e.g., “assess for stroke” when half of the patient’s face begins to droop). Team SA actions were those that described communicating with other members of the medical team (e.g., “contact cardiology”). Inappropriate actions
were any actions that did not functionally address the presenting concern (e.g., “assess for stroke” when the patient has converted back to atrial fibrillation).

Frequency counts of these coded responses are shown in Figure 30. VR simulation participants provided a comparable number of appropriate action steps to Direct simulation participants in pretest (2.1 & 2.3 averages, respectively). In posttest, VR simulation participants provided a comparable number of average appropriate action steps compared to Direct participants (2.6 & 3.0 averages, respectively). The average frequency of team SA action steps was significantly higher for Direct participants (M = 1.38, SD = 0.72) than VR participants (M = 0.71, SD = 1.01) in pretest; t = -2.22465, p < 0.10. The frequency of team SA action steps was comparable across groups in posttest (0.8 avg. VR, 0.9 avg. Direct). On average, both groups provided similar frequencies of inappropriate actions in pretest (1.2 in VR, 0.8 in Direct). During posttest, there was a larger discrepancy. VR participants provided significantly fewer inappropriate actions (M = 0.22, SD = 0.42) than Direct participants (M = 0.50, SD = 0.91) in posttest; t = -10.45126, p < 0.001.

Situational Awareness Across Successive Levels

As discussed above, the conceptual levels of situational awareness build upon one another. Stimuli must first be perceived (L1) to then be comprehended (L2). The accuracy of comprehension (L2) then sets the occasion for accuracy in predicting (L3). The relationships between successive levels are shown in Figure 31 & Figure 34. Figure 31 shows the relationship between measures of situational awareness L1 and L2. Participants’ scores on the brief patient
assessment questionnaire are shown for each of the six eye fixation patterns previously identified. During pretest, participants showing the Attentive and Observant patterns scored the highest on their questionnaires. During posttest, participants showing the Attentive and Inattentive patterns scored highest.

An overall correlation analysis reveals a small negative correlation between eye fixation rate and L2 accuracy scores in pretest (N = 20; r = -0.27) and an insignificant negative correlation between these measures in posttest (N = 23; r = -0.07). Further correlation analyses were conducted to assess for any more specific relations. A variability-based analysis is shown on scatterplots in Figure 32. The pretest scatterplot reveals a moderate correlation (N = 9; r = -0.48) between the rate of stable eye fixation patterns and L2 accuracy scores, along with an insignificant correlation (N = 11; r = -0.04) between the rate of variable eye fixation patterns and L2 accuracy scores. The posttest scatterplot reveals a small correlation (N = 13; r = -0.24) between the rate of stable eye fixation patterns and L2 accuracy scores, along with an insignificant correlation (N = 10; r = 0.06) between the rate of variable eye fixation patterns and L2 accuracy scores.

These same data are shown in Figure 33, separated by which AOI the participants fixated on at the highest rate, rather than by the stability or variability of their eye fixation patterns. Participants whose fixation rate was highest on the patient showed a large negative correlation with L2 accuracy scores (N = 5; r = -0.74) during pretest, and a large positive correlation with L2 accuracy scores (N = 7; r = 0.70) during posttest. Participants whose fixation rate was highest on
their partner’s avatar showed a small negative correlation with L2 accuracy scores (N = 11; r = -0.17) during pretest, and an insignificant positive correlation with L2 accuracy scores (N = 13; r = 0.08) during posttest. Finally, participants whose fixation rate was highest on any AOI other than their partner’s avatar or the patient showed an insignificant negative correlation with L2 accuracy scores (N = 4; r = -0.03) during pretest, and a large negative correlation with L2 accuracy scores (N = 3; r = -0.65) during posttest.

Figure 3 shows the relationship between measures of situational awareness L2 and L3; this scatter plot shows L2 brief assessment scores and the frequency of L3 responses. The L3 responses shown are summations of both 
appropriate actions and team SA actions for each participant. For instance, if a participant responded with three appropriate actions and two team SA actions, their L3 frequency score would be 5. Trend lines are also shown; these reveal a moderate positive correlation between L2 accuracy scores and L3 response frequencies in pretest (r = 0.63) that becomes insignificant during posttest (r = 0.19). For comparison, the bottom graph in Figure 3 shows this same relationship among Direct simulation participants’ data. These data reveal an insignificant positive correlation between L2 accuracy scores and L3 response frequencies in both pretest (r = 0.25) and posttest (r = 0.03).

Social Validity

All participants completed multiple assessments of social validity. The first assessment discussed below is a training assessment – specifically, the T-TAQ. The T-TAQ is part of the TeamSTEPPS® curriculum and is used to assess
attitudes towards working on interdisciplinary medical teams. This assessment was distributed twice, once after the pretest and once after posttest simulation sessions. The second assessment discussed below assesses the social validity of both simulation modalities – VR and Direct. Students responded to a series of questions regarding the simulation modality they participated in throughout this study. This assessment was distributed once, at the conclusion of posttest simulations.

**T-TAQ Results**

Average scores for each of the five assessment categories of the T-TAQ are shown in Figure 35. Average scores are separated by which training workshop the participants attended. The top graph shows pretest results, and the bottom graph shows posttest results. Scores range from 1 (low agreement with statements) to 5 (high agreement with statements). Items 20, 21, 24, and 30 are inversely scored (e.g., “strongly agree” = 1 instead of 5). Average scores closer to 5 indicate the higher level of importance placed upon that specific T-TAQ category, as indicated by participants’ level of agreement.

All group averages ranged between 4.49 – 4.84 across items and distributions of the assessment. During pretest, statistically significant differences were seen between groups for two of the five categories. Participants in the TeamSTEPPS® group rated *team structure* and *situation monitoring* significantly higher (p < 0.01) than participants in the ACE group. During posttest, these two differences maintained at the same level of significance, along with three additional statistically significant differences. *Leadership* and *communication*
were rated significantly higher (p < 0.01 and p < 0.10, respectively) by the TeamSTEPPS® group, while mutual support was rated significantly higher (p < 0.10) by the ACE group.

**Simulation Modality Survey Results**

Participants completed a survey regarding their simulation modality (VR or Direct). The results of this survey are shown in Table 4. 83% (19/23) of VR simulation participants indicated they would be extremely or somewhat likely to volunteer for a VR simulation again. In comparison, only 59% (13/22) of Direct simulation participants indicated the same. The majority of both groups indicated they would recommend their simulation modality to other students (83% VR, 82% Direct). The majority of both groups also indicated that their simulation modality would add educational value if it were used more frequently (78% VR, 86% Direct). The novelty, fun, or realism of the simulation was the most commonly cited factor that students enjoyed most about VR simulations, followed by the opportunity to work with a partner. Working with a partner was the most frequently cited factor that students enjoyed about Direct simulations, followed by the opportunity to practice effective communication skills.

Students also responded to VR-specific questions. Only 53% of VR simulation participants and 41% of Direct simulation participants had ever used VR before this study. All of these previous VR users had minimal time in VR (1-2 hours), except for one participant in the VR group who had 10+ hours of VR experience. Previous uses of VR were predominantly for playing video games for both groups, although some unique VR applications were reported: studying,
military training, and watching movies. The VR participants alone were asked about the realism and ease of use for their VR simulations. 87% (20/23) reported the VR environment was either very or somewhat realistic, and 96% (22/23) reported the VR simulation was either extremely or somewhat easy to use.

**Material Cost Analysis**

Table 5 shows an estimated cost analysis for a simulation-based study of this kind. The materials listed are based on the starting assumption that researchers have only a large enough space with which to work. The materials listed for VR simulations include all necessary hardware to run a simulation of this kind. Specialized software was not used for eye-tracking data analysis, and is therefore not included in this estimate. The materials listed for Direct simulations are the referent stimuli objects used in the current study (i.e., the real-world equivalents of the virtual stimulus objects tracked as AOIs in VR). These objects give a degree of external validity to the simulated patient handoffs. It should be noted that while the cost of materials for VR simulations ($6,193.84) is higher than Direct simulations ($2,034.99), there is a more significant opportunity to use those same pieces of VR hardware for a variety of simulations. On the other hand, the Direct simulation materials can only be used for similar simulations – that is, relatively simple patient exam room simulations.

A labor cost analysis was not conducted; labor hours will inevitably vary, given the complexity of simulation and depth of data analysis required. The current study utilized graduate students, undergraduate students, and working professionals on a volunteer basis. In other words, no labor costs were paid,
although anecdotally the time required was comparable across simulation modalities. Future research applications may be able to arrange a similar paradigm, while organizations looking to utilize VR simulations for training or evaluation will certainly accumulate varying labor costs. In-depth performance data analysis, such as that presented above, requires substantially more labor than simple data analyses. Selecting key performance metrics and automating the measurement processes to the fullest extent is generally recommended to minimize labor costs in future VR applications.

**Discussion**

This study provided a preliminary context for validating the use of VR simulations in place of Direct simulations within healthcare education. Student performance in both simulation modalities was assessed and compared using a number of behavioral measures described in detail above. Data yielded by these assessments, compared across simulation modalities, demonstrate the validity of using VR simulations in place of Direct simulations for simulated events of this nature. The similarities and differences in performance are discussed below. An overview of primary and secondary dependent measure findings is followed with a discussion of research implications, an analysis of the current limitations (both addressed and unaddressed during the study), and some recommendations for future research.

**Primary Dependent Measures**

The first set of primary dependent measures analyzed in this study pertained to communication accuracy of handoff providers and handoff receivers.
VR and Direct simulation participants completed two rounds of simulated patient handoffs with nearly identical handoff provider accuracy at the group level, measured using the previously described descriptive analysis to code statements as “Correct,” “Missing,” “Erred,” or “Omitted.” Group-level analyses indicate comparable average results, while conversion to rate of response reveals differences between simulation modalities. Participants in the VR simulations tended to emit Correct responses at a faster rate than Direct simulation participants. Individual participant change scores also reveal that more VR simulation participants tended to improve their performance from pretest to posttest, compared to Direct simulation participants.

Overall, handoff provider accuracy was comparable at the group level. VR simulation participants achieved this level of accuracy in significantly less time than Direct simulation participants. Nonverbal communication, such as facial expressions and gestures, likely influences the duration of patient handoffs. While gestures were approximated in VR (moving one’s arms while holding the VR controllers causes the respective avatar also to move their arms), facial expressions were not. Anecdotally, the majority of VR participants only moved their arms when raising or lowering the hand which held their virtual notecard. This discrepancy in handoff duration is worth noting for future applications of VR; if simulated tasks are time-sensitive, VR may be the better option to pursue.

Further performance differences across simulations were revealed when analyzing the verbal behavior of handoff receivers. Using the same descriptive analysis categories above, the frequency of synthesis by receiver comments
(information repeated back for clarification) was consistently higher in Direct simulations than in VR. Frequencies of “Missing” and “Erred” were relatively low in both groups during pretest and posttest, but Direct simulation participants tended to emit more “Correct synthesis by receiver comments. Furthermore, the number of participants emitting any synthesized comments was greater than VR – both in pretest and substantially in posttest. The lack of nonverbal communication in VR discussed above may play a role in this discrepancy, as may the general novelty of VR to most participants.

The next set of primary dependent measures pertain to situational awareness – levels 2 (comprehension) and 3 (prediction). Comprehension was assessed using a brief patient assessment after each simulation. Direct simulation participants scored, on average, higher than VR simulation participants in both pretest and posttest, although both groups showed improvement in posttest. An initial evaluation of predicting responses revealed similar performance again, as assessed by experts from within the respective healthcare education programs. Another way of evaluating the accuracy of predictive responses in future research is to test predictive responses against actual cases as they develop.

Table 6 shows a summarized overview of various performance comparisons between participants in VR simulations and Direct simulations. Specifically, this table indicates the prevalence of statistically significant differences in performance across both simulation modalities. For comparisons where performance in both simulation modalities is equivalent (i.e., no
statistically significant differences), each simulation modality receives a +0 score. For comparisons where one performance in one simulation modality is significantly better than in the other, that simulation modality receives a +1 score. Both VR and Direct simulations end with a score of 5. In other words, there are ten performance comparisons where performance significantly differs – five in favor of VR simulations and five in favor of Direct simulations. The remaining 20 comparisons show no statistically significant differences. This analysis supports the conclusion that overall, VR and Direct simulations evoke equivalent performance. The raw statistical data used are shown in Table 7.

