Spring 2014

The Effects of Serial Position, Evaluation Format, and Behavioral Isolate on Verbal and Nonverbal Clinical Cue Recognition and Performance Ratings

William Robert Turner Jr.
Old Dominion University

THE EFFECTS OF SERIAL POSITION, EVALUATION FORMAT, AND BEHAVIORAL ISOLATE ON VERBAL AND NONVERBAL CLINICAL CUE RECOGNITION AND PERFORMANCE RATINGS

by

Timothy Robert Turner, Jr.
B.A. May 2004, Mississippi State University
M.A. August 2006, Xavier University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

HUMAN FACTORS PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2014

Approved by:

Mark Scerbo (Director)

Ivan Ash (Member)

James Bliss (Member)

Gayle Gliva-McConvey (Member)

ABSTRACT

THE EFFECTS OF SERIAL POSITION, EVALUATION FORMAT, AND BEHAVIORAL ISOLATE ON VERBAL AND NONVERBAL CLINICAL CUE RECOGNITION AND PERFORMANCE RATINGS

Timothy Robert Turner, Jr.
Old Dominion University, 2014
Director: Dr. Mark W. Scerbo

Standardized patients are individuals trained to realistically portray specific physical and psychological symptoms and evaluate healthcare trainees on their patient interaction skills. Prior research suggests that individual differences among standardized patients often result in assessment variance. This study examined the effects of cue serial position and evaluation format on individuals' perceptual awareness and recognition accuracy of verbal and nonverbal clinical cues. It was predicted that implementing periodic evaluations would reduce participant working memory load and permit better awareness and recognition of relevant clinical cues than the traditional post-scenario evaluation format. The concurrent evaluation benefit was also expected to mitigate the well-documented serial position decrement for information occurring in the middle of a scenario. The results suggested that verbal and nonverbal cues appearing early or late in the scenario were generally more salient than those appearing mid-scenario, but observers were better able to recognize both when permitted to offload working memory through periodic evaluation. The study also investigated the impact of a single inconsistent, unprofessional behavior exhibited by the simulated healthcare provider (SHP) on participant ratings of the SHP's clinical competence. The behavioral isolate did not influence participants' overall rating regardless of where it occurred in the scenario. Further, the isolate affected the segmental ratings of both evaluation groups when

embedded early in the scenario and also affected the ratings of the concurrent evaluation group when embedded later in the scenario. This implies a reluctance on the part of retrospective participants to integrate new or conflicting information as the scenario progressed and further suggests that a successful SHP performance is unlikely to be negatively impacted by a single isolated act of unprofessionalism.

Pursuit of an advanced academic degree is a significant, life-altering experience that impacts not only the student but also his or her closest friends and family. Years of sacrifice pave the way to intellectual maturity, leaving an indelible mark that can be appreciated only by those who have witnessed it firsthand. For those who see it to completion, this shared experience will have been a source of strength and solidarity for life. It goes without saying that a number of my closest friends and family shared in these struggles and celebrations right alongside me. They offered encouragement when needed most, congratulations when scarcely justified, and (most importantly) a steadfast faith in my ability to persevere regardless of the circumstance. Without question they each deserve a share of the credit for this success. Although a brief dedication could never do them justice, I'll just have to trust that they each know how grateful I truly am.

This work is dedicated to my wife, Stacey, who has stood beside me throughout these years showing nothing but unwavering support, encouragement, and understanding. She never failed to prop me up when my confidence and motivation faltered—a true sign of her unshakeable faith in me and in the future.

ACKNOWLEDGEMENTS

First and foremost I would like to express most sincere gratitude to my academic mentor and dissertation director, Dr. Mark Scerbo. I cannot imagine having managed this project to completion without Dr. Scerbo's expertise, guidance, and seemingly limitless patience. His relentless dedication to the educational and professional development of his students has always been an inspiration, and I will consider him a friend and mentor deserving of the highest level of admiration for the rest of my life.

I also want to acknowledge the invaluable contributions of my committee members—Dr. Ivan Ash, Dr. Jim Bliss, and Mrs. Gayle Gliva-McConvey. Their thoughtful ideas, comments, and questions strengthened the quality of this work and guided me throughout the process. I also want to thank Dr. Bliss for his exceptional guidance and research mentorship during my early years in the program.

The Sentara Center for Simulation and Immersive Learning at Eastern Virginia Medical School also commands my greatest appreciation. I especially want to thank Dr. Thomas Hubbard, Gayle Gliva-McConvey, Amelia Wallace, Matt Weaks, Tiffany Chilcott, and Sterline Everette for the tremendous enthusiasm and support they held for this research project.

I extend my sincere thanks to CDR V. Andrea Parodi, NC, USN (Ret.), who through years of mentorship and friendship exemplified the consummate professional and provided the opportunity for invaluable field experience in the military healthcare domain.

Finally, acknowledgments would not be complete without mention of the overwhelming support from friends and family. This is especially true of my wife, Stacey, without whom this would neither have been possible nor held meaning.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The State of Modern Healthcare

In a well-publicized review of healthcare quality in the United States, the U.S. Institute of Medicine (IOM) estimated that between 44,000 and 98,000 Americans die each year as a result of preventable medical error (Kohn, Corrigan, & Donaldson, 1999). This rate of preventable loss would be considered unacceptable in other high-risk industries, yet it has now been ostensibly tolerated in healthcare for more than a decade.

In addition to tragic outcomes directly affecting patients and their families, preventable medical error also adds a significant financial burden to the most expensive healthcare system in the world. The annual cost associated with preventable medical error in the United States is estimated to be $17.1 billion (van den Bos et al., 2011). Since its publication, *To Err is Human* has inspired much discussion of systems-based initiatives and training programs designed to address perceived sources of preventable medical error. Despite all the attention that patient safety initiatives have enjoyed in recent years, several follow-up articles suggest that preventable errors leading to the unexpected death of a patient have actually increased since the original IOM report (Jewell & McGiffert, 2009).

In a recent survey of patient safety progress in the American healthcare system, Jewell and McGiffert (2009) conclude that preventable medical error now contributes to more than 100,000 deaths annually. The failure to mitigate preventable error is attributed to characteristics inherent in the culture of modern healthcare. Intended or not, these cultural characteristics often function as significant barriers to patient safety. The healthcare industry has been described as lacking transparency and often treating incident reporting as a strictly punitive rather than corrective process. There is no single national structure for patient safety accountability and no national entity with the authority to enforce best practices. For example, reporting major patient safety failures and sentinel events to an oversight body, while encouraged, is not mandatory. For frontline champions of patient safety initiatives, lack of institutional support as evidenced by scant funding, authority, and accountability has proven to be a significant obstacle (Wachter, 2010).

The Case for Nontechnical Skills Training

As one of the largest national healthcare accreditation entities, The Joint Commission accredits approximately 18,000 healthcare organizations and patient safety programs in the United States. Its primary mission is to promote quality patient care through periodic inspection and evaluation of healthcare organizations, based in large part on the self-reporting of its accredited member organizations. Among the data that The Joint Commission's member organizations are encouraged to report are sentinel events, defined by The Joint Commission as unexpected occurrences involving (actual or risk of) death or serious physical/psychological injury to a patient. For each sentinel event, the offending organization is expected to conduct a "thorough and credible" root cause analysis within 45 days to identify the circumstances under which the event occurred for the purpose of improvement (The Joint Commission, 2011).

The Joint Commission maintains a database of reported sentinel events, including subsequent root cause analyses and data pertaining to event resolution. Based on these data, leadership and communication failures account for the majority of root causes in documented sentinel events. Communication failures account for approximately 70-80%

of Joint Commission-reported sentinel event root causes annually (The Joint Commission, 2010; Salas, Diaz Granados, Weaver, & King, 2008). Further, 55% of offending organizations report that cultural characteristics such as hierarchy and intimidation act as barriers to promoting effective communication among healthcare professionals. Other authors have described a culture of fear in which open communication is devalued, thus contributing to the overall lack of progress in patient safety.

According to the National Patient Safety Foundation (NPSF, 2010), the healthcare industry's culture is wrought with hostility, abuse, intimidation, and professional disrespect. The NPSF states that the current culture actively stifles learning and threatens patient safety by discouraging open communication among all levels of healthcare professionals. It is a culture in which students, residents, nurses, and junior physicians are reluctant to question decisions or seek alternatives when a disagreement occurs; this is intensified by the reinforcing nature of a rigid, hierarchical power structure defined by job role (e.g., surgeon, intern, nurse), rank (i.e., for military personnel), and experience (NPSF, 2010; Shostek, 2007). Those personnel with less experience often fail to speak up or contribute due to intimidation or lack of confidence, regardless of the potential value that their perspectives may contribute (Shostek, 2007; Maxfield et al., 2005).

Dr. Lucian Leape has asserted that disrespect in healthcare increases the potential for mistakes by fostering anger, fear, and self-doubt (Leape, 2012; Leape et al., 2012). Regardless, nearly all healthcare professionals have witnessed or been subjected to disrespectful behavior. More alarming is the fact that most patients have been disrespected by their care providers as well, even if they do not recognize it as such (e.g., being made to wait unnecessarily). Leape went on to say that mutual respect and the fostering of multidisciplinary relationships through communication and teamwork are critical to achieving safety in any industry. Physician-colleague interactions are not the only interpersonal dynamics that impact patient safety. To the contrary, literature examining the physician-patient dynamic establishes links between physicians' interpersonal skills and patient satisfaction, adherence to treatment, clinical outcomes, and tendency to litigate (Duggan & Parrott, 2000; Flocke, Miller, & Crabtree, 2002; Stewart, Meredith, Brown, & Galajda, 2000; Wooford et al., 2004). Epstein et al. (2005) contended that patient-centered communication is a healthcare provider's moral obligation, and is exemplified by helping patients feel understood, attending to the patients' psychosocial context, and facilitating patients' involvement in their own healthcare through education and active decision-making.

Clearly, open communication and active patient involvement are outcomes that should be sought by healthcare professionals and modern educational curricula. All medical schools are now required to include some form of interpersonal skills training and evaluation as part of their medical curricula (Levinson, Lesser, & Epstein, 2010). However, the degree of emphasis placed on interpersonal skills training by medical schools varies (Hulsman, Ros, Winnubst, & Bensing, 1999), and the majority of medical professionals are not held formally accountable for interpersonal conduct once they leave medical school (Aggarwal et al., 2009; Jewell & McGiffert, 2009; Levinson & Roter, 1993).

The Accreditation Council for Graduate Medical Education (ACGME), which accredits approximately 9,000 postgraduate medical training programs across the United

States, has identified interpersonal/communication skills as part of its outcomes assessment project for core medical competencies (Swing, 2007). These skills are exemplified by effective information exchanges with patients, patient families, and other healthcare professionals. Despite the ACGME's integration of nontechnical skills development into medical assessment protocols for interns and residents, practicing physicians are typically still assessed only on technical competency (Aggarwal et al., 2009; Jewell & McGiffert, 2009; Levinson & Roter, 1993). Lack of formal nontechnical skills accountability coupled with skewed perceptions of their own interpersonal skill proficiency (Aggarwal et al.; Undre et al., 2007) may be contributing to the apparent disconnect among healthcare professionals regarding communication efficacy. Indeed, Moorthy et al. (2005; 2006) suggest that the ability of healthcare practitioners to reliably critique their own nontechnical performance is insufficient to support self-regulation and interpersonal skill development.

To date, relatively few nontechnical skills training programs exist for healthcare professionals. Of the programs that have published outcomes data, the majority do not provide enough descriptive information about the specific target skills and training methods to paint a clear picture of success for researchers (Cegala & Broz, 2002). Further, a general disconnect between training objectives and assessment protocols exists, suggesting that many programs place little emphasis on measuring training outcomes (Cegala & Broz, 2002; Salas et al., 2006). As a result, research investigating the efficacy of nontechnical skills training in healthcare is still in its infancy. However, Salas et al. (2008) argue that didactic training and task exposure alone are not enough to result in substantial nontechnical skill development. Rather, the training curriculum must incorporate hands-on learning and timely feedback to ensure interpersonal growth. To this end, simulation-based training (SBT) has been proposed as an ideal tool for enhancing technical and nontechnical skills training for healthcare professionals.

Simulation-Based Training in Healthcare

For centuries the prevailing method for acquiring advanced skill in medical and surgical practice has been the apprentice model, whereby learners observe senior medical staff performing a task or procedure and then attempt to replicate the procedure themselves on live patients (Cavusoglu, Tendick, & Sastry, 2002; Gorman, Meier, Rawn, & Krummel, 2000; Hyltander, Liljegren, Rhodin, & Lonroth, 2002). This model is a realistic yet expensive form of education that carries a significant degree of risk for patients and creates an unnecessarily stressful environment for the learner. As a result, the model has begun to face increasing scrutiny in light of recent technological advances making hands-on practice via simulation a practical alternative for early-stage skill acquisition (Gallagher et al., 2005; Leach & Philibert, 2006; Satava, 2001).

One of the greatest benefits of SBT is the flexibility it affords learners and educators (Epstein, 2007; Haluck et al., 2007). Simulation-based training can target anything from a single skill (e.g., basic manual dexterity) to a series of skills forming a complete procedure (e.g., laparoscopic cholecystectomy). Skills can be developed in a safe, realistic environment with tolerances designed to convert failures into a productive element of learning. It is also capable of presenting a variety of difficult or rare cases that may otherwise be neglected during the learner's normal training regimen (Cavusoglu et al., 2002; Moody, Baber, & Arvanitis, 2002; Wang, Burdet, Vuillemin, & Bleuler, 2005). Simulators can present cases any number of times over short intervals, permitting the learner to continually incorporate feedback and fine-tune critical cognitive and motor skills. Perhaps most importantly, SBT can be performed without direct supervision of senior medical faculty and therefore fosters an around-the-clock learning environment (Epstein, 2007).

Whereas certain types of simulators (e.g., manikin or virtual reality) have proven remarkably well-suited for honing technical and motor skills, another form of simulator is equally well-suited for realizing the benefits of SBT for interpersonal skill development. Standardized patients (SPs) provide the requisite social context of face-to-face human interaction for developing nontechnical components of clinical practice such as history taking, patient education, and communicating difficult news (Wallace, 1997). Further, SPs (like other types of simulators) are capable of monitoring learners' performance throughout the training scenario and providing real-time feedback to facilitate learning (Kripalani, Bussey-Jones, Katz, & Genao, 2006; Wallace, 1997; 2007).

Standardized Patients

Standardized patients are individuals who portray specific physical and psychological conditions for the purpose of training future healthcare professionals. Standardized patients may range in experience from laypersons with little or no formal theater training to veteran actors with years of professional experience. Some have been diagnosed with the medical conditions they portray, whereas others may be completely healthy (van der Vleuten & Swanson, 1990). Regardless of their unique backgrounds and experiences, SPs commit to mastering a wealth of information through extensive foundational training for the purpose of providing realistic, standardized experiences for learners.

The concept originated in 1963 when Dr. Howard Barrows developed a case for neurology students at the University of Southern California (Barrows, 1987; Barrows & Abrahamson, 1964). Prior to the introduction of standardized patients, medical students were typically evaluated at the end of a clerkship by faculty recollections of student professionalism rather than by direct observation of the students interacting with patients (Barrows, 1993). To improve the training and assessment of medical students' requisite skill sets, Barrows taught a model to portray symptoms of a paraplegic patient with impaired reflex function in both feet, dissociated sensory loss, and one blind eye. After each session, the model completed a short checklist and provided performance feedback to the students. Although SP methodology and assessment techniques have evolved substantially over the past four decades, the fundamental learning principles upon which they were originally developed continue to represent an innovative approach to medical education.

Modern SPs generally develop four unique skill sets to construct a meaningful simulated patient encounter for healthcare professionals. These include the ability to realistically portray specific physical and psychological symptoms, conduct detailed observations of learners' clinical behaviors, recognize pertinent information for post-encounter evaluations, and provide timely feedback to learners during a short post-encounter debrief (Wallace, 2007). Healthcare providers' evaluations are based on four clinical performance areas: history taking, the physical examination, interpersonal interaction, and patient information sharing.

When evaluating a medical students' competence in taking detailed patient histories, the SP first studies case facts made available to them before the encounter.

These facts establish the parameters for communicating to the medical student specific symptoms and lifestyle characteristics pertinent to a successful diagnosis. They also determine how forthcoming the patient should be while interacting with the medical student. Standardized patients should determine when and how much information to provide without inadvertently leading medical students to the case solution. After the encounter, the SP completes a brief patient history checklist to document which case facts were uncovered during the patient interview and which relevant facts were neglected (Wallace, 2007).

Standardized patients also study the physical manifestations of the target medical condition and realistically simulate these as appropriate. This provides medical students the opportunity to conduct a physical examination. Before an SP can evaluate medical students' performances on the physical examination, he or she should first understand the appropriate maneuvers. Only then will the SP be able to determine whether each element of the examination was performed correctly. As with the patient history evaluation, the SP will complete a brief checklist after the encounter to provide feedback about the physical examination (Wallace, 2007).

Physician-patient interaction reflects upon the medical provider's communication skills or "bedside manner," a set of skills related to a variety of critical patient outcomes discussed earlier (Duggan & Parrott, 2000; Greenfield, Kaplan, & Ware, 1985; Greenfield et al., 1988; Hall, Roter, & Katz, 1988; Levinson, 1994; Roter, 1989; Squier, 1990). Seven core competencies have been considered vital to achieving a high level of interpersonal connection with patients: physicians' ability to establish rapport and initiate dialogue with the patient, gather all necessary information, understand the facts from the patient's perspective, share information in such a way that it can be understood by the patient, reach agreement on both the problem and an appropriate plan of care, and provide closure to the patient (Kalamazoo, 2001). Arguably the most important component of physicians' interpersonal interactions with patients involves nonverbal behavioral cues.

Standardized patients often evaluate healthcare professionals' interpersonal skills differently than patient histories and physical examination components. In addition to dichotomous checklists reflecting success or failure for each item, the interpersonal interaction evaluation may contain anchored Likert-type scales, open-ended text boxes, or a variety of other methods in rating students' interpersonal interactions. The SP considers not only which interpersonal behaviors are expected of learners, but also how to discriminate the magnitude of the behaviors. The SP then evaluates and provides specific behavioral examples justifying the rating for each item.

Finally, healthcare providers will have ample opportunity to practice communicating diagnosis, treatment, and difficult information to the patient in a clear, concise manner. Standardized patients are coached in advance concerning how to handle the physician-patient encounter. Depending on the learning objectives and case details, an SP may or may not be instructed to probe for additional clarification or specific information when the opportunity arises. After the encounter, the SP will rate the learner's patient education skills using a checklist similar to those used for patient history and physical examination.

Standardized patients have been shown to improve students' confidence in their clinical abilities, demonstrating the face validity of SP programs (McGovern, Johnston, Brown, Zinberg, & Cohen, 2006). Undergraduate medical genetics students who were provided the opportunity to practice soliciting patient histories and communicating genetic information and testing procedures to SPs from a variety of backgrounds reported greater confidence in their abilities to draw a pedigree, assess genetic risks based on family history and pedigree, and communicate genetic risks than students who were not provided the same opportunity to practice with SPs.

Standardized patients produce generally accurate and reliable summative assessments and clinical skills ratings for healthcare professionals, provided they have received the proper training and supervision (Colliver & Reed, 1993; De Champlain, Margolis, King, & Klass, 1997; Elliott & Hickam, 1987; Furman, 2008; Heine et al., 2003; Pangaro et al., 1997; Williams, 2004). However, a sizeable portion of variance in medical students' scores has been attributed to the individual SPs with whom they interacted (van Zanten, Boulet, & McKinley, 2007). Although this variance was not considered a serious threat to overall interpersonal skills ratings, it is a noteworthy source of error. In this case, SPs were considered capable of discriminating between the extremes of medical student performance (overall high- and low-ability candidates) but struggled to incorporate the finer details of student performance.

In summary, SPs are tasked with memorizing numerous evaluation items and corresponding behavioral anchors to effectively rate medical students' competence in four primary clinical skill areas (Wallace, 2007). A typical simulated patient encounter can last from 5 minutes to over an hour, depending on the case's complexity and range of learning objectives. For the duration of the case, the SP assumes the role of an afflicted patient, interacts with the learner, attends to specific verbal and nonverbal behavioral cues indicative of medical students' clinical competence, and eventually evaluates relevant performance details (including specific behavioral examples) for a formal post encounter evaluation.

Benefits of Standardized Patients

Standardized patients create a training environment in which healthcare professionals can practice their diagnostic and communication skills without fear of doing harm to real patients (Lane & Rollnick, 2007). Errors are treated as learning opportunities and, although the SP may act as if they are suffering (e.g., taking bad news especially hard if the medical student fails to demonstrate empathy), no actual harm has been done. Similarly, the use of SPs is also an ideal method for training medical providers to disclose errors to patients (Chan, Gallagher, Reznick, & Levinson, 2005). Because poor performance with an SP will not result in actual malpractice litigation, learners are free to explore a range of approaches while learning to properly disclose key information.

As trained assessors, SPs also anticipate being examined numerous times by learners and are prepared to withstand substandard performance for the sake of learning and improvement (Barrows, 1993). Depending on the quality of students' clinical interactions, SPs may adopt a range of realistic patient responses in simulating the experience of an actual patient encounter. As a result, medical students are subjected to the simulated consequences of their individual clinical approaches and are afforded the opportunity to refine performance through repeated trials.

Standardized patients also contribute a sense of social and psychological realism to the simulated patient encounter in that medical students find themselves interacting with a living, breathing human being with the potential to be as complicated and unpredictable as an actual patient. The professional and logistical challenges associated with live human interaction prepare medical students for the encounters they are likely to face upon entering an actual clinical setting. Among the many additional benefits that SPs have to offer, this added social dimension reflects a component of healthcare training which has been labeled "veritable reality" (Wallace, 1997).

Potential Limitations of Standardized Patients

The overall demand placed on a skilled SPs' attention and working memory during the typical simulated patient encounter is significant (Baddeley, 1986; Wickens, 1984; Williams, Klamen, & McGaghie, 2003). Standardized patients learn the clinical details and role of a specific patient, assuming the unique symptomatology and personal nuances of the character as faithfully as possible. While maintaining the integrity of the role, they carefully observe the communication (verbal and nonverbal) and behavioral techniques of the student and respond to them appropriately. They determine whether and how much information should be disclosed so they do not lead the student, and be able to improvise within designated case bounds when unexpected questions or events arise. Upon completion of the encounter, the SP will need to accurately recall or recognize on an evaluation form key behavioral indicators and carefully evaluate the student on each clinical skill set using a combination of evaluation instruments. In many cases, the SP also conducts a verbal debrief with the student in which specific examples of the student's performance are discussed.

Standardized patients may be vulnerable to a variety of perceptual and cognitive limitations, impacting the information reported in post-encounter evaluations (Newlin-

Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013); this is a significant consideration given that SPs have become the most widely used method for clinical skills assessment among medical schools and residency programs (Langenau et al., 2011). A great deal of importance is placed on SP encounters as a component of medical students' Objective Structured Clinical Examination (OSCEs) and many current licensing and certification examinations (Adamo, 2003; Barrows, 1993; Boulet, Smee, Dillon, & Gimpel, 2009; Errichetti, Gimpel, & Boulet, 2002; Gimpel, Boulet, & Errichetti, 2003; Langenau et al., 2011; Langenau et al., 2010).

According to Newlin-Canzone, Scerbo, Gliva-McConvey, and Wallace (2013), the detection rate of nonverbal cues decreases and false memory reports increase when SPs perform multiple roles (such as portrayal and assessment) simultaneously. This suggests that in the absence of hard data SPs may depend on schematic formulas to fill in generic details consistent with their overall impressions (Endsley, 1988; Fracker, 1988; Rumelhart, 1984), which could significantly impact the reliability of communication scores and feedback.

A Situation Awareness Framework

The SP's evaluation-feedback role is fundamentally one that involves monitoring specific elements of the environment (i.e., healthcare professional's behavioral cues) to establish and maintain awareness of the student's clinical performance. Relevant behavioral cues are observed through careful allocation of attention and maintained in memory for the duration of the encounter. Endsley (1995a) describes these processes as the fundamental basis of situation awareness (SA)—situation perception. In fact, the SA framework provides a particularly useful taxonomy for describing both SP task requirements and limitations.

Situation awareness has been described by some theorists as the *process* of schema-guided exploration during which the activated schemata themselves are constantly updated by incoming data (Adams, Tenney & Pew, 1995; Smith & Hancock, 1995). Other theorists emphasize SA as the *product* of continually monitoring the environment, differentiating between perceptual and cognitive levels of end-state awareness (Endsley, 1995a; Hourizi & Johnson, 2003). In a recent survey of the SA literature, Salmon, Stanton, Walker, and Jenkins (2009) identified approximately 30 definitions and more than a half-dozen separate theories delineating the construct. In general, SA reflects an awareness and understanding of what is going on around an observer in a dynamic, constantly-changing environment. Although SA has received considerable attention from the research community over the past two decades, the detailed nature of its fundamental attributes is still subject to debate.

The most frequently cited definition is that of Endsley (1995a), describing SA as the end-state of perceiving elements in the environment within a volume of time and space (Level 1 SA), comprehending their meaning (Level 2 SA), and projecting their future states (Level 3 SA). This is accomplished by continuous scanning and integration of environmental data, a process designated "situation assessment." Endsley's conceptualization is based on a subset of structures within the information processing model, with primary emphasis placed on the interaction of attention, working memory (WM), and long-term memory (LTM). The 3-level model has enjoyed a lengthy tenure as the most popular among SA theorists and researchers because of its conceptual simplicity and the ease with which it can be measured and applied to systems design (Salmon, Stanton, Walker, & Jenkins, 2009).

Endsley's Situation Awareness Model

Endsley (1995a; Endsley & Garland, 2000) describes Level 1 SA as the basic perception of status, attributes, and dynamics of relevant elements in the environment. Perceiving a nonverbal performance cue (e.g., tone of voice) during a patient interview is an example of Level 1 SA. The situation perception process may be influenced by a number of key variables including task requirements, situation complexity, and operator characteristics such as specific goals, capabilities, and expectations (Endsley & Garland, 2000). For example, an SP may be instructed to pay attention only to the unique subset of clinical skills that first-year students would be expected to demonstrate. Similarly, expectations based on prior experience will inform the SP about which aspects of the environment are most critical to attend to and when they are likely to occur (Endsley, 1995a).

In this manner, selective sampling of the environment allows an SP to sort through and process complex environmental data more efficiently. However, the knowledge structures guiding this selective sampling process constitute a significant limitation if the SP is unaware of unanticipated or seemingly irrelevant (yet vital) information (Endsley, 1988; Fracker, 1988). According to Jones and Endsley (1996), Level 1 SA failures such as ignoring or misperceiving critical information, inadequately sampling the environment, and succumbing to information overload account for over three-quarters of all SA-related errors.

Level 2 SA involves processing situational elements beyond basic perception. Understanding how the medical student's tone of voice relates to their clinical performance (e.g., an abrupt tone indicates poor choice of interpersonal demeanor) constitutes one example of Level 2 SA. This includes synthesis of individual elements into meaningful patterns, matching such patterns to existing knowledge structures stored in LTM (interpretation and recognition), and comprehending the significance of situational elements as they relate to existing goals (Endsley & Garland, 2000). Level 2 SA is influenced by operator goals and expectations to the extent that fully-developed frameworks (mental models or schemata) exist in LTM. In other words, experienced SPs are more likely than novices to develop and maintain SA by drawing on their own internal mental models to help facilitate information acquisition and interpretation because such frameworks are more readily available to them.

Although top-down processing may reduce the overall demand on working memory for maintaining SA, it does not eliminate the influence of bottom-up processing (Endsley, 2001; Garsoffky, Schwan, & Hesse, 2002). Rather, Level 2 SA still primarily operates within the domain of working memory. Jones and Endsley (1996) report that a significant percentage of SA errors at this level can be attributed directly to working memory failures, and this is evident in novice operators as well as data-saturated veterans (Adams, Tenney, & Pew, 1995). To be sure, data integration, pattern matching, and comparison of data to established goals each draws heavily on working memory (Endsley & Garland, 2000).

Level 3 SA represents the combination of Levels 1 and 2 SA for the purpose of projecting future states (Endsley, 1995a). This level of awareness is largely dependent upon the individual operator's experience, from which a reasonable extrapolation of near-future situation status based on an accurate understanding of the current status may be generated. A SP exhibiting Level 3 SA not only perceives and understands the significance of a medical student's behaviors, but also projects how the student is likely to perform over the next few minutes as a result of understanding how the student has performed to that point.

Though related, SA is conceptually distinct from decision-making, performance, and LTM. Endsley (1995a; Endsley & Garland, 2000) contends that it is possible to have perfect situation awareness and still make an inappropriate decision due to lack of experience or limited decision choices. Likewise, it is possible to take correct actions (as a result of chance) under imperfect SA conditions. As mentioned earlier, several factors may influence a SP's ability to establish and maintain SA. Namely, heightened workload and stress have been identified as factors related to active SP task requirements such as observation and unscripted interaction (Endsley, 1988; 1995a; Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013). These factors are especially impactful on Levels 1 and 2 SA because focused attention is requisite yet high workload and stress can have detrimental effects on operator attention, memory, and decision-making processes (Hockey, 1986; Janis, 1982; Wright, 1974; Keinan, 1987). Additionally, complex situations demand greater mental resources to track and record dynamic changes, which in turn strain an individual's ability to maintain SA.

