The Effects of Mental Models and Expertise on Running Memory and Clinical Handoff Effectiveness

Brittany Lee Anderson-Montoya
Old Dominion University, b.l.anderson64@gmail.com

THE EFFECTS OF MENTAL MODELS AND EXPERTISE ON RUNNING MEMORY AND CLINICAL HANDOFF EFFECTIVENESS

by

Brittany L. Anderson-Montoya
B.S. December 2005, Old Dominion University

M.S. December 2008, Old Dominion University

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

DOCTOR OF PHILOSOPHY

PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2015

Approved by:

Mark W. Scerbo (Director)

James P. Bliss (Member)

Ivan K. Ash (Member)

Dana E. Ramirez (Member)

Ginger S. Watson-Papelis (Member)

ABSTRACT

THE EFFECT OF MENTAL MODELS AND EXPERTISE ON RUNNING MEMORY AND CLINICAL HANDOFF EFFECTIVENESS

Brittany L. Anderson-Montoya
Old Dominion University, 2015
Director: Dr. Mark W. Scerbo

The goal of the present study was to examine the effects of mental models and expertise on the ability to process handoffs of information. In addition, the role of active or passive processing was examined. Three groups of participants participated, differing in their level of clinical expertise to represent a novice, intermediate, and expert population. Participants performed an abstract running memory span task and two tasks resembling real world activities, an air traffic control (ATC) handoff task, and a clinical handoff task. For all tasks list length and the amount of information to be recalled was manipulated. Further, in the ATC and the clinical handoff tasks, information was presented in an organized or unorganized sequence. Recall scores decreased as list length increased on all tasks. Regarding processing strategy, all participants used passive processing for the running memory span and ATC tasks. The novices also used passive processing for the clinical task. The experts, however, appeared to use more active processing as they recalled more relevant than irrelevant items. Irrelevant information negatively impacted all participants, resulting in lower handoff scores and decreased recall of relevant items. Regarding organization, experts had lower handoff scores for the clinical unorganized lists while intermediates and novices were not significantly affected. There was no effect of organization on the groups for the ATC task. Overall, the results indicated that individuals with clinical expertise and a developed mental model rely more

on active processing of incoming information while individuals with little or no knowledge rely on passive processing. Further, presenting irrelevant information and unorganized information incongruent with a developed mental model can negatively impact performance.

This dissertation is dedicated to Mark W. Scerbo. Thank you for believing in me when I did not believe in myself.

ACKNOWLEDGEMENTS

I want to thank everyone who made this dissertation possible. First, I would like to express my deep gratitude to my advisor Mark W. Scerbo for all of your patience, unwavering support, and excellent guidance throughout my entire tenure as a graduate student. I would not be where I am today without your support. I want to thank my committee James P. Bliss, Dana E. Ramirez, Ivan K. Ash, and Ginger S. Watson-Papelis for your excellent insight and suggestions. Further, I would like to thank Thomas Hubbard for assisting in coordinating the research process at the medical school.

I have been fortunate to have the support of a number of friends who patiently listened to me as sounding boards as I worked out the logistics of what I was attempting to do. I would especially like to thank Margie Robinson, Tiffany Squyres, Stephanie Latigona, Amanda Fox, Liz Newlin-Canzone, Robert Turner, and Rebecca Kennedy.

I am grateful to my mom, Kathy, for always insisting “education, education, education”. Mom, I took that statement and ran with it. I must thank my mother-in-law Chi Chi and father-in-law Freddy. You helped in many ways and it was greatly appreciated. Chi Chi our long walks with Tucker were cherished and helped me clear my head on a number of occasions. I also want to acknowledge my sister-in-law Christy for always making the time to chat when I needed to hear a friendly voice.

Finally, I would like to tell my husband Noah that his love, humor, and support helped keep my spirits high and helped me focus. I am grateful that you stuck with me through the good and the bad and look forward to the next chapter in our lives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Communication of information is a fundamental component of human interactions. However, the successful communication of information can prove to be a challenging task and can be affected by a number of variables. Some components that might affect successful communication relate to an individual's memory and that individual's own internal understanding, or mental model, of what information needs to be passed along. Further, whether an individual is an expert in an area might also affect his or her ability to communicate critical information. Large amounts of information might tax an individual's memory leading to less successful communication while an inadequate understanding of the current situation might also compromise successful transmission of information.

The healthcare field is one domain where critical information needs to be shared in a timely and effective manner. About a decade ago, the Accreditation Council for Graduate Medical Education (ACGME) passed a mandate that restricted the number of hours that residents could work to 80 per week (ACGME, 2003; Poulose et al., 2005). An effect of this mandate was an increase in the number of transitions of patient care among healthcare providers (Chang, Arora, Lev-Ari, D'Arcy, & Keysar, 2010; Nemeth et al., 2006; Patterson & Wears, 2010), which have been associated with errors in patient care (Chu et al., 2010; Riesenberg, Leitzsch, Massucci, et al., 2009; Weinger, et al., 2010). In the attempt to address this issue, the ACGME (2011) has issued a new set of standards regarding transitions of patient care for medical programs that took effect July 1, 2011. Among the standards is a requirement for residents to be competent in communication with team members during the handoff of patient care. Therefore, the need to examine characteristics that affect communication during transitions of care has become paramount in the medical domain. Some of these include memory, mental models, and expertise.

CHAPTER II

BACKGROUND

Memory

Historical Background

Memory refers to an individual's ability to maintain and store information, from the past and the present, obtained through learning or experience (Cofer, 1976; Davis & Palladino, 1997; Hunt & Ellis, 2004). The concept of what constitutes human memory has a long history, dating back thousands of years to Plato and Aristotle who speculated on the nature of memory (Burnham, 1888). However, the modern concept of memory can be traced back to the first book published on the topic by Ebbinghaus in 1885 (Hunt & Ellis, 2004). Ebbinghaus conducted a series of experiments on memory, looking at learning and forgetting of novel material, using over 2,000 nonsense syllables created to avoid effects of learned associations with familiar words (Bower, 2000; Hunt & Ellis, 2004; Matlin, 2002; Tulving, 1979). Key findings from Ebbinghaus' research were that forgetting decreased with increased exposures to the same list and that memory for the list declined as more time elapsed from the last exposure.

William James (1890) was a philosopher who wrote about memory and was the first to suggest there are two components to memory, which he referred to as primary and secondary memory. James characterized primary memory as a state of mind where new information is held in conscious thought, and described secondary memory as information that was in an inactive state, but could be called to an active state because it had previously been experienced. However, the concept of multiple components of memory would not be revisited for over fifty years due to a shift in the focus of psychology in the United States and in Europe to other forms of research, specifically behaviorism (Hunt & Ellis, 2004; Matlin, 2002). During this time, research on memory stagnated with very little published (Tulving, 1979). The one major exception to this was the work done by Bartlett, a British psychologist, who extended the work of Ebbinghaus using real words and stories (Bartlett, 1932; Matlin, 2002).

The arrival of the cognitive renaissance in the 1960s refocused research in psychology and there was great renewed interest in memory (Bower, 2000; Hunt & Ellis, 2004; Matlin, 2002; Tulving, 1979). The concept of multiple components of memory was embraced and led to many theories and models. For example, Waugh and Norman (1965) utilized James's nomenclature and further suggested that primary memory is limited while secondary memory is larger and more stable. Around the same time several investigators suggested the existence of a third component of memory, which was largely influenced by the introduction of the information-processing model (Hunt & Ellis, 2004). The new model described three components of memory: sensory memory, short-term memory (formerly primary), and long-term memory (formerly secondary). Sperling (1960; 1963) was one of the first to describe a sensory memory, but he referred to it as visual information storage. As Hill and Bliss (1968) noted, the early three-stage models including sensory memory were largely based on visual research to show support for that stage of memory. By 1970, Norman concluded that most researchers accepted the three-stage model of memory, with a few exceptions (e.g., Bernbach, 1970; Murdock, 1970; Wickelgren, 1970). Atkinson and Shiffrin (1968) proposed the most fundamental three-stage model, which would later become known as the modal model of memory (Hunt & Ellis, 2004; Matlin, 2002).

Atkinson and Shiffrin Model of Memory

Atkinson and Shiffrin's (1968) model of memory is comprised of the sensory register, short-term register, and long-term register. They described the sensory register as a temporary storage compartment for incoming sensory information. The information is held for one or two seconds without rehearsal before deteriorating. However, some of the sensory information might pass into the short-term store. Atkinson and Shiffrin defined the short-term store as a person's working memory. They postulated that information in this store is also fragile but can last longer (up to 30s without rehearsal) than information in the sensory storage. However, if a person engaged in active rehearsal then a limited amount of information could be held in this store for an indefinite time period. Atkinson and Shiffrin also suggested that information could be transferred from the short-term store to the long-term store. Last, the short-term store could receive input from the long-term store. Atkinson and Shiffrin described the long-term store as a more permanent storage unit with a large capacity.

Baddeley's Working Memory Model

A few years after Atkinson and Shiffrin (1968) proposed their model of memory, Baddeley and Hitch (1974) noted that little was still known about the role of short-term memory in information processing and suggested there was a common working memory space for reasoning, learning, and comprehension. Further, they described how this system differed from the other proposed short-term memory systems previously described. The major departure from Atkinson and Shiffrin's model was that short-term memory was not a single unitary store (Baddeley, 1981; Baddeley, 1986; Baddeley & Hitch, 1974). Rather, based upon preliminary investigations, they hypothesized that short-term memory was composed of subsystems and that it should be referred to as working memory. Baddeley and Hitch suggested that working memory was a "work space" that performed two tasks: storage and processing.

The original working model consisted of three components: central executive, phonological loop, and the visuospatial sketchpad (Baddeley, 1981; Baddeley, 1986; Baddeley & Hitch, 1974). The central executive was considered the control component and was thought to integrate information from the other two systems, termed the slave systems. The phonological loop stored and maintained verbal material while the visuospatial sketchpad stored and manipulated visual and spatial information.

A major limitation of the original working memory model was that it did not address how information from the long-term store and subsidiary stores could be integrated together, a process termed binding, thus allowing for the formation of one cohesive representation and subsequent active manipulation of the integrated information (Baddeley, 2001; Baddeley & Wilson, 2002). Therefore, Baddeley (2000) proposed an additional component to the working memory model, the episodic buffer. This component is a temporary storage unit under the control of the central executive. Baddeley suggested that the buffer could integrate information from multiple sources, including the two slave systems and long-term memory. Further, the buffer is considered to be a "mental modeling space" (Baddeley, 2000; Baddeley, 2001; Baddeley & Wilson, 2002). The modeling space allows for manipulation and refinement of information, thus possibly resulting in new cognitive representations to aid in future decision making. Further research by Baddeley has led to more insight into the workings of the episodic buffer (see Allen, Baddeley, & Hitch, 2006; Baddeley, Hitch, & Allen, 2009; Baddeley, Allen, & Hitch, 2011). One of the features of the episodic buffer that is similar to other components of working memory, is it is considered to have a limited capacity (Baddeley, 2004).

Limited Capacity of Working Memory

A common component of any model of short-term/working memory is its limited capacity to hold information. In a classic paper, Miller (1956) suggested that the amount of information that an individual can maintain in immediate memory is limited. Miller suggested that individuals can hold roughly five to nine pieces of information in their immediate memory with an average of seven items. For example, someone can hold five to nine letters or numbers in their short-term memory. However, Miller also suggested that some of this information could be grouped into a single piece of information, which he referred to as a chunk. For example, the letters O, H, R, E, S could be presented and considered five pieces of information. However, the same letters could be presented as a word, horse, and be considered as one chunk of information; thus, allowing an individual to hold substantially more information in short-term memory through chunking. Recently, Cowan (2000) has argued that the amount of information that can be held in immediate memory might be limited more than what Miller (1956) suggested. Cowan reviewed several sources of data surrounding capacity limits using scenarios not examined by Miller and suggested that under different conditions the capacity limit was closer to four. He indicated that when an individual is prevented from engaging in rehearsal and from accessing long-term memory to recode information into more manageable chunks of information, the capacity limit is a smaller number than the original proposed range of 7 ± 2 items.

Another common theme of short-term/working memory models is that the information deteriorates rapidly if it is not rehearsed. Support for this notion can be found in research performed by Brown (1958) and Peterson and Peterson (1959). To examine short-term retention, Peterson and Peterson examined how well participants recalled presented items when half of the participants were instructed to verbally rehearse the items while the other half were prevented from rehearsing the items. They found that with rehearsal, retention improved, lending support to the notion that memories in the short-term store are susceptible to rapid deterioration if not rehearsed.

Running Memory

Occasionally, it is necessary to both hold and update items in working memory with or without the opportunity to engage in rehearsal. Pollack, Johnson, and Knaff (1959) first described running memory. The traditional running memory span test is a task where items must be recalled, but the individual does not know when the items will be recalled or the length of the list of items to be recalled. Therefore, the individual would need to maintain an ongoing list of items in working memory. However, Pollack et al. suggested that in applied scenarios, such as monitoring displays, the older information would no longer be relevant and therefore it would be necessary for old items to drop out of running memory while maintaining the newer items, therefore updating working memory. Pollack and colleagues presented digit spans of known and unknown lengths to participants and found that performance was better for the known length condition. However, in many real world scenarios, it is often necessary to maintain working knowledge of a number of dynamic items of unknown length that must be updated in working memory (Wickens & Hollands, 2000).

For example, Yntema and Mueser (1960) noted that in order to perform certain jobs, such as air traffic control (ATC), people must keep track of several changing variables at once. In a series of studies Yntema and colleagues (1960; 1962; 1963) examined how well individuals could keep track of many objects with changing variables. Yntema and Mueser (1960) compared how well people could keep track of information presented as either one object that had many traits or many objects that varied on the same trait. They found that increasing the number of variables that needed to be remembered resulted in poorer performance. Further, participants were able to keep track of one variable that differed along many attributes better than keeping track of multiple attributes that varied on the same trait.

In a follow-up study, Yntema (1963) examined how well participants could keep track of several objects with several varying attributes. This task is analogous to an air traffic controller keeping track of several different airplanes. Each airplane, assigned a flight number, would be tracked on a number of attributes such as altitude, airspeed, and estimated arrival time. Further, each attribute would differ for each plane. For example, one airplane might be maintaining an altitude of 25,000 feet and traveling at 450 miles per hour (MPH) with an arrival time of 20 min while another might maintaining an altitude of 30,000 feet and traveling at 500 MPH with an arrival time of 16 min. Yntema used nonsense objects and had participants listen to messages that consisted of a variable and its current state. At random points in time the participant would be asked to recall information about one of the variables. Yntema again found that as the number of objects increased, participants' retention declined. Further, a substantial number of mistakes were made with three or more variables, particularly when participants had to keep track of several objects that differed on the same attribute. Because it appeared that information about the same attribute interfered with memory for the different objects, Yntema suggested that each variable should have its own set of states. However, this is not always possible, necessitating the need to find a different approach to aid individuals when they must rely on running memory. Other dynamic running memory tasks have also found severe limits for the amount of information that could be recalled (Mackworth, 1959; Stein, Garland, Muller, 2010).

Running memory tasks require individuals to update their working memory (Morris & Jones, 1990). As previously mentioned, working memory is flexible and allows for updates of information thus providing a method for how new relations are formed from different sources of information, which is referred to as binding. The relational representations are mental models (Baddeley, 2001; Oberauer & Vockenberg, 2009) and will be addressed in more detail below. Oberauer and Vockenberg suggested that binding must occur quickly but must also be undone quickly in working memory as new information is presented altering the current state of the bind. As each bind is modified, memory is updated. Morris and Jones (1990) defined memory updating as the modification of a concept (mental model) in long-term memory in light of new incoming information.

Two strategies have been suggested for how individuals process information and update their memory in running memory span tasks: active and passive processing (Broadway & Engle, 2010; Bunting, Cowan, & Saults, 2006; Elosúa & Ruiz, 2008; Hockey; 1973; Morris & Jones, 1990). Active processing assumes that individuals are constantly updating their working memory by grouping incoming information and discarding old information as newer information is presented. On the other hand, passive processing suggests that individuals hold the presented information without trying to process it further and then they recall as much as possible when prompted.

Elosúa and Ruiz (2008) and Broadway and Engle (2010) examined the two processing strategies for running memory span tasks. Elosúa and Ruiz attempted to determine if individuals use an active approach when presented with a large amount of information. Broadway and Engle attempted to determine whether an active approach was adopted when the amount of time between presented items was varied. Elosúa and Ruiz varied the length of the word list (range 4 to 26); however, the number of words to be recalled was held constant at four. Broadway and Engle (2010) also varied the length of the stimulus list (range 4 to 10) along with the number of items to be recalled. They used whole recall trials, where all items had to be recalled and partial recall trials where only some of the targets had to be recalled. An advantage to using whole and partial recall trials is to prevent participants from immediately discarding the first items (Broadway & Engle, 2010).

All of these researchers expected evidence to support active processing, but in each instance the results supported a passive processing approach. However, in each study participants were presented with information that was abstract in nature, rather than concrete. In their running memory tests Yntema and colleagues (1960; 1962; 1963) modeled their tasks after the real world scenario of ATC, a concrete type of task. A limitation of the research by Yntema, however, was that even though the domain was ATC, the tasks themselves were still abstract in nature as they were not true ATC tasks (Stein et al., 2009). On the other hand, Venturino and colleagues (1994; 1997) extended the research by Yntema using a real world task of a fire department dispatcher, whose job it is to manage fire engines. However, a limitation of Venturino's research was that the participants were psychology students, not actual fire department dispatchers. If a real world task were examined using participants who worked in that field, it might be possible to examine whether information presented in a running memory format was actively or passively processed. For example, regarding ATC, the information the controller receives would be relevant to his or her overall ability to manage the aircraft in his or her airspace. A controller might be managing one airplane when he or she receives information that a storm is developing, a large flock of birds is in the airplane's path, and another airplane is entering the airspace 3,000 feet above the present aircraft. Because the incoming information is important to maintain control of the airspace (a concrete task) it might require active rather than passive processing of the information. However, an individual's understanding and expertise level with the task might also affect which approach is adopted. For an individual who is not familiar with ATC, the incoming information might not be meaningful resulting in a passive approach. If the individual has an understanding of ATC then an active approach might be adopted because the incoming information has meaning and affects how he or she continues to perform a task.