There are certain features of both simulation modalities that lead to more advantageous performance than the other. For example, the prevalence of nonverbal communication in Direct simulations likely contributed to the prevalence of synthesis by receiver comments. The novelty of VR simulations may have contributed to the higher rate of handoff providers’ “Correct” responding in VR. Technological considerations, along with the benefits and drawbacks of both simulation modalities, should factor into any decision to use VR for assessment. Training in VR was not assessed in the current study, but future research would benefit from evaluating the acquisition of new skills while training in VR simulations versus Direct simulations.

From a cost perspective, accumulating the necessary materials for VR may bear a higher initial cost; this depends on the simulation’s complexity. Simpler simulations, like patient handoffs in an exam room, are somewhat less expensive to set up and run outside of VR. However, there are advantages to
choosing VR simulations: the ability of people to work or train together when they are not co-located; simulations can be improved as time progresses without any physical relocation or addition of materials; the same hardware and software can be used for a nearly infinite variety of simulations as differing organizational needs arise and the technology continues to grow. It should also be noted that VR simulations may take more time to create initially, especially if more complex and interactive simulations are desired.

**Secondary Dependent Measures**

This study also provided several analyses of the interrelation between successive, conceptual levels of situational awareness. Six patterns of eye fixations during handoff receiver segments were classified using a) the rate of eye fixations on the participant’s highest-rate AOI, and b) the stability or variability of eye fixations on this AOI. This initial analysis of classifying eye fixation patterns should be replicated in the future. More data will either substantiate or refute the validity of these categories and any subsequent analyses. Correlations between eye fixation patterns and subsequent performance on brief patient assessment questionnaires revealed variability. Correlations ranged from insignificant to moderate when assessing based on stability or variability of eye fixations. A range of insignificant to strong correlations was seen when assessing based on participants’ highest fixation rate AOI. This range indicates the likelihood of factors in addition to eye fixations that determine situational awareness. Future research may, for instance, explore newly emerging *ear-tracking technology*, which has the capability of sensing and
recording micromovements of the ears (Nielsen, 2019). These micromovements, coupled with eye-tracking measures, may provide greater insight into measuring and improving situational awareness. Such measures may also address some current limitations in eye-tracking technology, which may include interference from prescription eyeglasses: “We’re really challenged to try to record all the things we’re trying to record, and what we find is that [for some subjects] we can’t get good eye-tracking signals. If a subject requires glasses, we can’t really get a good signal” (Noah, 2020).

**Implications**

This study demonstrates the substantial contribution of behavior analytic assessment towards evaluating healthcare student performance in simulations. Descriptive assessment methods were utilized to analyze and interpret the accuracy of verbal communication between handoff providers and handoff receivers. This study also assessed the relationship between communication accuracy during a patient handoff, and participants’ ability to recall patient information before predicting what should be done in light of case developments.

Participants in this study were a relatively homogeneous group, in terms of their performance. There is undoubtedly variation at the individual level, but the two simulation groups performed comparably; this further exemplifies the interchangeability of VR and Direct simulations within this specific population. As the complexity and interactivity of VR simulations increases, further empirical research can expand on the findings of this study. The “instructor mode” used to record virtual simulations can be used instead to deliver live performance
feedback during a simulation, rather than merely recording simulations. Coupled with the insertion of visual prompts and interactive checklists, behavioral skills training in VR is a feasible endeavor for future research and practical applications; this will be especially useful to shape healthcare students’ synthesis by receiver skillset.

The analyses of participants’ situational awareness in VR simulations are a step forward in analyzing the factors that exert influence on participants’ situational awareness. Most current eye-tracking research has not explored this relationship. Rather than using eye-tracking to measure behavior in any intervention-based study, the majority of eye-tracking research is “definitely basic science research from our perspective, at this moment” (Noah, 2020). Assuming these factors are reliably seen across future research studies (i.e., similar patterns of eye fixations are seen in similar contexts), best practices can be identified and worked into VR training programs for future and current healthcare professionals.

Furthermore, the findings of this study can be explored for replication in other industries similarly adopting CRM (Alavosius et al., 2017). As the aviation, nuclear power, and oil & gas industries also move forward in their respective digitalization efforts, VR is likely to play a role in training and assessment for these industries. VR has the potential to improve both the personalization of training and performance feedback, given the multitude of programmable simulations and behavioral measurements that can be captured. VR is also a viable solution to teams who are either not co-located or who seldom have
opportunities to train together. VR simulations support interdisciplinary collaboration as well – for both a research team and the participants involved. On the researcher end, various experts are needed to program simulations; populate the simulations with scenarios that are both internally and externally valid; and to measure and assess behavior. On the participant end, an interdisciplinary team (e.g., an entire patient healthcare team) can train together within VR simulations; the team dynamics and performance can then be evaluated in this highly controlled environment.

**Limitations**

The findings discussed above should be considered in light of some limitations. The group sizes in the current study were relatively small due to the number of available healthcare students as participants. While this does not invalidate the behavioral measures reported above, future research would benefit from larger group sizes; this will either replicate findings in the current study or reveal a greater spectrum of individual differences in performance that need to be accounted for.

The majority of the following limitations concern the technological features of this study. There was no opportunity to pilot test the VR simulation with one or two dyads before the first day of participant simulations, due to academic schedule constraints of participants and governmental restrictions introduced by the mandatory response to the COVID-19 pandemic. Such a pilot test would have revealed a small number of technical glitches earlier on. These glitches could then be addressed by the research team, minimizing variation during the
pretest and posttest phases. Instead, these glitches were identified during pretest simulations and remedied before posttest simulations.

First, participants identified a text wrapping issue with their virtual notecards. The VR software was programmed to wrap text and make the font size smaller as more characters were added. In this manner, students should be able to enter as much or as little patient information on their virtual notecards as they’d like. During pretest simulations, the text wrapping did not work; instead, text was cut off after a particular line or character length. Participants were told of this limitation during pretest once it was identified. They were advised to use abbreviations when possible, and the approximate line/character limit was communicated to them as well. This issue was resolved before posttest simulations, and participants were able to see as much or as little text on their virtual notecards as they had previously typed. While this issue may have had an impact on handoff providers’ communication accuracy in VR, the pretest results as shown in Figure 10 were not statistically different than those associated with participants’ performance in Direct simulations.

Second, the selection of customization options for their avatar (hair color, skin color, gender) required students to either walk or scroll their avatars over to buttons in the virtual simulation. These buttons had to be virtually “pushed” for the corresponding feature to change. Given the small size of the physical space students occupied during their VR simulations, this caused some difficulty. This selection process was altered for posttest simulations. During posttest, participants selected these features by using a virtual laser that emitted from their
right-hand controller. The laser need only point to one of these buttons; then, the participant could click the controller’s trigger button to select that virtual button and alter their avatar’s appearance accordingly.

Third, as previously noted by Maraccini et al. (2018), students did not have the opportunity to write any of their partner’s patient information while functioning as a handoff receiver. While this would have been a simple limitation to address in Direct simulations (i.e., provide pen & paper to take into the simulation room), the requisite hardware and software to write legibly and easily in VR were unavailable. Therefore, to keep simulation modalities equitable, students in both conditions were not permitted to take notes during their simulated patient handoffs. The inability to take notes was the only technological limitation that went unaddressed from pretest to posttest. Future research should explore the use of hardware and software that allows participants to write in VR, such that their writing is legible and useful in simulations. Having written notes will likely influence the number of items synthesized by the handoff receiver. More of these synthesized comments will probably be scored as “correct” when virtual, handwritten notes can be easily used.

While the transcription of simulations was completed by hand, this is a massive limitation regarding time allocation. Transcription software, especially software that can transcribe in real-time, would be invaluable to simulations such as these. This value becomes exponentially higher as the duration and complexity of future simulations increase. Such software integration would also greatly benefit non-academic organizations that often do not have extra
personnel on hand that can allocate large amounts of time to transcribing simulations. In both academic and non-academic cases, this would also allow for feedback delivery to occur either in real-time or with little delay following the simulation’s completion.

A minor non-technological limitation was identified and addressed between pretest and posttest simulations. A small number of students received the incorrect hypothetical case development questions used to measure their ability to predict what action steps are appropriate (SA L3). These students' SA L3 data were excluded from the group analysis, and a simple discriminative stimulus ($S^D$) was added to each survey before posttest simulations began. The $S^D$ used was a simple “For PA students only” or “For nursing students only” header on the first page of the survey. Following the use of this $S^D$ in the posttest assessment, no student was given the incorrect SA L3 questions.

Finally, it should be noted that the present study was conducted, as previously indicated, during the COVID-19 pandemic of 2020 in the US. The resulting governmental and university restrictions (e.g., limited numbers allowed to gather) introduced methodological challenges (e.g., a potential confound effect of COVID-19 training prior to experimental exposure) that could not be avoided. Moreover, due to scheduling constraints and curriculum requirements, students were unable to participate in a third round of simulations. A third simulation, following completion of a second training (e.g., TeamSTEPPS® participants completing ACE training after their second simulation), would have allowed for an opportunity to assess for maturation through repeated exposure to simulations.
Order effects of the trainings could have been assessed as well, and future research may benefit from this assessment. The opportunity to conduct pilot testing of VR simulation technology was also limited, resulting in some of the technical limitations discussed above.

**Future Research**

It is important to note that the findings presented above are *initial findings* in this area of research. This study is the third in a developing line of research; only one of these other two studies has utilized VR (Anbro et al., 2020). Potential limitations of comparison studies have been discussed critically (e.g., Johnston, 1988). Such limitations include an over-reliance on group-level data versus individual data, lack of generality of findings, and inappropriate comparisons between *no treatment* groups and some other procedure. Appropriate comparison studies, it is argued, answer “questions about the relative effects of two or more procedures for a particular combination of target behavior, client, and setting variables (Johnston, 1988, p. 7).” This study demonstrated that groups of similar skill level perform comparably in simulations, regardless of modality. Continued use of VR is therefore worth pursuing, as performance will not alter greatly; this will aid in HROs’ ongoing digitalization efforts. Digitalization helps reduce overall risk to employees, and are an increasing focus of HROs (Deloitte Center for Energy Solutions, 2017). Future research studies are poised to further explore behavioral assessment and training in VR. Several such suggestions are discussed below.
This study examined skills which are undoubtedly critical to working effectively within interdisciplinary medical teams, namely communication and situational awareness. However, technical skills within medicine were not assessed in the current simulations. These skills are plentiful: from foundational skills such as drawing blood or administering medication properly, to complex team-based procedures such as performing surgical operations. Future research can use increasingly interactive VR simulations to train and assess these skills among healthcare students. A highly controlled virtual environment can allow for mistakes to be made without consequence to human patients. Learning can be more precisely quantified for each individual. Ongoing training sessions in VR can potentially replace annual training and evaluation in physical simulators, which is standard practice within aviation. The example of error reduction that aviation has set can be generalized through such research.

As artificial intelligence (AI) becomes more prominent in today’s world, it can be integrated into VR simulations to function as team members, patients, or both. This integration will take time to refine, but the result would allow healthcare professionals to practice skills independent of their team members’ professional or personal schedules. As AI becomes increasingly sophisticated, it may be integrated within simulations for data analysis. AI can be developed to evaluate performance in real-time, giving corrective feedback or praise to the performer contingent upon their performance. Future research can compare the effectiveness of feedback delivered by AI to feedback provided by other humans.
Future research should also examine the performance measures used in this study in more detail. For instance, while the rates of “Correct,” “Missing,” “Omitted,” and “Erred” statements were assessed among handoff providers, there was no standard for comparison. Future research should examine the behavior of exemplary medical performers working in the industry. Comparing the performance of exemplars and students, or even typical performers, would allow for the generation of PIP measures (Gilbert, 2007); these can be used to measure typical performance relative to exemplary performance.