Although SPs perform a variety of tasks during a typical encounter, their ability to observe and later recall or recognize key performance indicators inevitably exercises the greatest direct impact on scoring reliability. To observe performance cues, the attention of SPs must be effectively directed to appropriate information sources to facilitate deeper cognitive processing (e.g., memory encoding). Recalling these cues during the post-encounter evaluation period presents a series of challenges as well, in that SPs must encode, maintain, and finally retrieve relevant information as efficiently as possible while simultaneously allocating attention to other task components. It is important to understand what SPs are capable of observing and recalling during the typical simulated patient encounter and whether known cognitive limitations necessitate any fundamental changes to the SPs' tasks.

Task 1: Observation

Observation is a function of bottom-up (attentional) features as well as top-down influences (LTM). Consistent with any task involving sustained attention, situation perception may be impeded by simple omission of cues (i.e., not looking at a piece of information) or attentional narrowing (due to heavy task load or distractions). Standardized patients track and maintain a large volume of dynamic information for an extended period of time, so it is not surprising that greater levels of mental workload and stress have been associated with patient encounters (Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013). In such a dynamic environment, it is essential to understand how attention is allocated and whether SPs are capable of attending to all necessary elements.

Several factors dictate how humans direct attention for the purpose of information acquisition. Learned scan patterns and information sampling strategies often result when operators begin to anticipate a system's behavior. Goals, expectations, and previous information also cause people to focus attention on specific aspects of the environment for selective processing (Endsley & Garland, 2000; Fracker, 1989). Experience directs expectations, which influence what we attend to and how we perceive the environment; prior information helps the operator understand the situation as it unfolds (Posner, Nissen, & Ogden, 1978; Palmer, 1975). Endsley and Smith (1996) report that fighter pilots' attention to certain targets on a display is directly related to the perceived importance of those targets to their identified task. Likewise, air traffic controllers shed attention to less important information (e.g., flight call number data) as task load increases (Endsley & Rodgers, 1998) and drivers pay more attention to cars in their immediate vicinity than those further away (Gugerty, 1998). In each of these examples, it is clear that attention is deployed in a manner best suited to support operator goals.

Attention

Attention has been defined simply as the concentration of mental activity or resources for processing external stimuli (Kahneman, 1973; Matlin, 1994). The concept was originally conceived of as a single-channel system (Welford, 1952; 1959; 1967), emphasizing central processing limitations to account for attentional overload. Later theories describe attention as a filtering mechanism to keep irrelevant stimuli from overloading processing and response capabilities (Broadbent, 1958; Deutsch & Deutsch, 1963; Norman, 1968; Treisman, 1964).

Alternately, capacity theories hold no assumptions that certain features (e.g., sensory processing or response organization) limit the allocation of attention. Rather, capacity models assume that attention itself is a limited resource and it is this limitation that inhibits performance (Kahneman, 1973; Wickens, 1984; 2002). Within this framework, attention is the process of allocating limited resources to any number of potential sensory inputs. The ability of an operator to carry out concurrent tasks that are competing for a limited amount of mental resources will inevitably depend on the rate of resource consumption. If a primary task demands a significant amount of attentional resources (e.g., an anesthetist monitoring vitals of a patient under general anesthesia), then the operator will have little spare attention to allocate to secondary tasks (e.g., troubleshooting a mechanical issue with the anesthesia machine).

Multiple resource theories (MRTs) are a subset of capacity theories which further assume that performance is influenced by the interaction of task demands, various capacity-limited resource pools, and the policy for allocating these resources (Wickens, 2008). A number of theoretical variations have been proposed (Boles, 2002; Boles et al., 2007; Kieras, 2007; Polson & Friedman, 1988; Ralph, Gray, & Schoelles, 2009; Salvucci & Taatgen, 2008; Wickens, 2002), each with their own unique emphases on individual elements of the theory. What these theoretical variants share is the common notion that attention is a limited-capacity resource that must be divided when multiple tasks are performed concurrently. To the extent that tasks are competing for the same pool of resources, overload may result and task performance will suffer. If concurrent tasks draw from different attentional resource pools, task interference may be mitigated and performance decrements will be minimal. However, the specific number and structure of these resource pools is still the topic of debate.

The Wickens model (1984; 2002) has proven to be one of the most influential MRT theories for human-system design applications (Horrey & Wickens, 2003; Liu & Wickens, 1992; Sarno & Wickens, 1995; Wickens & Colcombe, 2007; Wickens, Goh, Helleberg, Horrey, & Talleur, 2003; Wickens, Sandry, & Vidulich, 1983). The model suggests that attentional resources exist along four distinct dimensions: processing stages (perception/cognition, and action execution), perceptual modalities (auditory and visual), processing codes (visual and spatial), and visual channels (focal and ambient). In complex task environments, timesharing between two tasks may lead to degraded performance on one or both tasks because of the need to divide resources between them. The amount of degradation in either task is dependent upon whether common or separate resources are required for executing the tasks simultaneously. Thus, tasks that require an individual to share resources will be more difficult to perform than tasks drawing from separate resource pools.

Stages of processing are divided into perceptual/cognitive and action selection/execution. In other words, perceptual and cognitive activities (e.g., working memory) are theorized to draw from a separate resource pool than action execution (Isreal, Chesney, Wickens, & Donchin, 1980; Wickens, 1984). Similarly, visual processing is theorized to operate from a different set of underlying resources than verbal/linguistic processing. Tasks involving different sensory modalities (e.g., visual, auditory) also draw from separate resource pools, such that cross-modal task requirements and presentation can lead to improved task sharing and concurrent task performance. A number of studies have demonstrated the benefits of distributing sensory information across auditory and visual modalities (Aretz, 1983; Wickens, 1980; Wickens, Sandry & Vidulich, 1983). More recently, MRT has been expanded to distinguish between focal and ambient visual channels, predicting that multiple visual tasks could be supported concurrently as long as separate channels are involved in processing (Horrey, Wickens, & Consalas, 2006; Lenneman, Lenneman, Cassavaugh, & Backs, 2009; Weinstein & Wickens, 1992; Wickens, 2002).

To the extent that SPs must divide attention among separate tasks drawing from the same limited set of resources, performance decrements will likely occur. For example, a SP actively engaged in monitoring a medical student's verbal communication cues while simultaneously generating novel verbal responses will perform worse than a SP passively monitoring the situation without responding (Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013). Because timesharing among these tasks is an inherent aspect of the SP's professional role, it is important to understand how SPs allocate attention and whether these tasks can be redesigned to enhance performance.

Long-Term Memory

When task load exceeds human processing capabilities, important data are often ignored or quickly forgotten (Endsley & Rodgers, 1998; Jones & Endsley, 1996). The resulting information gaps lead to a potential performance decrement when the operator is later asked to recall specific task-relevant details. When such gaps arise, task operators are more likely to draw upon general knowledge structures (schemata) stored in long-term memory to "fill in" plausible details for the missing data (Endsley, 1988; Fracker, 1988; Rumelhart, 1984). In this manner, the information that we attend to and the details we ultimately recall are influenced by pre-existing mental structures known as schemata (Bartlett, 1932; Brewer & Treyens, 1981; Freeman, Romney, & Freeman, 1987). Schemata have been defined as clusters of knowledge representing general objects, perceptions, event sequences, or social situations (Thorndyke, 1984). Traits, attitudes, and goals may each be viewed as variations of schematic knowledge structures, each with the ability to influence perception and decision making (Anderson & Huesmann, 2003).

Schemata are stored in LTM and are easily accessible for aiding in the interpretation of sensory data, searching and reconstructing information from memory, guiding actions, and allocating mental resources (Rumelhart, 1984). They make information processing possible in a complex, dynamic world in which we are constantly inundated with external stimuli. When situational data from the environment begin to form patterns adhering to these specific knowledge structures, we are more likely to base situation assessments on internal rather than purely external data (Endsley, 1988; Fracker, 1988). This allows for more efficient situation processing because we no longer need to attend to every relevant detail, which constitutes a slow and effortful cognitive process (Fennema & Kleinmuntz, 1995; Gray & Fu, 2001; Wickens & Carswell, 1995).

Endsley (1988) describes a dual process whereby schemata influence which elements of the environment are attended, and the subsequently-attended environmental data are then activated (and are stored as) new schemata in LTM. However, schema-guided sampling comes at a cost, as situation perception is (to some degree) based on generalizations rather than case-specific details.

Schemata represent general conceptual templates rather than data-rich architectures. Although they are formulated from actual previous experiences, they do not retain the level of detail necessary for reconstructing specific examples from memory. Rather, they are automatically drawn upon to fill informational gaps with generalizations consistent with a mental exemplar, often eliciting false memories to fill the perceptual-cognitive void (Bartlett, 1932; Brewer & Treyens, 1981; Freeman, Romney, & Freeman, 1987). This can be useful to an overloaded operator if the information sought is rather generic, but it poses a significant challenge for those operators in need of detailed, case-specific recollections (as in the case of SPs).

Two studies conducted by Neuschatz et al. (2002) demonstrate the impact that underlying schemata have on details recalled from memory. In both studies, participants viewed a taped lecture in which the presenter exhibited a number of schema-consistent and schema-inconsistent behaviors. Schema-consistent behaviors include actions we would expect a lecturer to engage in, such as writing on a whiteboard, taking periodic drinks of water, or asking whether the audience has any questions. Schema-inconsistent behaviors would include any unusual actions for a presenter, such as dancing, smoking a cigarette, or taking a phone call. Participants were later asked to recall details from the lecture, and were found to recall schema-inconsistent behaviors more accurately and in more vivid detail than schema-consistent behaviors. However, the authors discovered that false memories were more likely to consist of schema-consistent actions—behaviors that participants expected to see from the presenter but in fact did not. This result is consistent with Sulin and Dooling (1974), who reported false memories for behaviors that never occurred but would be considered typical of the underlying schema. The number of false memories reported by Neuschatz et al. (2002) increased with retention interval, such that after a one-week delay the proportion of false memories nearly reached 50%.

Tuckey and Brewer (2003) reported that eyewitness memory for both schema-consistent and schema-inconsistent crime details suffers less decay over time and repeated eyewitness interview sessions than memory for schema-irrelevant details. Information that was not relevant to an underlying "bank robbery" schema decayed from memory earlier and to a greater degree than the other forms of information. The authors suggest that repeated interviewing may serve to strengthen associations of case details in memory as long as initial interviews are conducted before the memories start to decay.

If people believe that their LTM structures relating to the current situation are adequate, they tend to avoid seeking new information from the external environment in favor of this more generalized internal data. However, we often overestimate the accuracy of these internal knowledge structures, which can lead to overreliance on inaccurate data (Bjork, 1999). A relationship between working memory capacity and SA is therefore dependent on the completeness of LTM knowledge that the operator has stored. If this knowledge is sufficiently complete, the quality of SA should be less sensitive to working memory and more sensitive to the quality of LTM data; otherwise, one must attend to a larger amount of information in the environment, identify multiple schemata that may be appropriate, place info from these schemata into working memory, and integrate information into a single result.

Task 2: Recall and Recognition

Levels 1 and 2 SA also involve the temporary storage of information in memory (Jones & Endsley, 1996). Even if a relevant behavioral cue is initially perceived by the SP, he or she must still be able to recall or recognize the cue from memory during the post-encounter evaluation period. Instances in which a relevant cue was initially perceived but not successfully retained in the SP's memory would constitute a failure of Level 1 SA. According to Jones and Endsley (1996), Level 1 SA failures such as neglect or misperception of relevant data and memory loss constitute approximately three-quarters of all SA errors. An SP's ability to accurately maintain critical information in working memory will directly impact their ability to produce a faithful evaluation of learners' clinical performance. It is therefore important to discuss the nature and limitations of working memory to determine whether fundamental changes to the current simulated patient encounter paradigm would enhance SPs' recognition potential.

Short-Term (Working) Memory

Baddeley and Logie (1999) define WM as a set of cognitive components dedicated to processing one's environment, storing information and representations of environmental elements from the immediate past, manipulating information to support problem solving and knowledge acquisition, and developing and managing task-relevant goals. Because of this multifaceted relationship, WM has been designated the proverbial "workhorse" in acquiring and maintaining SA. Fracker (1988) and Endsley (1988) describe theories of situation awareness in which mental representations of the situation are developed and maintained in working memory (WM) as real-time environmental data accumulate. More specifically, the goals or objectives activated in WM serve as a framework for processing environmental stimuli (Endsley, 1988).

One of the most enduring WM theories, Baddeley's model (Baddeley, 1986; 2000; Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley & Wilson, 2002; Ericsson & Kintsch, 1995) is comprised of a limited-capacity attentional system called the central executive which coordinates and interacts with temporary stores of various information types (Baddeley, Hitch, & Allen, 2009). The central executive allocates limited attentional resources to three separate slave systems: the visuo-spatial sketchpad, phonological loop, and episodic buffer. It is also assumed to plan goal-relevant sub-tasks, control selective attention and inhibition, refresh the components of working memory, and code representations along a temporal continuum.

The visuo-spatial sketchpad represents the temporary storage and manipulation of visual patterns and spatial movement, and is therefore responsible for recording episodic data related to a medical student's nonverbal behavior. On the other hand, the phonological loop is responsible for processing speech-based information and supporting subvocal articulation. Medical students' verbal interactions with the SP would be processed and retained in this component. The phonological loop is involved in a range of speech-based processes from sentence comprehension (Lauro, Reis, Cohen, Cecchetto, & Papagno, 2010) to language acquisition and speech production (Adams & Gathercole, 1995; Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1993). The episodic buffer is a more recent addition to Baddeley's original WM model, representing a limited-capacity storage system in which information from short-term stores and LTM can be integrated into and temporarily stored as episodic chunks (Baddeley, Hitch, & Allen, 2009). It is completely dependent upon attentional control by the central executive.

Through the episodic buffer, the central executive can integrate information from each of the other two slave systems and bind it into unitary multidimensional representations. This is called "active binding," and is assumed to be highly demanding of the central executive's limited attentional resource capacity (Baddeley & Wilson, 2002). The buffer is assumed to be a centralized component for maintaining multidimensional (multimodal) episodic traces, bringing it in line with the theoretical contributions of Cowan (1988; 1993; 2005) and Engle (2002) which emphasize a simpler, common storage system for WM. Although information may be initially processed in either of the slave systems, it is ultimately integrated and maintained in the episodic buffer (Baddeley, Hitch, & Allen, 2009).

Consistent with MRTs of attention, two tasks drawing from the same WM components cannot be performed simultaneously due to overload (Baddeley, 2000). However, if the tasks draw from separate components, it should be possible to perform them concurrently without significantly impacting performance. SPs engaged in constructing speech-based dialogue as well as monitoring the same from medical students will be drawing on the same pool of limited attentional resources to support each task, and should therefore exhibit greater difficulty maintaining relevant performance data in working memory. However, information brought in through the visuo-spatial and articulatory subcomponents separately will inevitably be maintained in the episodic buffer (a limited capacity storage component) until it can be used. As information continues to accumulate in the buffer, it is expected that some loss will occur as a result of WM limitations (Brown, 1958; Loftus, Dark, & Williams, 1979; Melton, 1963; Moray, 1986; Peterson & Peterson, 1959).

Information must be continually rehearsed (i.e., through subvocal articulations) or reactivated for maintenance in WM; otherwise, it decays rapidly as a function of time and interference from subsequent information. According to Peterson and Peterson (1959), the likelihood of recalling a three-digit alphabetic stimulus from WM drops to around 50% after six seconds and is permanently lost within 18 seconds. Similar rates of decay for alphabetic digits have been reported by Brown (1958) and Melton (1963), and these results have been replicated for navigational information (Loftus, Dark, & Williams, 1979) and radar control data (Moray, 1986). Without continual rehearsal, information is unlikely to be retained in WM for more than a few seconds. It has therefore been recommended that operators act upon received information as promptly as possible (i.e., for tasks requiring an action or response to stimuli) to offload WM prior to decay or interference from subsequent stimuli (Hart & Loomis, 1980; Jacko, 1997).

Cowan (1993) reports that briefly focusing attention on task-specific items improves subsequent recall by 20% over unattended task-specific items when the delay interval is relatively short. However, as the recall delay is increased, participants will shift the focus of their attention to other task components and the benefit will quickly diminish. This result demonstrates the influence of attention in maintaining the contents of WM such that attended information is refreshed in memory and can be sustained for longer intervals. This is supported by evidence of improved recall for items in shorter event sets rather than longer event sets because individual items in longer sets cannot be refreshed as frequently (Baddeley, Thomson, & Buchanan, 1975; Cowan et al., 1992).

Cowan (1992) supports a decay-and-reactivation hypothesis of memory retrieval in which attention devoted to the recall of one item from memory inevitably subjects the remaining items to temporal decay. Once the initial item has been successfully reported, limited reactivation of the other items may occur as a result of scanning or rehearsing the remaining items prior to reporting the next sequential item. This process is repeated until all items have been reported or any remaining items have decayed completely. Thus, individual differences in recall span reflect the efficiency with which one can execute the retrieval and reactivation sequence during recall. Again, longer recall delay intervals (e.g., retrospective recall) will likely result in greater memory loss than shorter recall delay intervals (e.g., concurrent recall).

Working memory is limited not only by temporal retention, but also by the amount of information that can be stored at any given time. The capacity of WM has been estimated to be 7 ± 2 units of information (Miller, 1956), although this may be an overly optimistic estimate (Cowan, 2001). Miller's memory span estimate relates to the number of individual units of information that can be held in WM simultaneously, whether the information takes the form of individual alphanumeric digits or semantically meaningful patterns of information (e.g., words). The latter are known as "chunks," or bits of information bound together as a meaningful whole after having structure imposed upon them by existing knowledge in LTM. Chunking greatly increases the amount of information that can be maintained in WM at a given time (Baddeley, Hitch, & Allen, 2009).

As WM span approaches capacity, the information stored there will become more vulnerable to decay, as limited attention must be divided among a greater number of items for rehearsal/reactivation. Thus, the central executive would be required to allocate attention to the episodic buffer for maintenance of information as well as to the ancillary components for Levels 1 and 2 SA processing of incoming information. Greater WM span would be expected to translate into improved SA, a prediction that is indeed supported empirically (Barnett et al., 1987). However, this relationship does not fully explain the process of acquiring SA. Not all situation-relevant information is contained in the external environment, nor is it stored exclusively in working memory (Fracker, 1988).

Endsley (1988) asserts that the quality of SA is often moderated by a number of variables which include individuals' training, experience, and workload. These moderators reflect the top-down nature of situation awareness discussed previously, whereby knowledge structures in LTM influence attention to and interpretation of environmental data in conjunction with attentional and WM resources.

Serial Position Effects

When considering which details are likely to stand out in memory or be forgotten, it is important to discuss serial position effects. Research has demonstrated that recall accuracy depends in part on an item's serial position within a list of items or a series of events. During free recall, items that occur in the middle of a sequence are less likely to be recalled than items occurring early or late in the sequence. Items occurring late in the sequence are recalled with the highest frequency (Deese & Kaufman, 1957). These effects have been demonstrated in verbal tasks using probed recall and serial reconstruction (Avons, Wright, & Pammer, 1994; Nairne, Riegler, & Serra, 1991; Nairne, Whiteman, & Woessner, 1995) as well as in visual-spatial recall tasks (Avons, 1998; Hay, Smyth, Hitch, & Horton, 2007; Manning & Schreier, 1988; Smyth, Hay, Hitch, & Horton, 2005).

Serial position effects are commonly attributed to a combination of long-term and WM components (Atkinson & Shiffrin, 1968; Azizian & Polich, 2007). The assumption is that different memory operations are involved in the encoding of primacy and recency items. Items occurring early in the sequence are likely to be encoded in LTM as a result of increased rehearsal time and little interference from preceding items (Atkinson & Shiffrin, 1968; Bellezza & Cheney, 1973; Glanzer & Cunitz, 1966; Rundus, 1971), whereas items occurring late in the sequence are still active in WM (Atkinson & Shiffrin, 1968; Glanzer & Cunitz, 1966; Postman & Phillips, 1965).

By extension, the details that SPs recall during post-encounter evaluations may be influenced by their serial position within the simulated patient encounter. The last (or first) few minutes of a medical student's performance in the simulated patient encounter may receive disproportionate weight during the post-encounter evaluation due to greater availability of recall for performance indicators from those time periods, whereas the majority of performance which lies between these end points remains largely unacknowledged. Therefore, SPs may simply categorize medical students based on initial (or final) impressions rather than by employing the process of continuously-updated monitoring.

The Von Restorff (Isolation) Effect

Another phenomenon linked to memory retrieval is the isolation or von Restorff effect, named for German researcher Hedwig von Restorff. This researcher discovered that distinct items or events (isolates) are more likely to stand out in memory than other more common items or non-isolates (von Restorff, 1933). The effects of item distinctiveness on memory have often been attributed to the selective rehearsal of the distinct item (Bellezza & Cheney, 1973). Rundus (1971) demonstrated that participants naturally rehearse isolates more frequently than non-isolates because they are categorically different from background items in some way and are therefore more demanding of attention. However, this explanation is insufficient to account for all of the available data, including the results of von Restroff's own studies (Hunt, 1995).

Despite a longstanding emphasis on perceptual salience as the key mechanism for drawing attention to isolates (Green, 1956; Schmidt, 1991), researchers have also demonstrated that isolates presented at the beginning of a list, at which point no context for supporting perceptual salience can exist still result in a significant isolation effect (Dunlosky, Hunt, & Clark, 2000; Hunt, 1995). In this case, the isolate is theorized to become salient only after it has been retrieved and compared to subsequent items (i.e., it *emerges* as a conceptually salient feature). However, isolates presented later in the sequence are both conceptually *and* perceptually distinct from other items, and thus the resulting isolation effect will be even greater. Regardless of where they occur sequentially, isolates will receive more attention and processing than non-isolates.

The implication for SPs is that irregular events (e.g., exceptionally good or poor examples of clinical performance) may receive more consideration during the post encounter evaluation than other less-memorable performance cues. For example, a medical student may break from the patient interview for a moment to respond to a personal text message on her cellular phone. Assuming her or his clinical performance is otherwise generally commendable, neglecting the patient for several seconds to read and respond to a personal text message may carry significant weight in the SP's overall evaluation. This is problematic in that the isolate event, by definition, is not indicative of the medical student's general clinical performance. However, because of its salience and recall availability it is likely to result in an artificially deflated perception of the student's overall performance.

Directed Forgetting

It may be possible to mitigate the undue effect of an isolate on SP evaluations by utilizing a technique known as directed forgetting, in which individuals are instructed to forget a set of previous information in favor of new or more current information. When information from the past is no longer relevant to current or future goals, it can be beneficial to intentionally suppress or inhibit this information (Bjork, 1972; Block, 1971; Elmes, Adams, & Roediger, 1970; Muther, 1965; Sahakyan & Delaney, 2003; Titz & Verhaeghen, 2010; Zacks & Hasher, 1994). By suppressing the activation of previous information, working memory resources may be more efficiently focused on information most relevant to the current task (Hasher, Zacks, & May, 1999). As a result, irrelevant information from the past is less likely to interfere with encoding and retrieval of subsequent critical information (Lustig, May, & Hasher, 2001). Evidence from list-method studies of directed forgetting, in which a complete list of to-be-forgotten (TBF) items is studied and encoded by participants before they are instructed to forget the items, demonstrates participants' ability to ignore previously attended information in favor of improved recall of information from a subsequent list (Bjork, 1972; Block, 1971; Elmes, Adams, & Roediger, 1970; Muther, 1965).

Several underlying mechanisms have been theorized to influence directed forgetting, including cessation of item rehearsal, segregation of TBF information into a separate set within memory, and the inhibition of retrieval for irrelevant information (Bjork, 1989; Geiselman & Bagheri, 1985; MacLeod, 1989). The effects of directed forgetting have been demonstrated in terms of eliminating the encoding and recall interference characteristics of TBF information and poor participant recall of TBF items during later testing periods (Bjork, 1989). As a result of list-method cueing studies, an inhibitory mechanism has been theorized to actively suppress prior information and prevent its interference with subsequent working memory encoding and retrieval (Bauml, 2008; Basden, Basden, & Gargano, 1993; Geiselman, Bjork, & Fishman, 1983; Goernert

& Larson, 1994; Sahakyan & Delaney, 2003; Sahakyan, Delaney, & Waldum, 2008; Sahakyan & Kelley, 2002; Whetstone, Cross, & Whetstone, 1996).

Measuring Situation Awareness

The value of Endsley's (1995) model lies in its conceptual simplicity, allowing for identification and investigation of system requirements at each level (Salmon, Stanton, Walker, & Jenkins, 2009). As a framework for studying SP task requirements, it emphasizes the role of memory in both the observation and evaluation of medical students. More importantly, it establishes measurable outcomes that relate what information SPs are able to retain in memory and draw upon when they evaluate medical students' performance. For tasks requiring written and verbal assessment of medical student's clinical performance, this is of supreme importance. An assessment paradigm susceptible to influence from inaccurate recall and schematic generalizations may be inadequate when verbatim recall and justification of ratings are demanded of SP.

A variety of SA assessment techniques currently exist (see Stanton, Salmon, Walker, Baber, & Jenkins, 2005 for a review). The most economical assessment method involves pencil-and-paper self-reporting along a set of predetermined SA categories. For example, the situation awareness rating technique (SART; Taylor, 1990) is typically administered at the end of a scenario and requires the operator to rate the task along a series of SA dimensions (e.g., information quantity, arousal, spare capacity, etc.) using a Likert-type scale or other similar means. The SART method is non-invasive in that it does not interfere with task performance, and is simple both to administer and analyze. However, a number of limiting factors have been identified not just for SART, but SA assessment techniques in general that include self-reporting or post-scenario querying.

Aside from the well-documented phenomenon of socially desirable responding (van de Mortel, 2008), Endsley (1995b) argues that operators are largely unaware of their own SA limitations and are therefore unable to report that which they do not know. Further, the data are subjective and therefore filtered by the operator's own unique interpretation and understanding of the key SA dimensions. Because SART ratings are solicited only after the scenario has concluded, they are often spuriously correlated with task performance and workload (Stanton, Salmon, Walker, Baber, & Jenkins, 2005), influenced by rationalization and generalization (Nisbett & Wilson, 1977), and subject to memory limitations (Hart & Loomis, 1980; Jacko, 1997).

The situation present assessment method (SPAM; Durso, Hackworth, Truitt, Crutchfield, & Manning, 1998) is a SA assessment technique designed to avoid operators' recall limitations altogether. The SPAM method not only introduces real-time probes throughout a given task (thus avoiding post scenario-only reporting), but also requires operators to rapidly locate and report the status of specific information in the environment as the scenario continues to progress. The amount of time required for the operator to report the requested situational information serves as a proxy measure for SA. The benefit of SPAM is the concurrent recall format, thus avoiding common limitations of delayed recall techniques (i.e., WM decay and interference from subsequent information). However, the limitations of SPAM include intrusiveness (due to probing without freezing task elements) and the fact that it does not directly measure operators' SA (Stanton, Salmon, Walker, Baber, & Jenkins, 2005).

Memory is a central mechanism of SA—processing and integrating situational data to form dynamic situation representations (WM) and providing the necessary infrastructure for managing and interpreting data (Endsley & Garland, 2000). Measuring SA provides insight into which aspects of the environment an operator has available in memory at a given time, and by extension the situational elements that are likely to influence decision-making in tasks such as performance evaluation. Operators who establish SA at the cost of high working memory load are vulnerable to losing it if demands on working memory increase (Wright, Taekman, Endsley, 2004). It is therefore advisable to measure SA concurrently, rather than retrospectively after the task scenario has concluded.

The most frequently implemented and cited SA assessment method (Stanton, Salmon, Walker, Baber, & Jenkins, 2005) is the situation awareness global assessment technique (SAGAT; Endsley, 1995b), an objective technique for measuring SA based on Endsley's (1995a) three-level model. It is also one of the most well-validated techniques for assessing SA, having been determined reliable and valid in routine tasks among fighter pilots (Endsley & Garland, 2000), nuclear power plant operators (Collier & Folleso, 1995), and automobile drivers (Gugerty, 1997). Its predictive validity has also been established among pilots and air traffic controllers (Endsley, 1990a; 1990b), and its measurement sensitivity has been found to exceed that of other techniques like real-time probing without simulation freezes and subjective SA measures (Endsley & Garland, 2000; Endsley, Selcon, Hardiman, & Croft, 1998; Endsley, Sollenberger, & Stein, 2000).

The SAGAT method involves administering queries during a series of simulation freezes to assess operator SA at each of Endsley's levels (Endsley, 1995b). Each set of queries addresses only the situational information deemed relevant to the operator's task at that particular time. When a freeze occurs, all task dynamics cease until the simulation is allowed to resume and any environmental cues are blacked out (if feasible) so they cannot be referenced by the operator and thus bias results (Wright, Taekman, & Endsley, 2004). The SAGAT method therefore emphasizes the situation-relevant information that the operator is maintaining in memory at any given time throughout the scenario.