Running memory span tasks have been performed using a variety of stimuli including single digits (e.g., Bunting et al., 2006; Pollack et al., 1959), letters (Broadway & Engle, 2010), and disyllabic words (e.g., Elosúa & Ruiz, 2008; Postman, Turnage, & Silverstein, 1964). One limitation to using digits is that if the list exceeds nine items, then repetition starts to occur. Further, when presented verbally some of the digits are monosyllabic (e.g., one, two) while some are disyllabic (e.g., zero, seven). Using letters offers the ability to increase the string of information, however, there are still limits. First, letters that can be easily confused when presented verbally (e.g., “B” and “P”) may need to be excluded. Further, another potential issue concerns the use of abbreviations. For example, if the following string of letters were presented, H X N V D, for non-clinical individuals each letter would most likely be perceived as a separate item, resulting in 5 items of information. However, clinicians might chunk the letters into known medical abbreviations. H X could become one chunk because Hx is an abbreviation for history and physical while N V D could become one chunk because it is the medical abbreviation for nausea, vomiting, and diarrhea. Therefore, the clinicians might perceive the information as two chunks of information, resulting in a possible advantage for performing the running memory span task.

Mental Models

As humans acquire new information, our previously held conceptions about objects and processes might be altered. These conceptions are mental models. The concept of a mental model has been examined in multiple fields, resulting in some confusion about how to describe them (Brewer, 2006). Some of these fields include cognitive psychology, human factors, science education (Brewer, 2006) and more recently natural resource management (Jones, Ross, Lynam, Perez, & Leitch, 2011). Despite its perceived value, there is still not a single, unified concept of a mental model, although progress has been made to reach consensus and resolve some of the underlying confusion (Brewer, 2006; Rutherford & Wilson, 1991/2004). Brewer offered the following definition of a mental model based on common themes found from reviewing the literature from various fields: a mental model consists of an individual’s internal cognitive representation for domains that are mechanical-causal in nature and further allows the individual to explain the domain in question.

Development of Mental Model Construct

Craik (1943) has been credited with first describing the modern notion of what constitutes a mental model (Johnson-Laird, 1983; Jones et al., 2011; Moray, 1997/2004). Craik postulated that thought serves a definitive function, and a core property is to predict future events. This is accomplished through reasoning and Craik described three fundamental processes. The first process is translation in which external information is converted into internal words or symbols. The second process entails internal analysis and interpretation and the third process is retranslation. During this phase internal concepts are reconverted to an external process.

Craik (1943) suggested that we hold “small-scale models” in our thoughts of different processes. These models can assist in reasoning. The models allow individuals to imagine different occurrences and their outcomes without actually having them occur. Craik stated that models allow different actions to be tried and assessed thereby allowing the prediction of future events while drawing on the stored knowledge of past events in memory to aid in present and future decisions.

In the early 1980s the concept of an internal model suggested by Craik (1943) started to garner more attention, resulting in the publication of two books (Gentner & Stevens, 1983; Johnson-Laird, 1983). Rouse and Morris noted in 1986 that the term, mental model, was being used abundantly even though there was still no formal definition. Confusion surrounding the concept of a mental model continued into the 1990s (Moray, 1997/2004) during which time Rutherford and Wilson (1991/2004) acknowledged there was still not a single concept of a mental model, but began working together to address some of the confusion. More recently, Brewer (2006) identified some commonalties in the literature and provided the aforementioned description of a mental model.

Taxonomies of Mental Models

Norman (1983) proposed that there are three core components that must be addressed to examine mental models. The first element concerns an individual's beliefs about the physical system in question. The second component is based on the parameter of observability. Norman proposed that the components of a person's mental model should correlate with the actual observable components of the physical system. For example, if one has a mental model of a tree then their model should include a trunk, branches and leaves, which correlates with a physical tree. The final component Norman refers to is predictive power. Similar to Craik's (1943) notion of a model, Norman suggested that a person's mental model can be used to understand the components and operations of a physical system, thus allowing them to test different scenarios mentally.

Based on Norman's (1983) taxonomy, Moray (1997/2004) suggested that there are three types of mental models. Type 1 consists of a system designer's mental model. The model represents a system that is in the design phase that will be constructed at a future time. A Type 2 model refers to the operator's model of a system or device, which is often an imperfect representation of the system. An example provided by Moray is a user's model of how a calculator functions. Another example would be a physician's model of a patient's condition. Moray suggested a Type 2 model is a close match to Craik's description of a mental model. A Type 3 model is the researcher's model of another individual's model, typically a user or operator of a system or device. For example, an attending physician has a mental model of what constitutes their resident's mental model. However, these two models may not always correspond to each other.

Memory and Mental Models

There is some debate regarding where mental models are developed and held in memory. Craik (1943) described mental models as a product of long-term memory. Moray (1997/2004) supported this notion and suggested that mental models describe content found in the long-term memory store. On the other hand, Johnson-Laird (1983) and Brewer (1987) suggest that mental models are more transient constructs that are formed in working memory. Recently, however, Nersessian (2002) proposed that mental models occupy both long-term and working memory. The knowledge that an individual possesses concerning a particular topic is stored as a mental model in long-term memory, and that model can be called to working memory where it can be modified and updated when new pertinent information becomes available. This in turn will be stored back in long-term memory.

Handoffs in Medicine

Having an accurate mental model of a system might assist an individual in organizing their thoughts when they need to pass along information to another individual. Many industries, such as aviation and nuclear power, require effective handoffs of information (Arora & Johnson, 2008; Patterson, Roth, Woods, Chow, & Gomes, 2004). Handoffs consist of transferring information and responsibility from one individual to another. The healthcare field is also one domain where critical information needs to be shared in a timely and effective manner. Handoff strategies were first studied in the nursing domain (Raduma-Tomás, Flin, Yule, & Williams, 2011; Sharit, McCane, Thevenin, & Barach, 2005). Clair and Trussel (1969) were the first individuals to examine handoff communication, or what they referred to as a change of shift report. They asked nurses to state what they thought should be included in the change of shift report and then tape-recorded and reviewed actual handoffs. Clair and Trussel found that there were discrepancies between what the nurse's thought should be included and what was actually included in their reports. Further, they suggested that interventions such as providing guidelines, implementing instructions on how to perform a change of shift, and minimizing interruptions might result in better change of shift reports.

Despite recognizing that inadequacies existed during handoffs, it was not until recently that interest in improving handovers in healthcare increased in importance due to rising concerns over patient safety. International institutions, such as the British Medical Association and Australian Council for Safety and Quality, as well as medical governing bodies in the United States, have offered recommendations on how to perform patient handovers (Cleland, Ross, Miller, & Patey, 2009; Cohen & Hilligoss, 2010). For example, in 2006 the Joint Commission on the Accreditation of Healthcare Organizations (now known as the Joint Commission) made handoffs a national patient safety goal, requiring standardization (Arora & Johnson, 2008; Cohen & Hilligoss, 2010; Riesenberg, Leitzsch, Massucci, et al., 2009).

Another reason that handoffs have become a greater concern in recent years can be traced to the Accreditation Council of Graduate Medical Education (ACGME) mandate to restrict resident work hours to 80 per week (ACGME, 2003; Poulose et al., 2005). An unexpected consequence of this mandate was an increase in the number of transitions in patient care among healthcare providers (Arora & Johnson, 2006; Chang et al., 2010; Nemeth et al., 2006; Patterson & Wears, 2010), accompanied by an increase in errors in patient care (Chu et al., 2010; Riesenberg, Leitzsch, Massucci, et al., 2009; Weinger, et al., 2010). In an attempt to address the increase in errors, the ACGME has developed new standards for graduate medical programs. One requirement is that institutions must attempt to reduce the overall number of transitions of care (ACGME, 2011). A second requirement calls for standardized handoff protocols. Further, competency in one's ability to perform a handoff needs to be assessed. Therefore, the need to examine and improve communication has become paramount in the medical domain.

Challenges to Studying Clinical Handoffs

Poor handoffs can set the stage for events that may result in patient harm. Ineffective communication can result in longer hospital stays, repeated tests, and longer diagnosis and treatment times (Lawrence, Tomolo, Garlisis, & Aron, 2008; Patterson & Wears, 2010). All these events can also result in lower patient satisfaction and increase the cost of hospital stays. Therefore, it is important to address the challenges associated with performing handoffs to help determine more effective ways to transition patient care.

There are many factors that affect opportunities to study and improve clinical handoffs. Perhaps, the primary issue is a lack of agreement on the terminology used to describe this activity. Sign-outs, sign-overs, handovers, transitions of care, and turnovers have all been used in the literature (Arora & Johnson, 2008; Cleland, et al., 2009; Riesenberg, Leitzsch, Massucci, et al., 2009). Cohen and Hilligoss (2010) noted that there is no standard definition of a handoff and argued that this lack of consensus leads to confusion about how to improve handoffs. For the current paper, the definition provided by Patterson and Wears (2010) will be used to describe a patient handoff: "the process of transferring primary authority and responsibility for providing clinical care to a patient from one departing caregiver to one oncoming caregiver". A fundamental component of this definition, the complete transfer of responsibility for that patient, resonated in other definitions (Cohen & Hilligoss, 2010; Parush, Simoneau, Foster-Hunt, Thomas, & Rashotte, 2010; Sharit et al., 2005).

Another important factor is that there are many different types of handoffs (Patterson et al., 2004; Patterson, Roth, & Render, 2005; Patterson & Wears, 2010). For example, there are handoffs from physician to physician, nurse to physician, and nurse to nurse handoffs. Each of these handoffs might also differ depending on the experience level of the participants, such as attending physicians to other attendings, residents to residents, and residents to attendings. Further, there are handoffs within departments, from one department to another, and from outside institutions. These examples do not cover all possible types of handoffs, but they do highlight the difficulty of addressing how to best perform handoffs, as each type of handoff may require different kinds of information to be communicated and therefore make the standardization suggested by the Joint Commission difficult.

An additional challenge is that there are multiple methods of performing handoffs (Patterson, et al., 2005;Raduma-Tomàs, et al., 2011). For example, they can be done face-to-face, in writing, and with electronic sign-out systems. These methods fall into two classes of communication: synchronous and asynchronous (Horwitz & Detsky, 2011;

Parush, et al., 2010; Parker & Coiera, 2000). Synchronous communication occurs when the communicator and receiver are both present to perform the handoff, while asynchronous communication occurs without both parties present (Horwitz & Detsky, 2011; Parush, et al., 2010; Parker & Coiera, 2000). Synchronous communication usually occurs in face-to-face settings or over the phone. On the other hand, asynchronous communication can occur via written notes, e-mail, and voicemail. Both forms of communication have advantages and disadvantages. For example, synchronous communication allows for personal interaction and immediate clarification of ambiguous information, but requires a specific time for both parties to be present (Horwitz & Detsky, 2011). On the other hand, asynchronous communication allows for both parties to address the information at a time convenient to them, but also has a major disadvantage in that it prevents the receiver from being able to immediately clarify information.

Another challenge to improving handoffs is that people tend to overestimate their ability to communicate information effectively to another individual (Fay, Page, Serfaty, 2010; Fay, Page, Serfaty, Tai, & Winkler, 2008; Keysar & Henly, 2002). This bias in self-efficacy can be especially problematic when communicating critical information such as patient data to another staff member. For example, some information might be ambiguous to the receiver and the receiver might not perceive some information as critical. Chang and colleagues (2010) performed a study examining how pediatric interns viewed their effectiveness at communicating information during patient handoffs. They interviewed interns and asked the outgoing interns to rate the effectiveness of their communication, which was compared to the ratings of the oncoming interns. Chang et al. found that the outgoing interns believed they clearly communicated the most important piece of information to the oncoming interns; however, in 60% of the cases the oncoming interns did not view this as the most important piece of information.

Standardization and Training

As previously mentioned, the ACGME and the Joint Commission are requiring standardization of the handoff process. Standardization has potential benefits, but also has some limitations. Patterson (2008) suggested that a positive aspect of standardizing communication is that information could be conveyed more efficiently because it would be presented in an orderly fashion. Further, if no information is presented it can be assumed that there is nothing of importance to report.

However, Cohen and Hilligoss (2008, 2010) point out that presently there is little empirical evidence to support widespread standardization of handoffs, despite attempts to start developing standardized models and curricula (i.e., Arora & Johnson, 2006; Wayne, et al., 2008). Patterson (2008) also noted that other high-reliability organizations have not developed standardized verbal handovers, with the possible exception of transitions on nuclear submarines. Lack of proven standard verbal protocols from other industries prevents the medical industry from simply adopting a method already in place and modifying it, necessitating the development of a novel verbal protocol that needs to be tested and validated. However, Patterson and Wears (2010) performed a review of the handoff literature and determined that although there are a number of initiatives focused on improving handoffs, there is a lack of standard measurement tools available.

Currently, there is also a lack of training alternatives to formally teach proper handoffs (Arora & Johnson, 2006; Chu et al., 2009; Wayne et al., 2008), although Chu and colleagues (2009, 2010) have begun developing and testing a teaching curriculum.

Cleland et al. (2009) noted that it is common practice for doctors in training to perform many handoffs, even though they have never had any formal handoff training. Cleland et al. invited a number of physicians, including junior doctors, to participate in focus groups to gather information about handoffs. They reported that many of the junior doctors experienced high levels of stress during their first handoffs. The junior doctors reported feeling overwhelmed because they did not understand how to prioritize the patients and were uncertain about their exact roles and responsibilities. Further, the junior doctors embraced the idea of participating in formal handoff training. Cleland et al. suggested that an ideal time to teach handoffs might be during medical school.

Presently, the majority of the clinical handoff research remains largely descriptive (Manser, Foster, Gisin, Jaeckel, & Ummenhofer, 2010). Many of the studies offered observations that can help direct empirical research. Further, a number of systematic reviews of handoff research have been published recently (see Raduma-Tomàs et al., 2011; Riesenberg, Leitzsch, Massucci, et al., 2009; Riesenberg, Leitzsch, & Cunningham, 2010; Riesenberg, Leitzsch, & Little, 2009; Solet, Norvell, Rutan, & Frankel, 2005). Many barriers that affect handoffs have been identified. For example, Riesenberg et al. (2009, 2010) reviewed handoffs in the clinical literature and identified a number of potential barriers to effective communication including noisy environments, interruptions, lack of time, disorganized reports, and multitasking during the handoff.

Omission of important data has also been identified as a barrier to effective communication (Arora, Johnson, Lovinger, Humphrey, and Meltzer, 2005). Horwitz, Moin, Krumholz, Wang, and Bradley (2008) performed a prospective study of handoffs to examine the types of errors that arose due to omission of information. They audiotaped handoffs and then asked the receivers to note any problems that arose associated with the handoff. Horwitz et al. identified six types of omissions that compromised patient care. One was omitting the current clinical condition of the patient, which made it difficult to ascertain if the patient was deteriorating. The second was failing to include recent and scheduled events. Again, this led to confusion and lost time because the receiving doctor needed to go back and review charts to determine what had been done. The third, fourth and fifth omissions were all related to a lack of guidance for the oncoming doctor including the absence of anticipatory guidance, failure to assign tasks, and failure to provide a plan for how to complete the assigned task. The final omission identified was the absence of any rationale for what needed to be done, which also compromised the ability to care for the patient. It is important to note that Horwitz, Moin, Krumholz, Wang, and Bradley (2009) found that lack of familiarity with the patient was correlated with the omission of data.

Stages of Handoffs

One place to start addressing some of the challenges with handoffs is to analyze the individual components (Raduma-Tomàs, Flin, Yule, & Close, 2010). In general, there are three stages to a handoff: prehandover, handover, and posthandover. Although other descriptions that further differentiate the stages have been suggested (e.g., Lawrence, et al., 2008; Matthews, Harvey, Schuster, & Durso, 2002; Raduma-Tomàs, et al., 2011; Wears, et al., 2004), all share these general three stages. Also, an important point to consider is that there are always two individuals involved in a handoff: the outgoing caregiver and the oncoming caregiver; however, both parties are usually involved in only the handover stage.

Raduma-Tomàs et al. (2010) describe the prehandover stage as the time during which outgoing caregivers prepare for the transition of the patient to the oncoming shift. They suggest the most important step in this stage is ensuring that the patient list is fully updated. Other activities that are performed during this stage include checking what jobs have been completed, what tests have been ordered, determining where patients are located and if they are being transferred, placing patients notes in the handoff location, and determining what tasks the oncoming caregiver needed to perform. The handover stage consists of the actual transition from the outgoing caregivers to the oncoming caregivers. The posthandover stage occurs right after the handover. The primary activity performed during this stage is for the receiver to prioritize the tasks to be done for the patients who were received.

Expertise

Handoffs of patient care occur at multiple levels, from first-year residents to attending physicians, representing various levels of expertise. An individual's level of expertise might affect his or her ability to effectively communicate and pass along patient care information. Research regarding the development of expertise has grown significantly over the past half century (Ericsson, 2006). Expertise has been defined as the development of skills and comprehension based on extensive acquisition of knowledge (Chi, 2006). It has been suggested that to attain expertise in a particular area it takes a significant amount of time and engagement in the activity (Ericsson, 1998), as much as ten years, a phenomenon referred to as the ten-year rule (Ericsson, Prietula, & Cokely, 2007; Simon and Chase, 1973). However, the length of time can vary. Ultimately, many factors such as the specific domain (Ericsson, 2006), individual differences, and commitment to practice (Simonton, 2006) affect how long each individual takes to become an expert.

Norman, Eva, Brooks, and Hamstra (2006) point out that expertise in medicine is unique compared to other domains because it requires knowledge and competency at many levels such as motor skills and interpersonal skills. Further, medicine is always changing as new therapies become available and new techniques to perform procedures are developed, necessitating medical personnel to constantly update their skill set. For example, a surgeon might have been performing open surgeries for 15 years and considered an expert in his or her field; however, the introduction of laparoscopic surgery necessitates that the surgeon learn this new technique. Laparoscopic surgery, though, is fundamentally different from open surgery and the expert surgeon might be at the novice level on some of these skills.