Similarly, measures of situational awareness level 3 (perception) can be further investigated. Rather than asking participants to provide action steps for hypothetical patient cases, their predictive responses can be tested against actual cases – either current cases as they unfold or historical cases, where the participants are given pieces of information over time. Comparisons to exemplary performers can also be made in this regard, and research into which factors are essential for making accurate medical predictions should be conducted. Such research would allow the medical industry to refine its best practices continuously, informed by behavioral measures.

The majority of participants in the current study reported little to no prior use of VR. Future research would benefit from exploring this factor in greater detail. For instance, future research could program different, predetermined amounts of VR time for participants before they participate in a patient handoff simulation. By allotting time for participants to familiarize themselves within a VR space, and by varying that time across participants, researchers could determine
ideal levels of exposure to VR that would have an enhancing effect on performance. This may help address the infrequency of synthesis by receiver comments from handoff receivers in VR.

Finally, future research should examine the transfer of skills from VR simulations to clinical environments; this becomes critical when VR simulations are used for training purposes. While we may assume that these skills will transfer to clinical settings, robust performance evaluation will reveal whether this is the case. Eye-tracking glasses in research contexts may provide insight into whether patterns of observation (SA L1), for instance, are different in VR than in real life. This comparison may also help validate which measures of eye-tracking are most important. Are overall patterns, as identified above, most important? Or is a simpler metric like fixation count alone sufficient? Answers to these questions will likely have some bearing on shaping healthcare professionals’ situational awareness. At present, “it [eye-tracking] is wonderful because you get a rich set of measures, but it can be hard to pick which is the right one...we don’t have an answer, unfortunately, for which measure is best” (Arunachalam, 2020).

**Summary**

The purpose of this study was to assess the criterion validity of VR simulations in place of Direct (in-person) simulations, in the context of healthcare education. Results demonstrated that performance in both simulation modalities (VR and Direct) was comparable across several behavioral metrics, validating continued exploration and use of VR within the medical industry. Furthermore, initial analyses of the interrelationship between the successive, conceptual levels
of situational awareness revealed insights primed for further research exploration.

As this area of research develops, there are many experimental questions to be asked. Consequently, there are several recommendations that future research can incorporate. Some noteworthy recommendations include assessing and training technical skills in VR, integrating AI functions into simulations, comparing performance measures to those of identified exemplars in the field, and evaluating the transfer of skills from VR simulations to real-world clinical settings. The push for simulations to become more sophisticated and widely used has spanned decades. We are seeing the beginning stages of growth (likely, exponential growth) in VR’s capabilities. Continued research and application of VR simulation technology will conceivably play a role in systemically reducing the prevalence of medical error – in the United States and abroad. Let us remember the Institute of Medicine’s (1999) call to action on this matter, dating back to their landmark report, *To Err is Human*:

Health care organizations and teaching institutions should participate in the development and use of simulation for training novice practitioners, problem solving, and crisis management, especially when new and potentially hazardous procedures and equipment are introduced. Crew Resource Management techniques, combined with simulation, have substantially improved aviation safety and can be modified for health care use. (p. 179)
References


Moran, D.J. *Building safety commitment*. Joliet, IL: Valued Living Books, Inc.


### Tables

<table>
<thead>
<tr>
<th>Time Start</th>
<th>Time Stop</th>
<th>Physician (MD) Statements, Nurse (RN) Statements, &amp; In-Simulation Events</th>
<th>Sample Verbal Checkback</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td>0:13</td>
<td><em>(Instruction appears: “Please say START after the countdown;” 3-2-1 visual countdown; then “START”)</em></td>
<td>“Start.” (Used for video recording calibration)</td>
</tr>
<tr>
<td>0:13</td>
<td>3:26</td>
<td><em>(Simulation begins: RN briefs MD on patient vitals; MD asks the simulated patient intake questions)</em></td>
<td>N/A</td>
</tr>
<tr>
<td>3:27</td>
<td>N/A</td>
<td><em>(“CRISIS EVENT” - Simulated patient seizure begins)</em></td>
<td>N/A</td>
</tr>
<tr>
<td>3:28</td>
<td>3:30</td>
<td>MD: “She’s having an eclamptic seizure.”</td>
<td></td>
</tr>
<tr>
<td>3:31</td>
<td>3:35</td>
<td>RN: “Student, I’m going to need you to be a recorder and provide verbal feedback for us.”</td>
<td><em>optional</em> “Providing verbal checkbacks.”</td>
</tr>
<tr>
<td>3:38</td>
<td>3:41</td>
<td>MD: “Let’s have her in a left lateral decubitus.” <em>(MD &amp; RN move patient into a left lateral decubitus position)</em></td>
<td>1. <em>(1) Left lateral decubitus.</em></td>
</tr>
<tr>
<td>3:45</td>
<td>3:46</td>
<td>RN: <em>(moves bed rails up into position)</em> “Bed rails up.”</td>
<td>2. <em>(1) Bed rails up.</em></td>
</tr>
<tr>
<td>3:46</td>
<td>3:47</td>
<td>MD: <em>(moves bed rails up into position)</em> “Bed rails up.”</td>
<td></td>
</tr>
<tr>
<td>3:48</td>
<td>3:51</td>
<td>RN: “I’ve contacted the charge nurse and anesthesia.”</td>
<td>3. <em>(1) Charge nurse and (2) anesthesia contacted.</em></td>
</tr>
<tr>
<td>3:59</td>
<td>4:01</td>
<td>MD: “Oxygen on, airway is protected.” <em>(MD places oxygen mask on patient)</em></td>
<td>4. <em>(1) Oxygen on, (2) airway protected.</em></td>
</tr>
<tr>
<td>4:04</td>
<td>4:06</td>
<td>RN: “Oxygen is at 10 liters.” <em>(RN alerts charge nurse &amp; anesthesia)</em></td>
<td>5. <em>(1) Oxygen at (2) 10 liters.</em></td>
</tr>
<tr>
<td>4:07</td>
<td>4:09</td>
<td>MD: “Ok, let’s confirm she has no contraindications to magnesium sulfate.”</td>
<td></td>
</tr>
<tr>
<td>4:10</td>
<td>4:12</td>
<td>RN: <em>(checks patient chart)</em> “She has no contraindications.”</td>
<td>6. <em>(1) No contraindications to magnesium sulfate.</em></td>
</tr>
<tr>
<td>4:17</td>
<td>4:23</td>
<td>MD: “Let’s go ahead and give her 6 grams magnesium sulfate IV over 15 minutes, please.” <em>(RN holds medication up, showing label)</em></td>
<td>7. <em>(1) 6 grams (2) magnesium sulfate (3) IV (4) over 15 minutes.</em></td>
</tr>
<tr>
<td>4:24</td>
<td>4:28</td>
<td>RN: “6 grams magnesium sulfate IV over 15 minutes.” <em>(MD places oxygen mask on patient)</em></td>
<td></td>
</tr>
<tr>
<td>4:33</td>
<td>4:34</td>
<td><em>(RN begins administering medication; patient seizure ends)</em> MD: “The seizure has stopped.”</td>
<td>8. <em>(1) Seizure has stopped.</em></td>
</tr>
<tr>
<td>4:34</td>
<td>4:38</td>
<td><em>(MD continues administration of medication)</em> RN: “Magnesium sulfate is on board, 6 grams running.”</td>
<td></td>
</tr>
<tr>
<td>5:26</td>
<td>5:30</td>
<td>RN: “Magnesium sulfate is on board, 6 grams running.” <em>(MD briefs patient on what happened)</em></td>
<td>9. <em>(1) Magnesium sulfate (2) on board for (3) 6 grams running.</em></td>
</tr>
<tr>
<td>5:34</td>
<td>5:36</td>
<td>MD: “Fetal heart rate is returning to baseline.” <em>(MD continues administration of medication)</em></td>
<td>10. <em>(1) Fetal heart rate (2) returning to baseline.</em></td>
</tr>
<tr>
<td>N/A</td>
<td>5:42</td>
<td><em>(Simulation ends)</em></td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 1.** Timeline of simulation events (Anbro, et al., 2020). Sample verbal checkbacks meeting “correct” criteria are shown.
<table>
<thead>
<tr>
<th>SA Level</th>
<th>Behavioral Concepts</th>
<th>Examples in Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Perception)</td>
<td>Stimulus control</td>
<td>Doctor responds consistently to patient vitals displayed on-screen.</td>
</tr>
<tr>
<td></td>
<td>Conditional discrimination</td>
<td>Noises from various medical instruments: stable vs. unstable patient</td>
</tr>
<tr>
<td></td>
<td>Observing response</td>
<td>Nurse reading patient chart while handing off patient care to doctor.</td>
</tr>
<tr>
<td>2 (Comprehension)</td>
<td>Tacting stimuli function/features</td>
<td>Stating effects of various instruments/machines upon probe.</td>
</tr>
<tr>
<td></td>
<td>Interlocked behavioral chain</td>
<td>Verbal behavior in surgical team problem solving process.</td>
</tr>
<tr>
<td>3 (Prediction)</td>
<td>“Predicting” response</td>
<td>Team response(s) to problem (e.g., patient begins to have a seizure).</td>
</tr>
<tr>
<td></td>
<td>Generalization</td>
<td>Performance comparison: simulation vs. in-situ, novel event.</td>
</tr>
</tbody>
</table>

Table 2. A behavior analytic interpretation of situational awareness. This table expands on Endsley’s (1995a, 1995b) situational awareness (SA) model. The left column (“SA Level”) shows the three levels in Endsley’s model as written. The middle column (“Behavioral Concepts”) shows the relevant behavior analytic concept/principle that applies to each conceptualized level of SA. The right column (“Examples in Medicine”) show clinical in medicine for each of the behavioral concept/principle listed in the middle column.
<table>
<thead>
<tr>
<th></th>
<th>TeamSTEPPS® Tx</th>
<th>ACE Tx</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR Sims GROUP 1</td>
<td>N = 12 (original)</td>
<td>-1 (dropped)</td>
<td>N = 11 (PD, SV)</td>
</tr>
<tr>
<td>VR Sims GROUP 2</td>
<td>N = 12 (PD, SV)</td>
<td></td>
<td>N = 23 (PD)</td>
</tr>
<tr>
<td>VR Totals</td>
<td></td>
<td></td>
<td>N = 23 (SV)</td>
</tr>
<tr>
<td>Direct Sims GROUP 3</td>
<td>N = 12</td>
<td>-1 (no pretest)</td>
<td>N = 11 (PD)</td>
</tr>
<tr>
<td>Direct Sims GROUP 4</td>
<td>N = 10</td>
<td>-1 (video glitch)</td>
<td>N = 9 (PD)</td>
</tr>
<tr>
<td>Direct Totals</td>
<td></td>
<td></td>
<td>N = 20 (PD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 22 (SV)</td>
</tr>
<tr>
<td>Total TEAMSTEPPS® Totals</td>
<td>N = 22 (PD)</td>
<td></td>
<td>N = 43 (PD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 23 (SV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 45 (SV)</td>
</tr>
</tbody>
</table>

Table 3. Distribution of participants. Participants were quasi-randomly distributed into one of four groups: Group 1) VR-TeamSTEPPS®; Group 2) VR-ACE; Group 3) Direct-TeamSTEPPS®; Group 4) Direct-ACE. There were two technical issues in the direct simulation groups (one in group 3, one in group 4) that prevented two participants’ performance data (PD) from being used. No technical issues in VR groups prevented any students’ performance data from being used. All participants’ survey responses were used for social validity (SV) data.
### Participant Responses