Although routine simulation freezes coupled with multi-level SA queries is seemingly a more intrusive assessment method than retrospective and self-report techniques, Endsley et al. (Endsley, 1990a; 1990b; 1995b; Wright, Taekman, & Endsley, 2004) demonstrate that this concern is ultimately without empirical merit. Concurrent SA queries permit SA investigators to circumvent the cognitive limitations characteristic of retrospective querying (Endsley, Selcon, Hardiman, & Croft, 1998), and no adverse effects (e.g., memory decay or task interference) have been reported for simulation freezes of up to 6 minutes (Endsley, 1995b). In a more recent study describing the effects of real-time simulation-based training (SBT) feedback on learning among medical students, no loss of perceived realism as a result of periodic simulation freezes was reported (van Heukelom, Begaz, & Treat, 2010).

The SAGAT method appears to be an ideal assessment technique for investigating SP task performance for several reasons. First, the evaluation-feedback tasks that SPs perform may be intuitively categorized as Levels 1 and 2 SA (Endsley, 1995a) in that relevant information must be perceived, recognized and eventually recalled for the purpose of evaluating medical students' clinical competence. The SAGAT method emphasizes memory as the critical component of SA, bringing it into alignment with SP task requirements (in which memory plays a central role for both observation and evaluation of medical students). The technique is designed to mitigate overload and working memory limitations by incorporating periodic pauses; task-critical information is therefore offloaded prior to decay or retroactive interference from subsequent information (Hart & Loomis, 1980; Jacko, 1997).

Not only does the SAGAT methodology provide a practical means of assessing SPs' situation awareness, but it also represents a potential mechanism for reducing cognitive load and performance limitations currently associated with the simulated patient encounter. By implementing periodic breaks for evaluation and focusing scenario content on individual chunks or subcomponents of established clinical performance areas, SPs may be able to produce more accurate and thus reliable evaluation data. The concurrent recall paradigm may also reduce SP vulnerability to common recall phenomena such as serial position and isolation effects, further improving the accuracy of SP ratings and feedback. Short performance intervals result in less information to be recalled at each evaluation point, and directed forgetting during scenario breaks should mitigate the influence of previous information (including isolates) on subsequent evaluations.

Limitations of Situation Awareness

Despite its general appeal as a conceptual framework for organizing and studying specific cognitive elements of clinical performance evaluation, SA has received some criticism for remaining largely devoid of theoretical specifics (Banbury & Tremblay, 2004; Flach, 1995). Situation assessment and SA have been theorized to involve certain cognitive mechanisms, but SA theories largely fail to specify exactly how underlying mechanisms support the various aspects of SA and to what degree. This has resulted in considerable disagreement and confusion related to the formal definition, measurement, and study of SA (Banbury & Tremblay, 2004).

Additionally, some have argued that SA represents more than just what an operator is consciously aware of at any given point in time, but should also reflect that which an operator may be unconsciously or tacitly aware of as well (Banbury, Andre, & Croft, 2001; Smith & Hancock, 1995). Further, SA theories should offer insight into how information is gathered and maintained (i.e., the *process* of SA) as well as how external characteristics (e.g., the environment and/or task characteristics as opposed to focusing solely on operator characteristics) affect SA (Durso & Gronlund, 1999). Considering the presently unresolved state of SA's theoretical foundations, Endsley's SA model will serve merely as a conceptual framework from which to investigate task components characteristic of clinical performance evaluation (i.e., observation and recall). Although the model was not originally conceived for perceptual and memorial facets of interpersonal communication, it is believed that this framework can still provide a useful context for interpretation of this study's results.

Goals of this Research

A goal of this research was to better understand how a variety of cognitive factors may influence SP assessment abilities. This study explored whether a concurrent evaluation framework adapted from traditional SA research methodology reduced the amount of cognitive demand, and thus performance limitations, associated with the traditional retrospective evaluation technique for SP clinical competence evaluations. Further, this study investigated the impact of an atypical behavior or isolate on SP ratings, and whether a single isolate would significantly impact a learner's performance ratings across the entire encounter.

Study Description

Participants observed a video-recorded simulated patient encounter and rated the clinical performance of a simulated healthcare provider (SHP) playing the role of medical student. The encounter was divided into three qualitatively similar patient interview segments in accordance with the set of core skill areas defined by Wallace (2007). Each segment of the simulated patient encounter lasted approximately 5-7 minutes and included a subset of specific verbal and nonverbal behaviors to be exhibited by the SHP at pre-determined points. An independent, trained reviewer naive to the study's protocol and hypotheses was asked to evaluate each video recorded segment to ensure that the embedded (scripted) behavioral cues were in fact present, were presented only in the correct segment, and were clearly identifiable to the passive observer. The reviewer's evaluation of the scenario's content was in complete agreement with the scripted set of embedded behaviors (Appendix C).

Participants were asked to rate the SHP's verbal and nonverbal clinical performance. Half of the participants rated the SHP's performance at the end of each segment (i.e., using a concurrent evaluation format), during designed simulation pauses. The remaining participants rated the SHP's performance only once, after the entire simulated patient encounter has concluded (i.e., retrospective evaluation). To investigate isolation effects on participant ratings, half of the participants observed an example of unprofessional learner behavior (or isolate) at one point during the scenario (early or late) whereas the remaining participants did not.

Hypotheses

Hypotheses for Concurrent vs. Retrospective Evaluation Framework. Implementing a concurrent evaluation format for evaluating medical student clinical competence should have enabled participants to work from a smaller subset of performance criteria at any given time and to offload the contents of working memory more frequently. Therefore, information access and storage costs associated with working memory were expected to be reduced (Baddeley, Thompson, & Buchanan, 1975; Cowan et al., 1992; Endsley, Selcon, Hardiman, & Croft, 1998), resulting in more accurate recognition of nonverbal behaviors and more reliable scoring for verbal behaviors. Thus, Hypotheses 1a and1b were as follows:

1a. Participants in the concurrent evaluation group would generate more reliable verbal clinical performance ratings than participants in the retrospective evaluation group.

1b. Participants in the concurrent evaluation group would demonstrate more accurate nonverbal clinical performance evaluation than participants in the retrospective evaluation group.

Hypotheses for Serial Position Effects. Items appearing at the beginning or end of a list were more likely to be recognized as a result of working memory salience than items appearing between these endpoints (Atkinson & Shiffrin, 1968). Likewise, performance cues occurring early or late in the encounter were expected to be disproportionately weighted by participants during post-encounter evaluations. Performance cues were defined in this case as behaviors typical of any simulated patient encounter that demonstrate clinical competence (e.g., maintaining sufficient eye contact) or areas for needed improvement (e.g., neglecting to elicit critical information from the patient). However, the impact of serial position effects on participants' ratings should have been mitigated by reducing the time/content interval of encounter (i.e., via concurrent evaluation). This was because participants would need to maintain information in WM for shorter intervals, reducing potential for item decay and retroactive interference (Baddeley, Thompson, & Buchanan, 1975; Cowan et al., 1992; Hart & Loomis, 1980; Jacko, 1997). Thus, Hypotheses 2a and 2b were as follows:

2a. Evaluation of performance cues would be more accurate for items occurring in the early and late segments of the encounter than for the middle segment.

2b. Performance cues occurring in the middle segment were more likely to be recognized by participants in the concurrent evaluation condition than those in the retrospective evaluation condition.

Hypotheses for the Isolation Effect. Further, isolates have been shown to command attention and receive more thorough encoding in memory than typical events (Hunt, 1995). Isolates were defined here as unexpected, atypical behaviors that may reflect negatively on a medical student's clinical performance. As such, they were likely to stand out during post-encounter performance evaluation and potentially skew participants' ratings of the entire simulated encounter. However, previous research has demonstrated that participants are able to discount prior information during a decision-making task if they have been instructed to do so (Bjork, 1972; Block, 1971; Elmes, Adams, & Roediger, 1970; Muther, 1965). By instructing participants to discount performance from previous scenario segments (which have already been subject to evaluation), focusing only on rating performance during the current segment, the impact of an isolate on subsequent performance was expected to be mitigated. Therefore, Hypotheses 3a and 3b were as follows:

3a. Isolates would be granted disproportionate weight in participants' assessments as evidenced by lower overall competency ratings from participants in the isolate group than those in the control group.

3b. An isolation effect would impact overall competency ratings for the entire segment in which the isolate was contained. The impact of an isolate in the concurrent evaluation condition would be limited to the first or last segment (the segment in which it occurred). During each evaluation period, participants would be directed to forget events from earlier segments and focus only on the current segment. Therefore, directed forgetting should have mitigated the isolation effect on subsequent segments within the same scenario. Conversely, the isolate's impact on participants' ratings in the retrospective evaluation condition would affect the entire encounter because no segmental bounds with directed forgetting were employed.

This experiment was conducted in two studies. Study I participants were recruited from Old Dominion University's undergraduate student population whereas Study II constituted a replication of Study I drawing from a pool of trained SPs. Participants in both studies based their performance assessments on passive observation of a video-recorded simulated patient encounter to improve experimental control and the feasibility of data collection as well as to promote generalizability of results from Study I to Study II and additional cross-comparison of study results.

CHAPTER II

METHOD

Study I Design

A 3 encounter segment (patient history, substance abuse, and goals) x 3 isolate (early, late, or control) x 2 evaluation format (concurrent vs. retrospective) mixed design was used in Study I. The purpose of Study I was to determine which factors significantly influence participants' abilities to accurately recognize and record key performance indicators in the form of both verbal and nonverbal behaviors. Additionally, the study was designed to explore the degree to which overall performance ratings may be affected by a singular instance of unprofessional behavior.

Encounter segment was a within-subjects variable with three levels based on the SP case sequence. Encounter segment levels included patient history, substance abuse, and goals. Isolate was a between-subjects variable with three levels; participants observed an isolate behavior early in the scenario, late in the scenario, or not at all (control). Evaluation format was a between-subjects variable with two levels. Participants in the concurrent evaluation condition rated the learner's performance periodically throughout the scenario, whereas participants in the retrospective evaluation condition waited until the entire scenario had concluded to rate the learner. Participants were randomly assigned to the aforementioned between-subjects experimental conditions to help mitigate any potential confounds.

Participants based their performance assessments on passive observation of a video-recorded simulated patient encounter. This improved experimental control and feasibility of data collection. Study II was conducted for the purpose of generalizing any significant results from Study I to the relevant target population.

Study I Participants

Seventy-one undergraduate students from Old Dominion University were recruited to participate in the study. Recruitment was limited to individuals who were at least 18 years old and reported normal or corrected vision and hearing. One participant reported difficulty understanding and following instructions and was subsequently excluded from data analysis. The final sample consisted of 70 participants (17 male, 53 female). Mean age was 24.6 years (SD = 8.87). Fourteen participants reported previous experience with simulation in either an educational or training environment, with fewer reporting a general familiarization (4.3%) or specialized training (3%) with standardized patients. Each participant was awarded 2 units of course research credit for their participation. The study was approved by the Institutional Review Board at Old Dominion University.

Study I Materials & Apparatus

Informed Consent. An informed consent form provided participants with a description of the study in addition to any foreseeable risks or benefits (Appendix A).

Demographic Form. A demographic form (Appendix B) was used to collect information from participants related to their age, sex, relevant experience, and whether they had previous experiencing working as a trained SP.

Standardized Patient Introduction Video. A short 10-minute introductory video gave participants an overview of standardized patient roles, responsibilities, and task characteristics consistent with those presented by Wallace (2007). The video assumed no prior knowledge with regard to SPs or medical student training paradigms.

Standardized Patient Practice Case. A generic SP case video segment approximately 5 minutes in duration was selected for use as a practice trial for participants. The video was selected based on its representative sample of relevant, observable verbal and nonverbal medical student behaviors.

Standardized Patient Case. The SP case presented in this study was an expanded version of *Substance Abuse Painter* (Eastern Virginia Medical School, 2012), modified to include relevant behavioral cues through purposely scripted dialogue and events (Appendix C). The case was developed by Eastern Virginia Medical School's Theresa A. Thomas Professional Skills Center for reinforcing advanced medical students' interpersonal and communication skills. The case included pertinent patient medical history, learning objectives identified specifically for the case based on topics covered, scripted dialogue for both the patient and SHP, and evaluation materials in the form of behavioral checklists and a global clinical performance evaluation sheet. It was selected for use in this study because it emphasizes patient interviewing, discussion of substance abuse concerns, and assessing motivation for change, three qualitatively similar clinical skill areas that will support subsequent comparison of participants' evaluation data. It was also sufficiently complex to represent a typical simulated patient encounter and runs approximately 18-20 minutes.

Modified Master Interview Rating Scale (MIRS) Verbal Checklist. The Master Interview Rating Scale (MIRS) is a product of the Theresa A. Thomas Professional Skills Teaching and Assessment Center of the Eastern Virginia Medical School (MIRS, 2005) designed to assess medical student interviewing and interpersonal skills. It consists of 27 Likert-type items on a 5-point scale, each addressing a unique aspect of the medical student's interpersonal conduct. A modified MIRS verbal checklist (Appendix D) was developed specifically for this study and consists of 21 of the original MIRS items (e.g., addressing the patient by their surname or excessive use of medical jargon) with a dichotomous yes/no response format for each item. Participants were asked to indicate whether or not they observed each individual item during the course of the simulated patient encounter.

Nonverbal Behavior Checklist. A checklist of relevant nonverbal behaviors (Appendix E) was developed for this study based on a sampling of published SP practices and research (Collins, Schrimmer, Diamond, & Burke, 2011; Deladisma et al., 2007; Griffith, Wilson, Langer, & Haist, 2003; Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013). The instrument is comprised of 37 dichotomous items including nonverbal behaviors such as sufficient eye contact, clearing one's throat, and interrupting the patient. Participants were asked to indicate whether they observed each item during the course of the simulated patient encounter using a standard yes/no response format. This checklist was developed specifically for use in the present study.

Global Clinical Performance Evaluation. A global clinical performance assessment form (Appendix F) was used to elicit subjective ratings of the SHP's clinical competence during each individual video segment and for the encounter as a whole. This rating represents an overall SHP performance score based on participants' personal inclinations. The form consisted of four 7-point Likert-type items ranging from 1 (poor) to 7 (excellent). The global evaluation instrument was also developed specifically for use in this study.

Debriefing Form. A short debriefing form (Appendix G) summarized the nature of the study and all variables under investigation for participants. During the debrief, participants were encouraged to ask any questions they might have had before leaving the laboratory, and provided investigators' contact information in the event that any questions were to arise at a future time.

Study I Procedure

Upon arrival, participants read and signed the informed consent form prior to participating in the study. Next they were asked to complete a short demographic form. Participants were shown a 10-minute video introducing them to the roles and responsibilities of standardized patients and were permitted to ask any questions they had about SPs. After the introductory video, participants were briefed by the experimenter about the nature of the experimental task and the performance evaluation forms they will use to periodically rate the SHP's performance. The briefing included a review of basic case details, an overview of the case presentation format, and an item-by-item review of each assessment instrument's contents (verbal and nonverbal behavioral checklists, and global clinical performance assessment).

Following the case briefing, participants watched a 5-minute SP encounter video clip and completed a subsequent performance evaluation as an orientation to both the observation-evaluation task format and the various evaluation instruments. After completing practice evaluations, participants were provided feedback to ensure they understood each instrument and were comfortable with the assessment objectives.

Simulated Patient Encounter Video Segments and Evaluations. The video recorded patient encounter began with a patient waiting in an examination room. The SHP then entered the room and conducted a patient interview, discussed possible substance dependency issues with the patient, and helped the patient identify and organize personal goals in a manner consistent with the established case details. In the concurrent evaluation experimental condition, the encounter was frozen at three predetermined points to provide participants an opportunity to evaluate the SHP. Freezes coincided with transitions from one segment of the encounter to the next (e.g., between the patient interview and discussion of substance abuse), and participants were given 6 minutes (Endsley, 1995b) to complete evaluations of the preceding encounter segment. Evaluations included a combination of verbal behavioral checklist items, nonverbal behavior checklist items and global performance ratings. Participants were instructed to evaluate the SHP only on behaviors exhibited during the segment immediately preceding, and to disregard performance from any earlier segments.

In the retrospective evaluation condition, participants evaluated the SHP only after the entire simulated encounter was complete. In this condition, participants were asked to evaluate the SHP's performance throughout the entire scenario at a single point in time (post-scenario). In both experimental conditions, the SHP exhibited a number of scripted behavioral cues (desirable and undesirable) at various points throughout the encounter. Participants were expected to observe and later recognize these behavioral cues during the evaluation period.

Debrief. At the end of the experiment, the researcher read a debriefing statement to each participant fully explaining the goals and objectives of the study. The researcher then addressed any questions or concerns the participants might have had.

Dependent Measures. Participants were asked to report specific verbal and nonverbal behavioral indicators that they observed in the video-recorded case scenario. They used a modified MIRS verbal checklist to report verbal behaviors that they observed during the scenario. Nonverbal behaviors were recorded using a nonverbal behaviors checklist developed specifically for this study. Clinical performance ratings were recorded using a global clinical performance evaluation instrument, also developed specifically for this study.

Study II Design

Study II was a replication of Study I using real SPs as research participants. All relevant questionnaires, forms, and measures were the same as those used in the first experiment.

Study II Participants

Fifty-one SPs from the Sentara Center for Simulation and Immersive Learning at Eastern Virginia Medical School were recruited to participate in the study. All SPs were at least 18 years old and reported normal or corrected vision and hearing. Two participants were excluded from final analysis due to premature withdrawal from the study. One reported difficulty understanding the audio content due to a diagnosed hearing deficiency (despite reporting normal/corrected hearing on the demographic form), whereas the other could not complete the study due to a work-related scheduling conflict. The final sample consisted of 49 participants (23 male, 26 female). Mean age was 50.25

years (*SD* = 17.06). Twenty-one participants (42.9%) reported formal acting experience. Among these, formal acting experience constituted an acting class at school (24.5%), professional acting classes (32.7%), stage acting (38.8%), musicals (32.7%), improvisational classes (30.6%), and commercial or television acting (34.7%). Standardized patient professional work experience ranged from 2 months to 17 years, with a mean of 4.87 years (*SD* = 4.75). Eleven SPs reported advanced training as an SP trainer or administrator. Each participant was financially compensated for their time at a rate of $20/hour, which is consistent with their regular hourly wage as an SP. The study was approved by the Institutional Review Boards at Old Dominion University and Eastern Virginia Medical School.

Study II Procedure

Upon arrival at the data collection site, participants were provided informed consent to the researcher and were briefed about the nature of the scenario and evaluation instruments before completing the demographic form.

Simulated Patient Encounter. The videotaped encounter was the same as that used during Study I. Participants evaluated the SHP's clinical performance in accordance with Study I's procedure. In the concurrent evaluation condition, brief scenario halts coincided with transitions from one segment of the encounter to the next (e.g., between the patient interview and discussion of substance abuse). During each scenario halt, participants were given approximately 6 minutes (Endsley, 1995b) to complete evaluations of the preceding encounter segment. In the retrospective evaluation condition, participants evaluated the SHP only after the entire simulated encounter was completed.

Debrief. As in Study I, all participants were debriefed by the researcher after the scenario concluded.

CHAPTER III

RESULTS

Study I

Correct Identifications (Verbal). An ANOVA of the verbal correct identification scores revealed a significant Segment x Evaluation interaction, $F(2,136) = 17.59, p < .001$, *partial* $\eta^2 = .21$, power = 1.0 (see Table 1). The interaction was such that segment two verbal correct identification scores for participants in the concurrent evaluation group ($M = 4.85, SD = 1.12$) were significantly greater than those of participants in the retrospective evaluation group ($M = 3.00, SD = 1.31$), $t(68) = 6.3, p < .001, d = 1.53$. Mean verbal correct identification score differences between evaluation groups were not statistically significant for segments one and three. For concurrent evaluation participants, the mean verbal correct identification score for segment two was significantly better than for segments one ($M = 4.0, SD = .87$), $t(32) = 3.44, p < .01, d = .85$; and three ($M = 3.76, SD = .94$), $t(32) = -4.79, p < .001, d = -1.06$. Retrospective evaluation participants exhibited a lower mean verbal correct identification score for segment two than in segment one ($M = 3.92, SD = .92$), $t(36) = 4.17, p < .001, d = .81$. Retrospective evaluation participants' mean segment two score was also lower than that of segment three ($M = 3.51, SD = 1.28$), $t(36) = -2.29, p < .05, d = -.39$.

Table 1

Analysis of Encounter Segment and Evaluation Format on Verbal Correct Identifications

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	24.47	1	27.47	16.15	<.001	.19
Error	115.65	68	1.70			
Within-subjects						
Encounter Segment (ES)	4.41	2	2.20	2.33	.10	.03
ES x Evaluation Format (EF)	33.28	2	16.64	17.59	<.001	.21
ES x EF within-group error	128.65	136	.95			

False Memories (Verbal). ANOVA results indicate a significant main effect for segment on participants' verbal false memory reports, $F\ (2, 136) = 20.06, p < .001$, *partial* $\eta^2 = .23$, power = 1.0 (see Table 2). Participants in both evaluation groups reported a significantly greater number of false memories in segments two ($M = 5.16$, $SD = 2.62$) $t\ (69) = -4.7, p < .001, d = -.65$; and three ($M = 5.41$, $SD = 2.51$) $t\ (69) = -6.2, p < .001, d = -.78$; than in segment one ($M = 3.71$, $SD = 1.80$).

Table 2

Analysis of Encounter Segment and Evaluation Format on Verbal False Memories

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	5.16	1	5.16	.48	.49	.01
Error	731.60	68	10.76			
Within-subjects						
Encounter Segment (ES)	115.70	2	57.85	20.06	<.001	.23
ES x Evaluation Format (EF)	1.49	2	.74	.26	.77	<.01
ES x EF within-group error	392.29	136	2.89			

Correct Identifications (Nonverbal). Due to a significant Mauchly's test, corrected degrees of freedom are reported for the encounter segment variable. ANOVA results indicate significant main effects on participants' nonverbal correct identification scores for encounter segment, F (1.83, 124.20) = 19.34, p <.001, *partial* η^2 = .22, power = 1.0; and evaluation, F (1, 68) = 38.93, p < .001, *partial* η^2 = .36, power = 1.0 (see Table 3). With regard to encounter segment, participants' nonverbal correct identification scores were significantly higher in segments one (M = 3.76, SD = 1.30) t (69) = 4.55, p < .001, d = .65; and three (M = 3.86, SD = 1.17) t (69) = -5.94, p < .001, d = -.74; than in segment two (M = 2.77, SD = 1.71). With regard to evaluation format, concurrent participants' scores were higher for segments one ($M = 4.39$, $SD = 1.06$), $t(68) = 4.34$, $p < .001$, $d = 1.05$; two ($M = 3.70$, $SD = 1.70$), $t(58.05) = 4.86$, $p < .001$, $d = 1.2$; and three ($M = 4.33$, $SD = 1.22$), $t(68) = 3.46$, $p = .001$, $d = .84$; than retrospective participants' scores in segments one ($M = 3.19$, $SD = 1.24$), two ($M = 1.95$, $SD = 1.25$), and three ($M = 3.43$, $SD = 0.96$). Concurrent participants demonstrated higher nonverbal correct identification scores in segments one, $t(32) = 2.17$, $p < .05$, $d = .49$; and three, $t(32) = -2.35$, $p < .05$, $d = -.43$; than in segment two. Retrospective participants produced more nonverbal correct identifications in segments one, $t(36) = 4.27$, $p < .001$, $d = 1.0$; and three, $t(36) = -6.43$, $p < .001$, $d = -1.33$; than in segment two.

Table 3

Analysis of Encounter Segment and Evaluation Format on Nonverbal Correct Identifications

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	86.48	1	86.48	38.93	<.001	.36
Error	151.05	68	2.22			
Within-subjects						
Encounter Segment (ES)	48.28	1.83	26.44	19.34	<.001	.22
ES x Evaluation Format (EF)	6.47	1.83	3.54	2.59	.08	.04
ES x EF within-group error	169.78	124.20	1.37			

False Memories (Nonverbal). ANOVA results indicate a significant main effect for segment on participants' verbal false memory reports, F (2, 136) = 39.14, $p < .001$, *partial* $\eta^2 = .37$, power = 1.0 (see Table 4). Participants reported a significantly greater number of false memories in segment two ($M = 3.94$, $SD = 2.19$) than in segments one ($M = 2.20$, $SD = 1.89$), t (69) = -7.41, $p < .001$, $d = -.85$; and three ($M = 2.01$, $SD = 1.56$), t (69) = 7.92, $p < .001$, $d = 1.02$.

Table 4

Analysis of Encounter Segment and Evaluation Format on Nonverbal False Memories

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	.65	1	.65	.09	.76	.00
Error	469.11	68	6.90			
Within-subjects						
Encounter Segment (ES)	156.80	2	78.40	39.14	<.001	.37
ES x Evaluation Format (EF)	1.79	2	.90	.45	.64	.01
ES x EF within-group error	272.41	136	2.00			

Global Clinical Performance Ratings. Participants rated the SHP's general clinical performance for each of the individual scenario segments and the scenario as a whole on a 7-point Likert-type scale. A repeated measures ANOVA was conducted to explore the effects of the isolate and evaluation format on global clinical performance ratings for each encounter segment and the overall patient encounter rating. Due to a significant Mauchly's test, corrected degrees of freedom are reported. Results indicate a three-way interaction between isolate, evaluation format and encounter segment, F (4.5, 144.13) = 4.58, p = .001, *partial* η^2 = .13, power = .96 (see Table 5).

Isolate Effect on Concurrent Evaluation. Planned contrasts were used to explore the hypothesized isolate effects in study one, including pairwise comparisons within the early- and late-isolate groups across segments. Comparison of mean rating differences between encounter segments indicated that early-isolate (i.e., isolate was presented in segment one) concurrent participants rated the SHP's clinical competence in segment one ($M = 2.4$, $SD = .84$) lower than they rated the SHP in segments two ($M = 4.50$, $SD = .71$), $t\ (9) = -11.7, p < .001, d = -2.7$; and three ($M = 5.70$, $SD = .68$), $t\ (9) = -15.46, p < .001, d = -4.32$. Pairwise comparisons also revealed that late-isolate concurrent evaluation participants rated the SHP's clinical competence in that segment ($M = 3.58$, $SD = 1.38$) lower than they rated the SHP in segment two ($M = 4.33$, $SD = 1.37$), $t\ (11) = 2.46, p < .05, d = .55$. However, late-isolate concurrent participants' segment three ratings did not differ significantly from segment one SHP competence ratings ($M = 3.25$, $SD = .97$), $t\ (11) = .72, p > .05, d = .28$.

A series of one-way ANOVAs was also used to investigate whether an isolate effect existed segmentally between the isolate groups. A significant effect was detected in segment 3, $F\ (2, 32) = 10.0, p < .001$, *partial* $\eta^2 = .40$, power = .98 (Table 6). Tukey post hoc analysis of segment 3 ratings indicates that participants in the concurrent evaluation group rated the SHP's performance significantly lower in segment 3 when the isolate was present in that same segment ($M = 3.58$, $SD = 1.38$) than when the isolate was presented in segment 1 ($M = 5.7$, $SD = .68$) or not at all ($M = 5.09$, $SD = 1.22$), $p < .05$.

Isolate Effect on Retrospective Evaluation. Pairwise comparisons revealed that early-isolate retrospective participants rated the SHP's competence lower during segment one ($M = 3.45$, $SD = .69$) than they did in segments two ($M = 4.55$, $SD = .52$), $t\ (10) = -$

6.71, $p < .001$, $d = -1.8$; and three ($M = 5.09$, $SD = .70$), $t\ (10) = -8.05$, $p < .001$, $d = -2.36$. Late-isolate retrospective participants rated the SHP's competence higher in segment three ($M = 4.67$, $SD = 1.5$) than in segment two ($M = 3.75$, $SD = 1.22$), $t\ (11) = -2.93$, $p < .05$, $d = -.67$. Mean SHP competence ratings for this group did not differ significantly between segments three and one ($M = 3.5$, $SD = 1.31$), $t\ (11) = -2.08$, $p > .05$, $d = -.83$.