Although expertise in medicine has generated a lot of interest, the majority of research has been directed at differences across levels of experience for diagnostic decision-making (Norman et al., 2006), with little attention directed to how expertise affects clinical handovers. However, the findings from diagnostic studies might provide insight into how novices and experts differ in their clinical reasoning and subsequent handoff of patient information.

Memory and Medical Expertise

The role that memory plays in expert performance and how memory can be measured to study expertise has been a topic of interest to researchers since de Groot (1965) first described expert performance for chess masters (Norman, Brooks, & Allen, 1989). de Groot found that expert chess players were able to accurately recall up to 90%

of chess piece positions after a 5-sec exposure; however, this only occurred for legal positions. If the pieces on the board were in random positions that could not occur in the game, expert's recall was on the same level as novices.

Researchers have attempted to model medical expertise research based on the studies of chess, but with far less success (Eva, Norman, Neville, Wood, & Brooks, 2002; Norman, 2005; Norman et al., 2006). For example, Norman, Jacoby, Feightner, and Campbell (1979) and Coughlin and Patel (1987) examined expert versus novice performance for recalling typical and atypical cases. Experts recalled more information for the typical cases, but performed at the novice level for atypical cases. Coughlin and Patel, however, found no overall differences in the amount of information recalled based on expertise level. Other studies (e.g., Eva et al., 2002; Schmidt & Boshuizen, 1993) have found that individuals with an intermediate level of clinical expertise recall the most information after reading clinical protocols.

Muzzin, Norman, Feightner, Tugwell, and Guyatt (1983) examined how varying the time to study the cases affected recall performance. At an exposure of 2 minutes the level of expertise did not affect performance; however, at shorter exposure times experts recalled fewer details than novices. Upon closer inspection of the actual information recalled, though, it appeared the experts had processed the information more and picked out and grouped key elements while the novices had simply recalled the information verbatim.

Mental Models and Medical Expertise

Because recall experiments did not yield consistent results, research shifted to how experts and novices organize their knowledge (Norman, 2005; Norman et al., 2006).

Norman (2005) questioned whether gaining medical expertise is simply a matter of extensive practice; however, he suggested that this was unlikely. In his opinion, experts in medicine rely on a large network of mental representations to solve problems. Ericsson and Kintsch (1995) and Schmidt, Norman, and Boshuizen (1990) believe that regarding diagnostic decision-making, experts should have a different knowledge structure than novices. Consequently, Schmidt et al. proposed a theory of expertise based on the notion that mental models change with accumulated experience in medicine.

Schmidt et al. (1990) proposed a four-stage model. At Stage 1, medical students start to develop causal networks of disease. They learn the causes of certain diseases and the resulting consequences based on pathophysiological processes. However, they do not have a deeper understanding of how disease can manifest itself in different ways. The medical student's mental model is very basic and structured largely around what they have learned from textbooks rather than real experiences. In an earlier study, Schmidt, Boshuizen, and Hobus (1988) asked participants with different levels of expertise to describe the pathophysiological process of a disease. When novice students described a pathophysiologic process for a clinical case, the result was often a lengthy and inadequate description. On the other hand, advanced students were much more succinct, while also demonstrating a deeper understanding of the interrelations of all variables involved.

Schmidt et al. (1990) suggest that at Stage 2 the causal networks change into causal models. Signs and symptoms become clustered under diagnostic labels. They also claim that this does not occur until the students are exposed to real patients. The models develop through repeated exposure to patients and the synthesis of knowledge acquired during medical school. Schmidt et al. suggest that a medical student diagnosing his or her first patient requires extensive mental effort, but over time as the model develops for similar cases, heuristics become available that lessen the amount of knowledge that needs to be mentally activated to diagnose each case.

During Stage 3 knowledge becomes reorganized, from causal networks to list-like structures known as illness scripts (Schmidt et al., 1990). The illness scripts allow the students to develop simplified mental models, which in turn, allow them to understand how diseases manifest in different ways. An assumption is made that when a physician begins working on a diagnosis he or she searches memory for an illness script fitting the symptoms and then updates and modifies the script based on the current case. The scripts are also considered to be serial in structure. Schmidt et al. suggest that a physician uses the same structure to pass along information to another physician.

At Stage 4, which occurs almost simultaneously with Stage 3, individual patient encounters are stored in memory (Schmidt et al., 1990). These encounters enable the expert physician to rapidly recognize subsequent similar cases through pattern recognition.

There is evidence to support the idea that knowledge organization differs between novices and experts. Claessen and Boshuizen (1985) conducted a recall study similar to the memory studies discussed above, but instead focused on differences in knowledge organization across participants. They had medical students from two different universities participate. One group was taught in what was considered a traditional manor in its time. The students were taught theory and were not exposed to actual medical practice until fifth year. The other group was exposed to practical skills and also experienced encounters with simulated patients early in their training. They also had six doctors participate who served as experts. Claessen and Boshuizen had both student groups and the experts read a typical and an atypical case and sort the information into important and unimportant data. Then they had the participants attempt to reproduce the cases from memory.

Claessen and Boshuizen (1985) found that there were no differences for reproducing details based on whether the case was typical or atypical. Further, they found no difference among the three groups. However, they did find differences in how the experts recalled the cases; that is, the experts tended to cluster information. They concluded that as expertise develops doctors organize information of the illness and symptoms in a more patient-centered framework.

The model proposed by Schmidt et al. (1990) could help explain why novice physicians might struggle to perform handoffs. If they are at Stage 1, then they might have difficulty distinguishing the important information from the noncritical information. Indeed, Patel, Groen, and Frederiksen (1986) found that novices recall more irrelevant information compared to experts. Further, based on the proposed model of Schmidt et al., novices might present a basic interpretation of the disease, but might not be organized in their presentation. However, as they develop expertise and their mental models develop, they are able to start organizing information into related chunks. Muzzin et al. (1983) reported that novices recalled information from a clinical case sequentially, whereas the experts chunked the information. Further, one could assume that if the physician is on the receiving end, as expertise increases and mental models become more developed then he or she might be able to actively process the information being received in running memory and discard information deemed unnecessary.

CHAPTER III

THE PRESENT STUDY

The goal of the present experiment was to assess how well individuals with different levels of clinical expertise hand off information. As previously discussed, it has been suggested that people either actively or passively process information during running memory span tasks (Broadway & Engle, 2010; Bunting, et al., 2006; Elosúa & Ruiz, 2008; Hockey; 1973; Morris & Jones, 1990). Elosúa and Ruiz (2008) and Broadway and Engle (2010) hypothesized that individuals would actively process information; however, they found support for passive processing. One drawback to their research was that they used abstract tasks. Yntema and colleagues (1960; 1962; 1963) on the other hand modeled the tasks in their running memory span research on dynamically changing variables to represent more real world scenarios. Currently, there is little research that addresses how individuals process information in running memory when the information is needed to perform a task. An exception is work done by Hess and colleagues examining running memory performance in relation to monitoring changing information of graphic displays (1994; 1999). Therefore, one goal of the present research was to examine how individuals process information when it is relevant to performing a genuine task. It was hypothesized that individuals would engage in active processing when the presented information was relevant to performing another task; however, this was contingent upon them having knowledge about the task, which is discussed below.

Another goal of the present study was to examine the role of an individual's mental model and how it affects information processed in running memory tasks. Stein et al. (2010) noted that although the tasks by Yntema and colleagues were based on ATC, the experimental tasks themselves were still abstract in that the variables used held little or no meaning for the participant. Further, no work was performed using actual air traffic scenarios and air traffic controllers. Recall that Venturino et al. (1994; 1997) extended the work by Yntema and colleagues by using a realistic scenario of fire department dispatchers managing fire engines; however, the participants were psychology students, not trained dispatchers. If the information presented is needed to perform another task, but the task is unfamiliar, the individual might not have a mental model that can help distinguish which information needs to be retained and what can be discarded. For example, if an individual with no clinical training is presented with a list of information for patient care he/she might not be able to determine what information is truly relevant and what is extraneous. On the other hand, if an individual does have a mental model of the task, an active approach would be adopted. In clinical memory tasks Coughlin and Patel (1987) found that experts recalled significantly more critical details than novices. Patel et al. (1986) found that novices recalled more irrelevant details than experts, and Muzzin et al. (1983) found that experts grouped key items whereas novices tended to recall the information sequentially. Therefore, if a mental model helps to organize and distinguish the relevant from irrelevant information, then experts should actively process information in the clinical running memory tasks. Thus, the experts were expected to recall more critical information and regroup the key items. On the other hand, it was anticipated that the novices would passively process the information presented in the clinical running memory task resulting in sequential recall of relevant and irrelevant information and higher recall of irrelevant items compared to the experts.

An additional goal of the present research was to examine how expertise affects processing of information. As mentioned earlier, Schmidt et al. (1990) proposed a four-stage model of how expertise develops based on changing mental models. As medical students transition from novices to experts their mental models of clinical information also develop allowing for a better understanding of how diseases manifest as well as improved organization of their thoughts. The same model might also explain how individuals with different levels of clinical expertise process information and ultimately pass it along to another individual. Further, Stein et al. (2009) noted that experts in many fields show more effective use of memory than would be anticipated based on basic memory research.

To address these goals, three groups of participants representing different levels of expertise participated in the study. List length was manipulated for each of the tasks that the participants had to perform. It was hypothesized for all three tasks that as list length increased, correct detections would decrease. Regarding the disyllabic word running memory task it was anticipated that there would be no differences across expertise level for each of the measures and that the participants would adopt a passive approach, mirroring the results of Elosúa and Ruiz (2008) and Broadway and Engle (2010). Further, it was hypothesized that performance would be better on the partial lists compared to the whole recall lists.

One group was comprised of undergraduate students representing novices with no clinical background. It was hypothesized that when presented with clinical case scenarios in the form of a running memory task they would adopt a passive processing approach. Further, it was anticipated that there would be no differences in novice performance for recall length and organization.

Another group was comprised of expert clinicians. Recall that in the study by Muzzin et al. (1983) the experts actually recalled fewer details from clinical cases than less experienced individuals. However, upon further review of the data it appeared that the experts were picking out and grouping critical data. Therefore, it was hypothesized that the experts would rely on active processing. It was further hypothesized that experts would have higher handoff scores for the partial cases compared to the whole cases because there were fewer relevant items to recall in the partial cases. Regarding organization, it was hypothesized that the experts would exhibit better performance for the organized lists compared to the unorganized lists. In addition, it was anticipated that there would be an interaction effect for list length, recall length, and organization, with the combination of longer list lengths, whole recall, and unorganized lists resulting in poorer recall.

The third group was comprised of third-year medical students. Because they do have significant clinical experience, it was anticipated that they would be able to make some determinations about critical and noncritical information and it was hypothesized that they would exhibit an approach that falls between the novice and the expert performance.

To help establish whether differences in processing information are related to mental models and expertise a second running memory task unfamiliar to all participants (based on air traffic control) was also utilized. It was hypothesized that all groups would adopt a passive processing approach for this task recalling the data verbatim and in sequential order, similar to how the inexperienced individuals recalled the clinical data in the study by Muzzin and colleagues (1983). Further, it was anticipated that there would be no differences among the groups for list length or organization.

Method

Participants

Three groups of participants with different levels of expertise in medicine were asked to participate. All participants had normal or corrected-to-normal hearing and vision. Further, all participants were treated according to the American Psychological Association's ethical guidelines (2009) and IRB approval was obtained from Old Dominion University for the novice population and from Eastern Virginia Medical School for the intermediate and expert population. All groups were screened for experience in ATC or aviation. One intermediate had a background engineering related to aviation with limited ground school, but was still allowed to participate because he had no extensive knowledge of ATC or flight experience.

The novice group, which served as a control group, was comprised of undergraduate university students with no experience in the field of medicine. Thirty-six undergraduate students (27 females, 9 males) from Old Dominion University participated in this study through convenience sampling. They ranged in age from 18 to 26 (M = 20.81, SD = 1.91). An incentive to participate was credit offered that could be used to satisfy research participation requirements for Introductory Psychology or as extra credit for other psychology classes.

Nineteen undergraduate (4 second year, 15 third year) medical students (6 females, 13 males) from Eastern Virginia Medical School participated in this study. They ranged in age from 24 to 49 (*M* = 27.00, *SD* = 5.59). None of the medical students had formal handoff training, while eight indicated having some form of handoff experience. The handoff experience that participants had was handing off to other classmates, and receiving handoffs from other classmates and in some instances residents, especially while fulfilling their general surgery or pediatric clerkship rotations.

Twenty-one expert physicians (7 females, 14 males) from Eastern Virginia Medical School, Sentara Health System, and Children's Hospital of The King's Daughters participated in this study. They ranged in age from 28 to 72 (*M* = 42.57, *SD* = 12.53) and ranged in years of practice from 1 to 47 (*M* 13.55, *SD* =13.06). Participants came from six different specialty areas: pediatric emergency medicine (*n* = 4), neonatology (*n* = 4), pediatrics (*n* = 5), OBGYN/maternal fetal medicine (*n* = 3), emergency medicine (*n* = 4), and surgery (*n* = 1). Seven participants indicated having received or having been responsible for teaching and implementing formal handoff training.

Materials

Informed Consent Old Dominion University. Participant consent for the undergraduate students was obtained with an IRB approved consent form (see Appendix A). Participant consent for the third year medical students and the physicians was obtained with an Eastern Virginia Medical School IRB approved consent form (see Appendix B) and employee consent form (see Appendix C).

Background Information Form. Background information relevant to the study was gathered (see Appendix D). The form covered three areas: general background information, ATC/aviation experience, and clinical experience. Under the general background information section participants were asked to indicate their age, gender, and whether they had normal or corrected-to-normal vision and hearing. Further, they were asked to list any occupations they may have held involving handing off their position to an oncoming employee.

In the ATC/aviation experience section participants answered a set of questions to determine if they had any task relevant knowledge about aviation and ATC. Participants were asked to indicate if they had any formal training in ATC or aviation, taken any course covering either of these areas, and if they had any flight experience. Further, they were asked if they play flight simulation games or ATC games. If they answered, yes, to any of these questions they were excluded from the study. The participants were also asked if they had a family member or close friend who is a pilot or an air traffic controller. If they answered, yes, to this question and it was a family member with whom they have lived (or currently live with) or a close with whom they have discussed details of the aviation industry, they were also excluded from the study.

The clinical handoff experience section will differ for each population. The university students were asked if they had any clinical training or worked in a clinical setting. If they answered, yes, and had any clinical responsibilities, they were excluded from the study. In addition they were asked if they had a family member or close friend in the healthcare field. The same process used for the ATC question was followed to determine eligibility to participate. The clinical students and experienced physicians were asked to indicate if they were a medical student or an experienced physician. Students indicated their year in medical school and the experienced physician indicated how many years they had been in practice and their specialty. Both groups were asked if they had any formal training in handoffs, and if so to briefly describe the training. Further, they were asked to describe their handoff experience.

Tasks

All participants completed the same three sets of tasks: running memory span, ATC handoff, and a clinical handoff. Each task had its own instruction script and answer packet: running memory span (see Appendices F and G), ATC (see Appendices H and I), and the clinical task (see Appendices J and K). Further, the instructions and information in the tasks were prerecorded for consistent presentation.

Running Memory Span Task. The running memory span task was modeled on the tasks used in the studies performed by Elosúa and Ruiz (2008) and Broadway and Engle (2010). For the current study, disyllabic words from a master list (see Appendix L) were presented at a rate of one word every 2 sec. Recall that Elosúa and Ruiz and Broadway and Engle utilized different methods for the number of items to be recalled and whether whole recall trials were used. The approach for the current study was modeled after the technique used by Broadway and Engle. Lists of disyllabic words varying in length (and number of targets to be recalled) were presented. Further, whole and partial recall trials were used. Whole recall trials were used because Broadway and Engle noted that this procedure discourages individuals from ignoring the first items presented in anticipation that they will not need to recall them. Further, in the following two tasks (ATC handoff and clinical handoff) participants needed to determine what information to recall and the first items presented may or may not have been relevant.

Three list lengths were used: short, medium, and long. Elosúa and Ruiz (2008) suggested that short lists would minimize the number of updates that would need to be performed and would be less demanding. On the other hand, short list lengths conform to Miller's (1956) suggestion that individuals can hold 7 ± 2 chunks of information in their memory; therefore, it would be possible that participants might simply attempt to memorize the whole list. Elosúa and Ruiz (2008) suggested using longer list lengths to discourage participants from attempting to remember a whole list. Further, the amount of information to be recalled varied. In the whole recall condition, all the disyllabic words presented had to be recalled in sequential order. The partial recall length varied depending on the list length. The list lengths and number of items to be recalled are shown in Table 1. The short lists, medium lists, and long lists can be viewed in Appendix M, N, and O respectively. Three trials were performed for each task resulting in a total of 18 running memory span tasks.

Table 1

Design for Running Memory Span Task

	Short (8)	Medium (16)	Long (24)
Whole Recall	Recall all 8 words	Recall all 16 words	Recall all 24 words
Partial Recall	Recall last 6 words	Recall last 10 words	Recall last 14 words

Air Traffic Control Task. Recall that in the studies by Yntema and colleagues (1960, 1962, 1963), the stimuli were objects with different attributes requiring individuals to keep track of several dynamically changing variables that were supposed to simulate ATC; however, the tasks themselves were still abstract in nature. For the current study, real air traffic scenarios were modeled and participants were given information needed to maintain control of airspace.