<table>
<thead>
<tr>
<th>How likely is it that you would choose to participate in this same type of sim. again?</th>
<th>Direct Sim. Groups (N = 22)</th>
<th>VR Sim. Groups (N = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely likely = 6</td>
<td>Extremely likely = 10</td>
<td></td>
</tr>
<tr>
<td>Somewhat likely = 7</td>
<td>Somewhat likely = 9</td>
<td></td>
</tr>
<tr>
<td>No preference = 7</td>
<td>No preference = 1</td>
<td></td>
</tr>
<tr>
<td>Somewhat unlikely = 1</td>
<td>Somewhat unlikely = 2</td>
<td></td>
</tr>
<tr>
<td>Extremely unlikely = 1</td>
<td>Extremely unlikely = 1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Would you recommend this type of sim. to other students?</th>
<th>Yes = 18</th>
<th>Yes = 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>No = 4</td>
<td>No = 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Would this type of sim. add value to your education &amp; training if it was used more frequently?</th>
<th>Yes = 19</th>
<th>Yes = 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>No = 3</td>
<td>No = 5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What did you enjoy most about your simulation? (Comments sorted thematically)</th>
<th>Working with a partner/team (15)</th>
<th>Novelty/fun/realism of the VR experience (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practicing effective communication skills (6)</td>
<td>Working with a partner/team (4)</td>
<td></td>
</tr>
<tr>
<td>The activity was quick (1)</td>
<td>N/A (1)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Have you used VR before?</th>
<th>Yes = 9</th>
<th>Yes = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>No = 13</td>
<td>No = 15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>If “yes,” what is the max number of weekly hours you’ve used VR?</th>
<th>1-2 hours = 9</th>
<th>1-2 hours = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>10+ hours = 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>If you have used VR before, what did you use it for?</th>
<th>Video games = 7</th>
<th>Video games = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studying = 1</td>
<td>Movies = 1</td>
<td></td>
</tr>
<tr>
<td>Military training = 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VR participants only: How realistic did the VR environment look?</th>
<th>N/A</th>
<th>Very realistic = 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Somewhat realistic = 11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat unrealistic = 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not at all realistic = 1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VR participants only: How easy was the VR sim. To use?</th>
<th>N/A</th>
<th>Extremely easy = 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Somewhat easy = 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat difficult = 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extremely difficult = 0</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.** Simulation social validity responses. Both VR and Direct simulation participant responses (unless otherwise indicated) are shown.
### VR Simulation Costs

<table>
<thead>
<tr>
<th>Equipment Needed</th>
<th>Cost per Unit</th>
<th>Quantity Needed</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vive Pro Eye</td>
<td>$1,599.00</td>
<td>2</td>
<td>$3,198.00</td>
</tr>
<tr>
<td>XPS Computer Tower</td>
<td>$699.00</td>
<td>3</td>
<td>$2,097.00</td>
</tr>
<tr>
<td>Dell 23 Monitor</td>
<td>$149.99</td>
<td>3</td>
<td>$449.97</td>
</tr>
<tr>
<td>Dell Multi-Service wireless keyboard + mouse</td>
<td>$67.99</td>
<td>3</td>
<td>$203.97</td>
</tr>
<tr>
<td>Lighthouse stand (pack of 2)</td>
<td>$45.95</td>
<td>2</td>
<td>$91.90</td>
</tr>
<tr>
<td>Medical equipment virtual assets</td>
<td>$119.00</td>
<td>1</td>
<td>$119.00</td>
</tr>
<tr>
<td>Basic medical avatar models</td>
<td>$34.00</td>
<td>1</td>
<td>$34.00</td>
</tr>
<tr>
<td><strong>Total VR Sim. Materials Cost (est.)</strong></td>
<td><strong>$6,193.84</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Direct Simulation Costs

<table>
<thead>
<tr>
<th>Equipment Needed</th>
<th>Cost per Unit</th>
<th>Quantity Needed</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Simon (medical mannequin)</td>
<td>$695.00</td>
<td>1</td>
<td>$695.00</td>
</tr>
<tr>
<td>Clinton family practice table</td>
<td>$880.00</td>
<td>1</td>
<td>$880.00</td>
</tr>
<tr>
<td>HK surgical IV pole</td>
<td>$350.00</td>
<td>1</td>
<td>$350.00</td>
</tr>
<tr>
<td>Dome camera</td>
<td>$60.00</td>
<td>2</td>
<td>$60.00</td>
</tr>
<tr>
<td>Overbed table</td>
<td>$49.99</td>
<td>1</td>
<td>$49.99</td>
</tr>
<tr>
<td><strong>Total Direct Sim. Materials Cost (est.)</strong></td>
<td><strong>$2,034.99</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.** Estimated cost analysis. VR simulation costs are listed for the hardware and virtual assets needed to run a simulation similar to that used in the current study. Direct simulation costs are listed for the minimum materials needed to populate an empty room with the basic functional stimuli found in a patient exam room. Labor hours were volunteer-based and therefore not included in these calculations; anecdotally, the labor required to run both simulation types was comparable.
Performance Measurement Categories | Figure | VR | Direct
--- | --- | --- | ---

I. Handoff Provider: Communication Accuracy
1. “Correct” responses (most in pretest) | Fig. 10 | +0 | +0
2. “Correct” responses (most in posttest) | Fig. 10 | +0 | +0
3. “Missing” responses (least in pretest) | Fig. 10 | +0 | +0
4. “Missing” responses (least in posttest) | Fig. 10 | +0 | +0
5. “Erred” responses (least in pretest) | Fig. 10 | +0 | +0
6. “Erred” responses (least in posttest) | Fig. 10 | +1 | +0
7. “Omitted” responses (least in pretest) | Fig. 10 | +0 | +0
8. “Omitted” responses (least in posttest) | Fig. 10 | +0 | +0

Section I Subtotal | 1 | 0

II. Handoff Provider: Duration
1. Handoff duration (shortest in pretest) | Fig. 11 | +1 | +0
2. Handoff duration (shortest in posttest) | Fig. 11 | +1 | +0

Section II Subtotal | 2 | 0

III. Handoff Provider: Rate of Response
1. “Correct” rate (highest in pretest) | Fig. 12 | +0 | +0
2. “Correct” rate (highest in posttest) | Fig. 12 | +1 | +0
3. “Missing” rate (lowest in pretest) | Fig. 12 | +0 | +0
4. “Missing” rate (lowest in posttest) | Fig. 12 | +0 | +1
5. “Erred” rate (lowest in pretest) | Fig. 12 | +0 | +0
6. “Erred” rate (lowest in posttest) | Fig. 12 | +0 | +0

Section III Subtotal | 1 | 1

IV. Handoff Receiver: Synthesis by Receiver
1. “Correct” frequency (highest in pretest) | Fig. 17 | +0 | +1
2. “Correct” frequency (highest in posttest) | Fig. 17 | +0 | +1
3. “Missing” frequency (lowest in pretest) | Fig. 17 | +0 | +0
4. “Missing” frequency (lowest in posttest) | Fig. 17 | +0 | +0
5. “Erred” frequency (lowest in pretest) | Fig. 17 | +0 | +0
6. “Erred” frequency (lowest in posttest) | Fig. 17 | +1 | +0

Section V Subtotal | 1 | 2

V. Situational Awareness: Level 2 (Comprehension)
1. Average score (highest in pretest) | Fig. 29 | +0 | +1
2. Average score (highest in posttest) | Fig. 29 | +0 | +1

Section VII Subtotal | 0 | 2

VI. Situational Awareness: Level 3 (Prediction)
1. Highest avg. appropriate actions (pretest) | Fig. 30 | +0 | +0
2. Highest avg. appropriate actions (posttest) | Fig. 30 | +0 | +0
3. Highest avg. team SA actions (pretest) | Fig. 30 | +0 | +1
4. Highest avg. team SA actions (posttest) | Fig. 30 | +0 | +0
5. Lowest avg. inappropriate actions (pretest) | Fix. 30 | +0 | +0
6. Lowest avg. inappropriate actions (posttest) | Fig. 30 | +1 | +0

Section VIII Subtotal | 1 | 1

Total Calculations | VR | Direct
--- | --- | ---
Performance comparisons w/ statistically significant differences | 6 | 6
Performance comparisons w/ no statistically significant differences | 18 | 18

Table 6. Statistically significant differences in performance across simulations.
**Table 7.** Summary of statistical comparisons. The means, standard deviations, t-values, and p-values are shown for all thirty t-tests used for statistical comparison between simulation groups’ performance.