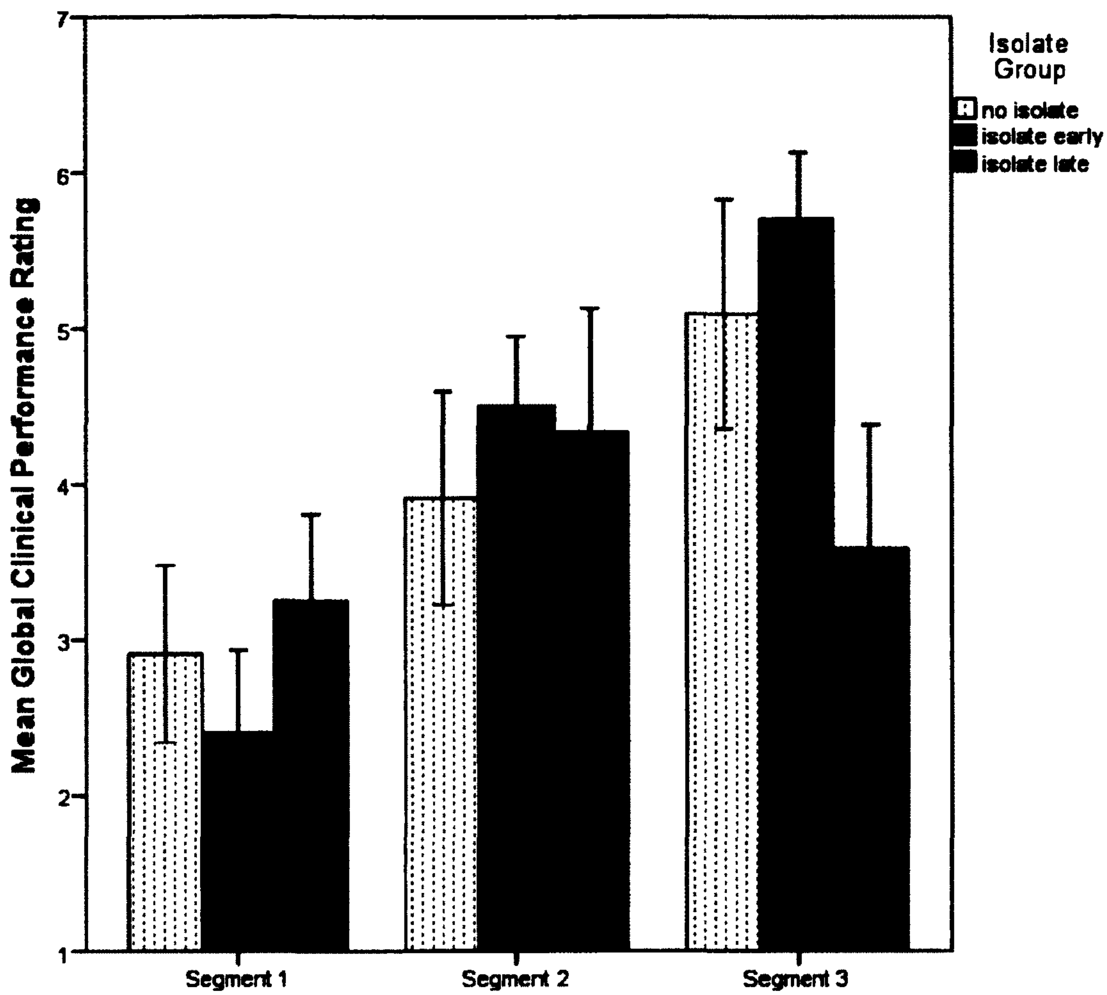

Figure 1. Participants' global clinical performance ratings by isolate condition and encounter segment for the concurrent evaluation group with error bars depicting +/- 2 SE.

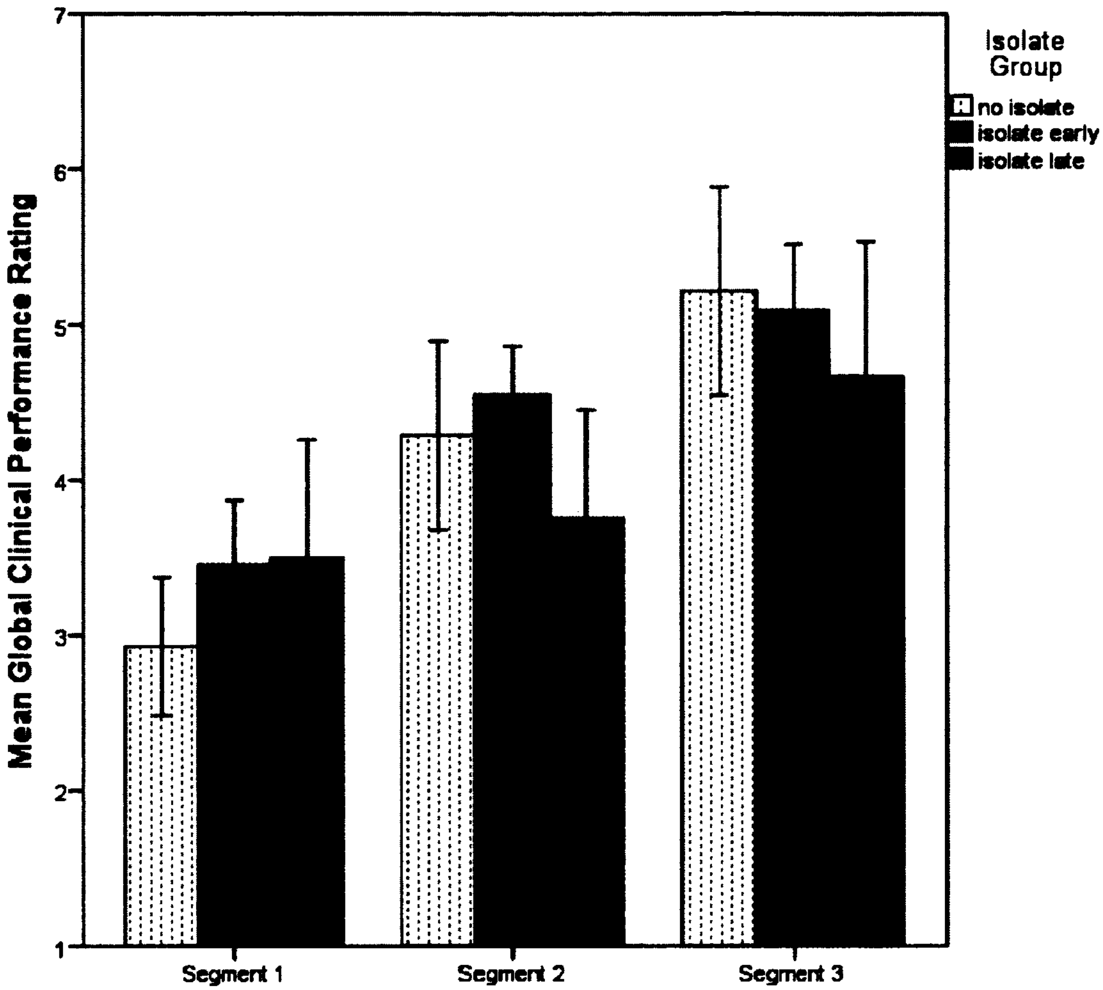

Figure 2. Participants' global clinical performance ratings by isolate condition and encounter segment for the retrospective evaluation group with error bars depicting +/- 2 SE.

Table 5

Analysis of Encounter Segment, Isolate, and Evaluation Format on Global Clinical Ratings

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Isolate	6.38	2	3.19	1.16	.32	.04
Evaluation Format	1.23	1	1.23	.45	.51	.01
I x R	.11	2	.06	.02	.98	.00
Error	176.29	64	2.75			
Within-subjects						
Encounter Segment (ES)	118.18	2.25	52.48	70.23	<.001	.52
ES x Isolate (I)	21.45	4.5	4.76	6.37	<.001	.17
ES x Evaluation Format (EF)	2.92	2.25	1.30	1.74	.18	.03
ES x I x EF	15.40	4.50	3.42	4.58	<.01	.13
Error	107.69	144.13	.75			

Table 6

Analysis of Isolate on General Clinical Ratings in Segment 3 for Concurrent Group

Source	*SS*	*df*	*MS*	*F*	*p*
Between subjects					
Between Groups	26.62	2	13.31	10.0	<.001
Within Groups	39.93	30	1.33		
Total	66.55	32			

Study II

Correct Identifications (Verbal). An ANOVA of the verbal correct identification scores revealed a significant Segment x Evaluation interaction, F (2, 94) = 3.73, $p < .05$, *partial* $\eta^2 = .07$, power = .62 (see Table 6). The interaction was such that segment two verbal correct identification scores for concurrent participants (M = 4.04, SD = 1.06) were significantly greater than those of retrospective evaluation participants (M = 2.75, SD = 1.54), t (47) = 3.43, p = .001, d = 1.0. Mean verbal correct identification score differences between evaluation groups were not statistically significant for segments one and three.

Table 7

Analysis of Encounter Segment and Evaluation Format on Verbal Correct Identifications

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	14.38	1	14.38	8.78	<.01	.16
Error	76.93	47	1.64			
Within-subjects						
Encounter Segment (ES)	7.77	2	3.89	2.85	.06	.06
ES x Evaluation Format (EF)	9.19	2	4.59	3.37	<.05	.07
ES x EF within-group error	128.04	94	1.36			

False Memories (Verbal). ANOVA results indicate a significant main effect for segment on participants' verbal false memory reports, F (2, 94) = 11.14, p < .001, *partial* η^2 = .19, power = .99 (see Table 7). Participants reported a significantly greater number of false memories in segments two (M = 4.86, SD = 2.77), t (48) = -3.53, p = .001, d = -.55; and three (M = 5.45, SD = 2.68), t (48) = -4.2, p < .001, d = -.80; than in segment one (M = 3.49, SD = 2.21). Concurrent evaluation participants reported fewer false memories during segment one (M = 2.84, SD = 1.65) than their retrospective comparators during segment one (M = 4.17, SD = 2.53), t (39.3) = -2.16, p < .05, d = -.64.

Table 8

Analysis of Encounter Segment and Evaluation Format on Verbal False Memories

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	29.46	1	29.46	2.79	.10	.06
Error	495.86	47	10.55			
Within subjects						
Encounter Segment (ES)	98.39	2	49.19	11.14	<.001	.19
ES x Evaluation Format (EF)	5.90	2	2.95	.67	.52	.01
ES x EF within-group error	415.15	94	4.42			

Correct Identifications (Nonverbal). ANOVA results indicate significant main effects on participants' nonverbal correct identification scores for encounter segment, F (2, 94) = 4. 17, p = .01, *partial* η^2 = .09, power = .78; and evaluation, F (1, 47) = 17.06, p < .001, *partial* η^2 = .27, power = .98 (see Table 8). With regard to encounter segment, participants' nonverbal correct identification scores were significantly higher in segments one (M = 3.31, SD = 1.46), t (48) = 2.41, p < .05, d = .33; and three (M = 3.43, SD = 1.29), t (48) = -2.86., p < .01, d = -.42; than in segment two (M = 2.78, SD = 1.72). With regard to evaluation format, scores of participants in the concurrent evaluation group were significantly better for segments one (M = 3.92, SD = 1.53), t (47) = 3.3., p < .01, d = .96; two (M = 3.56, SD = 1.64), t (47) = 3.65, p < .01, d = 1.06; and three (M = 3.80, SD = 1.16), t (47) = 2.13., p < .05, d = .62; than retrospective participants' scores in segments one (M = 2.67, SD = 1.09), two (M = 1.96, SD = 1.43), and three (M = 3.04, SD = 1.33). Participants in the retrospective evaluation condition exhibited significantly lower nonverbal correct identification scores in segment two than in either segment one, t (23) = 2.82., p = .01, d = .56; or three, t (23) = -3.47, p < .01, d = -.78. Concurrent evaluation participants did not exhibit this score decrement for segment two.

Table 9

Analysis of Encounter Segment and Evaluation Format on Nonverbal Correct Identifications

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation format (EF)	53.29	1	53.29	17.06	<.001	.27
Error	146.79	47	3.12			
Within-subjects						
Encounter Segment (ES)	12.07	2	6.04	4.71	<.01	.09
ES x Evaluation Format (EF)	4.40	2	2.20	1.72	.19	.04
ES x EF within-group error	120.46	94	1.28			

False Memories (Nonverbal). ANOVA results indicate a significant main effect for segment on participants' verbal false memory reports, $F(2, 94) = 23.24, p < .001$, *partial* $\eta^2 = .33$, power = 1.0 (see Table 9). Participants reported a significantly greater number of false memories in segment two ($M = 4.49$, $SD = 2.84$) than in segments one ($M = 2.69$, $SD = 2.15$), $t(48) = -4.33, p < .001, d = -.71$; and three ($M = 2.08$, $SD = 1.48$), $t(48) = 6.84, p < .001, d = 1.06$.

Table 10

Analysis of Encounter Segment and Evaluation Format on Nonverbal False Memories

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Evaluation Format (EF)	2.53	1	2.53	.30	.59	.01
Error	401.32	47	8.54			
Within-subjects						
Encounter Segment (ES)	153.03	2	76.52	23.24	<.001	.33
ES x Evaluation Format (EF)	1.03	2	.52	.16	.86	<.01
ES x EF within-group error	309.45	94	3.29			

Global Clinical Performance Ratings. Participants rated the SHP's general clinical performance for each of the individual scenario segments and the scenario as a whole on a 7-point Likert-type scale. A repeated measures ANOVA was conducted to explore the effects of the isolate and evaluation format on global clinical performance ratings for each encounter segment and the overall patient encounter rating. Due to a significant Mauchly's test, corrected degrees of freedom are reported. The analysis identified a Segment x Isolate interaction, F (4.33, 86.62) = 4.44, $p < .01$, *partial* $\eta^2 = .18$, power = .94 (see Table 10). A series of one-way ANOVAs was conducted to explore the mean ratings differences at each evaluation point by isolate condition. Results suggest that early-isolate participants rated the SHP lower during segment one (M = 2.53, SD = .64) than late-isolate (M = 3.47, SD = 1.25) and no-isolate participants (M = 3.37, SD = 1.20), $p < .05$. ANOVA results also indicated that participants rated the SHP lower in segment three (M = 3.53, SD = 1.30) when the behavioral isolate was present in that segment than when the isolate was present during segment one (M = 4.80, SD = 1.08) or not at all (M = 5.06, SD = 1.24), $p < .01$. A main effect was also discovered for evaluation, F (1, 40) = 13.82, $p < .05$, *partial* $\eta^2 = .12$, power = .64 (see Table 10). Concurrent evaluation participants generally rated the SHP lower during segment one (M = 2.64, SD = .81) than did their retrospective evaluation comparators (M = 3.52, SD = 1.31), $p < .01$.

Isolate Effect on Concurrent Evaluation. Planned contrasts were used to explore the hypothesized isolate effects. Pairwise comparisons revealed that early-isolate concurrent participants rated the SHP's clinical competence in segment one lower (M = 2.33, SD = .71) than they rated the SHP in segments two (M = 3.89, SD = 1.17), t (8) = -3.78, $p < .01$, d = -1.61; and three (M = 4.0, SD = .82), t (6) = -3.87, $p < .01$, d = -2.18. This effect was not observed when the isolate was presented in segment three, with late-isolate concurrent participants' segment three ratings showing no significant differences from segments one or two.

ANOVA results also indicated a significant between-groups isolate effect for segment 3, F (2, 22) = 8.58, $p < .01$, *partial* $\eta^2 = .46$, power = .94 (Table 11). Tukey post hoc analysis indicated that concurrent participants rated the SHP lower during segment 3 when the behavioral isolate was present in that segment ($M = 3.25$, $SD = .71$) than when the isolate was not presented at all in the scenario ($M = 5.25$, $SD = 1.28$), $p <.05$.

Isolate Effect on Retrospective Evaluation. Paired comparisons of mean ratings differences for retrospective participants across encounter segments indicated that early-isolate participants rated the SHP's segment one competence lower ($M = 2.5$, $SD = .76$) than segments two ($M = 4.38$, $SD = 1.19$), $t\ (7) = -3.42$, $p = .01$, $d = -1.88$; and three ($M = 5.5$, $SD = .76$), $t\ (7) = -7.94.$, $p < .001$, $d = -3.95$. However, this effect was not observed when the isolate was embedded in segment three; late-isolate retrospective participants' segment three competency ratings failed to differ significantly from segment one and two ratings. A series of one-way ANOVAs was conducted to explore the mean between-groups differences at each evaluation point by isolate condition. Results indicated a significant isolate effect for segment 1, $F\ (2, 22) = 5.18$, $p < .05$, *partial* $\eta^2 = .34$, power = .77 (Table 12). Tukey post hoc analysis revealed that retrospective participants rated the SHP lower during segment 1 when the behavioral isolate was present in that segment ($M = 2.5$, $SD = .76$) than when the isolate was not present at all ($M = 4.0$, $SD = 1.31$) or presented late ($M = 4.14$, $SD = 1.22$), $p <.05$.

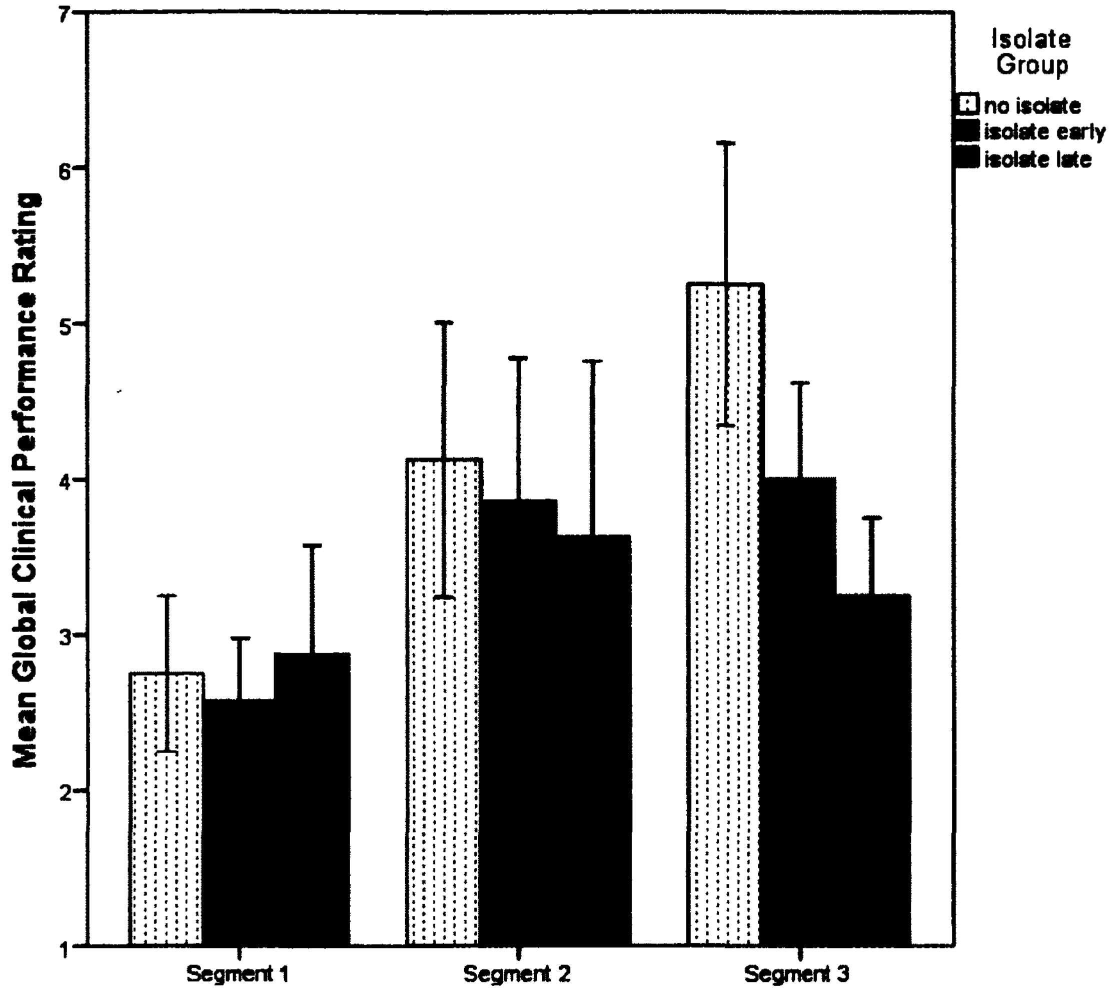

Figure 3. Participants' global clinical performance ratings by isolate condition and encounter segment for the concurrent evaluation group with error bars depicting +/- 2 SE.

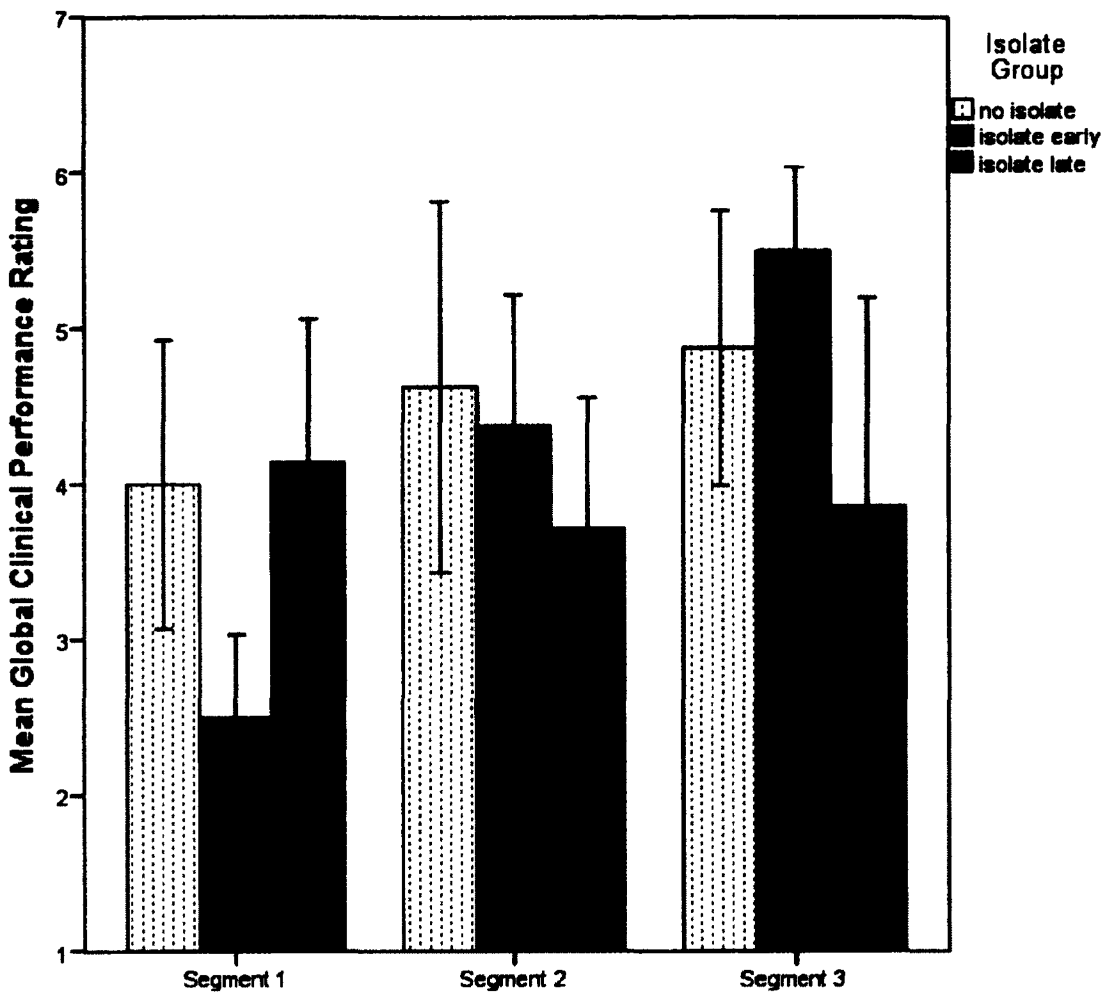

Figure 4. Participants' global clinical performance ratings by isolate condition and encounter segment for the retrospective evaluation group with error bars depicting +/- 2 SE.

Table 11

Analysis of Encounter Segment, Isolate, and Evaluation Format on General Clinical Ratings

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between-subjects						
Isolate (I)	10.97	2	5.48	2.25	.12	.10
Evaluation Format (EF)	13.82	1	13.82	5.66	<.05	.12
I x EF	.48	2	.24	.10	.91	.01
Error	97.70	40	2.44			
Within-subjects						
Encounter Segment (ES)	41.66	2.17	19.24	17.52	<.001	.31
ES x Isolate (I)	21.13	4.33	4.88	4.44	<.01	.18
ES x Evaluation format (EF)	1.34	2.17	.62	.57	.58	.01
ES x I x EF	11.43	4.33	2.64	2.40	>.05	.11
Error	95.10	86.62	1.10			

Table 12

Analysis of Isolate on General Clinical Ratings in Segment 3 for Concurrent Group

Source	*SS*	*df*	*MS*	*F*	*p*
Between subjects					
Between Groups	16.30	2	08.15	8.58	<.01
Within Groups	19.00	20	.95		
Total	35.30	22			

Table 13

Analysis of Isolate on General Clinical Ratings in Segment 1 for Retrospective Group

Source	*SS*	*df*	*MS*	*F*	*p*
Between subjects					
Between Groups	12.88	2	6.44	5.18	<.05
Within Groups	24.86	20	1.24		
Total	37.74	22			

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Study I

The first study was designed to explore the impact of evaluation format, serial position, and a behavioral isolate on verbal and nonverbal cue salience within the context of a simulated patient encounter. A set of behavioral cues representative of nontechnical clinical skills assessment was embedded into three discrete yet qualitatively similar scenario segments to serve as the basis of this study. The goal was to establish the aforementioned set of cognitive effects within a context-appropriate scenario prior to replication on a sample of standardized patient participants. Recent work has underscored the cognitive challenges associated with observation-evaluation tasks (Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013), highlighting the need for additional work aimed at enhancing clinical skills assessment.

Evaluation format and Encounter Segment on Verbal and Nonverbal Cue Reporting. It was predicted that participants who evaluated an SHP periodically throughout a simulated patient encounter (i.e., after each individual segment) would demonstrate more accurate verbal and nonverbal clinical cue recognition in general than their comparators who evaluated the SHP only after the entire encounter (i.e., all three segments) had concluded. This prediction is based on the idea that information stored in WM rapidly decays as a function of time and interference from subsequent information unless it is actively rehearsed and reactivated (Baddeley, Thompson, & Buchanan, 1975; Cowan et al., 1992). The rehearsal-reactivation process is highly demanding of limited attentional resources for information maintenance, therefore participants would only be able to attend to and later recognize a limited set of behavioral cues (Cowan 1992; 1993; 2001). Thus, to mitigate cognitive overload and subsequent information loss an operator can offload task-relevant information from WM more frequently (Hart & Loomis, 1980; Jacko, 1997).

Further, it was predicted that concurrent evaluation would be particularly beneficial for recognition of behavioral cues presented during the middle encounter segment, which represents a greater challenge for accurate evaluation due to serial position effects. It is well-established that information presented in a sequentially early or late position is more likely to be recalled by observers than information presented in the middle of a sequence, with a potential advantage for late presentation (Deese & Kaufman, 1957). These serial position effects are characteristic of long-term and working memory functions, and have been demonstrated for both verbal (Avons, Wright, & Pammer, 1994; Nairne, Riegler, & Serra, 1991; Nairne, Whiteman, & Woessner, 1995) and spatial tasks (Avons, 1998; Hay, Smyth, Hitch, & Horton, 2007; Manning & Schreier, 1988; Smyth, Hay, Hitch, & Horton, 2005). It was therefore predicted that participants would demonstrate greater difficulty in reporting accuracy for mid-segment behavioral cues, but that this effect would be mitigated for concurrent evaluation participants.

The results of Study I partially supported these hypotheses. Fewer verbal true-positive behaviors were correctly reported by retrospective evaluation participants during segment two than during segments one or three. By comparison, concurrent evaluation participants were not only able to correctly report a greater number of embedded verbal performance cues during segment two than their retrospective comparators, but also (unexpectedly) reported segment two cues more accurately than either segment one or three cues. Thus, a concurrent evaluation benefit for correct verbal cue reporting was apparent in segment two. Participants were expected to have greater difficulty recognizing information presented during segment two than in segments one and three as a result of event presentation sequence, so mitigation of the middle segment verbal recognition decrement for concurrent participants is an encouraging result because it suggests that ongoing evaluation may help offset some of the difficulty typically associated with accurate recognition of mid-sequence event sets.

Within the framework of situation awareness, Endsley (1995a) describes Level 1 SA as the basic perception of cues in the environment. Perceptual awareness in the current study was indicated by whether or not participants checked off the appropriate cues for each segment using the SHP evaluation forms provided. If a cue was both present in a segment and checked by the participant as having been present in that segment, then a basic perceptual awareness of that cue was assumed to have been established. If a cue was not perceived or initially perceived and subsequently forgotten prior to the evaluation, it was considered a Level 1 SA failure (Jones & Endsley, 1996).

Based on this assumption, Study I retrospective evaluation participants demonstrated a significant Level 1 SA decrement for segment two verbal cues that concurrent participants did not exhibit. Concurrent evaluation resulted in more accurate true-positive cue recognition than retrospective evaluation as a result of reducing the to-be-recognized information from 18 verbal and 21 nonverbal cues across the entire scenario to six verbal and seven nonverbal cues per segment. This placed the number of to-be-recognized cues for both verbal and nonverbal behaviors within the 7 ± 2 WM capacity range (Miller, 1956) for participants in the concurrent evaluation condition. Conversely, retrospective participants were required to maintain the total number of embedded verbal and nonverbal cues in WM until the scenario concluded (thus, negatively impacting the retrospective group's evaluation performance).

Once a cue has been perceived, it must be integrated via WM into a more global situation comprehension or Level 2 SA. Level 2 SA would result in formulating an impression of the SHP's segment-by-segment performance based on perceived behavioral cues. However, as observation time and mental effort associated with WM load increased with scenario progression, participants would have been more likely to rely on mental schemas to support situation comprehension. Thus, SHP evaluations would rely increasingly on generalizations in addition to direct observation (Endsley, 1988; Fracker, 1988; Gray & Fu, 2001; Freeman, Romney, & Freeman, 1987). It was therefore concluded that participants' increased reliance on a schematic framework for situation comprehension (e.g., as a result of data saturation) was evidenced by increases in the number of false memories reported and thus reflected a reduction of Level 2 SA.