Three list lengths were presented: short (8 items; see Appendix P), medium (16 items; see Appendix Q), long (24 items; see Appendix R). The number of items was consistent with the number of items presented in the running memory span task. For example, a list of five items that may or may not be needed to maintain airspace are indicated by the italicized words: “You are receiving flight *JFO290,* at *altitude 31,000 ft,* and you are receiving flight *HOU392,* whose next *reporting point is Chicago,* and is traveling at *43,000 ft”.*

Unlike the running memory span task, participants were not told the number of items that they must recall. Instead, they were informed that for the ATC handoff task they must decide what information is relevant to maintain in running memory. However, for each scenario presented the amount of relevant information to be recalled was manipulated to reflect the whole and partial recall trials in the running memory span task. For example, in the illustration above all items in italics would be relevant to maintaining control of the airspace and would need to be recalled. This represents a whole recall trial. On the other hand in the following message: “flight *ORF452* is traveling at a speed of *50 knots* and the weather is *sunny*”, only two items are relevant, flight number and airspeed. In this instance the weather does not need to be recalled, as it is not relevant for control of the aircraft. This represents a partial recall trial. The amount of relevant information was manipulated to reflect the partial recall trial parameters from the running memory span task (see Table 2). An additional variable addressed the organization of the information. Recall that Muzzin et al. (1983) found that experts regrouped items while novices recalled information sequentially. Therefore, organized and unorganized cases were used to determine if there are differences between the novices and the experts. In half the trials the information was organized while in the other half it was unorganized. Organized information contained information that was grouped together. For example, all weather information was presented together, all runway information was grouped, all notices to airmen (NOTAMS) were grouped, and so on. Unorganized scenarios presented information that was not grouped together. For example, an unorganized scenario might have presented a plane call sign, weather information, plane information, weather information, runway information, plane information, and so on. Three trials were performed for each task resulting in each participant performing a total of 36 ATC handoff tasks.

Table 2

Design for the Air Traffic Control Handoff Task and Clinical Handoff Task

		Short (8)	Medium (16)	Long (24)
Organized	Whole Recall	All 8 items	All 16 items	All 24 items
Unorganized	Whole Recall	should	should	should
		be recalled	be recalled	be recalled
Organized	Partial Recall	6 of the	10 of the	14 of the
Unorganized	Partial Recall	8 items should	16 items should	24 items should
		be recalled	be recalled	be recalled

Clinical Task. Participants also participated in a clinical handoff task. Again, this task represented a scenario where multiple items of information with different attributes are passed from one person to another. The clinical handoff task also represented a task that was familiar to some of the individuals participating in the study. The handoff scenarios were based on patients who had not been diagnosed because: a) it is likely that knowledge of different diseases would dictate how participants receive, organize, and determine relevant information during the handoff; and b) providing a diagnosis was likely to differentially affect novices and experts.

Three list lengths were again presented with the same numbers of presented information: short (8 items; see Appendix S), medium (16 items; see Appendix T), long (24 items; see Appendix U). The same parameters that were described for the ATC task were followed for the clinical task (see Table 2). Three trials were performed for each task resulting in each participant performing a total of 36 clinical handoff tasks.

Case Development

Subject matter interviews were conducted with expert physicians and air traffic controllers to develop the handoff cases. Further, a number of academic publications were referenced during case development to ensure realism of the information (Chabner, 2001; Collins, 2008; Dunn, 1998; Jarvis, 2000; Li, Kohrt, Caughey, 2007; Nolan, 2011; *Rapid Differential Diagnosis*, 2002; Strachan, Sharma, & Hunter, 2012) as well as the PubMed Health Diseases and Conditions, Mayo Clinic, and Federal Aviation Administration (FAA) websites.

Procedure

Upon arrival participants were asked to read and sign the informed consent form and complete the background information form. All participants performed the running memory task first. The order in which each participant completed the handoff tasks and the order of complexity for each task were assigned at random across participants.

The general instructions (see Appendix D) for the overall experiment were played from the prerecording. The participant was asked if he or she had any questions. The participant then listened to prerecorded instructions for the first task (see Appendix F for running memory span instructions; see Appendix H for ATC handoff instructions; see Appendix J for clinical handoff instructions). The participant was given an opportunity to ask questions about the task. Once all questions were addressed the participant was given a practice session for the first task.

Each practice consisted of an abbreviated version of the experimental task (see Appendix V). The running memory span task consisted of two lists of words, each of a different length. For one list the participant performed a whole recall trial and for the other list the participant performed a partial recall. The ATC and clinical handoff practice tasks each had two handoff scenarios of different lengths. One was organized and the other was unorganized. Further, one represented a whole recall trial where all information should be recalled while the other represented a partial recall trial where only some information was important to recall.

After each practice session the participant was again asked if he or she had any questions prior to beginning the experimental session. Once all questions were addressed the participant completed the experimental session for that task. Participants were given an opportunity to take a short break between tasks.

The participants were asked to record their responses in a paper packet provided to them after they listened to each vignette. They were monitored while listening to each message to prevent them from taking notes. However, *after* listening to the vignette, for the ATC and clinical handoffs they were provided with scratch paper, where they were allowed to write their initial responses while they gather their thoughts. They wrote their final answer in their answer packet. Further, the experimenter reviewed the responses at the end of the task for legibility and made any clarifications with the participant before starting the next session.

CHAPTER IV

RESULTS

All data were screened for outliers using stem-and-leaf plots and boxplots to protect against Type I and Type II error rate prior to running the analyses. There are different methods available to address outliers. The method used for this data set was to adjust outlying scores by replacing the score with a score that was one unit larger (or smaller) than the next most deviant score in the distribution (Field, 2009; Tabachnick & Fidell, 2001).

Mauchly's tests were used to check for any violations of sphericity. If there were violations of sphericity there were different corrections that could be applied to the data. Greenhouse-Geisser is the most conservative correction and there is some concern that for an epsilon value greater than .75 Greenhouse-Geisser might result in a Type I error (Field, 2009). On the other hand, the Huynh-Feldt correction is less conservative, but there is concern that the epsilon value can overestimate sphericity. A mixed approach has been suggested, where for an epsilon value of .75 and above the Hunyh-Feldt correction should be used and for values below .75 the Greenhouse-Geisser correction should be used. For the current data where sphericity was violated there were no significant differences among the tests, therefore the Greenhouse-Geisser correction was reported for all sphericity violations to remain consistent throughout the data.

Levene's tests were used to check for equality of variance. Regarding the running memory data there were no equality of variance violations. There were multiple cases of variance violations for the clinical total handoff score data: medium whole organized, medium whole unorganized, and long whole organized. There were also multiple cases of variance violations for the ATC total handoff score data: short partial organized, short partial unorganized, medium whole organized, medium whole unorganized, medium partial organized, long whole unorganized, and long partial organized. There are multiple approaches to take when variances differ (Keppel & Wickens, 2004). For both the clinical and ATC data a more stringent significance level of $\alpha = .025$ was applied to help protect against Type I error.

Regarding outliers, for the running memory data there were outliers that were identified across a number of conditions. Table 3 shows the number of outliers for correct detections, intrusions, transpositions, and location errors across expertise level and list type. For the intermediate's intrusion short partial list the outliers were not replaced. There were only three different scores in this particular condition for the intermediate group, fourteen of which were .04. The only other scores were .00 and .08 and therefore were appearing as outliers in the data set; however, due to the small amount of variance there were also issues with homogeneity of variance which increased if the outliers were adjusted. Therefore, in this particular instance it was determined that leaving the outlying scores was the appropriate course of action.

Table 3

Number of Outliers for Running Memory Span Task for Novices, Intermediates, and Experts

Scale	*SW*	*SP*	*MW*	*MP*	*LW*	*LP*
Novices						
Correct Detection	0	2	2	1	1	1
Intrusions	4	5	2	1	3	3
Transpositions	1	0	0	2	0	0
Location Errors	-	0	-	0	-	0
Intermediates						
Correct Detection	1	0	1	2	1	3
Intrusions	0	7*	0	1	0	2
Transpositions	1	1	0	0	0	0
Location Errors	-	1	-	1	-	2
Experts						
Correct Detection	2	0	1	3	0	0
Intrusions	0	0	1	0	1	3
Transpositions	0	0	0	0	1	1
Location Errors	-	1	-	2	-	2

*outliers not replaced

Regarding the clinical handoff data there were outliers that were identified across a number of conditions. Table 4 shows the number of outliers for handoff score organized and handoff score unorganized across expertise level and list length. There were no violations of normality or kurtosis.

Table 4

Number of Outliers for Clinical Handoff Task for Novices, Intermediates, and Experts

Scale	*SW*	*SP*	*MW*	*MP*	*LW*	*LP*
Novices						
Handoff Score Organized	0	0	0	0	0	0
Handoff Score Unorganized	2	1	1	1	0	1
Intermediates						
Handoff Score Organized	0	1	1	1	0	0
Handoff Score Unorganized	0	0	0	0	0	0
Experts						
Handoff Score Organized	0	0	0	0	0	0
Handoff Score Unorganized	3	0	0	0	4	1

The clinical relevant and irrelevant scores were also screened for outliers. Table 5 shows the number of outliers for these scores across expertise level.

Table 5

Number of Outliers for Clinical Relevant Scores, Irrelevant Scores, Partial Scores, Whole Scores, Organized Scores, and Unorganized Scores for Novices, Intermediates, and Experts

Scale	*Novice*	*Intermediate*	*Expert*
Relevant	0	0	0
Irrelevant	1	0	4
Partial	0	0	2
Whole	0	2	0
Organized	0	3	0
Unorganized	0	0	2

Regarding the ATC data there were outliers that were identified across a number of conditions. Table 6 shows the number of outliers for each type of handoff score across expertise level and list type. There were no violations of normality or kurtosis.

Table 6

Number of Outliers for Air Traffic Control Handoff Task for Novices, Intermediates, and Experts

Scale	*SW*	*SP*	*MW*	*MP*	*LW*	*LP*
Novices						
Handoff Score Organized	0	2	2	0	0	0
Handoff Score Unorganized	0	1	0	0	0	0
Intermediates						
Handoff Score Organized	0	0	0	0	2	0
Handoff Score Unorganized	0	0	0	0	2	1
Experts						
Handoff Score Organized	1	0	0	0	2	0
Handoff Score Unorganized	1	0	1	1	1	2

The ATC relevant and irrelevant scores were also screened for outliers. Table 7 shows the number of outliers for relevant scores and irrelevant scores across expertise level.

Table 7

Number of Outliers for Air Traffic Control Relevant Scores and Irrelevant Scores for Novices, Intermediates, and Experts

Scale	*Novice*	*Intermediate*	*Expert*
Relevant	0	1	0
Irrelevant	0	2	1

Running Memory

The same method used by Elosúa and Ruiz (2008) and Broadway and Engle (2010) to classify correct responses was employed in the current study for the running memory task. A correct detection was defined as a word recalled in the correct serial position. One point was assigned for each correct detection. Elosúa and Ruiz (2008) further categorized responses into response errors, which included intrusions (a word recalled that was not from the list) and position errors. The position errors were broken down into transpositions (a correct word recalled in the wrong position) and location errors (words presented early in the list that should not be recalled).

Regarding expertise level, confidence interval equivalence tests (Snow, Reising, Barry, & Hartstock, 1999; Wellek, 2010) were performed to determine whether the performance of the novices and intermediates was equivalent to the experts for each type of response. There are no standard criteria for equivalency ranges for memory; therefore, a range had to be established. Equivalency literature and expert/novice memory for clinical cases was consulted to help establish a range. Equivalency testing has been widely used for therapeutic bioequivalence (Snow et al., 1999) and the standard range is

±20% of the reference criteria (Lužar-Stiffler & Stiffler, 2002). Regarding the clinical literature, Norman et al. (1979) found that for typical medical cases the experts recalled 57% more items. Therefore, for the current study a range was selected that fell between the bioequivalence range of 20% and the experts recall of 57%. As a result, the criterion to determine acceptable performance was any score within 0.33 or 33% of the expert's scores set at a confidence interval of 90%. In addition, a 3 level of expertise (between variable; novice, intermediate, expert) x 3 list length (within variable; short, medium, long) x 2 recall length (within variable; whole, partial) mixed factorial ANOVA was used to analyze correct detections, intrusions, and transpositions as well as potential interactions of these variables with expertise. Bonferroni tests were used for post hoc analyses to protect against Type I errors. For the location errors a 3 level of expertise (between variable; novice, intermediate, expert) x 3 list length (within variable; short, medium, long) mixed factorial ANOVA was used to analyze the partial recall data. Elosúa and Ruiz (2008) and Broadway and Engle (2010) examined percentile scores for each type of response classification. For the present study the same method was followed, where all the scores were converted into proportions, to make comparisons between the list length conditions. Additionally, to examine whether participants would truly have fewer correct detections as list length increased the whole scores were also examined for list length. Descriptive statistics for the proportion data are shown in Table 8 and for the whole scores for list length in Table 9.

Table 8

Descriptive Statistics for Proportion Running Memory Span Task

Scale	*N*	*M*	*SD*	Skewness	Kurtosis
CD Short Whole	76	.34	.14	.42	-.38
CD Short Partial	76	.32	.19	.64	.56
CD Medium Whole	76	.13	.09	.63	.14
CD Medium Partial	76	.16	.09	.10	-.57
CD Long Whole	76	.09	.05	.14	-.71
CD Long Partial	76	.09	.06	.37	.31
I Short Whole	76	.03	.04	.52	-1.04
I Short Partial	76	.04	.04	.88	.77
I Medium Whole	76	.03	.03	.65	-.51
I Medium Partial	76	.02	.02	1.19	2.32
I Long Whole	76	.02	.01	.57	-.55
I Long Partial	76	.02	.02	.67	-.44
T Short Whole	76	.15	.11	.49	-.20
T Short Partial	76	.17	.12	.60	-.43
T Medium Whole	76	.16	.08	.30	-.48
T Medium Partial	76	.17	.10	.40	-.54
T Long Whole	76	.11	.07	.52	-.37
T Long Partial	76	.13	.07	.36	-.13
LE Short Partial	76	.31	.24	.37	-.73
LE Medium Partial	76	.12	.11	.64	-.80
LE Long Partial	76	.09	.08	.94	.84

Table 9

Descriptive Statistics for Whole Correct Detections Scores Running Memory Span Task

Scale	*N*	*M*	*SD*	Skewness	Kurtosis
CD Short	76	7.16	3.40	.63	.45
CD Medium	76	5.69	3.50	.44	-.14
CD Long	76	5.22	3.10	-.15	-.72

Expertise. The expert's mean score for correct detections was .21; therefore, the equivalency interval for correct detections was -.12 to .54. The 90% confidence interval for the novice score was .14 to .20 and for the intermediates it was .16 to .24. The confidence intervals for the novices and intermediates are shown against the expert's equivalency intervals in Figure 1. The expert mean score for correct detections is displayed as a diamond and there is also a solid black horizontal line extending across the figure indicating the expert's mean. The equivalency interval is indicated by an upper and lower horizontal dashed line extending across the figure along with a vertical arrow on the right side of the figure indicating the equivalency range. The novice and intermediate mean scores are displayed as a square and a circle, respectively. Error bars indicate the confidence interval range for the novices and intermediates. As can be seen in Figure 1, the novice and intermediate's confidence intervals fall within the equivalency interval indicating their performance was equivalent to the experts.

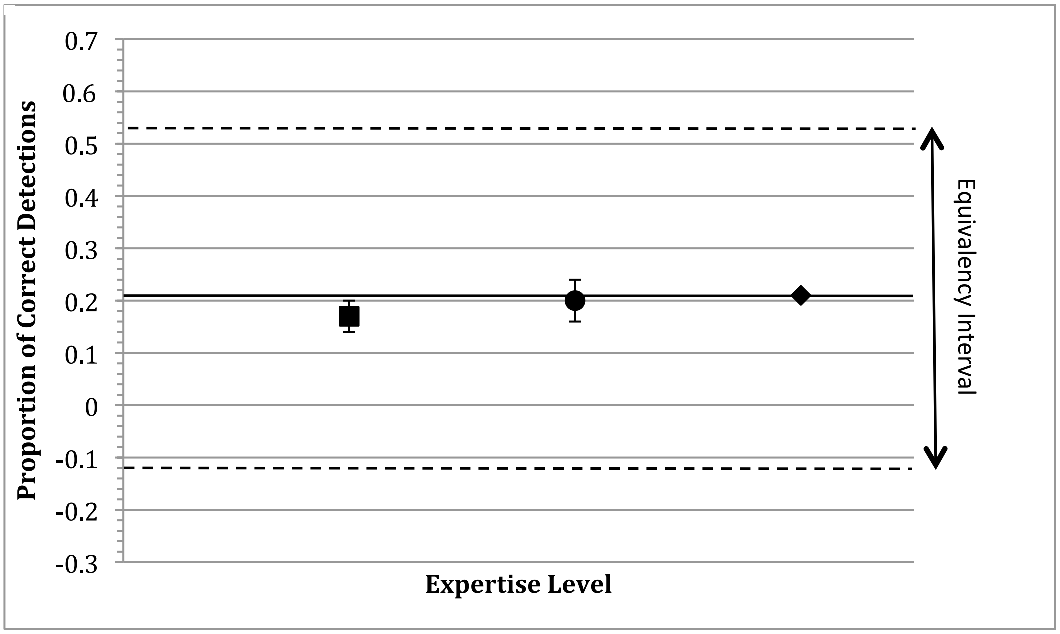

Figure 1. Running memory: proportion of correct detections for each level of expertise (90% confidence interval).

The expert's mean score for intrusions was .03; therefore, the equivalency interval for intrusions was -.30 to .36. The 90% confidence interval for the novice score was .02 to .04 and for the intermediates it was .02 to .04. As can be seen in Figure 2, the novice and intermediate's confidence intervals do not fall outside of the equivalency interval indicating their performance was equivalent to the experts.

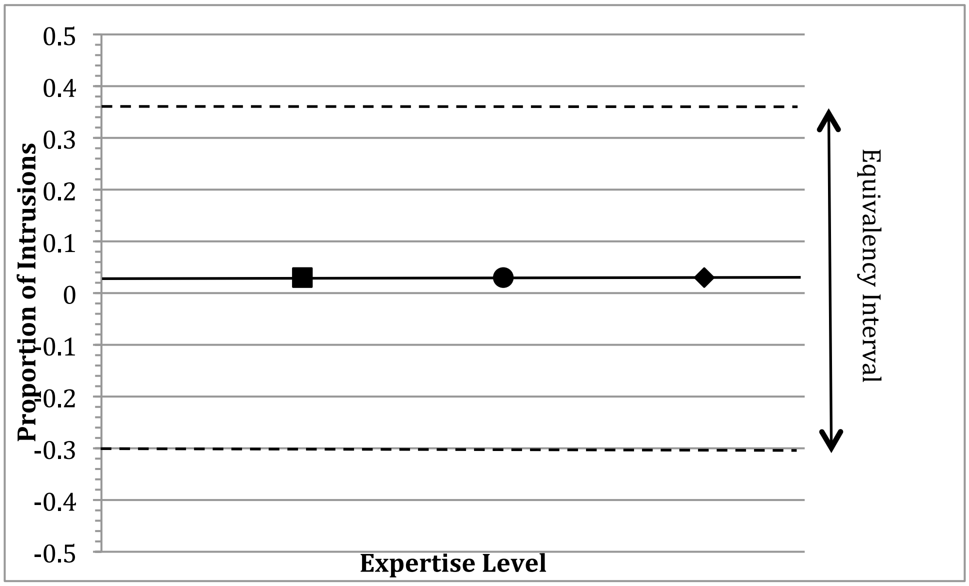

Figure 2. Running memory: proportion of intrusions for each level of expertise (90% confidence interval).