<table>
<thead>
<tr>
<th>Performance Comparisons</th>
<th>VR Simulations</th>
<th></th>
<th>Direct Simulations</th>
<th></th>
<th>t Value</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Std. Dev.</td>
<td>Average Std. Dev.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. Handoff Provider: Communication Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Correct” responses – pretest</td>
<td>12.78</td>
<td>2.84</td>
<td>13.38</td>
<td>2.16</td>
<td>-0.78063</td>
<td>0.219699</td>
</tr>
<tr>
<td>“Correct” responses – posttest</td>
<td>12.91</td>
<td>2.43</td>
<td>13.05</td>
<td>1.77</td>
<td>-0.20812</td>
<td>0.418072</td>
</tr>
<tr>
<td>“Missing” responses – pretest</td>
<td>1.13</td>
<td>1.46</td>
<td>1.10</td>
<td>1.18</td>
<td>0.08761</td>
<td>0.465303</td>
</tr>
<tr>
<td>“Missing” responses – posttest</td>
<td>1.09</td>
<td>0.99</td>
<td>0.95</td>
<td>0.80</td>
<td>0.01425</td>
<td>0.494351</td>
</tr>
<tr>
<td>“Erred” responses – pretest</td>
<td>0.30</td>
<td>0.47</td>
<td>0.29</td>
<td>0.46</td>
<td>0.13223</td>
<td>0.447717</td>
</tr>
<tr>
<td>“Erred” responses – posttest</td>
<td>0.22</td>
<td>0.42</td>
<td>0.48</td>
<td>0.60</td>
<td>-1.55035</td>
<td>0.064372</td>
</tr>
<tr>
<td>“Omitted” responses – pretest</td>
<td>2.74</td>
<td>2.16</td>
<td>2.19</td>
<td>1.97</td>
<td>0.87888</td>
<td>0.192234</td>
</tr>
<tr>
<td>“Omitted” responses – posttest</td>
<td>2.30</td>
<td>1.72</td>
<td>2.05</td>
<td>1.53</td>
<td>-0.74725</td>
<td>0.229589</td>
</tr>
<tr>
<td>II. Handoff Provider: Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handoff duration – pretest</td>
<td>1.77</td>
<td>0.62</td>
<td>2.05</td>
<td>0.73</td>
<td>-1.36782</td>
<td>0.089409</td>
</tr>
<tr>
<td>Handoff duration – posttest</td>
<td>1.60</td>
<td>0.50</td>
<td>2.18</td>
<td>0.60</td>
<td>-3.47465</td>
<td>0.000611</td>
</tr>
<tr>
<td>III. Handoff Provider: Rate of Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Correct” response rate – pretest</td>
<td>8.01</td>
<td>2.97</td>
<td>7.22</td>
<td>2.28</td>
<td>0.97086</td>
<td>0.168656</td>
</tr>
<tr>
<td>“Correct” response rate – posttest</td>
<td>8.71</td>
<td>2.63</td>
<td>6.35</td>
<td>1.72</td>
<td>3.41654</td>
<td>0.000722</td>
</tr>
<tr>
<td>“Missing” response rate – pretest</td>
<td>0.61</td>
<td>0.76</td>
<td>0.52</td>
<td>0.66</td>
<td>0.38456</td>
<td>0.351274</td>
</tr>
<tr>
<td>“Missing” response rate – posttest</td>
<td>0.72</td>
<td>0.64</td>
<td>0.44</td>
<td>0.37</td>
<td>1.70342</td>
<td>0.048029</td>
</tr>
<tr>
<td>“Erred” response rate – pretest</td>
<td>0.17</td>
<td>0.27</td>
<td>0.17</td>
<td>0.28</td>
<td>-0.09581</td>
<td>0.462069</td>
</tr>
<tr>
<td>“Erred” response rate – posttest</td>
<td>0.13</td>
<td>0.25</td>
<td>0.28</td>
<td>0.41</td>
<td>-1.54572</td>
<td>0.064929</td>
</tr>
<tr>
<td>IV. Handoff Receiver: Synthesis by Receiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Correct” frequency – pretest</td>
<td>1.70</td>
<td>0.82</td>
<td>3.62</td>
<td>2.14</td>
<td>-2.66796</td>
<td>0.007198</td>
</tr>
<tr>
<td>“Correct” frequency – posttest</td>
<td>2.62</td>
<td>1.85</td>
<td>5.05</td>
<td>3.04</td>
<td>-2.59322</td>
<td>0.007110</td>
</tr>
<tr>
<td>“Missing” frequency – pretest</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.60</td>
<td>-1.21136</td>
<td>0.119607</td>
</tr>
<tr>
<td>“Missing” frequency – posttest</td>
<td>0.31</td>
<td>0.48</td>
<td>0.48</td>
<td>0.68</td>
<td>-0.77945</td>
<td>0.220720</td>
</tr>
<tr>
<td>“Erred” frequency – pretest</td>
<td>0.10</td>
<td>0.32</td>
<td>0.08</td>
<td>0.28</td>
<td>0.18621</td>
<td>0.427035</td>
</tr>
<tr>
<td>“Erred” frequency – posttest</td>
<td>0.00</td>
<td>0.00</td>
<td>0.43</td>
<td>0.75</td>
<td>-0.250798</td>
<td>0.023907</td>
</tr>
<tr>
<td>V. Situational Awareness: Level 2 (Comprehension)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score – pretest</td>
<td>40.70</td>
<td>17.74</td>
<td>50.43</td>
<td>17.45</td>
<td>-1.83179</td>
<td>0.037040</td>
</tr>
<tr>
<td>Score – posttest</td>
<td>62.22</td>
<td>13.16</td>
<td>79.23</td>
<td>12.44</td>
<td>-2.88809</td>
<td>0.003081</td>
</tr>
<tr>
<td>VI. Situational Awareness: Level 3 (Prediction)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate actions – pretest</td>
<td>1.17</td>
<td>0.72</td>
<td>1.55</td>
<td>0.96</td>
<td>-1.47308</td>
<td>0.074007</td>
</tr>
<tr>
<td>Appropriate actions – posttest</td>
<td>2.65</td>
<td>1.47</td>
<td>3.05</td>
<td>1.56</td>
<td>-0.07444</td>
<td>0.470504</td>
</tr>
<tr>
<td>Team SA actions – pretest</td>
<td>0.39</td>
<td>0.58</td>
<td>0.18</td>
<td>0.39</td>
<td>1.40486</td>
<td>0.083622</td>
</tr>
<tr>
<td>Team SA actions – posttest</td>
<td>1.22</td>
<td>0.95</td>
<td>1.00</td>
<td>0.98</td>
<td>0.03506</td>
<td>0.486096</td>
</tr>
<tr>
<td>Inappropriate actions – pretest</td>
<td>0.13</td>
<td>0.34</td>
<td>0.09</td>
<td>0.29</td>
<td>0.41308</td>
<td>0.340800</td>
</tr>
<tr>
<td>Inappropriate actions – posttest</td>
<td>0.22</td>
<td>0.42</td>
<td>0.50</td>
<td>0.91</td>
<td>-1.82019</td>
<td>0.037848</td>
</tr>
</tbody>
</table>
Figure 1. TeamSTEPPS® model. The center of the model shows the four core skills or competencies that this iteration of CRM emphasizes: communication, leadership, situation monitoring, and mutual support. The core competencies are surrounded by the patient care team, made up of all primary and secondary healthcare providers (including the patient themselves). Each of the three points of the triangle highlights a desirable outcome (improvements in performance, knowledge, and attitudes) that can be achieved by the patient care team demonstrating the core competencies.
Figure 2. Participant communication accuracy. Data are taken from Anbro, et al. (2020). This figure shows participants' coded verbal responses across ten verbal checkback opportunities in pre-test & post-test. Black bars indicate the number of participant responses meeting "correct" criteria, white bars indicate the number of participant responses meeting "missed," "erred," or "omitted" criteria. Pre-test to post-test differences of correct responses are shown above the post-test bars.
Figure 3. Situational awareness pre-test vs. post-test measures. Data are taken from Anbro, et al. (2020). This scatterplot shows pre-test to post-test changes in 1) participant communication accuracy and 2) participant eye fixation duration on speakers while they are speaking. Descriptions are provided in italics for each quadrant.
Figure 4. Cumulative simulation-based research in behavior science. This figure shows the results of Ward & Houmanfar's original review (2011) of simulation-based research published in behavior analytic journals, shown to the left of the phase-change line. Anbro et al. (2020) extended this review through 2017; these data are shown to the right of the phase-change line.
Figure 5. History of virtual reality. This timeline shows the development of this technology from early conceptualization in sci-fi writings to modern day immersive hardware and software. Image retrieved from https://learn.g2.com/history-of-virtual-reality
Figure 6. I-PASS handoff tool. The colored alphanumeric annotations are added to illustrate the coding criteria used for scoring handoff accuracy. Specific criteria for each sub-section (e.g., 2a-2c: “summary statement”) vary across fictional patient cases used. All scoring criteria are shown in Appendix A.
Figure 7. Virtual simulation AOIs. The top image shows an overhead view of the virtual simulation used in the current study. Areas of interest (AOIs) are designated in red text within the image. The black bars protruding from the avatars' faces indicate where the participants are currently looking. The black bar must make contact with an AOI (more precisely defined within the simulation than in this figure) in order to register an eye fixation. Participants do not see these bars when they are in the simulation. The bottom image shows a first-person POV as the participant refers to their notes during a patient handoff.
Figure 8. Study methodology. This figure shows the primary phases of the study: Phase I) Pre-intervention assessment (“Pretest”); Phase II) Intervention; Phase III) Post-intervention assessment (“Posttest”). Phase IV was provided for all participants, though was not technically part of the present study. Lectures from Phase II were recorded and provided to students who attended the opposing lecture; TeamSTEPPS® students were given the ACE lecture, and ACE students were given the TeamSTEPPS® lecture.
Figure 9. Simulation process. This figure shows the three parts of the simulation process: Part 1) briefing, part 2) simulated handoffs, and part 3) debriefing.
Figure 10. Handoff provider performance during simulated patient handoffs. *Erred: VR participants made significantly fewer errors in posttest (p < 0.10)
Figure 11. Handoff durations. Average handoff time in VR simulations decreased by 10s from pretest (range: 1.1 – 3.9 min.) to posttest (range: 1.1 – 2.7 min.) Average handoff time in direct simulations increased by 8s from pretest (range: 1.1 – 3.9 min.) to posttest (1.1 – 3.1 min.)

*Pretest: Handoffs were significantly shorter in VR than Direct (p < 0.10)
***Posttest: Handoffs were significantly shorter in VR than Direct (p < 0.001)
Figure 12. Handoff provider response rates by simulation type.

*Posttest: Direct participants made “missing” responses at a lower rate (p < 0.10)

***Posttest: VR participants made “correct” responses at a higher rate (p < 0.001)
Figure 13. Pretest to posttest response rate change. Increases in “correct” response rate are an indication of greater communication accuracy. Decreases in “missing” and “erred” response rates also indicate greater communication accuracy.
Figure 14. Handoff provider response rates by training workshop attended. *Pretest: ACE participants demonstrated higher rates of “correct” responses and lower rates of “erred” responses (p < 0.10).
<table>
<thead>
<tr>
<th>Simulated Patient Handoff: Pretest</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant: 201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Multiplier</td>
<td>Score</td>
</tr>
<tr>
<td>“Correct” Comments</td>
<td>8</td>
<td>+1.00</td>
</tr>
<tr>
<td>“Missing” Comments</td>
<td>4</td>
<td>-0.25</td>
</tr>
<tr>
<td>“Omitted” Comments</td>
<td>6</td>
<td>-0.50</td>
</tr>
<tr>
<td>“Erred” Comments</td>
<td>0</td>
<td>-1.00</td>
</tr>
<tr>
<td>Total Possible Score</td>
<td>18</td>
<td>Raw Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Score</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulated Patient Handoff: Posttest</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant: 201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Multiplier</td>
<td>Score</td>
</tr>
<tr>
<td>“Correct” Comments</td>
<td>13</td>
<td>+1.00</td>
</tr>
<tr>
<td>“Missing” Comments</td>
<td>1</td>
<td>-0.25</td>
</tr>
<tr>
<td>“Omitted” Comments</td>
<td>2</td>
<td>-0.50</td>
</tr>
<tr>
<td>“Erred” Comments</td>
<td>0</td>
<td>-1.00</td>
</tr>
<tr>
<td>Total Possible Score</td>
<td>16</td>
<td>Raw Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Score</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pretest Score</th>
<th>Posttest Score</th>
<th>Change Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>22%</td>
<td>73%</td>
<td>+51%</td>
</tr>
</tbody>
</table>

**Figure 15.** Patient handoff scoring. This figure shows a sample of one participant’s data from the VR group. Change scores are calculated by subtracting pretest scores from posttest scores. Positive change scores indicate improved communication accuracy from pretest to posttest. Negative change scores indicate worsened communication accuracy from pretest to posttest.
Figure 16. Participant handoff change scores. Scores show the change from pretest to posttest communication accuracy, using the patient handoff scoring shown in Figure 15. The top graph shows both pretest and posttest accuracy scores for each participant, while the bottom graph shows only the pretest to posttest change for each participant.
Figure 17. Handoff receiver performance during simulated patient handoffs.

*Erred responses: Less “erred” synthesis in posttest VR simulations (p < 0.10)
**Correct responses: More “correct” synthesis in Direct simulations (p < 0.01)
Figure 18. “Filler” words during simulations. The rate of *filler words* (“um” and “uh”) are shown for participants during their pretest and posttest simulations. VR simulation participants are shown on the left and Direct simulation participants are shown on the right.
Figure 19. Communication accuracy in sample dyads. A dyad exhibiting exemplary performance is shown in the top graph, and a typically performing dyad is shown in the bottom graph.
Low Eye Fixation Rate (Slope < 0.50) | Medium Eye Fixation Rate (Slope = 0.50 – 0.99) | High Eye Fixation Rate (Slope ≥ 1.0)
---|---|---
Stable | “Inattentive” (see Figure 21) | “Attentive” (see Figure 22) | “Focused” (see Figure 23)
Variable | “Distracted” (see Figure 24) | “Divided” (see Figure 25) | “Observant” (see Figure 26)

**Figure 20.** Situational awareness level 1 pattern classifications. Visual patterns of eye fixations during handoff receiver simulation segments were classified based on two criteria: 1) the rate of eye fixations for the AOI with the highest fixation count, and 2) that AOI’s pattern variability with other, simultaneous AOI patterns.
Figure 21. Inattentive eye fixation pattern sample. During pretest, participant 106's eye fixation rate was low (slope < 0.5) for all AOIs during their handoff receiver segment. Their highest-slope AOI is “Partner Avatar”, which is stable during the handoff receiver segment.
Figure 22. *Attentive* eye fixation pattern sample. During posttest, participant 112’s eye fixation rate was medium (slope = 0.5 – 0.99) for at least one AOI during their *handoff receiver* segment. Their highest-slope AOI is “Partner Avatar”, which is stable during the *handoff receiver* segment.
Figure 23. Focused eye fixation pattern sample. During pretest, participant 101’s eye fixation rate was high (slope $\geq 1.0$) for at least one AOI during their handoff receiver segment. Their highest-slope AOI is “Patient”, which is stable during the handoff receiver segment.
Figure 24. *Distracted* eye fixation pattern sample. During pretest, participant 102’s eye fixation rate was low (slope < 0.5) for all AOIs during their *handoff receiver* segment. Their highest-slope AOI is “Notes”, which varies with “Partner Avatar” during the *handoff receiver* segment.
Figure 25. *Divided* eye fixation pattern sample. During posttest, participant 111’s eye fixation rate was medium (slope = 0.5 – 0.99) for at least one AOI during their *handoff receiver* segment. Their highest-slope AOI is “Partner Avatar”, which varies with “Notes” during the *handoff receiver* segment.
Figure 26. *Observant* eye fixation pattern sample. During pretest, participant 107’s eye fixation rate was high (slope $\geq 1.0$) for at least one AOI during their *handoff receiver* segment. Their highest-slope AOI is “Monitor”, which varies with “Partner Avatar” during the *handoff receiver* segment.
Figure 27. Eye fixation patterns during VR simulations. The number of students demonstrating each of the six eye fixation patterns is shown for pretest and posttest. Attentive and Distracted were the most frequent patterns during pretest. Attentive and Divided were the most frequent during posttest. One error occurred with eye-tracking data recording during pretest, none in posttest.
Figure 28. Eye fixation patterns in a sample dyad.
**Figure 29.** Situational awareness level 2 scores. Scores are determined via the post-simulation questionnaires. Participants’ correct responses earn points, which total 11 points per fictional patient case. Total correct responses are divided by 11, resulting in the scores shown.