All participants exhibited increases in verbal false memory reports for segments two and three. Thus, the concurrent evaluation approach yielded no significant advantage to participants in terms of mitigating false-positive reports for verbal behavioral cues. This result might be explained in part by a well-established modality effect in which verbal cues are more susceptible to false recognition and recall than visual or nonverbal cues as a result of the superior perceptual distinctiveness of the latter. In other words, mental images accompanying nonverbal behaviors perceived visually make them more salient in memory and thus more likely to be recalled than verbal behaviors that are not processed visually (Gallo, McDermott, Percer, & Roediger, 2001; Smith & Hunt, 1998). It is also assumed that participants encode both general and specific information while experiencing an event sequence, with later recall being a product of both "verbatim" and "gist" impressions (Reyna & Brainerd, 1995). If items probed during recall are consistent or plausible with a participants' general "gist" impression of an event sequence, they carry a greater probability of being falsely recalled.

A series of experiments by Payne, Elie, Blackwell, and Neuschatz (1996, Experiments 2 and 3 in particular) examined characteristics of false memory in recall, recognition, and recollection under a variety of conditions. The authors observed a false memory effect which increased in magnitude across repeated test trials under both free recall and forced recall test conditions. The authors also observed this effect in recognition-based recall and explained the increasing magnitude of false memories in terms of dependence upon an accumulation of gist-level extraction opportunities as opposed to shifting response criterion across test phases. The authors argue that subsequent testing serves to repeatedly present relevant list items to participants, thus permitting greater gist extraction of relevant information.

Further, it has been proposed that participants employ a distinctiveness heuristic whereby visual information that is unaccompanied by a "pictorial" representation in memory is considered unlikely to have actually been observed (Israel & Schacter, 1997). No such "pictorial" heuristic is available to scrutinize auditory information, which as a result is considered less perceptually distinct than visual information (Smith & Hunt, 1998). Thus, it is conceivable that as verbal cues accumulated throughout the subsequent encounter segments, participants experienced increasingly greater difficulty inhibiting sequentially plausible yet erroneous verbal cues.

Participants in Study I demonstrated significant Level 2 SA deficits during segments two and three for verbal comprehension as evidenced by an increasingly greater proportion of false memories in these segments. Increasing false memory reports suggests that as the scenario continued to unfold, relevant verbal cues became progressively less well-integrated into participants' situation comprehension for SHP performance. This was theoretically attributable to top-down schematic processing which, once activated, influenced participants' data sampling through lessening the influence of new and/or inconsistent information (Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009). Schematic processing can affect SA via perceptual screening of cues (Jones & Endsley, 1996) and information neglect for the purpose of preventing data saturation (Adams, Tenney, & Pew, 1995).

The strongest support for a concurrent evaluation benefit came from participants' nonverbal clinical cue performance data. All participants exhibited more accurate nonverbal cue reporting during segments one and three than during segment two, again supporting the idea that serial position affects the accuracy of skills assessment resulting from mid-sequence item presentation and memory characteristics. The effect was evidenced both in terms of an increase in true-positive nonverbal cue reports and fewer false-positive nonverbal cue reports during segments one and three. Regarding the latter, all participants exhibited a significant segment two spike in nonverbal false-positive reports with no benefit resulting from either evaluation format. However, concurrent participants demonstrated more accurate true-positive nonverbal cue evaluation than their retrospective comparators across all three encounter segments. The concurrent evaluation effect was therefore more robust for nonverbal than for verbal behavioral cues, an interesting result in light of recent work that identified nonverbal assessment as a particularly challenging aspect of live, interactive clinical skills assessment (Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013).

Again, participants' Level 1 SA in this study was indicated by embedded cue observations as verified against participants' behavioral observation check forms. If a cue was present and marked as such by the participant then perceptual awareness of that cue was assumed. All participants exhibited a Level 1 SA decrement for segment two nonverbal cues. Again, segment two theoretically represents a more challenging location for behavioral cue recognition as a function of its serial position. Thus, it was unsurprising that participants would be prone to erroneously reporting a greater number of false-positive cues as a result of increased reliance on schematic processing during this segment. Level 2 SA degradation for nonverbal cues was limited to segment two for all participants, supporting the notion that schematic frameworks did not play as active a role in making sense of nonverbal clinical cues except potentially during the middle segment (during which cues maintenance in WM was a greater challenge). Importantly, concurrent participants demonstrated better overall perceptual awareness of nonverbal cues than retrospective participants as evidenced by more accurate true-positive cue reporting across all three segments.

In summary, it was predicted that participants would demonstrate greater difficulty with verbal and nonverbal clinical skills assessment accuracy during segment two as a result of sequence-related salience vulnerabilities within long-term and working memory structures (Deese & Kaufman, 1957). These difficulties were expected to be most apparent in participants who were instructed to use the retrospective approach to clinical skills assessment as opposed to participants implementing a concurrent approach. The results of this study generally supported the theoretical predictions of a concurrent format. Concurrent participants did not exhibit the middle-segment accuracy decrement for verbal true-positive cue reporting characteristic of the retrospective group, and outperformed the retrospective group in terms of nonverbal true-positive cue reporting accuracy across all segments. Taken together, these results generally support the notion that periodic offloading of working memory may support a more accurate clinical skills assessment by mitigating the time- and interference-related decay of relevant performance information.

The Isolation Effect. The final set of hypotheses addressed the impact of a behavioral isolate on perceived SHP clinical competence. It was predicted that the behavioral isolate (i.e., an aggressive, unprovoked verbal patient reprimand) would be granted disproportionate weight in participants' overall clinical competency ratings of the SHP as evidenced by significantly higher ratings offered by those in the no-isolate condition as opposed to those in the isolate (early or late presentation) conditions. Interestingly, all three isolate groups rated the SHP's global clinical competence similarly regardless of the isolate's presence or location in the scenario. This suggests that a single instance of unprofessional behavior may be insufficient to significantly impact an evaluator's overall impression of clinical competence.

Further, it was predicted that participants in the isolate conditions would provide lower SHP clinical competence ratings for the individual segment in which the isolate was embedded (i.e., first or final segment) than in either of the two segments in which the isolate did not occur. It was not expected that the isolate penalty, if imposed, would extend to other segments of the scenario as a result of participants having been instructed to disregard all performance information from any other segments while rating a given segment (Hasher, Zacks, & May, 1999). The results generally supported this hypothesis. The early-isolate concurrent evaluation group rated the SHP's clinical competence lower during segment one than they did in segments two or three, suggesting that the isolate was both noticed and reflected in the group's ratings of segment one SHP clinical competence. Further, late-isolate concurrent evaluation participants rated the SHP's performance in segment 3 lower than their comparators who witnessed the isolate early or not at all. This suggests that a negative late-scenario behavioral cue is likely to have an immediate impact on the learner's evaluation score for that segment but is unlikely to carry over to affect the learner's overall score for the patient encounter scenario.

Early-isolate retrospective participants' segment one ratings were significantly lower than subsequent ratings for segments two and three, again supporting an isolate effect when the offending behavior was embedded in the first encounter segment. Late-isolate concurrent evaluation participants rated the SHP lower in segment three than they did in segment two, but not segment one. Late-isolate retrospective participants actually rated the SHP's clinical competence *higher* during segment three, despite the presence of the isolate, than they did in segment two. Contrary to the concurrent group, retrospective participants' segment 3 ratings also failed to support a significant isolate effect when compared to their retrospective early- and no-isolate comparators.

Although neither evaluation group rated the early-isolate SHP significantly lower in segment 1 than they rated the late-isolate and control SHPs, segment 1 early-isolate ratings were significantly lower than segments 2 and 3 early-isolate ratings for both evaluation conditions. When the isolate was presented early, all participants in the study penalized the SHP with a significantly lower competence rating that subsequently improved as the scenario continued to unfold. Alternately, an isolate effect was evident in the final segment only when participants evaluated the SHP using a concurrent evaluation format.

A potential explanation of this result is that retrospective participants, unable to offload working memory data throughout the scenario, were necessarily more dependent on schematic abstraction and reconstruction or gist impressions than on direct "verbatim" observation for evaluating the SHP's performance (Reyna & Brainerd, 1995; Rumelhart, 1984). Ultimately this may have resulted in the groups' ratings discrepancies for the late-isolate segment, as retrospective participants were less sensitive to the accumulation of new, and potentially conflicting, late-scenario information.

Another potential explanation resides in the nature of dispositional attribution. This process has been theorized to involve two distinct, serial stages: behavior identification and dispositional inference (Newman & Uleman, 1993; Trope, 1986). The first stage occurs early and without conscious effort (Uleman, 1989), reflecting an observer's automatic inclination to employ trait "terms" or categories as a means of identifying observed behaviors. It is not until the second stage that these trait categories are coupled via schematic activation to the observed target as relatively stable dispositional characteristics (Bassili, 1989). In other words, decisions about a target's more concrete dispositional features (e.g., clinical competence) are made only after the earlier process of behavioral identification has occurred and then only if sufficient motivation or attentional resources permit the more cognitively demanding inference process (Newman & Uleman, 1993). Otherwise, observers are unlikely to incorporate the observed behaviors into dispositional attribution.

Once schematic categories for interpreting observed behaviors are activated, information inconsistent with preliminary impressions is more likely to be disregarded or discounted rather than incorporated into a cognitively-demanding process of schematic updating or discrepancy resolution (Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009). Retrospective evaluation participants were theorized to be performing under conditions of greater working memory demand and therefore would be less likely than their concurrent evaluation comparators to revise their initial impressions of SHP clinical competence during the final encounter segment despite potentially conflicting information (Newman & Uleman, 1993). Reluctance to update initial impressions may account for the late-isolate retrospective evaluation group's higher competence ratings for the SHP in segment three than their concurrent comparators.

In the present study, Level 3 SA (i.e., projection of likely future states or events based on an understanding of preceding events) would have been indicated by differences among participants in the SHP's global clinical competence rating. The global clinical competence rating constituted an overall performance score based on a complete observation of the SHP's scenario performance and presumably represented a stable indication of patient interviewing competence. Participant ratings were expected to vary as a function of whether and where an isolate behavior was embedded in the scenario, although this result was not observed. It may have been the case that the global clinical competence rating as a measure of Level 3 SA was not quite sensitive enough to capture the more granular aspects of participants' SHP future clinical potential.

Study II

The goal of Study II was to replicate Study I findings in an operational environment within clinical education and training, drawing from a participant pool of trained standardized patients. The same videotaped encounter segments and embedded behavioral cues used in Study I served as the basis of this study. The primary research questions addressed by Study II were whether the effects obtained in Study I would generalize to a group of trained standardized patients and, if so, whether these effects contribute meaningful insight into the current medical education and training process.

Evaluation format and Encounter Segment on Verbal and Nonverbal Cue Reporting. It was again predicted that concurrent evaluation participants would demonstrate superior verbal and nonverbal clinical cue recognition than their retrospective evaluation comparators as a result of offloading task-relevant information from working memory more frequently (Hart & Loomis, 1980; Jacko, 1997). Further, it was predicted that concurrent recognition would be particularly beneficial for information presented during the middle encounter segment, which theoretically represents a greater challenge for participants in terms of memory encoding and later recall (Atkinson & Shiffrin, 1968; Azizian & Polich, 2007).

The results of Study II generally supported the stated hypotheses. Concurrent evaluation participants demonstrated more accurate true-positive verbal cue reporting during segment two than their retrospective comparators. However, participants' mean verbal true-positive cue evaluation accuracy did not differ significantly for segments one and three. Participants were expected to demonstrate greater difficulty in accurately reporting true-positive verbal cues during the second segment as a result of event serial position and memory effects (Atkinson & Shiffrin, 1968; Azizian & Polich, 2007). Thus, a concurrent evaluation benefit for segment two verbal true-positive reporting accuracy is again a very encouraging result because it suggests a potential means of offloading SP working memory to support more accurate evaluation of the more perceptually challenging mid-scenario behaviors.

As previously discussed, Level 1 SA involves basic cue perception (Endsley, 1995a) and was indicated in the current study by whether or not participants checked the appropriate segment-by-segment cues on the provided evaluation forms Study II retrospective evaluation participants experienced a Level 1 SA decrement for segment two verbal cues that concurrent evaluation participants did not exhibit. As in Study I, concurrent evaluation again resulted in more accurate true-positive cue recognition than retrospective evaluation as a result of reducing to-be-recognized information from scenario totals of 18 and 21 (verbal and nonverbal cues, respectively) to segment totals of six and seven. Concurrent participants exhibited a performance benefit during the more perceptually challenging middle segment as evidenced by improved true-positive recognition.

Level 2 SA theoretically represented participants' integrated impressions of the SHP's segment-by-segment performance based on the aforementioned behavioral cues. However, as observation time and mental effort associated with WM load increased with scenario progression, participants were more prone to activating mental schemas in support of situation comprehension. Thus, SHP evaluations relied increasingly on schematic generalizations of the SHP in addition to direct observation (Endsley, 1988; Fracker, 1988; Gray & Fu, 2001; Freeman, Romney, & Freeman, 1987). It was assumed that participants' increased reliance on a schematic framework for situation comprehension (e.g., as a result of data saturation) would be evidenced by an increase in the number of false memories reported and thus reflect a lower quality of Level 2 SA.

All Study II participants demonstrated significant Level 2 SA deficits during segments two and three for verbal comprehension as evidenced by an increasingly greater proportion of false memories in these segments. This increasing false memory rate suggests that relevant verbal cues became progressively less well-integrated into participants' situation comprehension for SHP performance as a result of schematic processing which, once activated, influenced data sampling by de-emphasizing new and/or inconsistent information (Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009). As discussed earlier, this finding may be partially attributable to a modality effect in which verbal behavioral cues have been associated with greater false memory report rates than nonverbal behavioral cues due to the inferior perceptual distinctiveness of verbal information (Gallo et al., 2001; Smith & Hunt, 1998).

Inferior perceptual distinctiveness for verbal information likely contributed to an increased reliance on "gist" impressions during subsequent recognition (Reyna & Brainerd, 1995), thus reflecting an increasingly greater challenge for participants in discounting false yet plausible items from the verbal checklist. This is supported by Payne et al. (1996), who argued that the increases in false memory magnitude over repeated test trials they obtained in multiple experiments was most likely attributable to gist-level situational processing resulting from repeated exposure to relevant list items.

Interestingly, concurrent participants exhibited fewer segment one verbal false-positive reports than their retrospective comparators, suggesting a potential concurrent evaluation benefit in terms of mitigating verbal false memory reports that was not observed in Study I. Prior training and experience may have afforded the SP participants an advantage in Study II by permitting them to scrutinize the SHP's segment one performance in greater detail than their Study I counterparts and thus maintain a higher degree of Level 2 SA for this segment than the retrospective participants. The SP participants were more familiar with the verbal checklist items used in this study than their Study I comparators, given that the items were drawn directly from the evaluation instrument commonly used within their SP cadre (i.e., the MIRS instrument). The increased level of familiarity may have enhanced verbal cue salience for SP participants in the concurrent evaluation condition, although this advantage apparently did not extend beyond segment one.

The strongest support for the concurrent evaluation benefit once again came from participants' recognition of nonverbal cues. Retrospective participants exhibited a Level 1 SA perceptual decrement for segment two nonverbal cues that their concurrent comparators did not exhibit. Concurrent participants demonstrated more accurate nonverbal true-positive cue reporting than retrospective participants across all three encounter segments. Importantly, concurrent participants did not exhibit the significant segment two dip in nonverbal true-positive cue evaluation accuracy that characterized the retrospective participants' performance. The absence of a mid-segment accuracy decrement for concurrent evaluation participants suggests that more frequent assessment of nonverbal clinical skills throughout a simulated encounter may improve evaluation accuracy overall, even for "middle segment" information that has traditionally proven challenging to accurately capture. As in Study I, concurrent participants again demonstrated better overall perceptual awareness of nonverbal cues than retrospective participants across all three segments.

Both groups exhibited a significant increase in false-positive nonverbal cues reported during segment two, with no significant mitigation of false memory reports for those in the concurrent group. Thus, the concurrent evaluation advantage for SPs was most evident in enhanced true-positive reporting accuracy for nonverbal cues throughout the entire simulated patient encounter as well as in a significant boost to true-positive recognition accuracy for verbal cues during segment two. The concurrent evaluation advantage was not as apparent in the mitigation of false memories, the only exception being for segment one verbal cues.

In summary, it was predicted that segment two would prove more challenging for participants in terms of accurate verbal and nonverbal behavioral cue reporting as a result of well-established serial position effects (Deese & Kaufman, 1957). Further, it was predicted that evaluation format would mitigate these effects, as participants in the concurrent evaluation group would be able to perform under conditions of lower cognitive demand and therefore would be able to allocate greater attention to situation perception and later recognition of specific behavioral cue data (Cowan, 1992; 1993; 2001). The results discussed here generally support these predictions.

For verbal clinical behavior cues, concurrent participants demonstrated more accurate true-positive recognition performance during segment two than their retrospective comparators. More impressively, concurrent participants outperformed their retrospective comparators across all three segments in terms of nonverbal clinical behavior cue reporting accuracy and did not exhibit the segment two accuracy decrement evidenced by retrospective participants. Again, these data support the notion that periodic offloading of working memory may support more accurate clinical skills assessment by mitigating time- and interference-related decay of relevant information.

The Isolation Effect. Again, it was predicted that the behavioral isolate would be granted disproportionate weight in participants' overall clinical competency ratings of the SHP as evidenced by significantly higher ratings offered by those in the no-isolate condition as opposed to those in the isolate conditions (Dunlosky, Hunt, & Clark, 2000; von Restorff, 1933). This was not the case, as all three isolate groups rated the SHP's global clinical competence in an equivalent manner regardless of the isolate's presence or location in the scenario. The lack of an appreciable isolate penalty as part of the SHP's overall scenario competency is consistent with Study I results and suggests again that a single instance of unprofessional behavior (regardless of where it occurs in an encounter) is unlikely to affect an SP's overall perception of an SHP's overall clinical competence.

Further, it was predicted that participants in the isolate conditions would provide lower SHP clinical competence ratings for the individual segment in which the isolate was embedded than in either of the two segments in which the isolate did not occur. It was not expected that the isolate penalty, if imposed, would extend to other segments of the scenario as a result of participants having been instructed to disregard all performance information from any other segments while rating a given segment (Hasher, Zacks, & May, 1999). The results generally supported this hypothesis.

Both evaluation groups rated the SHP's clinical competence lower in segment one than in subsequent segments when the behavioral isolate was presented early. For retrospective participants, early-isolate ratings were significantly lower than the late-isolate or no-isolate control groups, lending additional support for an early isolate effect with this group. Concurrent evaluation participants rated the SHP lower across all isolate conditions in segment 1 than their retrospective comparators, which may help explain the lack of an appreciable between-groups early isolate effect for concurrent evaluation.

Therefore, the hypothesis predicting an isolate effect for segment one was partially supported. Despite the fact that neither group rated the SHP significantly lower in segment 3 than the preceding segments when an isolate behavior was embedded late, a between-groups late isolate effect was observed for the concurrent evaluation group. Concurrent participants rated the late-isolate SHP's segment 3 performance significantly lower than the no-isolate control group. Interestingly, for concurrent participants segment 3 ratings did not differ significantly for the late- and early-isolate, suggesting that the isolate embedded in segment 1 may have enjoyed a carryover effect resulting in a segment 3 ratings penalty (albeit somewhat diminished). Regardless, neither early nor late isolate presence impacted the SHP's overall scenario evaluation rating and no late isolate effect was observed for retrospective participants.

The results suggest that the retrospective group was more reluctant than their concurrent comparators to modify impressions late in the scenario in spite of potentially conflicting data (Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009).

Retrospective impressions of SHP clinical competence were more heavily influenced by the early rather than late isolate, suggesting that impression formation is more dynamic in the initial segment than later in the encounter as a result of eventual schematic categorization and processing (Bjork, 1999). Conversely, concurrent participants penalized late-isolate SHP performance in segment 3, demonstrating a heightened perceptual awareness and sensitivity later in the scenario.

As in Study I, Level 3 SA would theoretically have been indicated by differences among participants in the SHP's global clinical competence rating. The global clinical competence rating constituted an overall performance score based on the entirety of the SHP's performance and represented a stable indication of patient interviewing competence. Participant ratings were expected to vary as a function of the presence and location of an embedded isolate behavior, but this result was in fact not observed in either study. Once again it may have been the case that as a measure of Level 3 SA the global clinical competence rating was insufficient to capture participants' projections of future SHP performance based on the limited observations provided by the current study.

General Discussion

The goal of this research was to explore the effects of evaluation format, the presence and location of a behavioral isolate, and event serial position on verbal and nonverbal clinical performance evaluations. Study I was designed to demonstrate the aforementioned effects with a convenience sample of undergraduate students, whereas Study II constituted a replication study drawing from a sample of trained standardized patients. Study II was intended to generalize results to a more contextually appropriate environment within the medical education and training domain. The hypotheses outlined for both studies were generally supported, but several unexpected results were obtained.

In both studies, verbal true-positive cue reporting generally adhered to the predicted response patterns. Participants performed equally well in segments one and three, likely the result of information from those segments being more salient and thus accessible (especially critical for retrospective participants) in long-term and working memory (Deese & Kaufman, 1957). Retrospective participants exhibited a mid-segment perceptual awareness decrement that was mitigated for concurrent participants. This decrement was attributed to the latter group's ability to periodically offload task-critical data held in working memory—information that became less salient to retrospective participants over time as a result of temporal interference and decay (Baddeley, Thompson, & Buchanan, 1975; Cowan et al., 1992).

Interestingly, both groups (across studies) produced an escalating pattern of verbal false-positive recollections that was not observed in nonverbal cue recognition. This suggests that as the scenario progressed, all participants became more reliant on schematic activation for processing verbal cues and thus less reliant on direct observation. As a result, participants' Level 2 SA was considered less well-integrated and reliable in terms of evaluative accuracy. A potential explanation for this finding is that verbal cues are perceptually less distinct than nonverbal cues and therefore represent a greater challenge for participants to discount as the number of plausible yet unobserved items accumulates over repeated trials (Gallo, McDermott, Percer, & Roediger, 2001; Smith & Hunt, 1998). Assuming that participants dual-encode information in verbatim and gist levels of detail (Reyna & Brainerd, 1995), participants were more likely to depend on gist reconstruction of lower-saliency verbal information and thus more prone to recall errors resulting from insufficient heuristic discrimination (Israel & Schacter, 1997).

Conversely, participants' nonverbal true-positive responses appeared to be more directly related to the observed false-positive patterns in both studies. Nonverbal true-positive cue reporting accuracy generally adhered to the predicted response pattern, with concurrent participants in both studies outperforming retrospective participants across all three segments. Study I participants in both groups and Study II retrospective participants exhibited a mid-segment nonverbal true-positive accuracy decrement that Study II concurrent evaluation participants did not exhibit. Thus, the concurrent evaluation format appears to be particularly effective for enhancing Level I SA for nonverbal cues as indicated by improved true-positive recognition.

Contrary to participants' verbal false-positive reports, the false-positive response pattern for nonverbal cues in both studies reflected an inverse of true-positive accuracy with a spike in false recollections during segment two. Thus, Level 2 SA for integration of nonverbal cues was superior for concurrent participants in both studies but was still affected to some degree by cue serial position. This would be expected if participants were relying more heavily on schematic or gist processing for mid-segment recollections. Challenges associated with mitigating false memories would be expected to play a greater factor in the middle encounter segment where data integrity is most vulnerable.

Verbal cues are likely more susceptible to schematic abstraction and reconstruction than nonverbal clinical cues as a result of the superior perceptual salience of the latter (Gallo, McDermott, Percer, & Roediger, 2001; Smith & Hunt, 1998), thus resulting in poorer Level 2 SA for verbal cues across evaluation groups in this study. Inaccurate but conceptually consistent events were more likely to be reported as potential for or reliance on schematic retrieval increased, with contextually inconsistent information more likely to be dismissed or disregarded (Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009).

Another contributing factor could be that most of the verbal cues presented in this study represented a natural, intuitive sequence of events within the context of a clinical interview such as introducing oneself to the patient and eliciting a timeline of the chief complaint (expected to be conducted early in an interview), discussing a variety of health issues (more likely occurring mid-sequence) and setting goals or providing closure (generally done immediately prior to concluding a typical medical interview). These events may have been more amenable to categorization in a manner conforming to an expected sequence during a clinical interview, thus supporting schematic abstraction and integration. These schematic encoding processes are not likely supported by nonverbal clinical cues (e.g., fist clenching, crossing arms or legs, and smiling at the patient), which are no more likely to occur early in the scenario as opposed to late. Thus, when asked to recognize critical performance cues for each segment, participants were less dependent on schematic abstractions for nonverbal cues than for verbal cues.

It was also predicted that participants would penalize the SHP for exhibiting an unexpected, unprofessional behavior which in this case happened to be an unjustified verbal tirade directed at the patient. When the isolate occurred early in the scenario, participants penalized the SHP for his performance in the first segment as expected. This effect was especially robust for Study II retrospective participants, in which SHP ratings for early isolate behavior registered significantly lower than retrospective late- and no-isolate control groups. However, when the isolate occurred later in the scenario, retrospective participants' ratings were largely unaffected. This reluctance on the part of retrospective participants to impose a late-isolate penalty is potentially due to the way we form impressions of others, in particular over a very brief period of time when the implied goal is to evaluate the other's ability or competence.

Dispositional attribution is a two-stage process involving automatic interpretation of behaviors through schematic categorization followed by the more cognitively demanding stage of binding behavioral interpretations to a more stable set of dispositional inferences (Newman & Uleman, 1993; Trope, 1986). Once schematic categories considered sufficient to explain observed behavior have been activated, information inconsistent with the observer's initial impressions is more likely to be disregarded or discounted rather than incorporated into a cognitively-demanding process of schematic updating or discrepancy resolution (Bjork, 1999; Bransford & Franks, 1971; Waggoner, Smith, & Collins, 2009). It is therefore conceivable that late-isolate presentation presents a greater challenge for those in a traditional retrospective evaluation framework to incorporate into their ratings as a result of the isolate behavior's lack of consistency with traits categorized under alternative schemata activated by earlier observations (e.g., competent, helpful, caring).

Study Limitations. There are several important limitations to this research that should be noted. First, the simulated patient encounter was primarily an observation task in which participants passively viewed a video-recorded patient interview. Participants did not interact with the SHP directly at any time. However, simulated patient encounters often require direct interaction between the learner and evaluator (i.e., standardized patient) in support of a portrayal-observation educational model. For example, a standardized patient would observe a learner's clinical performance while realistically portraying a sick patient. Thus, cognitive workload demands associated with active portrayal, observation, and improvisation tasks were not imposed in the current study. Although it is conceivable that the effects obtained here would be amplified with increased task load and mental demand, examining such a possibility was beyond the scope of the present study.

A second factor is that the presentation order of the encounter segments themselves was not counterbalanced. Randomizing the presentation order of encounter segments would have challenged the comprehensible "flow" of scripted events. Thus, segments were presented sequentially (patient history, substance abuse, future goals) for logical continuity. Lack of randomization could unintentionally introduce order effects into the response patterns of participants, but in this case a predetermined presentation order was used in order to preserve the integrity of a realistic patient encounter evolution. Additionally, all three encounter segments focused on one aspect of a patient encounter—the patient interview. This was intentionally done to ensure all three segments were qualitatively similar for investigative purposes, but effectively precluded examination of other equally important patient encounter areas such as a physical examination.

Another study limitation involves the complexity of detecting individual behavioral cues embedded in each segment of the scenario. Although the cues used in this study were not calibrated and balanced against an index of perceptual complexity, certain steps were taken to ensure that detection complexity for each cue was minimized.

Whenever possible, verbal cues were articulated by the SHP using the same language specified on the verbal behavior checklist to reduce ambiguity. For example, when the SHP was directed to demonstrate empathy toward the patient, the keyword "empathy" served as a marker of the corresponding verbal behavior *empathetic tone of voice* (SHP: "I can empathize"). For both verbal and nonverbal behavioral cues, an independent, naïve third-party reviewer was asked to view each video segment and verify the saliency of embedded cues by completing the corresponding checklists prior to data collection. The reviewer's checklist results were then used to validate the embedded behavioral cues by establishing a 100% agreement between the two lists. Thus, verbal and nonverbal clinical performance cues were comparably salient across segments.

Due to the nature of the evaluation formats studied, some participants were exposed repeatedly to critical verbal and nonverbal clinical behaviors by means of repeated evaluation throughout the scenario. All participants were briefed on the relevant checklist items and were given a review and explanation of each individual item prior to beginning the study, so all participants were provided the same basis for the evaluation criteria. However, the concurrent evaluation participants were asked to evaluate the SHP twice during the encounter, providing a working memory "refresher" in the form of multiple exposures to the evaluation instruments (and thus relevant clinical behaviors) that retrospective participants did not receive.

Regarding the evaluation instruments used in this study, lists of verbal and nonverbal clinical performance items were presented alongside dichotomous "yes/no" checklist options. Thus, performance was more reflective of participants' recognition processes than memorial search and retrieval processes; however, this response format is consistent with typical simulated patient encounter assessments in which evaluators are asked to conduct assessments using a list of relevant behaviors as a memory cue (e.g., Surgical Resident OSCE evaluation instruments; American College of Surgeons, 2008). For the sake of realism as well as practicality, the recognition-based evaluation aspect was maintained for this study. Additionally, simulated patient encounters also often incorporate more complex assessment instruments designed to elicit not only whether a behavior occurred but also the perceived quality of the behavior (e.g., using a Likert-type scale). This study was designed to address questions related to the salience of certain cues in memory rather than perceived quality or inter-rater reliability in the assessment of clinical behaviors and therefore contained a simpler set of evaluation instruments.