The expert's mean score for transpositions was .15; therefore, the equivalency interval for transpositions was -.18 to .48. The 90% confidence interval for the novice score was .12 to .18 and for the intermediates it was .12 to .20. As can be seen in Figure 3, the novice and intermediate's confidence intervals do not fall outside of the equivalency interval indicating their performance was equivalent to the experts.

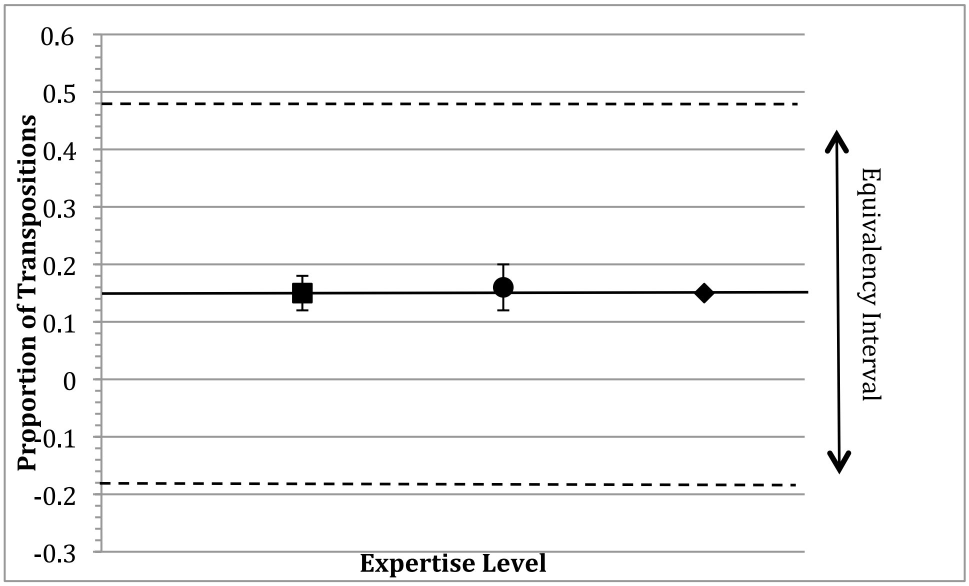

Figure 3. Running memory: proportion of transpositions for each level of expertise (90% confidence interval).

The expert's mean score for location errors was .03; therefore, the equivalency interval for location errors was -.14 to .53. The 90% confidence interval for the novice score was .11 to .19 and for the intermediates it was .12 to .22. As can be seen in Figure 4, the novice and intermediate's confidence intervals do not fall outside of the equivalency interval indicating their performance was equivalent to the experts.

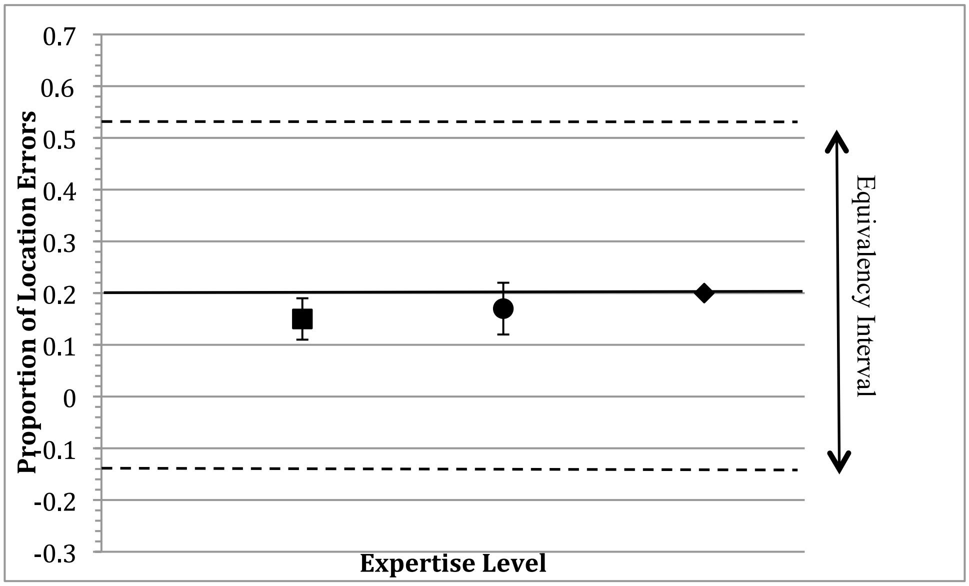

Figure 4. Running memory: proportion of location errors for each level of expertise (90% confidence interval).

Correct Detections. The results of the ANOVA for proportion of correct detections are shown in Table 10. A significant main effect for list length was observed, $F(1.32, 96.05) = 240.59$, $p < .001$, partial $\eta^2 = .77$. Participants had significantly more correct detections in the short list ($M = .34$, $SD = .17$) compared to the medium ($M = .15$, $SD = .09$) and long lists ($M = .09$, $SD = .06$) and participants had significantly more correct detections in the medium list compared to the long list. No other effects were significant. Regarding the whole scores there was a significant main effect for list length, $F(2, 146) = 27.79$, $p < .001$, partial $\eta^2 = .28$. Participants had significantly more correct detections for the short list ($M = 7.16$, $SD = 3.40$) compared to the medium ($M = 5.69$, $SD = 3.50$) and long lists ($M = 5.22$, $SD = 3.10$).

Table 10

Results of the Analysis of Variance for Running Memory Correct Detections

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	.20	2	.10	2.44	.09	.06
Error	2.94	73	.04			
Within Subjects						
List Length (L)	4.71	1.32	3.58	240.59	.00***	.77
L x E	.08	4	.02	1.97	.10	.05
Error (L)	1.43	96.05	.02			
Recall Length (R)	.01	1	.01	.60	.44	.01
R x E	.01	2	.00	.32	.73	.01
Error (R)	.57	73	.01			
L x R	.02	2	.01	1.82	.17	.02
L x R x E	.03	4	.01	1.79	.13	.05
Error (L x R)	.66	122.28	.01			

Note. ****p* < .001

Intrusions. The results of the ANOVA for intrusions are shown in Table 11. A significant interaction for list length and recall length was observed, $F(1.84, 134.55) = 4.98$, $p < .05$, partial $\eta^2 = 06$. A plot of the interaction is shown in Figure 5. Simple effects analyses for list length revealed that for the short list participants had significantly more intrusions for the partial list compared to the whole list, while for the medium list participants had significantly more intrusions for the whole list compared to the partial list. Additional simple effects analyses revealed a significant difference within the partial list, $F(2, 134.55) = 101.33$, $p < .001$, and the whole list, $F(2, 134.55) = 25.33$, $p < .001$. For the partial lists participants had significantly more intrusions for the short list compared to the medium and long list. For the whole list participants had significantly more intrusions for the short and medium list compared to the long list. Further, a significant main effect for list length was also observed, $F(1.57, 114.45) = 21.81$, $p < .001$, partial $\eta^2 = .23$. Participants had significantly more intrusions in the short list ($M = .04$, $SD = .04$) compared to the medium ($M = .03$, $SD = .02$) and long lists ($M = .02$, $SD = .02$) and participants had significantly more intrusions in the medium list compared to the long list. No other effects were significant.

Table 11

Results of the Analysis of Variance for Running Memory Intrusions

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	.00	2	.00	.36	.70	.01
Error	.11	73	.00			
Within Subjects						
List Length (L)	.03	1.57	.02	21.81	.00***	.23
L x E	.00	4	.00	1.09	.36	.03
Error (L)	.10	114.45	.00			
Recall Length (R)	.00	1	.00	1.86	.18	.03
R x E	.00	2	.00	.98	.38	.03
Error (R)	.05	73	.00			
L x R	.01	1.84	.00	4.98	.01*	.06
L x R x E	.00	4	.00	.20	.94	.01
Error (L x R)	.07	134.55	.00			

Note. *$p < .05$; ***$p < .001$

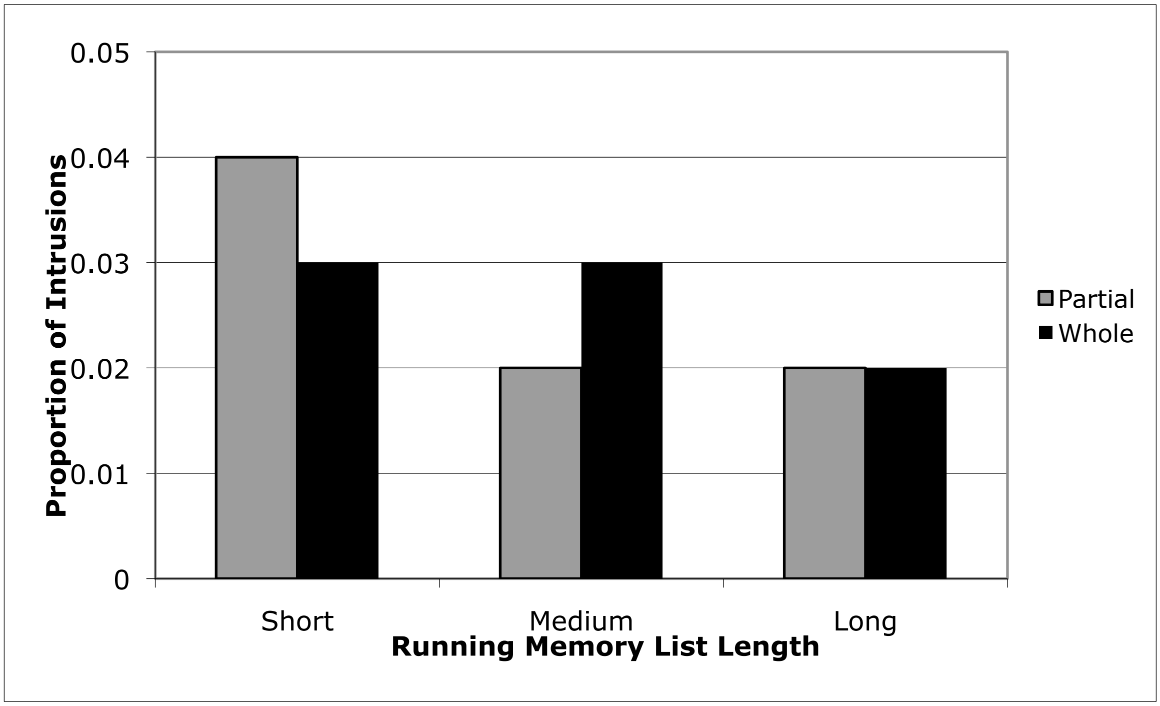

Figure 5. Proportion of intrusions for recall length as a function of list length.

Transpositions. The results of the ANOVA for transpositions are shown in Table 12. A significant main effect for list length was observed, $F(1.82, 133.14) = 13.72$, $p < .001$, partial $\eta^2 = .16$. Participants had significantly more transpositions in the short list ($M = .16$, $SD = .12$) and the medium list ($M = .17$, $SD = .09$) compared to the long list ($M = .12$, $SD = .07$). No other effects were significant.

Table 12

Results of the Analysis of Variance for Running Memory Transpositions

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	.02	2	.01	.44	.65	.01
Error	1.53	73	.02			
Within Subjects						
List Length (L)	.18	1.82	.10	13.72	.00***	.16
L x E	.06	4	.01	2.09	.09	.05
Error (L)	.97	133.14	.01			
Recall Length (R)	.03	1	.03	3.71	.06	.05
R x E	.00	2	.00	.25	.78	.01
Error (R)	.49	73	.01			
L x R	.00	1.73	.00	.36	.67	.01
L x R x E	.01	4	.00	.45	.77	.01
Error (L x R)	.92	126.17	.01			

Note. ****p* < .001

Location Errors. The results of the running memory location error ANOVA are shown in Table 13. A significant main effect for list length was observed, $F(1.29, 94.20) = 58.62$, $p < .001$, partial $\eta^2 = .45$. Participants had significantly more location errors in the short list ($M = .31$, $SD = .24$) compared to the medium ($M = .12$, $SD = .11$) and long lists ($M = .09$, $SD = .08$) and participants had significantly more location errors in the medium list compared to the long list. No other effects were significant.

Table 13

Results of the Analysis of Variance for Running Memory Location Errors

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	.07	2	.04	.89	.41	.02
Error	2.95	73	.04			
Within Subjects						
List Length (L)	2.07	1.29	1.60	58.62	.00***	.45
L x E	.04	4	.01	.53	.71	.01
Error (L)	2.57	94.20	.03			

Note. ***$p < .001$

Clinical and Air Traffic Control Handoff Task Scoring

A total handoff score was obtained for each clinical handoff and ATC handoff that was performed. All points resulting from the correct recall of information were added together and all negative points resulting from recalling an irrelevant piece of information in the partial recall lists or from an incorrect answer or intrusion, described below, were added together. The total negative point score was then deducted from the total correct response score to generate a total handoff score. The total scores for each type of handoff were then combined (e.g., all three short whole scores, all three short partial scores, etc.) for a total handoff score for each condition. For each handoff the number of relevant items recalled was obtained. Participants received a point for each relevant item recalled regardless of whether the whole item or only part of the item was recalled. Further, the participants received credit even if they incorrectly recalled an item of information. For example, for an ATC handoff if they presented with “no bird threat” but they recalled “there is a bird threat” they would receive a point for recalling a relevant item of information even though they incorrectly recalled the information. For the partial recall cases each irrelevant item that was recalled was assigned a point. For both the relevant and irrelevant scores a total was obtained in the same manner that a total handoff core was obtained for each condition.

Both strict and lenient criteria for scoring were applied to the clinical and ATC data. Regarding a correct detection, participants received one whole point for each item of relevant information that was correctly recalled in its entirety. If only part of the information was recalled a whole point was assigned for the lenient criterion, but a half point was assigned for the strict criterion. For example, if the participant had to recall “runway four inactive” and he or she recalled the whole statement, one point was assigned. On the other hand, if only “runway four” was recalled the participant still received a whole point for the lenient criterion while a half point was assigned for the strict criterion.

If the participant recalled a relevant item of information, but it was completely incorrect, a half point was assigned using the lenient criterion while no point was assigned under the strict criterion. For example, if the participant had to recall “lightning visible” and he or she recalled “no lightning” or “no lightning visible” there is a direct conflict of information. Although the participant correctly recalled that there was information pertaining to lightning, the ramifications of stating there is no lightning when it is present can have serious consequences; therefore, a half point was assigned for the lenient criterion while no point was assigned for the strict criterion. On the other hand, if an item was recalled that was partially correct/partially incorrect a whole point was assigned for the lenient criterion and a half point was assigned for the strict criterion. For example, if the participant had to recall "heavy cloud coverage" and he or she recalled "light cloud coverage" the participant was correctly recalling that there is cloud coverage, just not the amount of coverage; therefore, a whole point was assigned for the lenient criterion while a half point was assigned for the strict criterion.

If a participant recalled an item of information that was not in the handoff he or she would lose one point. Further, if the participant repeated information a point would be deducted. In some instances it was possible to receive no points and have no points deducted. For example, if a participant was informed that the weather is clear with sunny skies and no storms, two possible responses would be appropriate. One response would be to report the weather is fine or clear. The other response would be to report nothing about the weather. Either response was considered neutral and did not result in an addition or subtraction of points. If, however, all three items were listed (clear, sunny skies, no storms) three points would then be deducted as each individual item of information is not relevant to perform the task. Because the strict criterion minimized much of subjectivity in the data, it was chosen for data analysis; however, when differences between the two criteria emerged, data using both criteria are presented at the end of the results section. Further, although the whole score data were analyzed the scores were also converted to proportions to enable comparisons among the list length conditions. The results are presented in F-tables in Appendix W for the clinical data and Appendix X for the ATC data.

A sample of 36 cases was drawn from 8 different participants and used to calculate the inter-rater reliability on the strict criteria scoring. Half of the cases addressed the clinical scenarios and the other half, the ATC scenarios. The IRR across all cases was .85. There was a slight difference between the two types of scenarios. The IRR for the clinical scenarios (IRR = .89) was a little higher than for the ATC scenarios (IRR = .72).

Clinical Handoff Task

One-way ANOVAs were used to analyze relevant and irrelevant scores. The strict criterion clinical handoff scores were analyzed with a 3 level of expertise (between variable; novice, intermediate, expert) x 3 list length (within variable; short, medium long) x 2 recall length (within variable; whole, partial) x 2 organization (within variable; organized, unorganized) mixed factorial ANOVA. Descriptive statistics are shown in Table 14.

Table 14

Descriptive Statistics for Clinical Handoff Task

Scale	*N*	*M*	*SD*	Skewness	Kurtosis
HSO Short Whole	76	7.66	3.07	.33	-.24
HSO Short Partial	76	3.81	2.73	.34	-.27
HSO Medium Whole	76	9.36	3.84	.70	.09
HSO Medium Partial	76	-1.55	3.50	.66	-.09
HSO Long Whole	76	9.41	4.22	.08	.07
HSO Long Partial	76	-.59	4.40	.64	.20
HSU Short Whole	76	6.62	3.32	.60	-.04
HSU Short Partial	76	2.80	2.56	.07	-.93
HSU Medium Whole	76	9.13	3.82	1.01	1.09
HSU Medium Partial	76	1.51	3.51	.02	-.41
HSU Long Whole	76	7.36	3.85	.18	.08
HSU Long Partial	76	1.13	3.59	.03	-.38

Clinical Relevant Scores. The results of the ANOVA for relevant scores are shown in Table 15. There was a significant effect for expertise, $F(2, 75) = 43.62$, $p < .001$. The results of a Bonferroni post hoc test showed the experts ($M = 239.29$, $SD = 43.54$) and intermediates ($M = 211.89$, $SD = 47.14$), recalled significantly more relevant items compared to novices ($M = 143.64$, $SD = 32.35$).