*Pretest: Participants in Direct simulations scored significantly higher than participants in VR simulations (p < 0.10)

**Posttest: Participants in Direct simulations scored significantly higher than participants in VR simulations (p < 0.01)
Figure 30. Situational awareness level 3 response frequencies. Responses were scored into one of three categories: appropriate actions, actions to raise team situational awareness ("team SA"), and inappropriate actions.

*Pretest: Direct sim. participants provided more Team SA responses (p < 0.10)
***Posttest: VR participants provided fewer Inappropriate responses (p < 0.001)
Figure 31. Situational awareness in VR: Levels 1 and 2. Pretest and posttest brief patient assessment scores (L2) are shown for each of the six eye fixation patterns identified (L1).
Figure 32. Situational awareness levels 1 and 2, by stability. These scatterplots show the relationship between eye fixation patterns (L1) and brief patient assessment scores (L2). Pretest data are shown in the top graph, posttest data in the bottom graph. SA-L1 pattern labels described in Figure 20 are indicated.
Figure 33. Situational awareness levels 1 and 2, by AOI fixation. These scatterplots show the relationship between eye fixation patterns (L1) and brief patient assessment scores (L2). Pretest data are shown in the top graph, posttest data in the bottom graph.
Figure 34. Situational awareness in VR: Levels 2 and 3. The top scatterplot shows the relationship between brief patient assessment scores (L2) and the frequency of appropriate & team SA comments during hypothetical case developments (L3). The bottom graph shows the same relationship for Direct simulation participants, as a comparison.
Figure 35. T-TAQ survey data. Average scores are shown for each of the five categories assessed by the T-TAQ (shown in Appendix B).

*statistically significant differences in average rating scores (p < 0.10)
**statistically significant differences in average rating scores (p < 0.01)
# Appendix A
## Handoff Coding: Scoring Criteria

### Pretest

**Case 1 Patient: John Kim**  
**Handoff provider: PA students. Handoff receiver: Nursing students.**

<table>
<thead>
<tr>
<th>ILLNESS SEVERITY</th>
<th>1. Stable condition</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing” Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) “Stable”</td>
<td>Erred: Unstable/Critical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PATIENT SUMMARY</th>
<th>2a. 52 year old male</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) 52</td>
<td>Erred: Wrong age/gender</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2) Male</td>
<td>-Missing: 1 component left out</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2b. No significant medical history</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) No sig. (med.) hist. (“previously healthy” is ok)</td>
<td>Erred: Any novel medical history elements -Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2c. Admitted for new onset atrial fibrillation (a-fib) 4 hours ago</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) Admitted for a-fib (new onset) 2) 4 hours ago (“times 4hr.”)</td>
<td>Erred: Any other reason for admittance (NOTE: this is different than “presented with”; that is covered in items 3a-3c below); wrong time -Missing: Missing diagnosis or time</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3a. 12 hour history of shortness of breath...</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) 12 hour history (this will apply for 3a-3c and can be stated once for all 3, at any point) 2) shortness of breath</td>
<td>Erred: Wrong time duration, lists one wrong symptom not covered in 3a-3c -Missing: Leaves “12 hour hist.” out of 3a-3c altogether</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3b. Lightheadedness...</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) Lightheadedness</td>
<td>Erred: Describes a second wrong symptom not covered in 3a-3c -Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3c. and palpitations</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) Palpitations (or, “arrhythmia”)</td>
<td>Erred: Describes a third wrong symptom not covered in 3a-3c -Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4a. EKG was done in the ER</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) EKG done (in the ER)</td>
<td>Erred: EKG not yet done -Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4b. Patient was a-fib w/ (ventricular)</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>1) Patient was (diagnosed) a-fib</td>
<td>Erred: Wrong diagnosis or heart rate range (e.g., 100s, 120s, “elevated”)</td>
<td></td>
</tr>
<tr>
<td>Heart rate in the 130s</td>
<td>2) Heart rate in 130s (just “130” is ok)</td>
<td>-Missing: Missing diagnosis or heart rate range</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>4c. Patient is currently in normal sinus rhythm w/ rate of 84</td>
<td>1) Normal (sinus) rhythm 2) Rate of 84</td>
<td>-Erred: Abnormal rhythm, wrong rate -Missing: Missing “normal” rhythm or rate</td>
<td></td>
</tr>
<tr>
<td>5a. Watch (e.g., “concerned,” “keep an eye on,” “if patient converts,” etc.) for conversion back to a-fib</td>
<td>1) Any statement indicating concern about converting or to watch for it (“patient is on a heart monitor” is ok too)</td>
<td>-Erred: Any statement indicating no concern that patient may convert back to a-fib -Missing: N/A</td>
<td></td>
</tr>
<tr>
<td>6a. (IV) Cardizem has been given (in the ER)</td>
<td>1) Cardizem given</td>
<td>-Erred: Wrong medication -Missing: N/A</td>
<td></td>
</tr>
</tbody>
</table>

**ACTION LIST**

| 7. Provide teaching/education to patient regarding diagnosis (a-fib) | 1) Provide teaching/education 2) Regarding diagnosis (a-fib) | -Erred: Wrong diagnosis, “inform” about diagnosis (e.g., not “teach”) -Missing: “Provide teaching/education.”...about what? |

| 8. Give Lovenox as ordered | 1) Give/start Lovenox | -Erred: Wrong medication -Missing: N/A |

**SITUATIONAL AWARENESS AND CONTINGENCY PLANNING**

| 9. If patient converts back to a-fib, notify cardiology & prep for synchronized cardioversion | 1) If patient converts (back to a-fib) 2) notify cardiology 3) prep for (synchronized) cardioversion | -Erred: No “if” statement (e.g., “if patient converts back to a-fib...”) -Missing: Missing the notification of cardiology, prep for cardioversion, or both |

| 10. If there are any changes in rhythm (“sinus,” “heart rate/monitor”), obtain another EKG | 1) If there are changes in rhythm (PRN – “as needed”) 2) obtain (another) EKG | -Erred: Any action other than obtaining EKG if rhythm changes -Missing: “Obtain EKG” without criteria (changes in rhythm) |

**SYNTHESIS BY RECEIVER**

General instructions: See the above criteria to score each “synthesis by receiver” interaction as C/E/M/O.
## Pretest
### Case 2 Patient: Kate Delahantey
Handoff provider: Nursing students. Handoff receiver: PA students

### ILLNESS SEVERITY

<table>
<thead>
<tr>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing” Guidelines</th>
</tr>
</thead>
</table>
| 1. Stable condition | 1) “Stable” | -Erred: Unstable/Critical  
- Missing: N/A |

### PATIENT SUMMARY

<table>
<thead>
<tr>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
</table>
| 2a. 34 year old female | 1) 34  
2) Female | -Erred: Wrong age/gender  
- Missing: 1 component left out |
| 2b. Admitted today for an emergency cholecystectomy | 1) “Emergency”  
2) Cholecystectomy | -Erred: Wrong procedure  
- Missing: left out “emergency” |
| 2c. Procedure was to treat her acute cholecystitis | 1) “Acute”  
2) Cholecystitis | -Erred: “Chronic” instead of “acute”  
- Missing: Left out “acute” |
| 3a. She presented to the ER with 9/10 RUQ (right upper quadrant) pain and nausea | 1) 9/10 pain  
2) Pain in RUQ  
3) Nausea | -Erred: Wrong pain rating number (e.g., 8/10) or wrong location stated  
- Missing: pain rating, pain location, or nausea are not stated |
| 4a. Surgery was consulted and a laparoscopic cholecystectomy (“lap. chole.”) was performed | 1) “Laparoscopic”  
2) Cholecystectomy | -Erred: “Open” instead of “laparoscopic,” or wrong procedure instead of cholecystectomy  
- Missing: Left out “laparoscopic” |
| 4b. She has been out of surgery and on the floor now for 4 hours. | 1) Out of surgery (“on the floor;” “post-op”)  
2) 4 hours | -Erred: Wrong amount of time or states patient is in surgery  
- Missing: Left out time |
| 4c. She is complaining of nausea, has had one episode of emesis, and required a straight catheter (“cath.”) due to inability to void. | 1) Nausea  
2) One episode...  
3) Emesis (vomiting)  
4) Required a straight cath...  
5) Inability to void (retaining urine) | -Erred: Wrong number of emesis episodes, wrong type of cath. (e.g., Foley cath.)  
- Missing: Left out “nausea,” didn’t state # of episodes (or left out “emesis” entirely), left out the need for a cath. and/or why |
| 4d. She is currently on a clear liquid diet. | 1) Clear liquid diet | -Erred: Regular diet, “clear diet” (doesn’t specify liquids)  
- Missing: N/A |
| 5a. Ongoing pain assessment | 1) Pain assessment (“watch/monitor for pain”) | -Erred: No ongoing assessment needed  
- Missing: N/A |
| 5b. Ongoing post-op assessment | 1) Post-op assessment (“watch for...” e.g., nausea) | -Erred: No ongoing assessment needed  
- Missing: N/A |
| 6a. Patient is receiving 1-2 Percocet as needed (PRN) for pain | 1) 1-2 pills 2) Percocet | - Erred: Wrong amount and/or wrong medication  
-Missing: Left out amount or type of med |
| 6b. Meds given no more often than every 4 hours ("q4"). | 1) Every 4 hours (q4) | - Erred: Wrong time duration (e.g., every 2 hrs.)  
-Missing: N/A |
| 6c. IV antibiotics and normal saline (NS) | 1) IV antibiotics 2) IV saline (NS) | - Erred: Left out "IV" for either antibiotics, saline, or both  
-Missing: Left out antibiotics, saline, or both |

**ACTION LIST**

| 7. Order for IV anti-emetics to control nausea and vomiting | 1) IV anti-emetics | - Erred: Wrong method of delivery and/or wrong type of medication; left out “IV” (one “IV” can cover 7 & 8 but if they leave out IV, it’s erred)  
-Missing: "IV meds" |
| 8. Order for IV pain medication since she is throwing up | 1) IV pain medication | - Erred: Wrong method of delivery and/or wrong type of medication; left out “IV” (one “IV” can cover 7 & 8 but if they leave out IV, it’s erred)  
-Missing: "IV meds" |

**SITUATIONAL AWARENESS AND CONTINGENCY PLANNING**

| 9. Order for Foley catheter (“Foley cath.,” “Foley”) if the patient does not void in the next 6 hours | 1) Foley cath. 2) If patient does not void 3) Time: next 6 hrs. | - Erred: Wrong type of cath. (or does not specify), leaves out requirement for patient not voiding on their own, and/or wrong time frame  
-Missing: Leaves out order for Foley cath. and/or leaves out time requirement. |
| 10. Once nausea is well controlled, advance diet as she tolerates it ("P.O." – by mouth) | 1) Advance diet (solid food/PO) as tolerated (i.e., vomiting and nausea are well controlled) | - Erred: Remain on liquid diet  
-Missing: N/A |

**SYNTHESIS BY RECEIVER**

General instructions: See the above criteria to score each “synthesis by receiver” interaction as C/E/M/O.
### ILLNESS SEVERITY

<table>
<thead>
<tr>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stable condition</td>
<td>1) “Stable”</td>
<td>Erred: Unstable/Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing: N/A</td>
</tr>
</tbody>
</table>