Similarly, the study's methodology emphasized aspects of participant memory function and did not permit investigation of the more basic perceptual level of performance. That is, participants who failed to check off an appropriate checklist item were assumed to have forgotten the event or lacked confidence in their recollection as opposed to having not perceived it at all. However, the study was not designed to discriminate between the levels of performance. Participants were instructed to pay careful attention to the video and audio content while the videos were running, and appeared to do so, to promote the best possible conditions for item perception.

Regarding scenario and task realism, all participants were given a brief overview of the clinical evaluation instruments and case parameters immediately prior to participation. However, an actual simulated patient encounter would likely involve more preparation time on behalf of the evaluators. This would include sufficient time to memorize all relevant case materials and evaluation instruments, and to discuss in-depth the case, goals, objectives, and any symptoms/conditions to be portrayed with responsible training personnel. They might also be given the opportunity to conduct some dry-run rehearsals. Due to time limitations, experimental logistics and resource availability, this level of advanced preparation was not practical. However, participants were given a brief overview and some preparation time before the study began, and all participants indicated that they fully understood the task requirements and evaluation items prior to observing the videotaped scenario. Further, participants were responsible for both identifying specific clinically-relevant behaviors as well as the segment in which they occurred. Generally it would not be critical for evaluators to maintain the temporal aspects of item occurrences in memory, but this was nonetheless a requirement of the present study.

Last, the actors used to portray the SHP and patient being interviewed were drawn from the same cadre of healthcare training professionals as the SP participants in Study II. As a result, the majority of SP participants polled indicated that they recognized the actors (at least vaguely) on screen. This is not a factor expected to significantly influence performance, as Study II participants were trained healthcare evaluators and commonly find themselves evaluating each other for practice and internal quality control. During such instances, they would be required to set aside their own knowledge of the other's basic mannerisms and personality features in order to view them as a viable patient character for the purpose of evaluation. Nevertheless, the familiarity of most Study II participants with the on-screen actors constitutes a necessary consideration and potential study limitation.

Conclusions

In recent years, healthcare educators have increasingly begun to emphasize nontechnical skills training and accountability in light of data linking poor professional communication and leadership to extreme rates of preventable medical errors (Jewell & McGiffert, 2009; Kohn, Corrigan, & Donaldson, 1999) in addition to a range of negative patient outcomes (Duggan & Parrott, 2000; Flocke, Miller, & Crabtree, 2002; Stewart, Meredith, Brown, & Galajda, 2000; Wooford et al., 2004). Leaders in the medical community have called for a culture shift in which mutual respect and open communication among healthcare professionals will be an established standard for both educational settings and practice (Leape et al., 2012; NPSF, 2010; Shostek, 2007). Thus, existing nontechnical skills programs for healthcare professionals will continue to expand and evolve along with the needs of the healthcare community.

One of the most uniquely well-suited and widely adopted nontechnical clinical skills training paradigms in recent years has been the simulated patient encounter drawing on the expertise of trained SPs (Wallace, 2007). Standardized patients are optimal for simulation-based nontechnical skills training because they represent a high degree of psychological fidelity for learners honing requisite communication skills for activities such as clinical interviews, communicating difficult diagnoses, and disclosing mistakes. Further, the feedback generated by SPs is critical for continued development and growth of the medical student's nontechnical clinical skills.

However, certain aspects of the SP's roles within a simulated patient encounter make them susceptible to a variety of psychological effects which may influence the accuracy of feedback and evaluations. Previous research has addressed the reliability of

SP evaluations and found it to be quite high (Colliver & Reed, 1993; De Champlain, Margolis, King, & Klass, 1997; Elliott & Hickam, 1987; Furman, 2008; Heine et al., 2003; Pangaro et al., 1997; Williams, 2004); yet evidence also suggests that some of the scoring variance detected is directly attributable to the individual SPs (van Zanten, Boulet, & McKinley, 2007) and that SPs may be susceptible to cognitive challenges associated with the nature of observation and the conditions under which clinical behaviors are presented (Newlin-Canzone, Scerbo, Gliva-McConvey, & Wallace, 2013). It therefore makes sense to continue investigating factors known to affect the various cognitive processes involved in SPs' multifaceted role in medical education.

This study was designed to explore the influence of serial position and evaluation format on verbal and nonverbal clinical cue recognition accuracy. It was also designed to explore the isolation effect on participants' overall ratings of SHP clinical performance. A video-recorded simulated patient encounter served to have participants (naïve undergraduate students in Study I and trained standardized patients in Study II) observe three scenario segments and attempt to recognize critical performance cues from each segment during evaluation. Participants in the concurrent evaluation framework demonstrated an advantage in terms of more accurate mid-segment recognition for both verbal and nonverbal embedded cues than their retrospective comparators.

Middle segment verbal cues were more accurately detected and subsequently reported by participants who were able to offload working memory information periodically as opposed to participants who retained the information throughout the scenario. Although a mid-segment decrement was observed for concurrent participants' embedded nonverbal cue reporting accuracy, they were still able to report cues more accurately in all three segments than their retrospective comparators. Thus, a potential means of enhancing the traditional simulated patient encounter model would be to stop scenarios at predetermined points to allow for periodic SP evaluation.

As previously discussed, all participants demonstrated a pattern of increasing verbal cue false memories across the scenario. This was interpreted as increasing reliance on mental schemata for processing SHP performance details as the scenario unfolded. Trained SPs would be expected to hold sophisticated frameworks in LTM for scenario-based patient encounters, so it is unsurprising that categorization of SHP performance would occur as scenario duration and SHP data increase. Yet even naïve undergraduate students would be expected to hold a mental schema for a typical doctor's visit, so it was also unsurprising that Study I participants exhibited the same pattern for verbal false memories.

For nonverbal cues, all participants demonstrated a pattern of false memories inverse to that of the correctly identified nonverbal embedded cues, suggesting that actual observations were employed more consistently across the segments than mental schemata. This suggests that nonverbal clinical cues may be less prone to schematic encoding and subsequent integration. Relevant nonverbal clinical cues included behaviors such as checking wristwatch, yawning, or taking an aggressive tone with the patient. Although important items for clinical performance evaluation, there is an inherent randomness to the potential ordering of nonverbal items that does not hold for verbal items. For example, a doctor would be expected to introduce himself (verbal cue) at the beginning of an appointment, though he would be no more likely to clench his fist (nonverbal cue) at the beginning of the appointment as opposed to the middle or end of the appointment. Nonverbal behaviors represent cues that are unlikely to benefit from use of a schematic framework for recognition, and thus require greater observer effort to maintain performance throughout the scenario. As a result, false memories for nonverbal behaviors did not increase as scenario duration and experience with the SHP increased, but instead reflected the inverse of participants' embedded nonverbal cue reporting accuracy across all segments.

This study demonstrated how periodic evaluation may result in more accurate verbal and nonverbal behavioral reports from observers. It was also designed to explore the potential of an isolation effect on overall and segment-by-segment SHP clinical competence ratings. With regard to the former, an embedded behavioral isolate had no appreciable impact on participants' global SHP competence scores. This suggests that an otherwise successful performance on behalf of an SHP is unlikely to be negatively impacted by a single isolated act of unprofessionalism. Such behaviors are more likely to show up in an SHP's segment-by-segment performance ratings, particularly when the behavior occurs early rather than late in the segment for those using a traditional retrospective format for evaluation. Further, the appearance of a robust late-segment isolate effect was present for concurrent evaluation participants in both studies, suggesting that retrospective evaluation is generally less sensitive to a late-isolate effect than concurrent evaluation. For the formative evaluation, it is important that negative behaviors (e.g., inappropriate comments) be addressed. To be sure, addressing unprofessionalism, intimidation, and fear-based communication early on in a medical professional's career is now considered one of the most significant issues for the emerging culture shift in healthcare (Leape et al., 2012; NPSF, 2010; Shostek, 2007).

With this in mind, the present study demonstrates a potential method for enhancing reporting accuracy for clinical performance indicators within the context of a simulated patient encounter. Shorter encounters or segmented scenarios with periodic evaluation opportunities could promote a higher degree of behavioral specificity (and thus increased training efficacy) for formative evaluation and facilitate targeting specific behaviors and skills for improvement. On the other hand, a potential benefit of lengthier, uninterrupted simulated patient encounters is that they may promote a more stable summative evaluation emphasizing the healthcare professional's general clinical competence and interpersonal demeanor. The retrospective evaluation format could therefore be instituted when global assessment or an overall impression of healthcare professionals' clinical capabilities are of prime importance. Regardless, the specific encounter format implemented should be aligned with the purpose and goals of the simulation exercise.

One of the most important allowances of a simulation-based training and education program is a safe environment in which mistakes lead to learning opportunities rather than negative consequences. In healthcare, this principle applies not only to protection of the patient, but also to permitting learners to practice without fear of penalty for moments requiring further process improvement. This study demonstrated that in the case of isolated unprofessionalism, which is to be expected of learners trying out various approaches to effective interpersonal communication, this benefit is maintained through minimal impact on observers' ratings. Along these lines, another promising area for future research may include introducing a variety of isolate behaviors, both verbal and nonverbal, into clinical training scenarios to determine what effect specific types of negative behaviors have on otherwise positive SHP performance. Additionally, the effects of a positive isolate on otherwise poor SHP performance may yield informative results regarding the nature and resilience of schematic processing during clinical evaluation.

REFERENCES

Adamo, G. (2003). Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Medical Teacher, 25*, 262-270. doi: 10.1080/0142159031000100300

Adams, M.J., Tenney, Y.J., & Pew, R.W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104.

Aggarwal, R., Leong, J., Leff, D., Warren, O., Yang, G.-Z., & Darzi, A. (2009). New technologies for the surgical curriculum. *World Journal of Surgery, 32*, 213-216.

American College of Surgeons (2008). *Surgery Resident Objective Structured Clinical Examination Manual with Examination Materials.* DVD and CD-ROM available from: https://web4.facs.org/ebusiness/ProductCatalog/product.aspx?ID=124

Anderson, C. A., & Huesmann, L. R. (2003). Human aggression: A social-cognitive view. In M. A. Hogg & J. Cooper (Eds.), *Handbook of social psychology* (pp. 296–323). London: Sage Publications.

Aretz, A.J. (1983). A comparison of manual and vocal response modes for the control of aircraft subsystems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 97-101. doi: 10.1177/154193128302700127

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation: II*, 249.

Avons, S.E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology, 89*, 285-308. doi: 10.1111/j.2044-8295.1998.tb02685.x

Avons, S.E., Wright, K.L., & Pammer, K. (1994). The word-length effect in probed and serial recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 47*, 207-231. doi: 10.1080/14640749408401151

Azizian, A. & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of Cognitive Neuroscience, 19*, 2071-2081. doi: 10.1162/jocn.2007.19.12.2071

Baddeley, A.D. (1986). *Human Memory.* Oxford: Clarendon Press.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A.D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173. doi: 10.1037/0033-295X.105.1.158

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A.D., Hitch, G.J., & Allen, R.J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61*, 438-456.

Baddeley, A.D. & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589. doi: 10.1016/S0022-5371(75)80045-4

Baddeley, A.D. & Wilson, B.A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia, 40*, 1737-1743.

Banbury, S. P. & Tremblay, S. (2004). *A cognitive approach to situation awareness: theory and application.* Ashgate Publishing.

Banbury, S. P., Andre, A.D., & Croft, D.G. (2001). Do we "know" all that we "know"? The role of implicit knowledge in situation awareness. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society,* San Diego.

Barnett, B., Stokes, A., Wickens, C.D., Davis, T., Rosenblum, R., & Hyman, F. (1987). A componential analysis of pilot decision-making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 842-846. doi: 10.1177/154193128703100735

Barrows, H.S. (1987). *Simulated (standardized) patients and other human simulations.* Chapel Hill, NC: Health Sciences Consortium.

Barrows, H.S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine, 68*, 443-451.

Barrows, H.S. & Abrahamson, S. (1964). The programmed patient: A technique for appraising student performance in clinical neurology. *Journal of Medical Education, 39*, 802-805.

Bartlett, F.C. (1932). *Remembering: An experimental and social study.* Cambridge: Cambridge University Press.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 603–616.

Bassili, J. N. (1989). Trait encoding in behavior identification and dispositional inference. *Personality and Social Psychology Bulletin, 15*(3), 285-296. doi: 10.1177/0146167289153001

Bäuml, K.-H. (2008). Inhibitory processes. In J. H. Byrne & H. L. Roediger III (Eds.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 195-220). Amsterdam: Elsevier.

Bellezza, F.S. & Cheney, T.L. (1973). Isolation effect in immediate and delayed recall. *Journal of Experimental Psychology, 99*, 55-60. doi: 10.1037/h0034756

Bjork, R.A. (1972). Theoretical implications of directed forgetting. In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217-235). Washington, DC: Winston & Sons.

Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application. Attention and performance* (pp. 435-459). Cambridge, MA: The MIT Press.

Block, R.A. (1971). Effects of instructions to forget in short-term memory. *Journal of Experimental Psychology, 89*, 1-9. doi: 10.1037/h0031190

Boles, D. B. (2002). Lateralized spatial processes and their lexical implications. *Neuropsychologia, 40*, 2125–2135.

Boles, D. B., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007). Predicting dual-task performance with the multiple resource questionnaire. *Human Factors, 49,* 32–45.

Boulet, J.R., Smee, S.M., Dillon, G.F., & Gimpel, J.R. (2009). The use of standardized patient assessments for certification and licensure decisions. *Simulation in Healthcare, 4*, 35-42.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology*, *2*(4), 331-350. doi: http://dx.doi.org/10.1016/0010-0285(71)90019-3

Brewer, W.F. & Treyens, J.C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207-203.

Broadbent, D. E. (1958). *Perception and communication.* New York: Oxford University Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21. doi: 10.1080/17470215808416249

Cavusoglu, Tendick, & Sastry, 2002- Simulation-Based Training in Healthcare header

Cegala, D.J. & Broz, S.L. (2002). Physician communication skills training: A review of theoretical backgrounds, objectives and skills. *Medical Education, 36*, 1004-1016.

Chan, D.K., Gallagher, T.H., Reznick, R., & Levinson, W. (2005). How surgeons disclose medical errors to patients: A study using standardized patients. *Surgery, 138*, 851-858.

Collier S.G. & Folleso, K. (1995). SACRI: a measure of situation awareness for nuclear power plant control rooms. In: D.J. Garland & M.R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 115-122). Daytona Beach, FL: Embry-Riddle University Press.

Collins, L.G., Schrimmer, A., Diamond, J., & Burke, J. (2011). Evaluating verbal and nonverbal communication skills in an ethnogeriatric OSCE. *Patient Education and Counseling, 83*, 158-162.

Colliver, J.A. & Reed, W.G. (1993). Session two: Technical issues: Test application. *Academic Medicine, 68*, 454-460.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163-191. doi: 10.1037/0033-2909.104.2.163

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition, 21*, 162-167. doi: 10.3758/BF03202728

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-114.

Cowan, N. (2005). *Working memory capacity: Essays in cognitive psychology.* New York: Psychology Press.

Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1-17. doi: 10.1016/0749-596X(92)90002-F

De Champlain, A.F., Margolis, M.J., King, A., & Klass, D.J. (1997). Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine, 72*, S85-S87.

Deese, J. & Kaufman, R.A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54*, 180-187. doi: 10.1037/h0040536

Deladisma, A.M., Cohen, M., Stevens, A., Wagner, P., Lok, B., et al. (2007). Do medical students respond empathetically to a virtual patient? *The American Journal of Surgery, 193*, 756-760.

Deutsch, J.A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80-90.

Duggan, A.P. & Parrott, R.L. (2000). Research note: Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. *Human Communication Research, 27*, 299-311.

Dunlosky, J., Hunt, R.R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 649-657. doi: 10.1037/0278-7393.26.3.649

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In: Durso, Francis T. (Ed), *Handbook of applied cognition* (pp. 283-314). New York, NY, US: John Wiley & Sons Ltd.

Durso, F.T., Hackworth, C.A., Truitt, T., Crutchfield, J., & Manning, C.A., (1998). Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Quarterly, 6*, 1-20. doi: 10.1207/s15327108ijap0802_6

Eastern Virginia Medical School (2012). *Substance Abuse Painter*. Unpublished training scenario and case details.

Elliott, D.L. & Hickam, D.H. (1987). Evaluation of physical examination skills: Reliability of faculty observers and patient instructors. *Journal of the American Medical Association, 258*, 3405-3408.

Elmes, D.G., Adams, C., & Roediger, H.L. (1970). Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology, 86*, 103-107.

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Aerospace Systems* (pp. 97-101). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M.R. (1990a). A methodology for the objective measurement of situation awareness. In *Situation awareness in aerospace operations* (AGARD-CP-478; pp. 1/1-1/9). Neuilly-Sur-Seine, France: NATO—Advisory Group for Aerospace Research and Development.

Endsley, M.R. (1990b). Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 41-45). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M.R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Endsley, M.R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65-84.

Endsley, M.R. (2001). Designing for situation awareness in complex systems. In *Proceedings of the Second international workshop on symbiosis of humans, artifacts and environment* (pp. 1-14). *Kyoto, Japan.*

Endsley, M.R. & Garland, D.J. (2000). *Situation awareness: Analysis and measurement.* Mahwah, NJ: Lawrence Erlbaum.

Endsley, M.R. & Rodgers, M.D. (1998). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. *Air Traffic Control Quarterly, 6*, 21-44.

Endsley, M.R., Selcon, S.J., Hardiman, T.D., & Croft, D.G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M.R. & Smith, R.P. (1996). Attention distribution and decision making in tactical air combat. *Human Factors, 38*, 232-249.

Endsley, M.R., Sollenberger, R., & Stein, E. (2000). Situation awareness: A comparison of measures. In *Proceedings of the Human Performance, Situation Awareness, and Automation: User Centered Design for the New Millennium.* Savannah, GA: SA Technologies, Inc.

Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. doi: 10.1111/1467-8721.00160

Epstein, R.M. (2007). Assessment in medical education. *New England Journal of Medicine, 356,* 387-396.

Epstein, R.M., Franks, P., Fiscella, K., Shields, C.G., Meldrum, S.C., Kravitz, R.L., & Duberstein, P.R. (2005). Measuring patient-centered communication in patient-physician consultations: Theoretical and practical issues. *Social Science & Medicine, 61*, 1516-1528.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245. doi: 10.1037/0033-295X.102.2.211

Errichetti, A.M., Gimpel, J.R., & Boulet, J.R. (2002). State of the art in standardized patient programs: A survey of osteopathic medical schools. *Journal of the American Osteopathic Association, 102*, 627-631.

Fennema, M.G. & Kleinmuntz, D.N. (1995). Anticipations of effort and accuracy in multiattribute choice. *Organizational Behavior and Human Decision Processes, 63*, 21-32. doi: 10.1006/obhd.1995.1058

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*(1), 149-157.

Flocke, S.A., Miller, W.L., Crabtree, B.F. (2002). Relationships between physician practice style, patient satisfaction, and attributes of primary care. *The Journal of Family Practice, 51*, 835-840.

Fracker, M.L. (1988). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Aerospace Systems* (pp. 102-106). Santa Monica, CA: Human Factors and Ergonomics Society.

Fracker, M.L. (1989). Attention allocation in situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Visual Performance* (pp. 1396-1400). Santa Monica, CA: Human Factors and Ergonomics Society.

Freeman, L.C., Romney, A.K., & Freeman, S.C. (1987). Cognitive structure and informant accuracy. *American Anthropologist, 89*, 310-325.

Furman, G.E. (2008). The role of standardized patient and trainer training in quality assurance for a high-stakes clinical skills examination. *The Kaohsiung Journal of Medical Sciences, 24*, 651-655.

Gallagher, A.G., Ritter, E.M., Champion, H., Higgings, G., Fried, M.P., Moses, G., Smith, D., Satava, R.M. (2005). Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery, 241*, 364-372.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger III, H. L. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 339. doi: 10.1037/0278-7393.27.2.339

Garsoffky, B., Schwan, S., & Hesse, F.W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1035-1050.

Gathercole, S. & Baddeley, A.D. (1993). *Working memory and language: Essays in cognitive psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Geiselman, R.E. & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition, 13*, 57-62. doi: 10.3758/BF03198444

Gimpel, J.R., Boulet, J.R., & Errichetti, A.M. (2003). Evaluating the clinical skills of osteopathic medical students. *Journal of the American Osteopathic Association, 103*, 267-279.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, *5*, 351-360.

Goernert, P.N. & Larson, M.E. (1994). The initiation and release of retrieval inhibition. *The Journal of General Psychology, 121*, 61-66. doi: 10.1080/00221309.1994.9711172

Gorman, P.J., Meier, A.H., Rawn, C., & Krummel, T.M. (2000). The future of medical education is no longer blood and guts, it is bits and bytes. *The American Journal of Surgery, 180*, 353-356.

Gray, W. D., & Fu, W.-t. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. *CHI Letters, 3*, 112-119.

Green, R.T. (1956). Surprise as a factor in the von Restorff effect. *Journal of Experimental Psychology, 52,* 340-344.

Greenfield, S., Kaplan, S.H., & Ware, J.E. (1985). Expanding patient involvement in care: Effects on patient outcomes. *Annals of Internal Medicine, 102*, 520-528.

Greenfield, S., Kaplan, S.H., Ware, J.E., Yano, E.M., & Frank, H.J. (1988). Patients' participation in medical care: Effects on blood sugar control and quality of life in diabetes. *Journal of General Internal Medicine, 3*, 448-457.

Griffith, C.H., Wilson, J.F., Langer, S., & Haist, S.A. (2003). House staff nonverbal communication skills and standardized patient satisfaction. *Journal of General Internal Medicine, 18*, 170-174.

Gugerty, L. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*, 42-66. doi: 10.1037/1076-898X.3.1.42

Gugerty, L. (1998). Evidence from a partial report task for forgetting in dynamic spatial memory. *Human Factors, 40*, 498-508.

Hall, J.A., Roter, D.L. & Katz, N.R. (1988). Meta-analysis of correlates of provider behavior in medical encounters. *Medical Care, 26*, 657-675.

Haluck, R.S., Satava, R.M., Fried, G., Lake, C., Ritter, E.M., Sachdeva, A.K., Seymour, N.E., Terry, M.L., & Wilks, D. (2007). Establishing a simulation center for surgical skills: What to do and how to do it. *Surgical Endoscopy, 21*, 1223-1232.

Hart, S.G. & Loomis, L.L. (1980). Evaluation of the potential format and content of a cockpit display of traffic information. *Human Factors, 22*, 591-604. doi: 10.1177/001872088002200508

Hasher, L., Zacks, R.T., & May, C.P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application,* (pp. 653-675). Cambridge, MA: The MIT Press.

Hay, D.C., Smyth, M.M., Hitch, G.J., & Horton, N.J. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition, 35*, 176-190. doi: 10.3758/BF03195953

Heine, N., Garman, K., Wallace, P., Bartos, R., & Richards, A. (2003). An analysis of standardized patient checklist errors and their effects on student scores. *Medical Education, 37*, 99-104.

Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.) *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance,* (pp. 1-49). Oxford, UK: John Wiley & Sons.

Horrey, W. J., & Wickens, C. D. (2003). Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. In *Proceedings of the Second International Driving Symposium on Human Factors in Driver*

Assessment, Training, and Vehicle Design (pp. 7–12). Iowa City: Human Factors Research Program, University of Iowa.

Horrey, W.J., Wickens, C.D., & Consalus, K.P. (2006). Modeling drivers visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*, 67-78.

Hourizi, R. & Johnson, P. (2003). Towards an explanatory, predictive account of awareness. *Computers & Graphics, 27*, 859-872.

Hunt, R.R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2*, 105-112. doi: 10.3758/BF03214414

Hyltander, A., Liljegren, E., Rhodin, P.H., & Lonroth, H. (2002). The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surgical Endoscopy, 16*, 1324-1328.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*(4), 577-581. doi: 10.3758/BF03214352

Isreal, J., Chesney, G., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for a multiple capacity view of attention. *Psychophysiology, 17*, 259–273.

Jacko, J.A. (1997). An empirical assessment of task complexity for computerized menu systems. *International Journal of Cognitive Ergonomics, 1*, 137-147.

Janis, I.L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes.* Boston, MA: Houghton Mifflin.

Jewell, K. & McGiffert, L. (2009). *To err is human: To delay is deadly.* Austin, TX: Consumers Union.

Jones, D.G. & Endsley, M.R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine, 67*, 507-512.

Kahneman, D. *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kalamazoo (2001). Essential elements of communication in medical education: The Kalamazoo consensus statement. *Academic Medicine, 76*, 390-393.

Keinan, G. (1987). Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology, 52*, 639-644.

Kieras, D. (2007). Control of cognition. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 327–355). New York: Oxford University Press.

Kohn, L.T., Corrigan, J.M., & Donaldson, M.S. (Eds.) (1999). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Kripalani, S., Bussey-Jones, J., Katz, M., Genao, I. (2006). A prescription for cultural competence in medical education. *Journal of General Internal Medicine, 21*, 1116-1120.

Lane, C. & Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. *Patient Education and Counseling, 67*, 13-20.

Langenau, E.E., Dyer, C., Roberts, W.L., De Champlain, A.F., Montrey, D.P., & Sandella, J.M. (2011). Relationship between standardized patient checklist item accuracy and performing arts experience. *Simulation in Healthcare, 6*, 150-154.

Langenau, E.E., Dyer, C., Roberts, W.L., Wilson, C.D., & Gimpel, J.R. (2010). Five year summary of COMLEX-USA level 2-PE examinee performance and survey data. *Journal of the American Osteopathic Association, 110*, 114-125.

Lauro, L.J.R., Reis, J., Cohen, L.G., Cecchetto, C., & Papagno, C. (2010). A case for the involvement of phonological loop in sentence comprehension. *Neuropsychologia, 48*, 4003-4011.

Leach, D.C. & Philibert, I. (2006). High-quality learning for high-quality health care. *The Journal of the American Medical Association, 296*, 1132-1134.

Leape, L. (2012). Health care meets human factors: Flirtation or meaningful relationship? Keynote address presented at the 2012 Symposium on Human Factors and Ergonomics in Healthcare, Baltimore, MD. Retrieved from HFES website: http://www.hfes.org/web/HFESmeetings/HCSpresentations/leapekeynoteppt.pdf

Leape, L.L., Shore, M.F., Dienstag, J.L., Mayer, R.J., Edgman-Levitan, S., Meyer, G.S., & Healy, G.B. (2012). A culture of respect, part 1: The nature and causes of disrespectful behavior by physicians. *Academic Medicine, 87*, 1-8. doi: 10.1097/ACM.0b013e318258338d

Lenneman, J.K., Lenneman, J., Cassavaugh, N., & Backs, R. (2009). Differential effects of focal and ambient visual processing demands on driving performance. *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 306-312).

Levinson, W. (1994). Physician-patient communication: A key to malpractice prevention. *Journal of the American Medical Association, 273*, 1619-1620.

Levinson, W., Lesser, C.S., & Epstein, R.M. (2010). Developing physician communication skills for patient-centered care. *Health Affairs, 29*, 1310-1318.

Levinson, W. & Roter, D. (1993). The effects of two continuing medical education programs on communication skills of practicing primary care physicians. *Journal of General Internal Medicine, 8*, 318-324.

Liu, Y., & Wickens, C. D. (1992). Visual scanning with or without spatial uncertainty and divided and selective attention. *Acta Psychologica, 79,* 131–153.

Loftus, G.R., Dark, V.J., & Williams, D. (1979). Short-term memory factors in ground controller/pilot communication. *Human Factors, 21*, 169-181. doi: 10.1177/001872087902100204

Lustig, C., May, C.P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207. doi: 10.1037/0096-3445.130.2.199

MacLeod, C.M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 13-21. doi: 10.1037/0278-7393.15.1.13

Manning, S.K. & Schreier, H. (1988). Recency and suffix effects in pictures as a function of recall method. *The American Journal of Psychology, 101*, 97-109.

Master Interview Rating Scale (2005). *The clinical conversation and interview: The training manual for standardized patients.* (Available from the Theresa Thomas Professional Skills Teaching and Assessment Center, Norfolk, VA)

Matlin, M.W. (1994). *Cognition* (3rd ed.). New York: Harcourt Brace.

Maxfield, D., Grenny, J., McMillan, R., Patterson, K., & Switzler, A. (2005). *Silence kills: The seven crucial conversations for healthcare.* Retrieved March 01, 2011, from http://www.silenttreatmentstudy.com/silencekills

McGovern, M.M., Johnston, M., Brown, K., Zinberg, R., & Cohen, D. (2006). Use of standardized patients in undergraduate medical genetics education. *Teaching and Learning in Medicine, 18*, 203-207.

Melton, A.W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1-21. doi: 10.1016/S0022-5371(63)80063-8

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Moody, L., Baber, C., & Arvanitis, T.N. (2002). The role of haptic feedback in the training and assessment of surgeons. In *Proceedings of Eurohaptics 2001* (pp. 170-173). Birmingham, UK: University of Birmingham.

Moorthy, K., Munz, Y., Adams, S., Pandey, V., Darzi, A. (2005). A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. *Annals of Surgery, 242,* 631-639.

Moorthy, K., Munz, Y., Adams, S., Pandey, V., Darzi, A. (2006). Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theatre. *The American Journal of Surgery, 192*, 114-118.