Table 15

Results of the Analysis of Variance for Clinical Relevant Scores

Source	*SS*	*df*	*MS*	*F*	*p*
Between Subjects					
Expertise	136870.30	2	68435.15	43.62	.00***
Error	251408.68	75			

Note. ***$p < .001$

Clinical Irrelevant Scores. The results of the ANOVA for irrelevant scores are shown in Table 16. There was a significant effect for expertise, $F(2, 75) = 5.04$, $p < .01$. The results of a Bonferroni post hoc test showed the experts recalled significantly fewer irrelevant items ($M = 10.71$, $SD = 6.44$) compared to the intermediates ($M = 18.05$, $SD = 7.49$) and novices ($M = 16.28$, $SD = 8.69$).

Table 16

Results of the Analysis of Variance for Clinical Irrelevant Scores

Source	*SS*	*df*	*MS*	*F*	*p*
Between Subjects					
Expertise	618.97	2	309.48	5.04	.01
Error	5101.42	75			

Clinical Total Handoff Score. The results of the clinical strict criterion total handoff score ANOVA are shown in Table 17. A significant main effect for expertise was observed, $F(2, 73) = 48.05$, $p < .001$, partial $\eta^2 = .57$. Novices had significantly lower handoff scores ($M = 7.87$, $SD = 3.44$) compared to the intermediates ($M = 12.51$, $SD = 4.47$) and experts ($M = 14.95$, $SD = 4.43$), and intermediates had significantly lower handoff scores compared to the experts.

Table 17

Results of the Analysis of Variance for Clinical Handoff Scores

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	8686.02	2	4343.01	48.05	.00***	.57
Error	6598.06	73	90.38			
Within Subjects						
List Length (L)	4654.44	1.79	2597.56	193.27	.00***	.73
L x E	628.97	3.58	175.51	13.06	.00***	.26
Error (L)	1758.07	130.81	13.44			
Recall Length (R)	12935.18	1	12935.18	495.01	.00***	.87
R x E	904.64	2	452.32	17.31	.00***	.32
Error (R)	1907.57	73	26.13			
Organize (O)	28.63	1	28.63	3.73	.06	.05
O x E	95.59	2	47.80	6.22	.00**	.15
Error (O)	560.70	73	7.68			
L x R	1805.23	2	902.61	81.60	.00***	.53

Table 17 Continued

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
L x R x E	61.53	4	15.38	1.39	.24	.04
Error (L x R)	1615.01	146	11.06			
L x O	14.01	1.76	7.95	.94	.39	.01
L x O x E	65.42	3.53	18.56	2.20	.07	.06
Error (L x O)	1084.62	128.66	8.43			
R x O	161.54	1	161.54	19.14	.00***	.21
R x O x E	.04	2	.48	.06	.95	.00
Error (R x O)	616.25	73	8.44			
L x R x O	451.57	2	225.79	27.35	.00***	.27
L x R x O x E	84.33	4	21.08	2.55	.04*	.07
Error (L x R x O)	1205.20	146	8.26			

Note. *$p < .05$; **$p < .01$; ***$p < .001$

A significant main effect for list length was observed, $F(1.79, 130.81) = 193.27$, $p < .001$, partial $\eta^2 = .73$. Participants had significantly lower handoff scores for the short list ($M = 8.92$, $SD = 2.78$) compared to the medium ($M = 11.75$, $SD = 4.32$) and long lists ($M = 14.67$, $SD = 5.23$) and significantly lower scores for the medium compared to the long list. A significant main effect for recall length was observed, $F(1, 73) = 495.01$, $p < .001$, partial $\eta^2 = .87$. Participants had significantly higher handoff scores for the whole recall list ($M = 15.69$, $SD = 4.57$) compared to the partial recall list ($M = 7.86$, $SD = 3.66$).

Further, a number of significant interactions emerged which can be seen in Table 16. Lower order interactions that were predicted will be reported in detail along with the highest order interaction.

A significant interaction for list length and expertise was observed, $F(1.79, 130.81) = 193.27$, $p < .001$, partial $\eta^2 = .73$. A plot of the interaction is shown in Figure 6. Simple effects analyses for list length and expertise revealed significant differences for the short list, $F(2, 203.81) = 3.93$, $p < .05$, medium list, $F(2, 203.81) = 8.96$, $p < .001$, and long list, $F(2, 203.81) = 15.51$, $p < .001$. For each list length expert's scores were significantly higher than intermediates and novices, and intermediate's scores were significantly higher than novices. Additional simple effects analyses revealed significant differences within the novice group, $F(2, 130.81) = 6.69$, $p < .01$, intermediate group, $F(2, 130.81) = 26.66$, $p < .001$, and expert group, $F(2, 130.81) = 40.65$, $p < .001$. For each level of expertise, handoff scores were higher for the long list compared to the medium and short list, and the medium list was significantly higher compared to the short list.

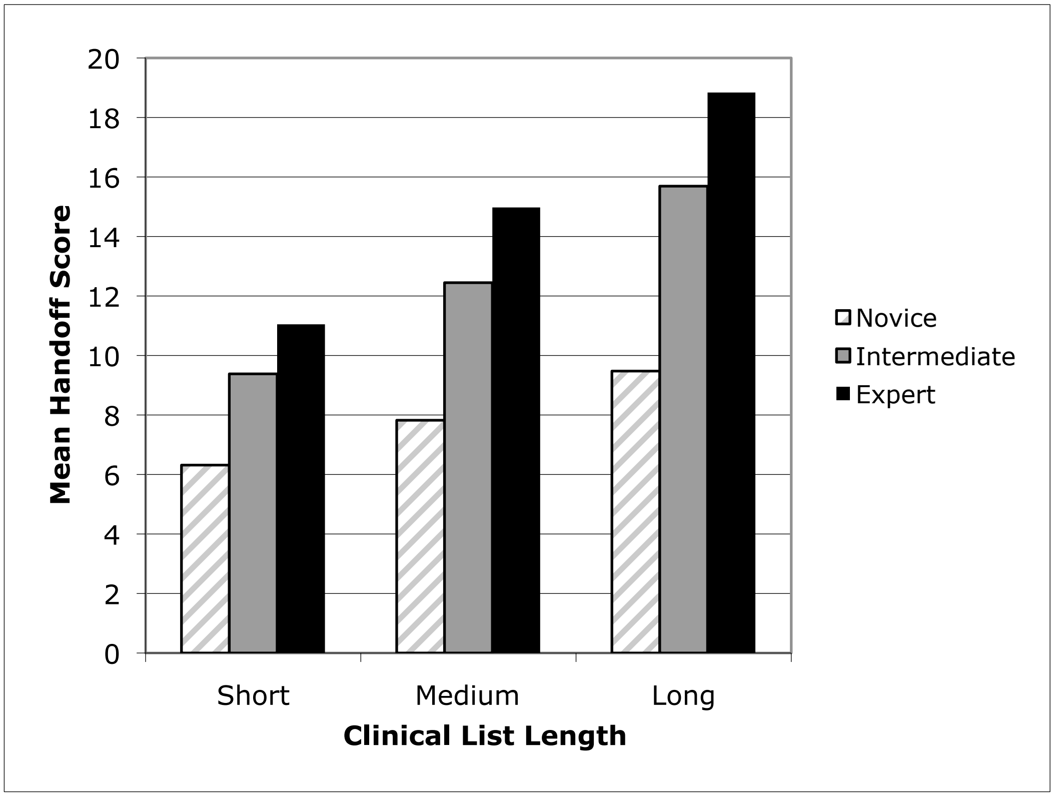

Figure 6. Mean clinical handoff score for expertise level as a function of list length.

A significant interaction for recall length and expertise was observed, $F(2, 73) = 17.31$, $p < .001$, partial $\eta^2 = .32$. A plot of the interaction is shown in Figure 7. Simple effects analyses for recall length and expertise revealed significant differences for the whole list, $F(2, 146) = 10.86$, $p < .001$. The expert's scores were significantly higher than intermediates and novices, and intermediate's scores were significantly higher than novices. Additional simple effects analyses revealed significant differences within the novice group, $F(1, 73) = 19.99$, $p < .001$, intermediate group, $F(1, 73) = 23.60$, $p < .001$, and expert group, $F(1, 73) = 40.59$, $p < .001$. For each level of expertise, handoff scores were higher for the whole list compared to the partial list.

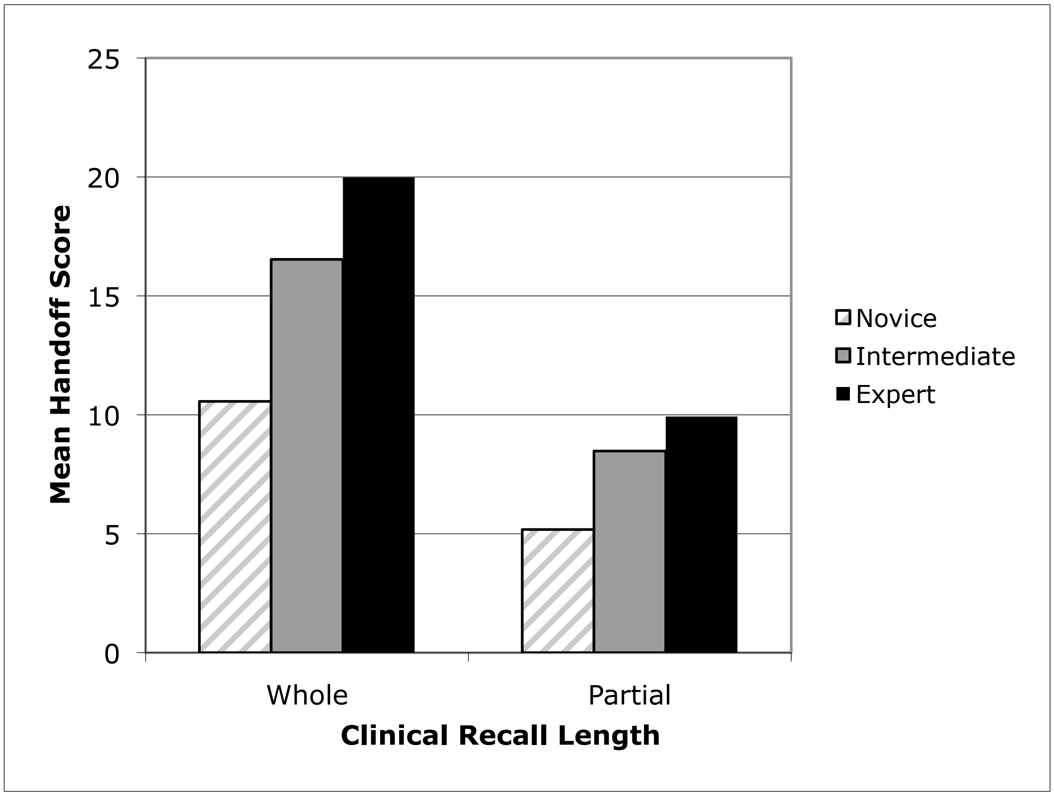

Figure 7. Mean clinical handoff score for expertise level as a function of recall length.

A significant interaction for organization and expertise was observed, $F(2, 73) = 6.22$, $p < .01$, partial $\eta^2 = .15$. A plot of the interaction is shown in Figure 8. Simple effects analyses for organization and expertise revealed significant differences for the organized list, $F(2, 146) = 8.71$, $p < .01$, and the unorganized list, $F(2, 146) = 6.20$, $p < .05$. For the organized list expert's scores were significantly higher than intermediates and novices, and intermediate's scores were significantly higher than novices. For the unorganized list the experts and intermediate's scores were significantly higher compared to the novices.

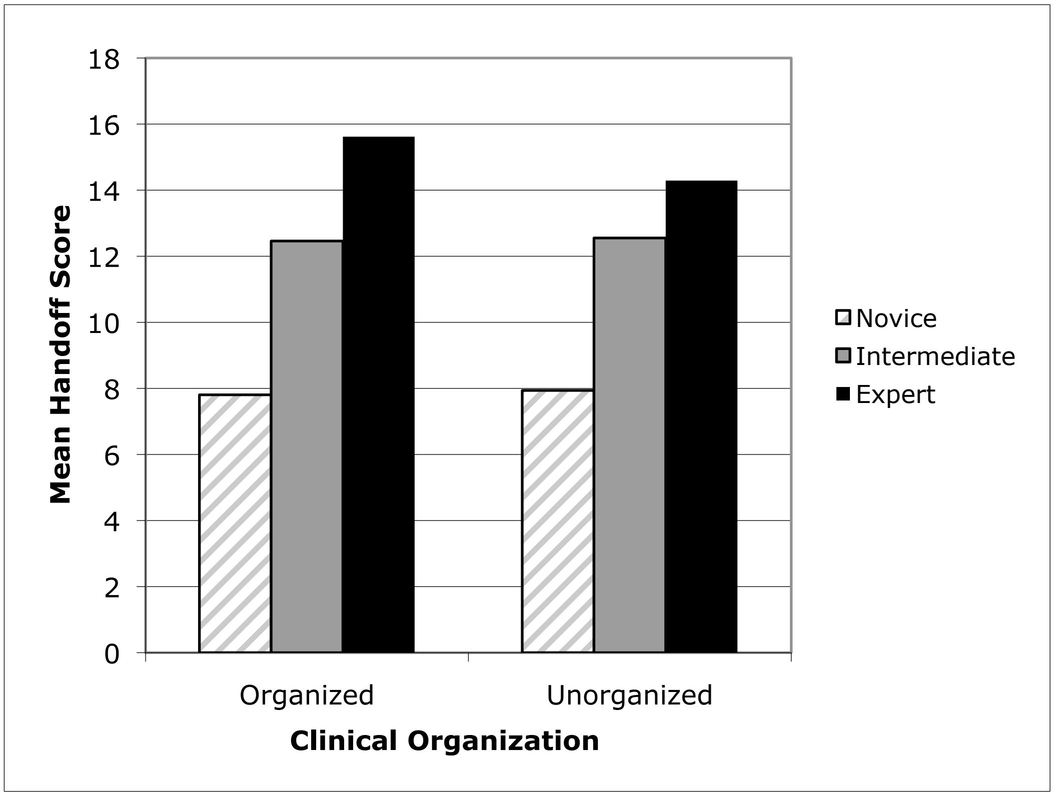

Figure 8. Mean clinical handoff score for expertise level as a function of organization.

A significant interaction for list length, recall length, organization, and expertise was observed, $F(4, 146) = 2.55$, $p < .05$, partial $\eta^2 = .07$. A plot of the interaction is shown in Figure 9. Simple effects analyses for list length by recall length by organization revealed significant differences for the short whole organized list, $F(2, 219) = 9.74$, $p < .001$, short whole unorganized list, $F(2, 219) = 8.13$, $p < .001$, medium whole organized list, $F(2, 219) = 19.45$, $p < .001$, medium whole unorganized list, $F(2, 219) = 15.41$, $p < .001$, medium partial organized list, $F(2, 219) = 7.77$, $p < .001$, medium partial unorganized list, $F(2, 219) = 3.59$, $p < .05$, long whole organized list, $F(2, 219) = 35.45$,

$p < .001$, long whole unorganized list, $F(2, 219) = 26.84$, $p < .001$, long partial organized list, $F(2, 219) = 11.83$, $p < .001$, and long partial unorganized list, $F(2, 219) = 6.05$, $p < .01$. For all conditions, novice's handoff scores were significantly lower compared to the intermediates and experts. Further, for the short whole unorganized, medium whole organized, medium partial organized, and long whole organized conditions intermediate's handoff scores were significantly lower compared to the experts.

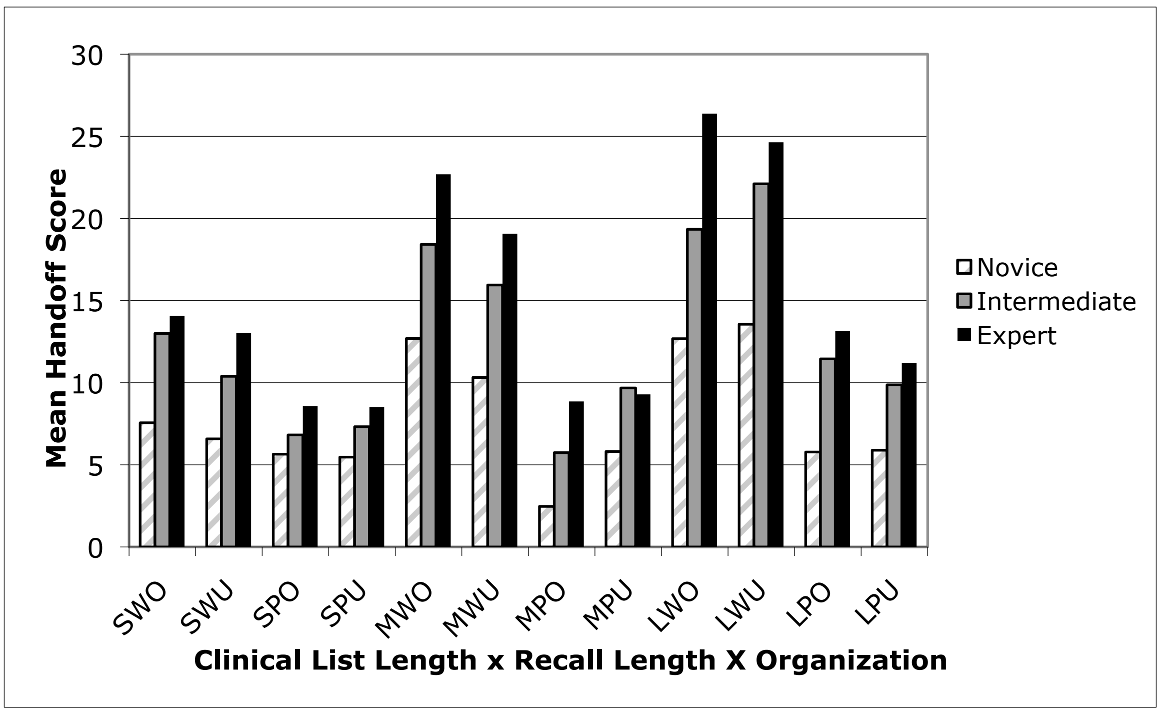

Figure 9. Mean clinical handoff score for expertise level as a function of list length, recall length, and organization.