### PATIENT SUMMARY

<table>
<thead>
<tr>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a. 53 year old female</td>
<td>1) 53 2) Female</td>
<td>Erred: Wrong age/gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: 1 component left out</td>
</tr>
<tr>
<td>2b. History of right MCL &amp; right ACL repair</td>
<td>1) Right MCL repair 2) Right ACL repair (can also say “right MCL &amp; ACL repair”)</td>
<td>Erred: “Left” or unspecified side for MCL/ACL repair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: Left out MCL or ACL repair</td>
</tr>
<tr>
<td>2c. Admitted for a right total knee replacement</td>
<td>1) Right 2) Total knee replacement</td>
<td>Erred: “Left,” “knee surgery” or another non-specific description of procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: Did not specify “right” (2b’s “right” does not count for 2c too)</td>
</tr>
<tr>
<td>3a. Patient was evaluated on an outpatient basis</td>
<td>1) Outpatient evaluation before surgery</td>
<td>Erred: Evaluated today/inpatient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: N/A</td>
</tr>
<tr>
<td>4a. Admitted for her surgery this morning</td>
<td>1) Admitted this morning</td>
<td>Erred: Gives another admitting time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: N/A</td>
</tr>
<tr>
<td>4b. Patient will be staying overnight for observation</td>
<td>1) Staying overnight 2) for observation</td>
<td>Erred: Discharge today</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: Leaves out “overnight” or the reason for staying (for observation)</td>
</tr>
<tr>
<td>4c. Patient has been able to ambulate 3x today</td>
<td>1) Patient has ambulated/been able to ambulate (today/post-op) 2) 3 times</td>
<td>Erred: States patient has not been able to ambulate, or states the wrong number of times</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: Leaves out the number of times the patient has ambulated post-op</td>
</tr>
<tr>
<td>5a. Ongoing pain/nausea assessment</td>
<td>1) Tolerating post-op recovery well / no complications</td>
<td>Erred: Not tolerating recovery well/complications present</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: N/A</td>
</tr>
<tr>
<td>6a. Receiving IV morphine, 2-5mg</td>
<td>1) IV 2) Morphine 3) 2-5mg.</td>
<td>Erred: Wrong method of distribution (e.g., “oral”), wrong drug, and/or wrong amount</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: Left out method of dist., what drug was given, and/or amount given.</td>
</tr>
<tr>
<td>6b. The morphine is every 4 hours</td>
<td>1) Every 4 hours (“q4 hours,” “q4”)</td>
<td>Erred: Wrong time period (e.g., q3 hours)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Missing: N/A</td>
</tr>
<tr>
<td>6c. The morphine is controlling pain well</td>
<td>1) “Well controlled,” etc.</td>
<td>-Erred: Pain is not well controlled on IV morphine -Missing: N/A</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------------------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>6d. Patient is not currently complaining of nausea</td>
<td>1) No nausea/no complaining of nausea</td>
<td>-Erred: Patient is experiencing nausea -Missing: N/A</td>
</tr>
</tbody>
</table>

### ACTION LIST

<table>
<thead>
<tr>
<th>7. Set up outpatient physical therapy (PT)</th>
<th>1) Outpatient 2) PT</th>
<th>-Erred: “Inpatient” -Missing: Missing “outpatient”</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Encourage patient to transition from IV to oral (PO) pain meds (Norco)</td>
<td>1) Transition from IV to PO pain meds (Norco)</td>
<td>-Erred: Wrong med stated -Missing: N/A</td>
</tr>
</tbody>
</table>

### SITUATIONAL AWARENESS AND CONTINGENCY PLANNING

<table>
<thead>
<tr>
<th>9. If the patient begins to develop any nausea, they can start oral (PO) Zofran</th>
<th>1) If nausea develops 2) Start PO Zofran</th>
<th>-Erred: Leaves out the “if” criteria, wrong med, and/or wrong/unspecified method of administration (e.g., IV Zofran or just Zofran) -Missing: N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. If patient has a bowel movement and transitions to PO pain med, she can be discharged</td>
<td>1) Patient has BM 2) Patient transitions to PO pain med (Norco) 3) Patient can be discharged</td>
<td>-Erred: “Patient can be discharged” without specifying BOTH BM and med transition -Missing: “Patient can be discharged,” while missing either BM OR med transition</td>
</tr>
</tbody>
</table>

### SYNTHESIS BY RECEIVER

General instructions: See the above criteria to score each “synthesis by receiver” interaction as C/E/M/O.
<table>
<thead>
<tr>
<th>ILLNESS SEVERITY</th>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stable condition</td>
<td>1) “Stable”</td>
<td>Erred: Unstable/Critical</td>
<td>Missing: N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PATIENT SUMMARY</th>
<th>Description</th>
<th>“Correct” criteria</th>
<th>“Erred” vs. “Missing”</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a. 66 year old male</td>
<td>1) 66  2) Male</td>
<td>-Erred: Wrong age/gender</td>
<td>-Missing: 1 component left out</td>
</tr>
<tr>
<td>2b. History of hypertension and smoking</td>
<td>1) History of hypertension  2) History of smoking</td>
<td>-Erred: No significant medical history</td>
<td>-Missing: Leaves out either history of hypertension or smoking</td>
</tr>
<tr>
<td>2c. Admitted for chest pain (“unstable angina” is also acceptable)</td>
<td>1) Chest pain/unstable angina</td>
<td>-Erred: Any other reason for admittance</td>
<td>-Missing: N/A</td>
</tr>
<tr>
<td>3a. Patient admitted yesterday</td>
<td>1) Admitted / presented yesterday</td>
<td>-Erred: Any wrong time (e.g., this afternoon)</td>
<td>-Missing: N/A</td>
</tr>
<tr>
<td>3b. 2 hours of chest pain</td>
<td>1) 2 hours (&quot;chest pain&quot; already covered in 2c)</td>
<td>-Erred: Another duration (e.g., 3 hours)</td>
<td>-Missing: N/A</td>
</tr>
<tr>
<td>3c. Chest pain was unrelieved by antacid use</td>
<td>1) (pain) unrelieved by antacid use</td>
<td>-Erred: Pain was relieved by antacid use</td>
<td>-Missing: N/A</td>
</tr>
<tr>
<td>4a. initial EKG showed no ST elevation or depression</td>
<td>1) EKG  2) No ST elevation  3) No ST depression (&quot;no abnormalities&quot; ok for no elevation or depression)</td>
<td>-Erred: left out that the results are EKG results, states there is either ST elevation or depression</td>
<td>-Missing: Left out no ST elevation, no ST depression, or both</td>
</tr>
<tr>
<td>4b. Chest pain was relieved by sublingual nitroglycerin (nitro)</td>
<td>1) Chest pain relieved  2) Sublingual  3) Nitro(glycerin)</td>
<td>-Erred: Chest pain not relieved, wrong medication or route of delivery (e.g., IV instead of sublingual)</td>
<td>-Missing: Given meds without stating they relieved pain, left out method (sublingual) and/or specific med (nitro)</td>
</tr>
<tr>
<td>4c. Cardiac enzymes are normal</td>
<td>1) Cardiac enzymes (markers) are normal (not elevated)</td>
<td>-Erred: Enzymes are abnormal, negative</td>
<td>-Missing: N/A</td>
</tr>
<tr>
<td>5a. Alcohol use assessment: Patient revealed he drinks 4-</td>
<td>1) 4-6 beers per day</td>
<td>-Erred: Wrong amount/range</td>
<td>-Missing: “Consumes alcohol” (does not specify amount).</td>
</tr>
</tbody>
</table>
| 6 beers per day (lied previously) | | | 1) Stress test scheduled  
2) This afternoon (later today) | -Erred: Test not yet scheduled, wrong time for test is specified (e.g., tomorrow, this evening) 
-Missing: Time of stress test not specified |

**ACTION LIST**

| 7. Make patient NPO (no food) for stress test | 1) Make patient NPO | -Erred: Patient can maintain normal diet 
-Missing: N/A |
| 8. Patient is requesting a nicotine patch | 1) Nicotine patch order request | -Erred: States patient is requesting something besides a nicotine patch 
-Missing: N/A |

**SITUATIONAL AWARENESS AND CONTINGENCY PLANNING**

| 9. No current order for a follow-up EKG; ask for EKG in case of any repeat chest pain | 1) Repeat/another EKG  
2) If chest pain returns / for repeat chest pain | -Erred: EKG no matter what (doesn’t specify “if chest pain returns”) 
-Missing: N/A |
| 10. Patient is at risk for alcohol withdrawal; alcohol withdrawal protocol ready as needed | 1) Alcohol withdrawal protocol  
2) As needed (PRN) | -Erred: States that no withdrawal protocol will be needed 
-Missing: Left out “as needed” or some variant of that condition |

**SYNTHESIS BY RECEIVER**

General instructions: See the above criteria to score each “synthesis by receiver” interaction as C/E/M/O.
Appendix B
TeamSTEPPS® Teamwork Attitudes Questionnaire (T-TAQ)

Please respond to the questions below by placing a check mark in the box that corresponds to your level of agreement from Strongly Disagree to Strongly Agree. Please select only one response for each question.

NOTE: Response option boxes are not shown below, but include: Strongly Disagree, Disagree, Neutral, Agree, & Strongly Agree.

TEAM STRUCTURE
1. It is important to ask patients and their families for feedback regarding patient care.
2. Patients are a critical component of the care team.
3. This facility’s administration influences the success of direct care teams.
4. A team’s mission is of greater value than the goals of individual team members.
5. Effective team members can anticipate the needs of other team members.

LEADERSHIP
6. High performing teams in healthcare share common characteristics with high performing teams in other industries.
7. It is important for leaders to share information with team members.
8. Leaders should create informal opportunities for team members to share information.
9. Effective leaders view honest mistakes as meaningful learning opportunities.
10. It is a leader’s responsibility to model appropriate team behavior.
11. It is important for leaders to take time to discuss with their team members plans for each patient.
12. Team leaders should ensure that team members help each other out when necessary.

SITUATION MONITORING
13. Individuals can be taught how to scan the environment for important situational cues.
14. Monitoring patients provides an important contribution to effective team performance.
15. Even individuals who are not part of the direct care team should be encouraged to scan for and report changes in patient status.
16. It is important to monitor the emotional and physical status of other team members.
17. It is appropriate for one team member to offer assistance to another who may be too tired or stressed to perform a task.
18. Team members who monitor their emotional and physical status on the job are more effective.
MUTUAL SUPPORT

19. To be effective, team members should understand the work of their fellow team members.
20. Asking for assistance from a team member is a sign that an individual does not know how to do his/her job.
21. Providing assistance to team members is a sign that an individual does not have enough work to do.
22. Offering to help a fellow team member with his/her individual work tasks is an effective tool for improving team performance.
23. It is appropriate to continue to assert a patient safety concern until you are certain that it has been heard.
24. Personal conflicts between team members do not affect patient safety.

COMMUNICATION

25. Teams that do not communicate effectively significantly increase their risk of committing errors.
26. Poor communication is the most common cause of reported errors.
27. Adverse events may be reduced by maintaining an information exchange with patients and their families.
28. I prefer to work with team members who ask questions about information I provide.
29. It is important to have a standardized method for sharing information when handing off patients.
30. It is nearly impossible to train individuals how to be better communicators.
Appendix C
Pre-Study Screening Criteria

The following questions will ask about your past experiences. Please answer these questions to the best of your ability, as your answers will help us assign you to one of four simulation groups. All responses will be kept anonymous.

1. What is your full name? We will only use this to non-randomly sort you into a non-VR group if you indicate that preference below.

2. In some cases, virtual reality may evoke motion sickness in people who are susceptible. Have you ever experienced motion sickness? Select one:
   a. Yes
   b. Maybe
   c. No

3. If you answered “yes” or “maybe” to the previous question, how severe were your symptoms?
   a. Severe
   b. Moderate
   c. Mild

4. Have you used virtual reality before?
   a. Yes
   b. No

5. If you answered “yes” to the previous question, how would you categorize your virtual reality use at its peak (a time period when you used it the most)? Select one:
   a. 1-2 hrs./wk.
   b. 3-4 hrs./wk.
   c. 5-6 hrs./wk.
   d. 7-8 hrs./wk.
   e. 9-10 hrs./wk.
   f. 10+ hrs./wk.