Moray, N. (1986). Monitoring behavior and supervisory control. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance*

(Vol. 2): Cognitive processes and performance (pp. 40.1-40.51). New York: Wiley-Interscience.

Muther, W.S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science, 3*, 429-430.

Nairne, J.S., Riegler, G.L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 702-709. doi: 10.1037/0278-7393.17.4.702

Nairne, J.S., Whiteman, H.L., & Woessner, L. (1995). Symmetrical cuing effects for item and position information. *The American Journal of Psychology, 108*, 345-358.

National Patient Safety Foundation (2010). *Unmet needs: Teaching physicians to provide safe patient care.* Boston, MA: National Patient Safety Foundation.

Neuschatz, J.S., Lampinen, J.M., Preston, E.L., Hawkins, E.R., & Toglia, M.P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology, 16*, 687-708. doi: 10.1002/acp.824

Newlin-Canzone, E.T., Scerbo, M.W., Gliva-McConvey, G., & Wallace, A. (2013). The cognitive demands of standardized patients: Understanding limitations in attention and working memory with the decoding of nonverbal behavior during improvisations. *Simulation in Healthcare, 8*(4), 207-214. doi: 10.1097/SIH.0b013e31828b419e

Newman, L. S., & Uleman, J. S. (1993). When are you what you did? Behavior identification and dispositional inference in person memory, attribution, and social judgment. *Personality and Social Psychology Bulletin, 19*, 513-513.

Nisbett, R.E. & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259. doi: 10.1037/0033-295X.84.3.231

Norman, D.A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*, 522-536. doi: 10.1037/h0026699

Palmer, S.E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D.A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 279-307). Hillsdale, NJ: Erlbaum.

Pangaro, L.N., Worth-Dickstein, H., MacMillan, M.K., Klass, D.J., & Shatzer, J.H. (1997). Performance of "standardized examinees" in a standardized-patient examination of clinical skills. *Academic Medicine, 72*, 1008-1011.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*(2), 261-285. doi: http://dx.doi.org/10.1006/jmla.1996.0015

Peterson, L. & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198. doi: 10.1037/h0049234

Polson, M. C., & Friedman, A. (1988). Task-sharing within and between hemispheres: A multiple-resources approach. *Human Factors, 30,* 633–643.

Posner, M.I., Nissen, M.J., & Ogden, W.C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H.L. Pick & E.J. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137-157). Hillsdale, NJ: Erlbaum.

Postman, L., & Phillips, L. W. (1965). Short term temporal changes in free recall. *Quarterly Journal of Experimental Psychology, 17*, 132-138.

Ralph, J., Gray, W.D., & Schoelles, M.J. (2009). The functional resource hypothesis as a basis for understanding cognitive workload in immediate interactive behavior. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 404-408). Santa Monica, CA: Human Factors and Ergonomics Society.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1-75. doi: http://dx.doi.org/10.1016/1041-6080(95)90031-4

Roter, D.L. (1989). Which facets of communication have strong effects on outcome—A meta-analysis. In M. Stewart & D.L. Roter (Eds.), *Communicating with medical patients* (pp. 183-196). Newbury Park, CA: Sage Publications.

Rumelhart, D.E. (1984). Schemata and the cognitive system. In R.S. Wyer & T.K. Skrull (Eds.), *Handbook of social cognition* (pp. 161-188). Hillsdale, NJ: Lawrence Erlbaum.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63-77.

Sahakyan, L. & Delaney, P.F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language, 48*, 195-206. doi: 10.1016/S0749-596X(02)00524-7

Sahakyan, L., Delaney, P.F., & Waldum, E.R. (2008). Intentional forgetting is easier after two "shots" than one. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 408-414. doi: 10.1037/0278-7393.34.2.J

Sahakyan, L. & Kelley, C.M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1064-1072. doi: 10.1037/0278-7393.28.6.1064

Salas, E., Diaz Granados, D., Weaver, S.J., & King, H. (2008). Does team training work? Principles for health care. *Academic Emergency Medicine, 15*, 1002-1009.

Salas, E., Wilson, K.A., Burke, C.S., & Wightman, D.C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors, 48*, 392-412.

Salmon, P.M., Stanton, N.A., Walker, G.H., & Jenkins, D.P. (2009). *Distributed situation awareness: Theory, measurement and application to teamwork.* Surrey, UK: Ashgate Publishing.

Salvucci, D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115,* 101–130.

Sarno, K. J., & Wickens, C. D. (1995). The role of multiple resources in predicting time-sharing efficiency: An evaluation of three workload models in a multiple task setting. *International Journal of Aviation Psychology, 5*, 107–130.

Satava, R.M. (2001). Surgical education and surgical simulation. *World Journal of Surgery, 25*, 1484-1489.

Schmidt, S.R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*, 523-542.

Shostek, K. (2007). Developing a culture of safety in ambulatory care settings. *Journal of Ambulatory Care Management, 30*, 105-113.

Smith, K. & Hancock, P.A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors, 37*, 137-148.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*(4), 710-715. doi: 10.3758/BF03208850

Smyth, M.M., Hay, D.C., Hitch, G.J., & Horton, N.J. (2005). Serial position memory in the visual-spatial domain: Reconstructing sequences of unfamiliar faces. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 58*, 909-930. doi: 10.1080/02724980443000412

Squier, R.W. (1990). A model of empathic understanding and adherence to treatment regimens in practitioner-patient relationships. *Social Science & Medicine, 30*, 325-339.

Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., & Jenkins, D.P. (2005). *Human factors methods: A practical guide for engineering and design.* Hampshire, UK: Ashgate Publishing Limited.

Stewart, M., Meredith, L., Brown, J.B., & Galajda, J. (2000). The influence of older patient-physician communication on health and health-related outcomes. *Clinical Geriatric Medicine, 16*, 25-36.

Sulin, R.A. & Dooling, D.J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology, 103*, 255-262. doi: 10.1037/h0036846

Swing, S.R. (2007). The ACGME outcome project: Retrospective and prospective. *Medical Teacher, 29*, 648-654.

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situation awareness in aerospace operations*

(AGARD-CP-478; pp. 3/1-3/17). Neuilly-Sur-Seine, France: NATO—Advisory Group for Aerospace Research and Development.

The Joint Commission (2010). Sentinel event data: Root causes by event type 2004 - Fourth Quarter 2010. Retrieved March 14, 2011, from http://www.jointcommission.org/Sentinel_Event_Statistics

Thorndyke, P.W. (1984). Applications of schema theory in cognitive research. In J.R. Anderson & S.M. Kosslyn (Eds.), *Tutorials in learning and memory* (pp. 167-191). New York: W.H. Freeman & Co.

Titz, C. & Verhaeghen, P. (2010). Aging and directed forgetting in episodic memory: A meta-analysis. *Psychology & Aging, 25*, 405-411. doi: 10.1037/a0017225

Treisman, A.M. (1964). Verbal cues, language and meaning in selective attention. *American Journal of Psychology, 77*, 206-219.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*(3), 239. doi: 10.1037/0033-295X.93.3.239

Tuckey, M.R. & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9*, 101-118. doi: 10.1037/1076-898X.9.2.101

Uleman, J.S. (1989). *A framework for thinking intentionally about unintended thought.* In J.S. Uleman & J.A. Bargh (Eds.), *Unintended thought* (pp. 425-449). New York: Guilford.

Undre, S., Koutantji, M., Sevdalis, N., Gautama, S., Selvapatt, N., Williams, S., Sains, P., McCulloch, P., Darzi, A., & Vincent, C. (2007). Multidisciplinary crisis simulations:

The way forward for training surgical teams. *World Journal of Surgery, 31*, 1843-1853.

van de Mortel, T.F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing,* 25(4), 40-48.

van den Bos, J., Rustagi, K., Gray, T., Halford, M., Ziemkiewicz, E., & Shreve, J. (2011). The $17.1 billion problem: The annual cost of measurable medical errors. *Health Affairs, 30*, 596-603. doi: 10.1377/hlthaff.2011.0084

van der Vleuten, C. & Swanson, D.B. (1990). Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine, 2*, 58-76.

van Heukelom, J.N., Begaz, T., & Treat, R. (2010). Comparison of postsimulation debriefing versus in-simulation debriefing in medical simulation. *Simulation in Healthcare, 5*, 91-97.

van Zanten, M., Boulet, J.R., & McKinley, D. (2007). Using standardized patients to assess the interpersonal skills of physicians: Six years' experience with a high-stakes certification examination. *Health Communication, 22*, 195-205.

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld (The effects of field formation in the trace field). *Psychologie Forschung*, *18*, 299–342. doi:10.1007/BF02409636

Wachter, R.M. (2010). Patient safety at ten: Unmistakable progress, troubling gaps. *Health Affairs, 29*, 1165-1173.

Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology*, *45*(4), 1028-1031. doi: http://dx.doi.org/10.1016/j.jesp.2009.04.026

Wallace, P. (1997). Following the threads of an innovation: The history of standardized patients in medical education. *Caduceus, 13*(2), 5-28.

Wallace, P. (2007). *Coaching standardized patients: For use in the assessment of clinical competence.* New York: Springer Publishing.

Wang, F., Burdet, E., Vuillemin, R., & Blueler, H. (2005). Knot-tying with visual and force feedback for VR laparoscopic training. In *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society* (pp. 5778-5781).

Weinstein, L.F. & Wickens, C.D. (1992). Use of nontraditional flight displays for the reduction of central visual overload in the cockpit. *The International Journal of Aviation Psychology, 2*, 121-142.

Welford, A.T. (1952). The "psychological refractory period" and the timing of high-speed performance. *British Journal of Psychology, 43*, 2-19.

Welford, A. T. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quarterly Journal of Experimental Psychology, 11*, 193–210.

Welford, A. T. (1967). Single-channel operation in the brain. *Acta Psychologica, 27*, 5–22.

Whetstone, T., Cross, M.D., & Whetstone, L.M. (1996). Inhibition, contextual segregation, and subject strategies in list method directed forgetting. *Consciousness and Cognition: An International Journal, 5*, 395-417. doi: 10.1006/ccog.1996.0025

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York: Academic Press.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Sciences, 3,* 159-177.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-455. doi: 10.1518/001872008X288394

Wickens, C.D. & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37,* 473-494. doi: 10.1518/001872095779049408

Wickens, C. D. & Colcombe, A. (2007). Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Human Factors, 49,* 839–850.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors, 45,* 360–380.

Wickens, C. D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors, 25,* 227–248.

Williams, R.G. (2004). Have standardized patient examinations stood the test of time and experience? *Teaching and Learning in Medicine, 16,* 215-222.

Williams, R.G., Klamen, D.A., & McGaghie, W.C. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine, 15,* 270-292.

Wooford, M.M., Wofford, J.L., Bothra, J., Kendrick, S.B., Smith, A., & Lichstein, P.R. (2004). Patient complaints about physician behaviors: A qualitative study. *Academic Medicine, 79*, 134-138.

Wright, M.C., Taekman, J.M., & Endsley, M.R. (2004). Objective measures of situation awareness in a simulated medical environment. *Quality and Safety in Healthcare, 13*, i65-i71.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology, 59*, 555-561.

Zacks, R.T. & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241-264). San Diego, CA: Academic Press.

APPENDIX A

ODU INFORMED CONSENT DOCUMENT

PROJECT TITLE: THE EFFECTS OF SITUATION AWARENESS, EVENT SERIAL POSITION, AND THE ISOLATION EFFECT ON STANDARDIZED PATIENTS' SCORING RELIABILITY IN A SIMULATION-BASED TRAINING SCENARIO

INTRODUCTION

The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. Your participation in the study titled: THE EFFECTS OF SITUATION AWARENESS, EVENT SERIAL POSITION, AND THE ISOLATION EFFECT ON STANDARDIZED PATIENTS' SCORING RELIABILITY IN A SIMULATION-BASED TRAINING SCENARIO (located in the Center for Simulation & Immersive Learning) is completely voluntary. It is your right and responsibility to inform the researcher if you wish to cease participation at any time.

RESPONSIBLE PROJECT INVESTIGATOR

Mark W. Scerbo, PhD
Professor
College of Sciences
Department of Psychology
Old Dominion University

PROJECT CO-INVESTIGATOR

Thomas W. Hubbard, MD, MPH, JD
Director, Center for Simulation & Immersive Learning
Eastern Virginia Medical School

GRADUATE RESEARCHER

T. Robert Turner, MA
Research Associate
Virginia Modeling, Analysis, and Simulation Center
Old Dominion University

DESCRIPTION OF RESEARCH STUDY

Several studies have been conducted on the subject of using simulators to improve medical student training and education. One type of simulator commonly used to train medical students is live human beings trained to act sick and let the medical student diagnose them, while at the same time evaluating the professional and technical performance of the medical student. Previous research has shown that trained actors are highly beneficial for medical student training, but relatively few studies have explored how reliably the actors can perform all of the complex tasks involved in performing and evaluating. None of the previous studies have explored how event sequences and training scenario length affect the actors' reliability.

If you decide to participate, then you will join a study involving research of how event sequences and time affect our ability to recall key events and report them accurately for the purpose of critiquing a medical student's interpersonal skills. If you say YES, then your participation will last for two hours at the Center for Simulation & Immersive Learning. Approximately 60 individuals will be participating in this study.

EXCLUSIONARY CRITERIA

To the best of your knowledge, you should not have any diagnosed vision or hearing deficits that would keep you from participating in this study. If you do have any of these deficits, you must wear the required corrective lenses or hearing aids. You must be at least 18 years of age to participate. If you have any questions regarding your ability to participate, please ask the researcher for clarification at this time.

RISKS AND BENEFITS

RISKS: If you decide to participate in this study, then you may face a risk of mild anxiety related to the evaluation of another individual's performance. The researcher tried to reduce these risks by limiting the amount of time required for performance assessment. And, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: The main benefit to you for participating in this study is you will gain some knowledge of how simulation-based training is used to enhance the clinical development of medical professionals.

COSTS AND PAYMENTS

The researchers want your decision about participating in this study to be absolutely voluntary. Yet they recognize that your participation may pose some inconvenience. The researchers are unable to give you any payment for participating in this study. If you decide to participate in this study, you will be compensated according to your regular wage as a standardized patient.

NEW INFORMATION

If the researchers find new information during this study that would reasonably change your decision about participating, then they will give it to you.

CONFIDENTIALITY

The researchers will take reasonable steps to keep private information, such as questionnaires, and video/audio recording confidential. The researcher will remove identifiers from the information and store information in a locked filing cabinet prior to its processing. The results of this study may be used in reports, presentations, and publications; but the researcher will not identify you. Of course, your records may be subpoenaed by court order or inspected by government bodies with oversight authority.

WITHDRAWAL PRIVILEGE

It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study -- at any time. Your decision will not affect your relationship with Eastern Virginia Medical School or the Center for Simulation & Immersive Learning. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation.

COMPENSATION FOR ILLNESS AND INJURY

If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of harm, injury, or illness arising from this study, Old Dominion University, Eastern Virginia Medical School, and the researchers will NOT be able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact Dr. Mark Scerbo at (757) 683-4217, Dr. Thomas Hubbard at (757) 446-7093, Dr. George Maihafer the current IRB chair at (757) 683-4520 at Old Dominion University or at the Office of Research at (757) 683-3460, who will be glad to review the matter with you.. If you have any questions pertaining to your rights as a research subject you may contact a member of the Institutional Review Board through the ODU Institutional Review Board office at (757) 446-8423.

VOLUNTARY CONSENT

By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researchers should be able to answer them:

Dr. Mark Scerbo at (757) 683-4217
T. Robert Turner at (757) 638-4440 or (513) 254-7105

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. George Maihafer, the current IRB chair, at (757) 683-4520, or the Old Dominion University Office of Research, at (757) 683-3460.

And importantly, by signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

Subject's Printed Name & Signature	**Date**

INVESTIGATOR'S STATEMENT

I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

Investigator's Printed Name & Signature	**Date**

APPENDIX B

EVMS SUBJECT CONSENT FORM

Eastern Virginia Medical School (EVMS) Institutional Review Board

STUDY TITLE

THE EFFECTS OF SITUATION AWARENESS, EVENT SERIAL POSITION, AND THE ISOLATION EFFECT ON STANDARDIZED PATIENTS' SCORING RELIABILITY IN A SIMULATION-BASED TRAINING SCENARIO

INVESTIGATORS

Thomas W. Hubbard, MD, MPH, JD
Director, Center for Simulation & Immersive Learning
Eastern Virginia Medical School

Mark W. Scerbo, PhD
Professor
College of Sciences
Department of Psychology
Old Dominion University

T. Robert Turner, MA
College of Sciences
Department of Psychology
Old Dominion University

WHY IS THIS STUDY BEING DONE?

Several studies have been conducted on the subject of using simulators to improve medical student training and education. One type of simulator commonly used to train medical students is live human beings trained to act sick and let the medical student diagnose them, while at the same time evaluating the professional and technical performance of the medical student. Previous research has shown that trained actors are highly beneficial for medical student training, but relatively few studies have explored how reliably the actors can perform all of the complex tasks involved in performing and evaluating. None of the previous studies have explored how event sequences and training scenario length affect the actors' reliability.

The purpose of this study is to determine how event sequences and time affect our ability to recall key events and report them accurately for the purpose of critiquing a medical student's interpersonal skills. This is not a sponsored study.

WHY ARE YOU BEING ASKED TO TAKE PART?

You are being asked to participate in this research project because you are a trained Standardized Patient between the ages of 18-85.

This is a research study. This study includes only people who choose to take part. Please take your time to make your decision and feel free to ask any questions you might have.

What Are Some Important Details about this Study?

At this local site about 60 people will take part in this study. We will need no more than 2 total hours of your time for this research project. This includes up to one hour for case preparation/review prior to on-site data collection, and one hour of on-site participation on the day of data collection.

When Should You Not Take Part?

If you meet any of the following conditions, you should not take part in this study:

- You are under 18 years old
- You are over 85 years old

What Is Involved in the Study?

You will be "randomized" into one of the study groups described below. This means that you will be assigned into a group by chance. It is like flipping a coin. A computer program may do this - neither you nor the investigator will choose what group you will be in. You will have a 1 in 6 chance of being placed in any group.

- Early/Late Isolate vs. No Isolate
- Concurrent vs. Retrospective Recall

The following are standard procedures that will be done because you will be in this study:

You will be asked to review and memorize a set of patient characteristics and symptoms for the purpose of portraying them to a "medical student" confederate. The confederate will ask you a series of interview questions, and you will be tasked with answering them in a manner consistent with that of the character and case you have prepared for in advance. When you have completed the interview, you will be asked to evaluate the confederate's verbal and nonverbal behaviors using a set of clinical competency evaluation instruments designed specifically for this research study. The encounter will be video recorded in order to ensure the consistency of presentation of all verbal and nonverbal behaviors of interest, but the video content will not be published or otherwise be made publicly available. They will be used for quality control and analysis purposes only, and will not be presented at conferences or used to promote this research.

The following are experimental procedures that are being tested in this study:

We are studying the function of event sequences and time in affecting standardized patients' ability to recall key events and report them accurately for the purpose of critiquing a medical student's interpersonal skills.

What Are the Risks of the Study?

There are very few known risks to you, beyond what we would normally expect from your daily activities as a standardized patient.

There also may be other risks that are unknown and we cannot predict.

A risk associated with allowing your data to be saved is the release of personal information from your study record. We will strive to protect your records so that your personal information (like name, address, social security number and phone number) will remain private.

ARE THERE BENEFITS TO TAKING PART IN THE STUDY?

You may gain a sense of accomplishment from contributing to a study examining the psychological characteristics of those in your discipline and occupational area. However, you will receive no direct benefit from participating in the study.

WHAT ABOUT CONFIDENTIALITY?

Information learned from this research may be used in reports, presentations and publications. None of these will personally identify you.

WHAT WILL PARTICIPATION IN THE STUDY COST OR PAY?

There are no additional costs to you associated with taking part in this study.

You will be compensated for your time in accordance with your normal hourly wages policy through Eastern Virginia Medical School's standardized patient program.

WHAT IF YOU GET INJURED?

Eastern Virginia Medical School and Old Dominion University will not provide free medical care for any sickness or injury resulting from being in this study. Financial compensation for a research related injury or illness, lost wages, disability, or discomfort is not available. However, you do not waive any legal rights by signing this consent form.

WHAT ARE YOUR RIGHTS AS A PARTICIPANT?

Taking part in this study is your choice. If you decide not to take part, your choice will not affect any benefits to which you are entitled. You may choose to leave the study at any time. If you leave, the study it will not result in any penalty or loss of benefits to you.

WHOM DO YOU CALL IF YOU HAVE QUESTIONS OR PROBLEMS?

For questions about the study, contact the investigator, Dr. Thomas Hubbard, at (757) 446-7093. You may also contact Dr. Mark Scerbo at (757) 683-4217 or Robert Turner at (513) 254-7105.

For questions about your rights as a research participant, contact a member of the Institutional Review Board through the Institutional Review Board office at (757) 446-8423.

If you believe you have suffered an injury as a result of your participation in this study, you should contact the principal investigator, Dr. Mark Scerbo, at (757) 683-4217. You may also contact Dr. Robert Williams, an employee of Eastern Virginia Medical School, at (757) 446-8423.

SIGNATURE			
You will get a copy of this signed form. You may also request information from the investigator. By signing your name on the line below, you agree to take part in this study and accept the risks.			
____________________ Signature of Participant	________________ Typed or Printed Name	________________ Relationship to Subject	___/___/___ MM/ DD/ YY

STATEMENT OF THE INVESTIGATOR OR APPROVED DESIGNEE	
I certify that I have explained to the above individual the nature and purpose of the study, potential benefits, and possible risks associated with participation in this study. I have answered any questions that have been raised and have witnessed the above signature. I have explained the above to the volunteer on the date stated on this consent form.	
____ Signature of Investigator or Approved Designee	_____/_____/_____ MM/ DD/ YY

Sufficient space for the IRB stamp should be included on the 1st page or on the last page of the consent form.

APPENDIX C

EVMS EMPLOYEE/STUDENT ADDENDUM CONSENT FORM

Eastern Virginia Medical School (EVMS) Institutional Review Board

Study Title:	The effects of situation awareness, event serial position, and the isolation effect on standardized patients' scoring reliability in a simulation-based training scenario
Name of Investigator:	**Thomas W. Hubbard, MD, MPH, JD**
Sponsor:	N/A
Name of Subject:	For participants less than 18 years old, all references to "you" in this consent form are referring to "you", "your child" or a "minor for whom you are a legally appointed representative".[1]

You are being asked to participate in the above research study, which is being conducted at Eastern Virginia Medical School (EVMS), where you are an employee or student. The research study has been described to you, in writing, on the attached consent form. You have also had the opportunity to ask the investigators conducting this study any questions that you may have regarding participation in this study.

The purpose of this addendum consent form is to inform you that you have the right to choose not to participate in this research study. If you choose not to participate, or to withdraw at any time, it will not affect your standing as an employee or student.

If you are an employee, your participation will not place you in good favor with the investigator, your supervisor, or EVMS (e.g., increase in salary, promotion, extra vacation, or the like). Not participating will not adversely affect your employment with EVMS, in particular the position that you currently hold. If you are a student, your participation will not place you in good favor with the investigator or other faculty (e.g., receiving better grades, recommendations, employment). Also, not participating in this study will not adversely affect your relationship with the investigator or other faculty.

If you suffer a physical injury or illness as a result of participating in this research study, you will not receive a financial payment. Treatment for such injury or illness is not covered under Workmen's Compensation. Any immediate emergency medical treatment you may need as a result of participating in this study will be provided as outlined in the attached consent form. Eastern Virginia Medical School provides no compensation plan or free medical care plan to compensate you for such injuries. If you believe you have suffered an injury as a result of your participation in this study, you should contact the principal investigator, Dr. Thomas Hubbard, at (757) 446-7093. You may also contact Dr. Robert Williams, an employee of Eastern Virginia Medical School, at (757) 446-8423. If you have any questions pertaining to your rights as a research subject you may contact a member of the Institutional Review Board through the Institutional Review Board office at (757) 446-8423.

SIGNATURE			
You will get a copy of this signed form. You may also request information from the investigator. By signing your name on the line below, you agree to take part in this study and accept the risks.			
____________________ Signature of Participant	____________________ Typed or Printed Name	____________________ Relationship to Subject	___/___/___ MM/ DD/ YY

STATEMENT OF THE INVESTIGATOR OR APPROVED DESIGNEE	
I certify that I have explained to the above individual the nature and purpose of the study, potential benefits, and possible risks associated with participation in this study. I have answered any questions that have been raised and have witnessed the above signature. I have explained the above to the volunteer on the date stated on this consent form.	
____ Signature of Investigator or Approved Designee	___/___/___ MM/ DD/ YY

Sufficient space for the IRB stamp should be included on the 1st page or on the last page of the consent form.

APPENDIX D

ODU & EVMS DEMOGRAPHIC FORMS

Study I ODU Undergraduate Participants

1. Age: ___________

2. Sex: Male Female

3. Do you have normal or corrected-to-normal vision? Yes No

4. Do you have normal or corrected-to-normal hearing? Yes No

5. Do you have any previous experience working with simulation in an educational or training environment?

   Yes No

If so, please describe:

____

____

____

____

6. Do you have any previous experience working with standardized patients?

   Yes No

If so, please describe:

____

____

____

____

7. Have you ever worked or received training as a standardized patient?

   Yes No

Study II EVMS Standardized Patient Participants

1. Age _________
2. Sex _________ (M/F)
3. Do you have any formal acting experience? ________ (Yes/No)

   If yes, please indicate the type of acting experience in the space provided
   (Check all that apply)
   - ❑ Acting classes in school
   - ❑ Professional acting classes
   - ❑ Acting in plays
   - ❑ Acting in musicals
   - ❑ Improvisational classes
   - ❑ Commercial or TV acting
   - ❑ Other ______________________

4. Have you ever been diagnosed as having a deficiency in your vision? ____ (Yes/No)

   If yes, do you have correction with you? (i.e., glasses, contacts, etc.)? _____ (Yes/No)

5. Have you ever been diagnosed as having a deficiency in your hearing? ____ (Yes/No)

   If yes, do you have correction with you? (i.e., hearing aid, etc.)? ____ (Yes/No)

6. Approximately how long have you been a SP? __________________
7. Have you received special training as an SP to be a SP trainer? _____ (Yes/No)

APPENDIX E

STANDARDIZED PATIENT CASE & SUPPLEMENTAL MATERIALS

Standardized Patient Protocol

Institution/Case Author: EVMS
Case Title: Substance Abuse
History X

PATIENT DEMOGRAPHICS: to be used for recruiting the Standardized Patient

a) **age range 25-40**
b) **gender .... male preferred**
c) **race .... non-specific**
d) **socioeconomic ....blue collar**
e) **education ....college**
f) **affect to be simulated...cooperative, some discomfort**

SUMMARY OF CASE

Opening Statement:
"My shoulder is still bothering me."

Agenda:
#1. Shoulder pain
#2. Needs refill of pain meds

History of Present Illness:
One week ago the patient fell off a ladder at work and landed on his right side. He wasn't in too much pain at the time, but the next morning he found that his shoulder was quite painful (level 7), had some swelling, and was bruised. No radiation. He describes the pain as an ache. The patient was seen later that day at Urgent Care. X-rays showed no broken bones. The patient was told to rest it, apply ice, was prescribed pain medication, and told to follow-up with his doctor.

He stayed home from work for the rest of the week. Rest and meds help alleviate the pain, movement aggravates it. The shoulder is better than it was, but still hurts, especially when raising the arm. The patient returned to work yesterday. Presently, the pain is usually a level 3, but was as high as 5 or 6 after working yesterday.

He is here today for his follow-up appointment.

Past Medical History:
General health: good
Illnesses: bronchitis 2 yrs. ago
Hospitalizations/Surgeries: none
Injuries/Accidents: broken ankle 5 yrs. ago playing football
Medications: Lortab 5mg- 2 every 4 hrs.
Allergies: none

Family History:
Grandparents: Paternal grandfather- arthritis, grandmother- HTN
Maternal grandparents- healthy
Parents: healthy
Siblings: healthy

Social History:
The patient is a painter for a small company. He graduated from Princeton with a degree in Business Administration. Throughout high school and college he maintained an A average. After college, he returned to the area and took the painting job, intending it to be temporary until he could find work in his field. The patient married his college girlfriend soon after graduation. They separated almost a year ago, and are planning to divorce. He left his wife because *"I got sick of her always nagging me."* Since then, he has dated a few women (4 or 5) but nothing serious. He practices safe sex, although he is not currently sexually active.

The patient does not use tobacco, drinks beer occasionally (2-3 per week), and smokes marijuana daily. No other drugs. He started smoking in college. His drug use has steadily increased over the past few years to where he now spends $100.00 per week on pot (was occasional joints prior to the painting job- now ½ ounce per week). He doesn't see a problem with it as: all the guys he works with smoke, *"it's not like I drive around messed up"*, and he finds it relaxing. He's never had any legal problems related to his drug use. **If asked:** he will admit to his pot use as a point of contention in his marriage.