Additional simple effects analyses revealed significant differences within the novice group, $F(11, 146) = 85.79$, $p < .001$, intermediate group, $F(11, 146) = 101.57$, $p <$

.001, and expert group, $F(11, 146) = 170.99$, $p < .001$. For the novice group the long whole organized and long whole unorganized conditions resulted in significantly higher handoff scores compared to all other conditions except for the medium whole organized condition. The medium whole organized and medium whole unorganized conditions resulted in significantly higher handoff scores compared to all short conditions, the medium partial conditions, and the long partial conditions. Further, the medium whole organized condition resulted in significantly higher handoff scores compared to the medium whole unorganized condition. The short whole organized condition resulted in significantly higher handoff scores compared to the short partial, medium partial, long partial, and short whole unorganized conditions. The short whole unorganized condition resulted in significantly higher handoff scores compared to the short partial unorganized and medium partial organized conditions. The short partial unorganized condition resulted in significantly higher handoff scores compared to the medium partial organized condition. Finally, the medium partial organized condition resulted in significantly higher handoff scores compared to the long partial conditions and the medium partial unorganized condition.

For the intermediate group the long whole organized and long whole unorganized conditions resulted in significantly higher handoff scores compared to all other conditions except for the long whole organized condition compared to the medium whole organized condition. Further, the long whole unorganized condition resulted in significantly higher handoff scores compared to the long whole organized condition. The medium whole organized and medium whole unorganized conditions resulted in significantly higher handoff scores compared to all short conditions, the medium partial conditions, and the long partial conditions. Further, the medium whole organized condition resulted in significantly higher handoff scores compared to the medium whole unorganized condition. The short whole organized condition resulted in significantly higher handoff scores compared to the short partial, medium partial, short whole unorganized, and long partial unorganized conditions. The short whole unorganized condition resulted in significantly higher handoff scores compared to the short partial and medium partial organized conditions. The long partial organized condition resulted in significantly higher handoff scores compared to the short partial and medium partial conditions. The long partial unorganized condition resulted in significantly higher handoff scores compared to the short partial and medium partial organized conditions. The medium partial unorganized condition resulted in significantly higher handoff scores compared to the short partial conditions. Finally, the short partial unorganized condition resulted in significantly higher handoff scores compared to the medium partial organized condition.

For the expert group the long whole organized and long whole unorganized conditions resulted in significantly higher handoff scores compared to all other conditions except for or the long whole organized condition compared to the medium whole organized condition. The medium whole organized and medium whole unorganized conditions resulted in significantly higher handoff scores compared to all short conditions, the medium partial conditions, and the long partial conditions. Further, the medium whole organized condition resulted in significantly higher handoff scores compared to the medium whole unorganized condition. The short whole organized and the short whole unorganized conditions resulted in significantly higher handoff scores compared to the short partial, medium partial, and long partial unorganized conditions.

The long partial organized condition resulted in significantly higher handoff scores compared to the short partial and medium partial conditions. The long partial organized and long partial unorganized conditions resulted in significantly higher handoff scores compared to the short partial and medium partial organized conditions. Finally, the long partial organized condition resulted in significantly higher handoff scores compared to the medium partial unorganized condition.

Lenient Criterion/Proportion Data for Clinical Handoff

For the clinical data, there were no differences in the pattern of results and statistically significant effects for the lenient and strict criteria. Regarding the proportion data, it was revealed that the list length effect was still significant, $F(1.80, 131.67) = 255.64$, $p < .001$, partial $\eta^2 = .78$; however, unlike the whole scores, as list length increased the handoff scores decreased. The proportion of correct data recalled decreased significantly from the short list ($M = .42$, $SD = .16$) to the medium list ($M = .29$, $SD = .13$) and from the medium to the long list ($M = .25$, $SD = .12$). The rest of the proportion results are presented in a table in Appendix X.

Air Traffic Control Handoff Task

Regarding expertise level, confidence interval equivalence tests were performed on the strict criterion total handoff score to determine whether the novice's and intermediate's performance was equivalent to the expert's performance. In addition, confidence interval equivalence tests were performed on the relevant scores and irrelevant scores for expertise. The criterion to determine acceptable performance was again set at any score within 1/3 of the expert's scores set at a confidence interval of 90%. Further, a 3 level of expertise (between variable; novice, intermediate, expert) x 3 list length (within variable; short, medium, long) x 2 recall length (within variable; whole, partial) x 2 organization (within variable; organized, unorganized) mixed factorial ANOVA was used to analyze the strict criterion total handoff score performance data as well as potential interactions of these variables with expertise. Descriptive statistics are shown in Table 18.

Table 18

Descriptive Statistics for Air Traffic Control Handoff Task

Scale	*N*	*M*	*SD*	Skewness	Kurtosis
HSO Short Whole	76	10.72	4.29	.20	-.83
HSO Short Partial	76	6.75	2.85	.21	-.44
HSO Medium Whole	76	16.89	6.49	.72	.74
HSO Medium Partial	76	5.05	4.46	.16	-.27
HSO Long Whole	76	18.13	8.55	.64	-.11
HSO Long Partial	76	9.23	5.86	.08	-.52
HSU Short Whole	76	9.32	3.88	.39	-.21
HSU Short Partial	76	6.78	2.82	.22	-.38
HSU Medium Whole	76	14.14	5.81	.76	.06
HSU Medium Partial	76	7.74	4.34	.16	-.61
HSU Long Whole	76	18.76	7.12	.02	-1.11
HSU Long Partial	76	8.35	4.77	.14	-.13

The ATC relevant and irrelevant scores were also checked for equality of variance. Descriptive statistics are shown in Table 19.

Table 19

Descriptive Statistics for Air Traffic Control Relevant and Irrelevant Scores

Scale	*N*	*M*	*SD*	Skewness	Kurtosis
Relevant	76	119.75	40.26	.41	-.29
Irrelevant	76	26.45	8.50	.38	.40

Air Traffic Control Expertise. The expert's mean score for the strict criterion ATC handoff score was 6.16; therefore, the equivalency interval for the handoff score was 4.11 to 8.21. The 90% confidence interval for the novice score was 2.54 to 4.06 and for the intermediates it was 4.27 to 7.39. As can be seen in Figure 10, the intermediate's confidence intervals fall within the equivalency interval indicating their performance was equivalent to the experts; however, the novice's confidence intervals fall outside the equivalency interval range indicating the novice recall of relevant items is not equal to that of experts.

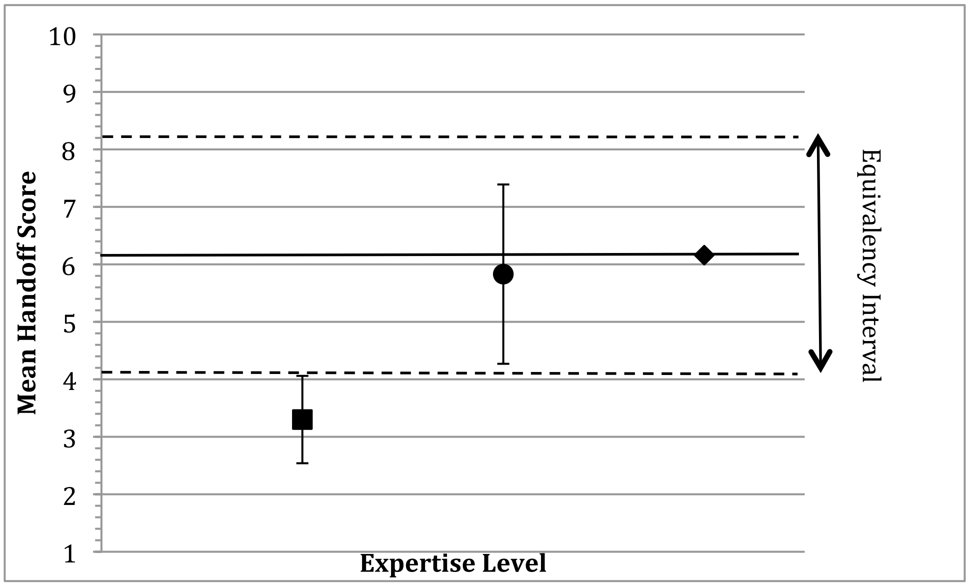

Figure 10. ATC handoff scores for each level of expertise (90% confidence interval).

Air Traffic Control Relevant Scores. The expert's mean score for the ATC relevant items was 150.43; therefore, the equivalency interval for the relevant score was 100.29 to 200.57. The 90% confidence interval for the novice score was 85.09 to 100.79 and for the intermediates it was 125.12 to 148.14. As can be seen in Figure 11, the intermediate's confidence intervals fall within the equivalency interval indicating their performance was equivalent to the experts; however, the novice's confidence intervals fall outside the equivalency interval range indicating the novice recall of relevant items is not equal to that of experts.

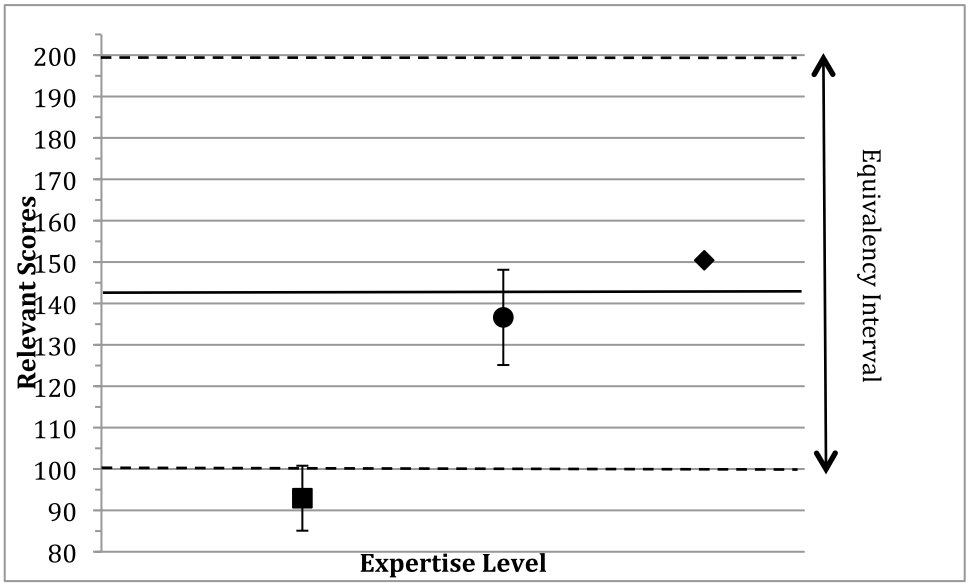

Figure 11. ATC relevant scores for each level of expertise (90% confidence interval).

Air Traffic Control Irrelevant Scores. The expert's mean score for the ATC irrelevant items was 28.57; therefore, the equivalency interval for the irrelevant score was 19.05 to 38.09. The 90% confidence interval for the novice score was 22.09 to 25.69 and for the intermediates it was 26.64 to 31.26. As can be seen in Figure 12, the novice and intermediate's confidence intervals do not fall outside of the equivalency interval indicating their performance was equivalent to the experts.

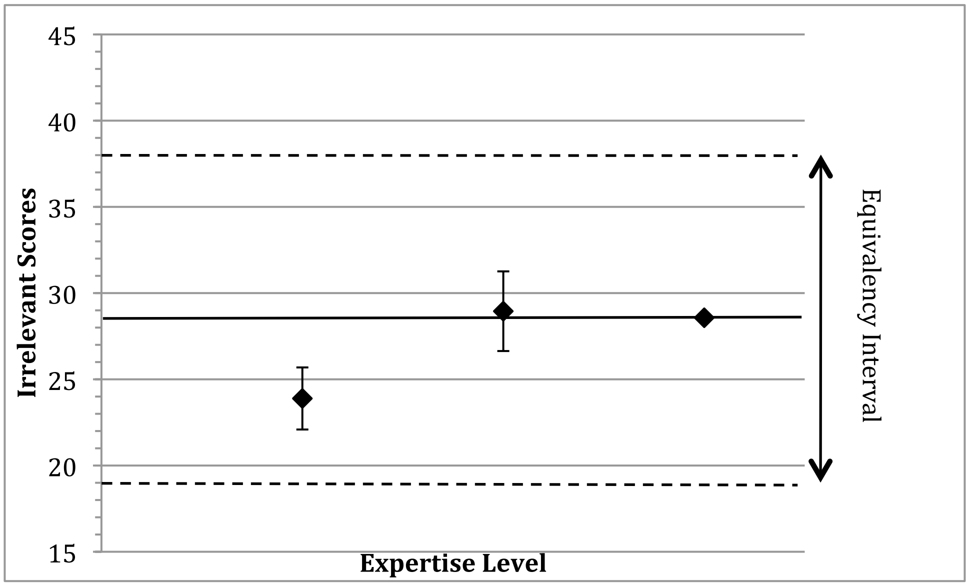

Figure 12. ATC irrelevant scores for each level of expertise (90% confidence interval).

Air Traffic Control Total Handoff Score. The results of the ATC strict criterion total handoff score ANOVA are shown in Table 20. A significant main effect for expertise was observed, $F(2, 73) = 17.32$, $p < .001$, partial $\eta^2 = .32$. Novices had significantly lower handoff scores ($M = 3.30$, $SD = 2.69$) compared to the intermediates ($M = 5.83$, $SD = 3.62$) and experts ($M = 6.16$, $SD = 3.60$).

Table 20

Results of the Analysis of Variance for Air Traffic Control Handoff Scores

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	1668.31	2	834.16	17.32	.00***	.32
Error	3516.86	73	48.18			
Within Subjects						
List Length (L)	147.39	2	73.70	10.83	.00***	.13
L x E	44.14	4	11.04	1.62	.17	.04
Error (L)	993.20	146	6.80			
Recall Length (R)	11105.25	1	11105.25	643.91	.00***	.90
R x E	100.34	2	50.17	2.91	.06	.07
Error (R)	1259.01	73	17.25			
Organize (O)	.09	1	.09	.01	.91	.00
O x E	8.83	2	4.41	.64	.53	.02
Error (O)	506.13	73	6.93			
L x R	1179.09	2	589.55	85.11	.00***	.54
L x R x E	32.90	4	8.23	1.19	.32	.03
Error (L x R)	1011.34	146	6.93			
L x O	249.10	2	124.55	23.51	.00***	.24
L x O x E	139.64	4	34.91	6.59	.00***	.15
Error (L x O)	773.52	146	5.30			
R x O	323.16	1	323.16	55.33	.00***	.43
R x O x E	38.28	2	19.14	3.28	.05*	.08
Error (R x O)	426.33	73	5.84			
L x R x O	143.63	2	71.81	11.00	.00***	.13

Table 20 Continued

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
L x R x O x E	15.41	4	3.85	.59	.67	.02
Error (L x R x O)	953.46	146	6.53			

Note. $*p < .05$; $***p < .001$

A significant main effect for list length was observed, $F(2, 146) = 10.83$, $p < .001$, partial $\eta^2 = .13$. Participants had significantly higher handoff scores for the short list ($M = 5.65$, $SD = 2.50$) compared to the medium ($M = 4.98$, $SD = 3.51$) and long lists ($M = 4.65$, $SD = 3.91$). A significant main effect for recall length was observed, $F(1, 73) = 643.91$, $p < .001$, partial $\eta^2 = .90$. Participants had significantly higher handoff scores for the whole recall list ($M = 8.72$, $SD = 3.37$) compared to the partial recall list ($M = 1.47$, $SD = 3.24$).

Further, a significant interaction for list length, organization, and expertise was observed, $F(4, 146) = 6.59$, $p < .001$, partial $\eta^2 = .15$. A plot of the interaction is shown in Figure 13. Simple effects analyses for list length by organization revealed significant differences for the short organized list, $F(2, 219) = 4.12$, $p < .05$, short unorganized list, $F(2, 219) = 6.14$, $p < .01$, medium unorganized list, $F(2, 219) = 4.39$, $p < .05$, and the long organized list, $F(2, 219) = 5.89$, $p < .01$. For all four conditions, novice's handoff scores were significantly lower compared to the intermediates and experts. Additional simple effects analyses revealed significant differences within the novice group, $F(5, 146) = 2.94$, $p < .05$, intermediate group, $F(5, 146) = 3.70$, $p < .01$, and expert group, $F(5,$

146) = 3.80, $p < .01$. For the novice group handoff scores were significantly higher for the short organized list compared to the short unorganized, medium organized, and long organized list, the medium unorganized score was significantly higher compared to the short unorganized, medium organized, and long organized score, and the long unorganized score was significantly higher compared to the medium organized and long organized score. For the intermediate group handoff scores were significantly higher for the short organized list compared to the short unorganized, medium organized, and long unorganized list, the short unorganized score was significantly higher compared to the medium unorganized score, and the medium unorganized score was significantly higher compared to the medium organized and the long unorganized score. For the expert group handoff scores were significantly higher for the short organized list compared to the medium organized, long organized, and long unorganized list, the short unorganized score was significantly higher compared to the medium organized and long unorganized score, the medium unorganized score was significantly higher compared to the medium organized and the long unorganized score, and the long organized score was significantly higher compared to the long unorganized score.