6. If you have used virtual reality before, describe what you used it for. Did you play video games, browse the internet, watch tv/movies, participate in online chats, etc.?

7. Is there any reason (motion sickness or otherwise) that you do not want to participate in a virtual reality simulation? Note that if you answer “yes” to this question, we will place you in a group that will complete simulations in-person, rather than in virtual reality. You do not need to elaborate on your reasoning. All responses are all confidential.
   a. Yes
   b. No
Appendix D
Fictional Patient Cases for Simulated Handoffs

Case 1. Delivered by PA students during pre-test
Name: John Kim
DOB: 6/13/1968
Code Status: Full

John Kim is a 52 y/o male in stable condition with no significant medical history admitted for new-onset atrial fibrillation 4 hours ago. He presented with 12-hour history of shortness of breath, lightheadedness, and palpitations. An EKG was done in the ER, show that the patient was in A-fib with ventricular rate in the 130’s. In the ER, IV Cardizem has been given. The patient has now converted back to a normal sinus rhythm with a rate of 84. Right now, you want to ask the RN to provide teaching to Mr. Kim regarding his new diagnosis of atrial fibrillation. You also alert the nurse that you have put in a Lovenox order and ask them to begin administering it. As you think ahead for this patient, you have some additional orders for the nurse. You are concerned that Mr. Kim may convert back into atrial fibrillation. You tell nurse that if this happens, they should notify cardiology and prepare for synchronized cardioversion. Additionally, this patient is currently on a heart monitor, so you ask the nurse to please perform an EKG if the patient’s rhythm changes from its current normal sinus rhythm.

Case 2. Delivered by Nursing students during pre-test
Name: Kate Delahantey
DOB: 12/27/1985
Code Status: Full

Ms. Delahantey is a 34 y/o female in stable condition, admitted today for an emergency cholecystectomy to treat her acute cholecystitis. She presented to the ER with 9/10 RUQ pain accompanied by nausea. Surgery was consulted and a laparoscopic cholecystectomy was performed. She has been out of surgery and on the floor now for 4 hours. She is c/o nausea and has had one episode of emesis. She also required a straight cath. due to inability to void. She is currently on a clear liquid diet. She is currently receiving IV antibiotics and IV normal saline. For pain control, she is able to get 1-2 Percocet 5/325 given as needed, no more than every 4 hours. However, you want to ask the doctor to write an order for IV pain medication since she is throwing up. You also want to ask for an order for IV anti-emetics to control the nausea and vomiting. Because this patient has already required straight catheterization for post-op urinary retention, you are concerned that she will require this again, and want to ask the doctor if you can insert a Foley catheter if she does not void in the next 6 hours. Additionally, based on your experience with post-op lap chole patients, you know that once her nausea is well controlled she will likely want to start eating, and so you want to ask the doctor for if you can advance the patient’s diet as she tolerates it.
**Case 3. Delivered by PA students during post-test**
Name: Annie Samproni  
DOB: 9/23/1966  
Code Status: Full

Ms. Samproni is in stable condition. She is a 53 y/o female tennis player with a history of a R MCL repair & R ACL repair admitted for a right total knee replacement. Pt was evaluated on an outpatient basis and admitted for surgery this morning and will be staying the night for observation. For her pain, she is receiving IV morphine 2-5 mg q 4 hours. She reports this is controlling her pain well. She is not complaining of nausea at this time. She has been able to ambulate 3x today. Right now, you want to ask the RN to set up outpatient PT for this patient prior to discharge. Additionally, you have put an order in for PO Norco and want to ask the nurse to help encourage the patient to transition from IV to PO pain medication. As you think ahead for this patient, you have some additional orders for the nurse. You realize that pain medication may cause nausea in some patients, so you plan to tell the nurse that if they patient begins to develop any nausea, they may start PO Zofran. Additionally, you see that Ms. Samproni is tolerating her post-op recovery well. Because of that, you want to tell the nurse that once Ms. Samproni has a bowel movement and has effectively transitioned to PO pain medication, she can be discharged home.

**Case 4. Delivered by Nursing students during post-test**
Name: Dave Johnson  
DOB: 10/3/1953  
Code status: Full

Mr. Johnson is a stable 66 y/o male with a history of HTN and smoking admitted for chest pain. He presented yesterday with 2 hours of chest pain unrelieved by antacid use. Upon admission to the ER, his EKG showed no ST elevation or depression, although his chest pain was relieved with sublingual nitroglycerin. Serial cardiac enzymes are normal. He was diagnosed with unstable angina. He is scheduled to have a stress test later this afternoon. During your morning assessment, Mr. Johnson revealed to you that he was not truthful with the doctor about his alcohol use, and that he actually drinks 4-6 beers a day. Right now, you know that you need to ask the doctor to make this patient NPO so that he is able to have his nuclear medicine stress test later today. Mr. Johnson is also requesting a nicotine patch during his hospital stay, so you make a note to ask the doctor for an order for that as well. After reviewing his orders, you notice that an order was never entered to do an EKG for any repeat chest pain. Because you are aware that this is standard of care for chest pain patients, you plan to ask the doctor for this order. Lastly, you are aware that Mr. Johnson is at a risk for alcohol withdrawal based on his admitted alcohol use. For this reason, you want to ask the doctor if alcohol withdrawal treatment protocol can be implemented if patient begins to show signs of alcohol withdrawal.
Appendix E
Patient Handoff: Comprehension (SA L2) & Projection (SA L3) Questions

**SA Level 2 Questions (answered by all students)**
1. What was the patient’s first name?
2. What condition was the patient in at the time of handoff?
3. What is the reason the patient is in the hospital today?
4. What was done to address this?
5. What medication, if any, was ordered?
6. What action step(s) were discussed for this patient?

**SA Level 3 Questions**

Case 1: John Kim (answered by nursing students)
1. Your patient begins to have lightheadedness, change in mental status, and drooping on the left side of his face. What do you do?
2. Your patient converted back into atrial fibrillation. What do you do first?

Case 2: Kate Delahantey (answered by PA students)
1. Despite starting MSO4 15mg IV q 6 hours, your patient continues to have RUQ pain that is worsening. What do you do?
2. Your patient has good pain control but is still unable to eat due to nausea and vomiting despite being given IV Zofran 4 mg q 6 hours. It is post-op day 2. What do you do first?

Case 3: Annie Samproni (answered by nursing students)
1. Your patient is started on Percocet 5/325mg po q 8 hours. She develops nausea and vomiting after the first dose. What do you do?
2. Your patient lives in an apartment with no elevator and has to walk up 5 flights of stairs to get to her apartment. What can you do to help her?

Case 4: Dave Johnson (answered by PA students)
1. You find the patient outside the hospital smoking with his nicotine patch on. What do you do?
2. Your patient’s chest pain gets worse and repeat EKG shows changes concerning for an acute MI. What do you do?
Appendix F
Simulation Feedback Questions

The following questions pertain to your simulation experience. There are no “correct” answers; all responses will remain anonymous.

NOTE: Response options are not shown below, but include: Strongly Disagree, Disagree, Neutral, Agree, & Strongly Agree.

1. How likely is it that you would choose to participate in this same type of simulation again?

2. Would you recommend this type of simulation to other students?

3. Would this type of simulation add value to your education & training if it was used more frequently?

4. What did you enjoy most about your simulation experience?

VR simulation participants only:

5. How realistic did the VR environment (minus the avatars) look?

6. How easy was the VR simulation to use/navigate?
### Appendix G
Brief Patient Assessment Questionnaire: Scoring Criteria

**Case 1: John Kim**  
Handoff Receivers: Nursing students

<table>
<thead>
<tr>
<th>Comprehension (Recall)</th>
<th>Answer(s) &amp; Point Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What was the patient’s first name?</td>
<td>+1: John</td>
</tr>
<tr>
<td>2. What was the patient's condition at the time of your handoff?</td>
<td>+1: Stable</td>
</tr>
</tbody>
</table>
| 3. What is the reason this patient is in the hospital today? | +1: Shortness of breath  
+1: Lightheadedness  
+1: Palpitations |
| 4. What has been done so far to address this? | +1: EKG was done (in the ER); **and**  
+1: IV Cardizem given (in the ER) |
| 5. What medication (if any) has been ordered? | +1: Lovenox |
| 6. What action steps were discussed for this patient? | +1: Provide teaching for afib diagnosis; **and**  
+1: Notify cardiology & prep for synchronized cardioversion if patient reverts to afib; **and**  
+1: Perform EKG if patient’s rhythm changes from normal sinus rhythm |

Max. Possible Score: 11

**Case 2: Kate Delahantey**  
Handoff Receivers: PA students

<table>
<thead>
<tr>
<th>Comprehension (Recall)</th>
<th>Answer(s) &amp; Point Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What was the patient’s first name?</td>
<td>+1: Kate</td>
</tr>
<tr>
<td>2. What was the patient’s condition at the time of your handoff?</td>
<td>+1: Stable</td>
</tr>
<tr>
<td>3. What is the reason this patient is in the hospital today?</td>
<td>+1: Right upper quadrant pain (9/10)</td>
</tr>
</tbody>
</table>
| 4. What has been done so far to address this? | +1: Cholecystectomy (laparoscopic); **and**  
+1: Straight cath. to allow voiding; **and**  
+1: Placed on a clear liquid diet; **and**  
+1: Given IV antibiotics & saline |
<table>
<thead>
<tr>
<th>Case 3: Annie Samproni</th>
<th>Handoff Receivers: Nursing students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comprehension (Recall) Question</strong></td>
<td><strong>Answer(s) &amp; Point Values</strong></td>
</tr>
<tr>
<td>1. What was the patient’s first name?</td>
<td>+1: Annie</td>
</tr>
<tr>
<td>2. What was the patient’s condition at the time of your handoff?</td>
<td>+1: Stable</td>
</tr>
<tr>
<td>3. What is the reason this patient is in the hospital today?</td>
<td>+1: Right total knee replacement; or +0.5: Total knee replacement</td>
</tr>
<tr>
<td>4. What has been done so far to address this?</td>
<td>+1: Surgery; and +1: Morphine (IV) for pain</td>
</tr>
<tr>
<td>5. What medication (if any) has been ordered?</td>
<td>+1: Norco (PO); and +1: Zofran (PO, PRN)</td>
</tr>
<tr>
<td>6. What action steps were discussed for this patient?</td>
<td>+1: Set up outpatient PT; and +1: Transition IV to PO pain meds; and +1: Zofran given if nausea presents; and +1: Discharge after BM &amp; successful transition to PO pain meds</td>
</tr>
<tr>
<td><strong>Max. Possible Score</strong></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case 4: Dave Johnson</th>
<th>Handoff Receivers: PA students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comprehension (Recall) Question</strong></td>
<td><strong>Answer(s) &amp; Point Values</strong></td>
</tr>
<tr>
<td>1. What was the patient's first name?</td>
<td>+1: Dave (“David” also accepted)</td>
</tr>
<tr>
<td>2. What was the patient’s condition at the time of your handoff?</td>
<td>+1: Stable</td>
</tr>
<tr>
<td>3. What is the reason this patient is in the hospital today?</td>
<td>+1: Chest pain/unstable angina</td>
</tr>
<tr>
<td>4. What has been done so far to address this?</td>
<td>+1: EKG administered; and +1: Nitroglycerin (sublingual)</td>
</tr>
<tr>
<td><strong>Max. Possible Score</strong></td>
<td><strong>11</strong></td>
</tr>
<tr>
<td></td>
<td>What medication (if any) has been ordered?</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------</td>
</tr>
</tbody>
</table>
|   | What action steps were discussed for this patient? | +1: Stress test scheduled; and  
+1: Make patient NPO; and  
+1: Order nicotine patch; and  
+1: Order EKG if repeat chest pain; and  
+1: Order alcohol withdrawal protocol if patient shows symptoms |

Max. Possible Score 11