A typical weekday for him starts with 2 cups of coffee at home, then a thermos of coffee, which is usually gone by noon. After that, he may have 2 or 3 Red Bulls to keep him going. He smokes marijuana on and off throughout the workday, with more in the evening. CAGE question responses:

Cut Down- sometimes, it's expensive
Annoy- yes
Guilty- no
Eye opener- yes, weekends (wakes and bakes)

Negotiation Phase: if the learner is skillful in explaining the correlation between substance abuse and aggravation of problems, the patient will agree to <u>decrease</u> marijuana and caffeine

"Symptoms of addiction affect not only individual but also friends and family around you. Family show signs of symptoms such as anxiety, self-blame, isolation from outside friends and family to hide turmoil, lost personal identity because of focusing on other, and shame of situation.

Patient Behavior, Affect, Mannerisms:
The patient is pleasant, although somewhat in pain (rubs shoulder occasionally), casual dress (jeans, t-shirt)

F- feels stupid, should've been more "focused"; accident would've been prevented
I- healing too slowly, maybe more serious than originally thought
F- shoulder: limiting ability to do all job requirements, pot use: decreasing motivation
E- shoulder: wants pain meds refill

PATIENT INFORMATION

Josh Claypool comes into the office today for a follow-up appointment for a shoulder injury.

TASKS

1. Elicit:
   - HPI (description of patient illness)
   - PMH (past medical history)
   - FMH (family medical history)
   - SH/Patient Profile (social history)

2. Practice:
   - Elicitation of Concerns
   - Negotiation of the Agenda
   - Questioning Skills
   - Elicitation of the Narrative Thread (Pt.'s story)
   - Pt.'s perspective (FIFE)
   - Summarization
   - Transitional Statements
   - Motivation to Change

Study Guide Checklist

Chief Complaint – shoulder pain
1. Onset – 1 week ago
2. Duration - constant
3. Frequency – daily

4. Location – right shoulder
5. Radiation- none
6. Quality – aches
7. Intensity- 4 out of 10
8. Associated symptoms - none
9. Alleviating factors – pain meds, rest
10. Aggravating factors- movement

Past Medical History

11. General health- good
12. Illnesses – bronchitis 2 yrs ago
13. Hospitalizations / Surgeries- none
14. Injuries –broken ankle 5 yrs ago
15. Medications – Lortab
16. Allergies - none

Family History

17. Grandparents- Pat. GF- arthritis, Pat. GM: HTN, Mat. GF and GM- healthy
18. Parents- healthy
19. Siblings- healthy

Social History

20. Marital status - separated
21. Children – none
22. Occupation- painter
23. Caffeine use – thermos plus 2 cups of coffee in AM, 2-3 Red Bulls in PM
24. Tobacco use – no tobacco hx
25. Alcohol use – 2-3 beers per week
26. Recreational drug use – yes, pot, ½ oz. per week
27. CAGE- C: sometimes, A: yes, G: no, E: yes

MIRS:

Elicitation of Concerns

Negotiation of the Agenda

Questioning Skills
Elicitation of the Narrative Thread (Pt.'s story)
Pt.'s perspective (FIFE)
Summarization
Transitional Statements
Motivation to Change

Index of Behavior Type and Occurrence

Segment 1: Patient History	
[illegible]	[illegible]
Introduces Self	Backward Lean, Slouching
Eliciting Complaint Timeline	Interrupting
Use of Multiple Question	Leaning Far Forward
Eliciting Concerns	Leaning to Side, Propping
Medical Jargon	Taps Clipboard w/ Pen
Assessing Patient Perspective	Clenching Fist
	Tone of Voice
Segment 2: Substance Abuse	
[illegible]	[illegible]
Patient Surname	Crossing Arms
Medical Jargon	Smirk
Verifying Patient Info	Pressed Lips Together, Concern
Empathy/Acknowledging	Small, Quick Hand Gestures
Lack of Knowledge/Experience	Touch Mouth with Finger/Pen
Address Impact on Family	Cleared Throat
	Large, Fluid Hand Gestures
Segment 3: Goals	
[illegible]	[illegible]
Assess Motivation	Smirk
Assess Understanding	Smiling
Setting Agenda	Crossing Legs
Assess Support System	Scratch Face with Pen
Outlining Intervention	Tone of Voice
Medical Jargon	Head Toss
	Shaking Patient's Hand

COLOR KEY:

Verbal cues coded GREEN (verbal behaviors incorporated from MIRS items)

Nonverbal cues coded ORANGE (nonverbal behaviors integrated from nonverbal checklist)

Isolates coded RED

SEGMENT 1 – PATIENT HISTORY (~5 mins.)

DS: "Good afternoon. I'm Dr. Severs (Introduction of self). I'm covering all of Dr. Adams' patients today. How are you feeling?"

JC: "My shoulder is still bothering me from my accident last week."

DS *(looking at patient chart):* "Your chart says that you fell and landed on it at work last Monday, can you tell me what happened after that?" (Eliciting timeline of chief complaint)

JC: "Well, my shoulder was pretty swollen and bruised up pretty badly. The next day I could barely lift my arm above my chest, which made it hard to work. I was afraid I'd broken it or something, so I stopped in and saw Dr. Adams about it last Wednesday."

DS: "Dr. Adams prescribed Lortab for the pain?"

JC: "Yeah, it wasn't broken, so he just said I should ice it, rest, and take the pain medication every few hours."

DS: "And it's still hurting you right now?"

JC: "Yeah, it's gotten a little better but it still hurts."

DS *(looking at patient chart):* "Okay, have you ever experienced any similar injuries in the past?"

JC: "Not like this. I broke my ankle a few years back playing intramural football, but I've never hurt myself at work before. I've never really been all that clumsy, so I guess--" *(DS interrupts)*

DS *(looking at patient chart, interrupting JC):* "That's fine. Any major illnesses or other injuries I should know about?" (Use of a multiple question)

JC: "I had bronchitis a couple of years ago, but that wasn't anything serious. Nothing I can think of other than that, everything else is fine."

DS *(looking at patient chart):* "Any family history of illness?"

JC: "Uh…well, my grandfather has pretty bad arthritis and my grandmother has high blood pressure."

DS: *(leans forward, looking up at patient, bracing torso with elbow on top of knee but not invading JC's personal space)* "Okay. Well, tell me a little bit more about your accident."

JC: "If I'd been more focused on what I was doing it wouldn't have happened I guess. One minute I was on top of the ladder painting and the next I was on the ground and the ladder was on top of me. It didn't really hurt too badly when I fell, but it got worse over the next couple of days so I went in to Urgent Care and had it looked at. They just told me to rest, keep some ice on it, and take my pain meds."

DS *(removes elbow from knee and sits up straight again, looking at patient chart)*: "Were you able to do that?"

JC: "Yeah I stayed home from work the rest of the week and kept ice on it, took the meds, which helped. Still, when I move it hurts. I went back to work yesterday, and it was okay for most of the day, but last night it was killing me."

DS *(still looking at patient chart)*: "I see. Do you have any worries or concerns about anything else (Eliciting spectrum of patient's concerns)?"

JC: "Well, just that it's not healing fast enough. I need to get back to 100% as soon as possible so I can do my job. That's my primary concern I guess."

DS leans backward, away from patient and into a slouching position; looks up at patient

---ISOLATE---

DS: (snapping, angry) "Frankly, I don't think you're taking this seriously enough. I can assure you that problems don't just magically fix themselves overnight." OR "I know that you're concerned about getting better as quickly as possible. However, I think you need to remember that these things take time and you'll get there eventually."

---ISOLATE---

DS *(looking at patient chart):* "Compared to last week, just after the fall, does it hurt about the same, or is it getting better? Worse?"

JC: "Compared to last week it doesn't hurt as much, but it still hurts pretty bad."

DS *(looking at patient chart):* "Okay. I see here that you told Dr. Adams last week the pain was about a 7 out of 10. How would you describe it today?"

DS taps clipboard with finger/pen as JC responds

JC: "Probably about a 3 today, but after work last night it was about a 5 or 6. I've got a pretty physical job, constantly lifting heavy equipment and climbing. That's probably why it's hurting again. I can't think of anything other than that, except maybe it was worse than I originally thought. It's just not healing fast enough."

DS: "Okay, well it's going to take some time to get back to normal. I would still advise as much rest as possible, especially while you're on the Lortab. As I'm sure Dr. Adams told you, Lortab contains a combination of acetaminophen and hydrocodone and has a number of side effects you'll want to remain aware of. The most common side effects are light-headedness, dizziness, drowsiness, nausea, and dysphoria. It can also impair mental and physical activity, and you don't want to risk another fall or hurt yourself while operating heavy machinery at work. You shouldn't drink alcohol while taking Lortab. The acetaminophen in Lortab undergoes sulphation and glucuronidation prior to being eliminated from the body by the liver. An excessive amount of acetaminophen in the liver can saturate the sulphation and glucuronidation pathways, leading to the formation of a toxic metabolite known as NAPQI (Use of medical jargon)."

JC: "What does that mean?"

DS clenches fist, suggesting impatience with JC, while explaining the following:

DS: "Well, eventually a buildup of toxic compounds could result in serious liver damage. Consuming alcohol or taking other acetaminophen-based medications while on Lortab might exacerbate the problem."

JC: "Yeah, well, I took most of last week off, but I have to get back to work this week. I'm already in hot water with my boss. With everything else going on, I don't need to lose my job on top of it."

DS: "Okay. How has your injury affected your daily life? Have you had to change your routine any? (Eliciting patient's perspective, specifically patient function)"

JC: "Well I try to take it easy at work when I can, which isn't often. If I slack off too much I get in trouble with my boss, plus it isn't fair to the other guys who have to pull the extra weight. After work it just hurts. I've been taking it easy at home too, last weekend I skipped out on my softball league because I knew I'd pay for it later if I played. Otherwise it's pretty much the same, just trying not to do too much until it heals."

SEGMENT 2 – SUBSTANCE ABUSE (~5 mins.)

DS: "Okay Mr. Claypool (Using patient surname), that all sounds fine. What were you hoping I could do for you today?"

JC: "I just want to make sure my shoulder is healing the way it should and get a refill on my pain meds. Those seem to help more than anything."

DS *(using quick, erratic, large (aggressive) hand gestures):* "Well I can certainly take a look at it for you. I understand your need to get back to work, but I really do want you to continue to take it easy until the pain eventually subsides. Sometimes injuries like this tend to add a lot of stress and frustration to our already hectic lives. Is there anyone else in your household who might be stressed out over your injury as well? (Addressing impact of injury on family)"

JC: "My wife."

DS: "How are things with your wife?"

JC: (*laughing*) "Things at home are never good."

DS: "What do you mean?" *(crosses arms)*

JC: "Just my wife, bills, and everything else. Especially my soon-to-be ex-wife. We got married right after college, things were good back then. I got a degree in Business from Princeton, good GPA and all that, got married…then we moved back here and I couldn't find a job except in construction and things just went downhill from there. I've been split up from my wife for about a year, we're getting divorced. She just nagged too much."

DS presses lips together, demonstrating empathy or concern

DS: "What did your wife nag you about?"

JC: "Everything. The job, bills, not making enough money. Mostly she wouldn't leave me alone about smoking. I started smoking pot in college, but she didn't really care about it until after we got married. She said it was why I wasn't doing anything with myself, it was holding me back. So what, I drink a couple of beers a week and smoke some weed. It's not the end of the world."

DS *(looking at patient chart):* "How much marijuana do you smoke per week?"

JC (*thinks for a few seconds*): "I don't know, it depends. Usually around half an ounce. Most of the guys I work with smoke about the same, who cares? It's not like I drive around messed up, it just helps me relax and forget about all the crap in my life. She was always going on about how much I spend on pot, which is only like a hundred bucks a week, and how I don't care about anything but smoking weed with my buddies."

DS: "Do you take any other recreational drugs?"

JC: "No, I just smoke weed. And not even that much of it."

DS: "What about other stimulants?"

JC: "Yeah I drink coffee, I have to get up at like 4:30 every morning. Who wouldn't drink coffee?"

DS: "How much coffee do you drink?"

JC *(thinks for a few seconds)*: "Um, a couple of cups in the morning. I take a thermos with me to work that lasts until around lunch, then I have a couple of Red Bulls in the afternoon. Maybe 3 or more if it's a really long day. And yeah, I smoke some weed at work here and there, but I usually wait until I get home at night to do most of that."

DS: "So during a typical day you drink a couple of cups of coffee in the morning, a thermos of coffee throughout the day, a couple of Red Bulls in the afternoon, and smoke marijuana. Is that correct (Verifying patient information)?"

JC: "Yeah."

DS: "Have you ever felt the need to cut back on any of the caffeine or the marijuana?"

JC: "With the weed, yeah. Sometimes. I mean it's only $100 a week, but still that adds up. It gets expensive, so sometimes I do cut back a little to save money. But when I've got the money, I don't think about it really. I've never thought about cutting back on the coffee and Red Bull, except sometimes it makes working harder because I'm sorta shaky and drop tools."

DS: "Has anyone close to you talked to you about your habits?"

JC: "Yeah, my wife did all the time. I already mentioned that. Oh, and one time my sister asked me to stop smoking pot, and I told her to mind her own business. I know she was just worried and didn't mean anything by it, but still it pissed me off and I sorta yelled at her for it."

DS: "I can empathize (Empathy & acknowledging patient cues). Sometimes our friends and family members have a hard time expressing the way they feel to us, especially about something like substance dependence, and their concerns often come across to us as criticisms. Have you ever felt guilty about smoking or drinking too much?"

JC: "I've gone too far with it, if that's what you mean, and regretted it after. Sure, I've had lots of hangovers and stuff. But I don't beat myself up over it constantly."

DS: "Do you ever smoke marijuana or drink excessively first thing in the morning to steady yourself or to get rid of a hangover from the previous night?"

JC: "Yeah well the coffee is first thing in the morning to wake me up, otherwise I just can't get going at all. Weekends I usually sleep in and then smoke some weed first thing after waking up because I don't have anything going on with work. Just smoke and hang out at the house all day, kinda wake and bake if you know what I mean."

DS (*Using slow, fluid, small (calm) hand gestures)*: "Well, this isn't really my area of specialization (admitting lack of knowledge/experience), but I'm a little concerned that you're relying so heavily on substances like caffeine and marijuana to get you through the day. These substances may also be having a greater impact on your quality of life, possibly more than you're aware of. Symptoms of substance dependence and abuse manifest not only in the user, but also in friends and family members who are close to the individual. For example, I often see people close to a substance abuser suffering from conditions like severe anxiety and fear, which can lead to more complicated social dynamics like you mentioned with your wife. Friends and family sometimes feel like they're to blame for not being able to effectively communicate their fears to you and help you see the problem more clearly. They also tend to become more isolated from the outside world because they're afraid of how they and you may be perceived by other friends, family, that sort of thing. In extreme cases they might begin to lose some of their own self-identity because they become so focused on others' problems and on fixing them. So when your wife was nagging you about your marijuana use, she was probably just worried and trying to help you in her own way."

DS clears throat

JC: "Yeah I know that. I just didn't think it was that big a deal. I think she was making more out of it than was there, you know?"

DS: "Well, again, she might have gone overboard with it because she was worried and didn't know how else to communicate her anxiety to you. Family is our basic source of strength, and it's important to reinforce that with a sense of stability and mutual support. *(looking at patient chart)* Research has shown that women who are not drug abusers are affected by male users in their life such that they have difficulty maintaining interpersonal relationships, suffer from economic insecurity, and sometimes resort to violence out of fear and frustration."

DS touches pen to mouth briefly

DS: "Your increased irritability and difficulty making it through the day without constant pick-me-ups are likely side-effects of this substance abuse. Not to mention the real possibility that your dependence on these substances could be contributing to your inability to find a better job and develop in a professional sense."

JC: "Yeah, I get that. I do."

DS: "I know it's a lot to think about. Just remember that I'm here to help. Do you have any questions for me?" (Encouraging questions)

SEGMENT 3 – GOALS (~5 mins.)

JC: "Suppose I want to cut back on some of this stuff, how would I start off? Should I just cut back some or try to cut stuff completely out, cold turkey?"

DS *(nodding head)*: "Well, let's talk about that. First I'd like to get a sense of your motivation to cut these substances out of your life right now. What are your highest priorities when it comes to changing your lifestyle?"

JC: "I guess I didn't really think about how much I was using this stuff as a crutch. I mean, I still don't know how I'm going to make it sun-up to sun-down without it, but when you take into account how much I've been using and how little I've accomplished with myself, it kind of scares me. Plus I think some of my family would be happy to see me cutting out all the marijuana. That's why I want to cut back I guess. I don't want to be dependent on this stuff to live my life day to day."

DS: "Would you consider changing some of your habits?" (Assessing patient motivation to change)

JC: "Yeah, I definitely want to do something about it."

D looks at patient chart, recording JC's answer

DS *(smiling)*: "I'm glad to hear that. Have you ever tried to cut back on your substance use in the past?"

JC: "I cut out the marijuana for a couple of months right after I moved back from Princeton, but I started on it again after I started working construction. Those guys smoke a little, and I just fell back into it."

DS: "Okay. When you cut out the marijuana before, was it an intentional effort to quit smoking?"

JC: "Not really. I moved and didn't really know where to go to buy around here, but I didn't miss it too much either while I wasn't smoking it."

DS: "Do you think your coworkers would support you if you tried to stop smoking, or would they present an obstacle to your quitting?"

JC: "I don't think they'd care if I quit smoking."

DS: "And if they smoke at work, could you remove yourself from that?"

JC: "Yeah, well not all of them smoke. We take breaks at the same time, but I could just take a break with some of the other guys who don't smoke and I'd be okay."

DS: "What about cutting back on coffee and Red Bull?"

JC: "I could do that. Cut back to a cup in the morning and a Red Bull in the afternoon, maybe?"

DS *(crossing legs):* "I think that's a good place to start. So the plan is to cut back your coffee to one cup a day, cut back to one Red Bull a day, and stop smoking marijuana completely. Take your breaks with the non-smokers and put some distance between yourself and those smoking marijuana."

JC: "Yes, I think I can do that."

DS: "Good. Just for the sake of clarity, I'd like you to repeat back to me in your own words what immediate changes you plan to make in your life." (Assessing patient understanding)

JC: "Immediately? I'll cut the coffee down to a cup a day and cut back on my Red Bulls to one a day. Stop smoking marijuana completely and stop hanging out around other people that do it."

DS: "Good. How confident are you right now that you can succeed with this plan, on a scale of 0 to 10?"

JC: "I'd say 7 or 8. It'll be hard to get going that early without coffee, but I'll manage. The rest should be okay."

D looks at patient chart, recording JC's answer

DS: "Good. So let's talk about a few other points that will come into play. I'd like to get a sense of what your expected gains are with respect to changing your habits, as well as what your short- and long-term goals should be. We need to ascertain what kind of support network you have available, and also what the plan going forward should be. Once we've discussed all of that, we can see whether you have any questions about anything (Setting agenda)."

JC: "Sounds good."

DS *(scratching face with pen/fingers):* "First, what do you expect to gain from adjusting your lifestyle? How do you expect your life to change for the better?"

JC: "Well in the short-term, hopefully I'll start to feel better, healthier. Get some focus back, maybe not feel like I'm dragging through each day. Longer-term hopefully I can focus on getting my social life back in order. It just seems like everything is so far out of reach right now. Like everything is messed up, I just don't see how just cutting back on coffee and marijuana is going to fix things. It all feels kinda hopeless right now. I don't know."

---ISOLATE---

DS: (snapping, angry) "Frankly, I don't think you're taking this seriously enough. I can assure you that problems don't just magically fix themselves overnight." OR "I know that you're concerned about getting better as quickly as possible. However, I think you need to remember that these things take time and you'll get there eventually."

---ISOLATE---

DS: "You have quite a few friends and family members, possibly even coworkers, who would like to see you move in this direction, don't you?"

JC: "Yeah, I guess."

DS: "It's important to identify individuals who truly want to see you succeed so that you know where to turn for moral support when the time comes. Who can you talk to if you start to feel overwhelmed or stressed? (Assessing support network)"

JC: "Well my sister, for one. She already knows some of my problems and like I already said, she wants to see me stop using. So I could call her for support, I know she'd be there for me. My parents would be there for me too. I don't talk to them very often, but they've always been there for me. I've got some guys I play softball with, friends of mine, they're a good group. I bet they'd be willing to help out too if I needed it."

DS: "Good. Well, let's discuss the plan and see where we are (Discuss intervention/plan). I want you to discuss your goals with your friends and family, both short-term and long-term goals. Ask them to provide support however they can, and stay in regular contact with them to update them on your progress. That way if you start having trouble, they'll be in a position to see it and help. Let them know you consider them an important part of your support network, and that you might call on them from time to time to help you stay motivated to change. But remember, ultimately you're the one who is responsible for changing your lifestyle. They can help keep you on track, but you have to hold yourself accountable too. This may be difficult, but having a strong support network is going to make a real difference."

JC: "Yeah."

DS briefly looks at watch

DS: "I also want to schedule a follow-up appointment for next month to check on your arm and your progress with the caffeine and drug use. If anything changes in the meantime, I want you to give me a call. (Providing closure)"

JC: "Okay, sounds good." *(Shakes hands with DS)*

APPENDIX F

STANDARDIZED PATIENT VERBAL COMMUNICATION ASSESSMENT

Behavior	☑
Introduced self	
Addressed the patient by his/her surname (last name)	
Asked patient to state their own concerns, if they had any	
Set an agenda or sequence of discussion topics	
Asked the patient for their narrative concerning key events	
Requested information to help establish a timeline of the chief complaint	
Refrained from using technical or medical jargon	
Verified information that the patient provided by stating it back to them	
Attempted to learn the patient's perspective and/or beliefs about the injury	
Inquired about the patient's feelings about the injury and if/how it has changed the patient's life	
Addressed the impact of the injury on the patient's family	
Attempted to determine what financial and/or emotional support systems the patient could depend on during treatment	
Used supportive comments to demonstrate empathy and acknowledge the patient's situation	
Encouraged the patient to ask questions	
Admitted lack of knowledge or experience	
Attempted to determine whether the patient fully understood the information provided about injury, prognosis, and/or treatment options	
Assessed the patient's motivation to change behavior, mindset, or personal habits	
Explained any relevant investigations, tests, or interventions to the patient	
Provided closure to the patient by discussing next steps, future goals, and/or when next meeting will occur	
Invited the patient to contribute thoughts, ideas, suggestions, and/or preferences in determining the plan of care	
Used a multiple or double-barreled question touching on more than one issue	

APPENDIX G

STANDARDIZED PATIENT NONVERBAL COMMUNICATION ASSESSMENT

Category	Behavior	☑
Eye Contact	Maintained sufficient eye contact by looking directly at the patient more often than not	
	Looked at watch	
	Looked at a pager or cell phone	
Body	Leaned forward (toward patient), bracing torso with elbow on top of knee	
	Leaned backward (away from patient) into a slouching position	
	Crossed arms	
	Crossed legs	
	Crossed ankles	
Head	Nodded head to affirm patient's statements	
	Shook head, as if telling the patient "no"	
	Cocked head to one side	
Facial Expressivity	Smiled at the patient, demonstrating acceptance	
	Pressed lips together, demonstrating empathy or concern	
	Frowned at patient, demonstrating condescension or judgment	
	Yawned	
Hand Gesturing	Used slow, fluid, small (calm) hand gestures	
	Used quick, erratic, large (aggressive) hand gestures	
	Rubbed ear with hands	
	Scratched self on nose	
	Rubbed mouth	
	Scratched self on face	
	Shook patient's hand	
	Ran fingers through hair	
	Pointed index finger at patient	
	Touched mouth with finger/pen	
	Clenched fist	
Touch	Touched patient on the arm for encouragement or empathy	
	Touched patient on the shoulder for encouragement or empathy	
	Touched patient on the leg for encouragement or empathy	

	Touched patient on the back for encouragement or support	
Voice	Tone of voice was judgmental or condescending at times	
	Tone of voice was empathetic at times	
	Interrupted the patient	
	Coughed	
	Cleared throat	
Tapping	Tapped hands, indicating impatience toward the patient	
	Tapped feet, indicating impatience toward the patient	

APPENDIX H

STANDARDIZED PATIENT OVERALL PROGRESS ASSESSMENT

Please indicate how you felt the doctor performed by selecting an overall performance rating for each of the patient encounter segments listed below.

Patient History Segment: (Circle One)	1 (Poor)	2	3	4 (Good)	5	6	7 (Excellent)
Substance Abuse Segment: (Circle One)	1 (Poor)	2	3	4 (Good)	5	6	7 (Excellent)
Goals Segment: (Circle One)	1 (Poor)	2	3	4 (Good)	5	6	7 (Excellent)
Overall Rating: (Circle One)	1 (Poor)	2	3	4 (Good)	5	6	7 (Excellent)

APPENDIX I

DEBRIEFING FORM

Thank you for participating in this study, titled "The effects of situation awareness, event serial position, and the isolation effect on standardized patients' scoring reliability in a simulation-based training scenario." The purpose of this research is to better understand how a variety of factors may impact standardized patient evaluations of medical students and to explore different ways of designing simulation-based training scenarios for medical student education. Your participation in this study is helping to broaden our understanding of the cognitive demands placed on standardized patients.

Again, thank you for participating in this study!

APPENDIX J

Table 1

Means and Standard Deviations by Encounter Segment and Evaluation Format for Studies 1 and 2

Study 1	Concurrent Evaluation Format (N = 33)						Retrospective Evaluation Format (N = 37)					
	Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Verbal Correct ID	4.0	0.87	4.85	1.12	3.76	0.94	3.92	0.92	3.0	1.31	3.51	1.28
Verbal False Memories	3.67	1.9	4.97	2.31	5.15	2.25	3.76	1.72	5.32	2.9	5.65	2.72
Nonverbal Correct ID	4.39	1.06	3.7	1.7	4.33	1.22	3.19	1.24	1.95	1.25	3.43	0.96
Nonverbal False Memories	2.39	1.71	3.91	1.93	2.03	1.53	2.03	2.03	3.97	2.43	2.0	1.6

Study 2	Concurrent Evaluation Format (N = 25)						Retrospective Evaluation Format (N = 24)					
	Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Verbal Correct ID	4.0	1.0	4.04	1.06	3.92	1.08	3.92	1.14	2.75	1.54	3.42	1.35
Verbal False Memories	2.84	1.65	4.68	2.53	4.96	2.65	4.17	2.53	5.04	3.04	5.96	2.66
Nonverbal Correct ID	3.92	1.53	3.56	1.64	3.8	1.16	2.67	1.09	1.96	1.43	3.04	1.33
Nonverbal False Memories	2.6	2.12	4.44	3.18	1.84	1.68	2.79	2.23	4.54	2.5	2.33	1.24

Table 2

Means and Standard Deviations for Isolate Groups by Encounter Segment and Evaluation group for Study 1 and 2

Study 1	Concurrent Evaluation Format						Retrospective Evaluation Format					
	Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Early Isolate SHP Rating	2.40	0.84	4.50	0.71	5.70	0.68	3.45	0.69	4.55	0.52	5.09	0.70
Late Isolate SHP Rating	3.25	0.97	4.33	1.37	3.58	1.38	3.50	1.31	3.75	1.22	4.67	1.50
No Isolate SHP Rating	2.91	0.94	3.91	1.14	5.09	1.22	2.93	0.83	4.29	1.14	5.21	1.25

Study 2	Concurrent Evaluation Format						Retrospective Evaluation Format					
	Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Early Isolate SHP Rating	2.33	0.71	3.89	1.17	4.00	0.82	2.50	0.76	4.38	1.19	5.50	0.76
Late Isolate SHP Rating	2.88	0.99	3.63	1.60	3.25	0.71	4.14	1.22	3.71	1.11	3.86	1.77
No Isolate SHP Rating	2.75	0.71	4.13	1.25	5.25	1.28	4.00	1.31	4.63	1.69	4.88	1.25

VITA

T. Robert Turner, Jr.
rturner@facs.org

Education

Old Dominion University, Norfolk, VA

Doctor of Philosophy May 2014
Major: Human Factors Psychology

Xavier University, Cincinnati, OH

Master of Arts August 2006

Major: Industrial/Organizational Psychology

Mississippi State University, Starkville, MS May 2004

Bachelor of Arts
Major: General Psychology

Bio Sketch

Robert Turner is Assistant Director of Simulation-Based Surgical Education and Training for the American College of Surgeons (Chicago, IL). His role with the ACS involves leveraging state-of-the-art technologies, techniques, and institutions in support of advanced simulation-based surgical education and training. Before joining the ACS, Robert worked as a project scientist for the Virginia Modeling, Analysis, and Simulation Center (Suffolk, VA) and was involved in developing a simulation-based educational system for military nurse corps pre-deployment training. His research interests include multi-sensory architecture and display, perception and performance in extreme environments, and human systems integration issues within the field of biomedical (biomechanical) engineering.

Selected Presentations/Publications

Turner, T.R., Scerbo, M.W., Gliva-McConvey, G., & Wallace, A. (2013). Evaluating physician-patient communication: The effects of concurrent and retrospective assessment on the recall of verbal and nonverbal behaviors. Paper presented at the 2013 International Annual Meeting of the Human Factors and Ergonomics Society, San Diego, CA.

Scerbo, M.W., **Turner, T.R.**, Padilla, M.A., Gliva-McConvey, G., & Walker, P. (2012). Teaching communication skills using standardized patients in a team scenario: A G theory analysis. *Simulation in Healthcare, 6*(6), 482.

Turner, T.R. & Parodi, V.A. (2012). Theoretically-driven infrastructure for supporting healthcare teams training at a military treatment facility. *Military Medicine, 177*(2), 139-144.