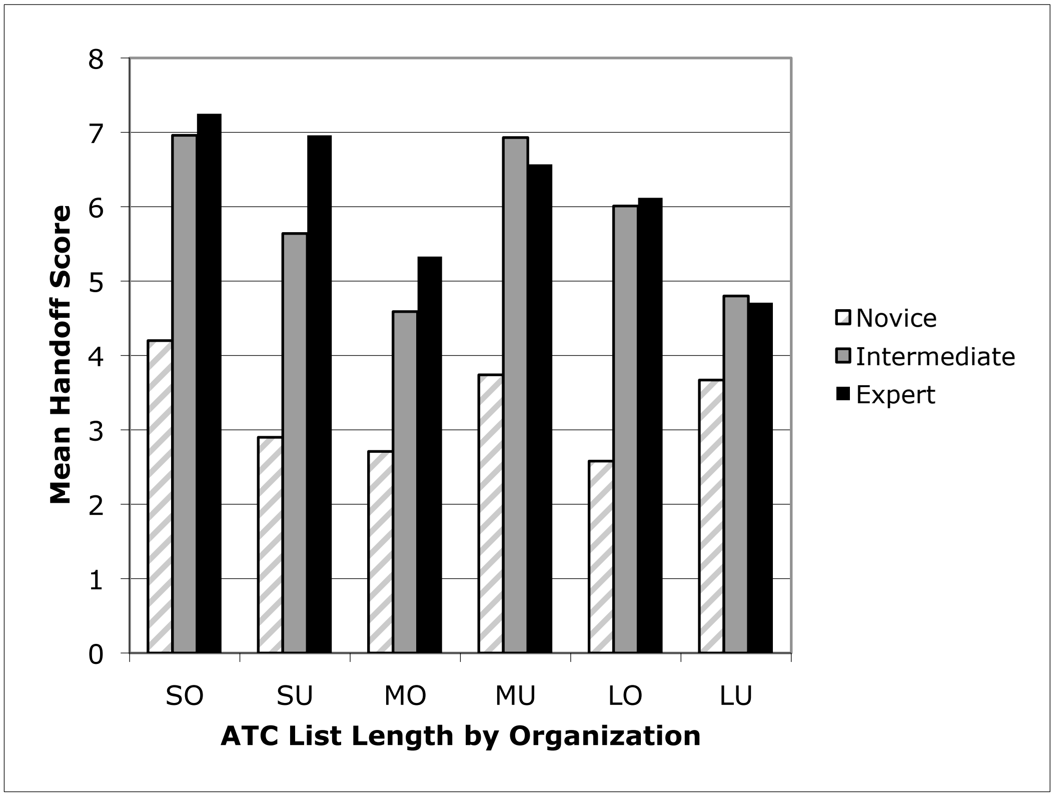

Figure 13. Mean ATC handoff score for expertise level as a function of list length and organization.

A significant interaction for recall length, organization, and expertise was observed, $F(2, 73) = 3.13$, $p < .05$, partial $\eta^2 = .08$. A plot of the interaction is shown in Figure 14. Simple effects analyses for recall length by organization revealed significant differences for the whole organized list, $F(2, 146) = 4.77$, $p < .05$, whole unorganized list, $F(2, 146) = 3.37$, $p < .05$, and the partial organized list, $F(2, 146) = 3.77$, $p < .05$. For all conditions, novice's handoff scores were significantly lower compared to the intermediates and experts except compared to the intermediate's partial organized score. Additional simple effects analyses revealed significant differences within the novice group, $F(3, 73) = 105.07$, $p < .001$, intermediate group, $F(3, 73) = 68.76$, $p < .001$, and expert group, $F(3, 73) = 71.19$, $p < .001$. For the novice group, intermediate group, and expert group participant's handoff scores were significantly higher for the whole organized list compared to all other lists, the whole unorganized score was significantly higher compared to the partial scores, and the partial unorganized score was significantly higher compared to the partial organized score except for the experts.

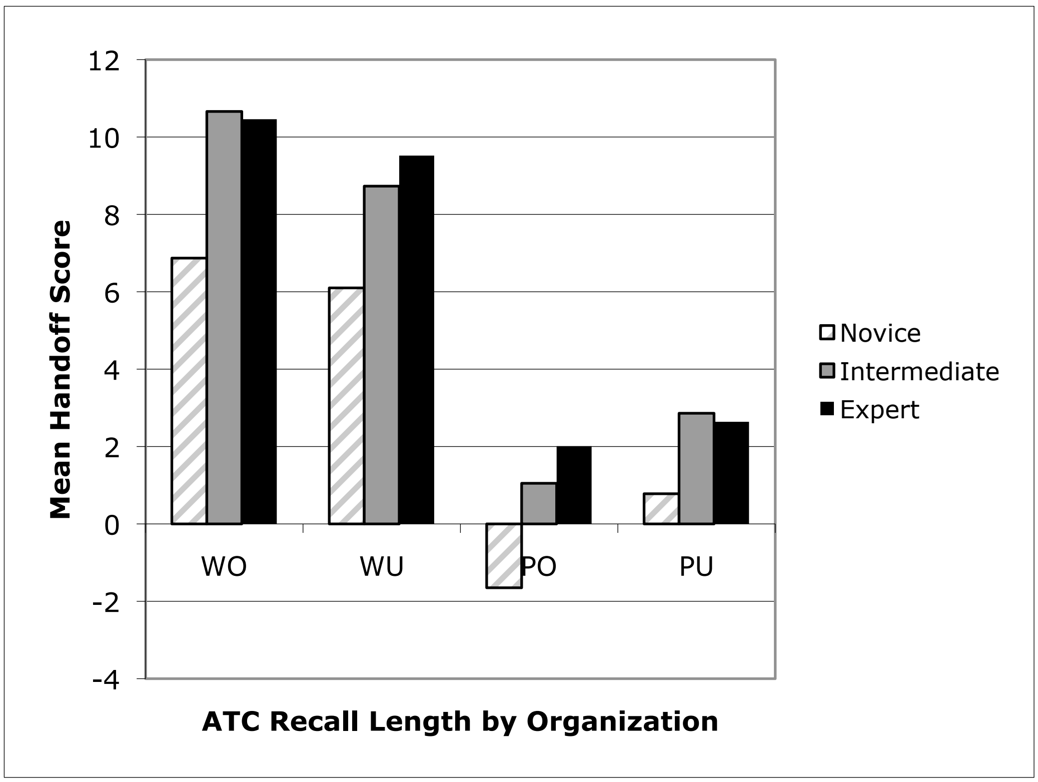

Figure 14. Mean ATC handoff score for expertise level as a function of recall length and organization.

A significant interaction for list length, recall length, and organization was observed, $F(2, 146) = 11.00$, $p < .001$, partial $\eta^2 = .13$. A plot of the interaction is shown in Figure 15. Simple effects analyses for recall length by organization revealed significant differences for the whole organized list, $F(2, 219) = 3.60$, $p < .05$, whole unorganized list, $F(2, 219) = 6.19$, $p < .01$, and the partial organized list, $F(2, 219) = 30.41$, $p < .001$. For the whole organized list participant's handoff scores were higher for the medium and long list length compared to the short list length. For the whole unorganized list participant's handoff scores were higher for the medium list compared to the short and long lists. For the partial organized list participant's handoff scores were higher for the short list compared to the medium and long list. Additional simple effects analyses revealed significant differences within the short list, $F(3, 146) = 61.12$, $p < .001$, medium list, $F(3, 146) = 349.40$, $p < .001$, and long list, $F(3, 146) = 269.33$, $p < .001$. For the short list participant's handoff scores were significantly higher for the whole organized list compared to all other lists, the whole unorganized score was significantly higher compared to the partial lists, and the partial organized score was significantly higher compared to the partial unorganized list. For the medium list participant's handoff scores were significantly higher for the whole organized list and the whole unorganized compared to the partial lists, and the partial unorganized score was significantly higher compared to the partial organized list. For the long list participant's handoff scores were significantly higher for the whole organized list compared to all other lists, the whole unorganized score was significantly higher compared to the partial lists, and the partial unorganized score was significantly higher compared to the partial organized list.

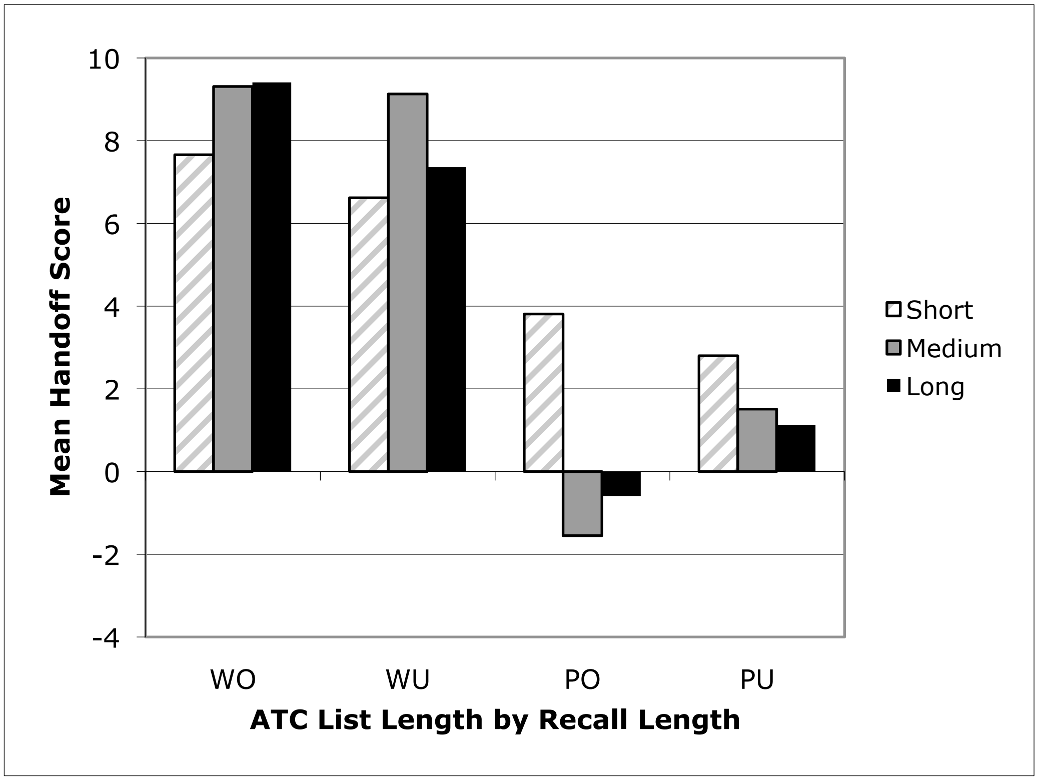

Figure 15. Mean ATC handoff score for list length as a function of recall length and organization.

Lenient Criterion/Proportion Data for Air Traffic Control Handoff

For the ATC data, there were a few slight differences between the strict and lenient criteria. For the list length data, the list length main effect failed to reach significance (p = .096) with the lenient scoring; however, the trend was the same. As list length increased, overall handoff scores decreased. There was also a significant interaction for recall by expertise (p = .044). However, all the higher order interactions remained the same.

Regarding the proportion data, it was revealed that the list length effect was still significant, $F(2, 146) = 392.15$, $p < .001$, partial $\eta^2 = .84$. Consistent with the whole scores and as might be expected as list length increased the proportions decreased. The proportion of correct data recalled decreased from the short lists (M = .26, SD = .14) to the medium lists (M = .11, SD = .10) and to the long lists (M = .07, SD = .08). The rest of the proportion results are presented in Appendix W.

CHAPTER V

DISCUSSION

The purpose of the present research was to examine the effects of mental models and expertise on the ability to process handoffs of information. Participants performed three types of tasks: a running memory task, a clinical handoff task, and an ATC handoff task. Further, the role of active or passive processing was examined.

Running Memory Task

Recall that it has been suggested that people process information in either a passive or active manner when performing running memory span tasks (Broadway & Engle, 2010; Bunting, et al., 2006; Elosúa & Ruiz, 2008; Hockey; 1973; Morris & Jones, 1990). Elosúa and Ruiz used lists of disyllabic words and Broadway and Engle used lists of letters to examine whether individuals would engage in active or passive processing of the information. Both sets of researchers predicted that individuals would actively process information; however, they found support for passive processing. Therefore, one goal of the present study was to replicate the work performed by Elosúa and Ruiz and Broadway and Engle using lists of disyllabic words. It was hypothesized that participants would passively process the word lists, consistent with the results of Elosúa and Ruiz and Broadway and Engle. Further, another goal was to establish that all three groups of participants (novice, intermediate, and expert) have the same basic memory capacity for an abstract memory task; therefore, it was hypothesized that performance would be equivalent across expertise level. Also recall that three different list lengths were used: short, medium, and long. It was anticipated that as list length increased, correct detections would decrease.

Expertise. For the current study a series of equivalence tests were performed on the running memory data. Equivalence testing is useful when the goal is to establish that groups or conditions do not differ from one another (Snow et al., 1999; Wellek, 2010). Indeed, for the current study it was anticipated that all three groups would have equivalent performance on the running memory tasks. Regarding correct detections, the equivalence test confirmed that the performance of the novices and intermediates was equivalent to that of the experts suggesting that all three groups had the same basic memory capacity for an abstract task. These results indicate that highly trained individuals, such as the clinicians in this study, do not have better memory capacities for abstract memory tasks compared to other groups of individuals without such specialized training and years of experience.

Regarding intrusions, transpositions, and location errors equivalence testing again confirmed that performance was comparable among all three groups. Novices, intermediates, and experts all exhibited a passive processing strategy regarding the disyllabic word running memory data.

Correct Detections. Consistent with expectations, the results from the current study showed that as list length increased the proportion of correct detections decreased. Participants had an average of .34 correct detections for the short list, which fell to .15 for the medium lists, and fell even further to an average of .09 hits for the long lists. Recall that whole scores were also examined to determine whether participants would truly have fewer correct detections as list length increased. Consistent with the proportion results as list length increased handoff scores decreased. The mean score for the short list was 7.16, for the medium list it was 5.69, and for the long list it was 5.22. These results are similar to those of other studies showing that as list length increases, the ability to recall items from the list decreases (Miller, 1956; Ratcliff, Clark, & Shiffrin, 1990; Robinson & Darrow, 1924).

Intrusions. Regarding intrusions, or words that were recalled that were not part of the list, there was a significant effect for list length. Participants had fewer intrusions as the list lengths increased. The higher number of intrusions in the shorter lists compared to the longer lists might seem unusual considering more errors might be expected for a longer list. A possible explanation is that overall, participants made more responses (both correct and incorrect) in the shorter list condition but made fewer responses overall as list length increased. Elosúa and Ruiz (2008) did not report any effects for intrusions based on list length; however, their participants were only asked to recall the last four words from each list regardless of list length. Therefore, the participants could anticipate having to recall only four words for every list, which could also limit the number of errors they made. In the present study, the method used by Broadway and Engle (2010) was adopted where the amount of information that needed to be recalled varied across lists, thus preventing the participants from anticipating how many words they would need to recall. However, Broadway and Engle only classified responses into correct detections; therefore, comparisons with their data cannot be made.

There was also a significant interaction of list length with recall. Participants had more intrusions for the short partial list compared to the short whole list. On the other hand, the reverse was observed for the medium list. In general, for both the whole list and partial list conditions participants had more intrusions in the short list compared to the longer list.

Although intrusion errors were committed, and there was a significant effect for list length, it must be pointed out that overall, there were very few intrusions. Participants made 4%, 3%, and 2% intrusions for the short, medium, and long lists, respectively. Thus, intrusion errors represent a very small portion of the data and had minimal impact regarding responses.

Transpositions. The results from the current study also showed a list length effect for transpositions. A transposition error occurred when a word was recalled correctly from the target list, but in the wrong serial position. Participants had more transpositions in the short and medium lists compared to the long lists. Again, as was the case for intrusions, participants made more overall responses in the shorter conditions compared to the longer conditions.

Location Errors. Location errors, or words in the partial recall lists that were presented early in the list and not supposed to be recalled, are integral to help interpret the results and determine how the information was processed. Both Elosúa and Ruiz (2008) and Broadway and Engle (2010) predicted that for their running memory span tasks participants would use an active processing approach, and they even went so far as to try and elicit active processing. However, both sets of researchers found that participants relied on passive processing. This was evident in the partial recall trials where participants recalled words presented earlier in the list. These words should not have been recalled but rather discarded from working memory if an active approach was indeed being used.

For the current study the results are consistent with Elosúa and Ruiz (2008) and Broadway and Engle (2010). Participants had location errors for all three list lengths, suggesting that they were relying on passive processing of the information. Further, as expected, as list length increased the proportion of location errors decreased. The medium and long list lengths were 16 and 24 items long, respectively, well beyond the normal memory capacity described by Miller (1956). Again there were fewer overall responses made for the longer lists; therefore, there were fewer opportunities for location errors to occur. However, the fact that they did occur even in the long lists is evidence that the participants were not actively trying to discard older items, suggesting they used a passive approach.

Clinical and Air Traffic Control Handoff Tasks

Recall that a limitation of the studies by Broadway and Engle (2010) and Elosúa and Ruiz (2008) was their use of an abstract task. Further, although Yntema and colleagues (1960; 1962; 1963) modeled their tasks after the real world scenario of ATC, the tasks themselves were still rather abstract in nature. Venturino and colleagues (1994; 1997) actually used a real world fire department dispatcher task. The drawback to their research was that it was limited to psychology students who had no knowledge or expertise as a fire dispatcher. Therefore, another goal of the current study was to examine how individuals process information when it is relevant to performing a genuine task. It was hypothesized that participants would engage in active processing when presented with information that was relevant to performing another task; however, this was contingent upon them having requisite knowledge about the task.

Two tasks were chosen for the participants to complete: a clinical handoff task and an ATC handoff task. For the clinical task, the participants had different levels of knowledge and expertise. Thus, it was expected that the mental models would differ among the groups. It was anticipated that the novices, who had no clinical experience, would have to rely on passive processing of the information. On the other hand, it was anticipated that the expert physicians would have developed mature mental models for the task and could actively process the information. Intermediates were expected to fall in between the novices and experts. On the other hand, the ATC task served as a control condition because none of the participants had knowledge or expertise in this domain and therefore should not have an appropriate mental model. It was anticipated that all three groups would process the information in a passive manner and that performance would be comparable among all three groups.

Clinical Handoff Task

Clinical Expertise. Regarding the total handoff score for the clinical handoff task it was anticipated that expert's handoff scores would be superior to novice handoff scores because they would have developed mental models for the clinical task. Further, because the intermediates also had some clinical knowledge and training, it is was anticipated that their handoff scores would fall between the expert and novice data. Consistent with expectations, novices had the poorest overall handoff score (7.87), intermediates fell in between (12.51), and experts had the best handoff score (14.95).

Clinical Relevant/Irrelevant Scores. Another goal of the present study was to examine whether a mental model would help distinguish between relevant and irrelevant information. Past research indicates that clinical experts do not necessarily recall more information compared to novices, but rather they recall more relevant items of information and fewer irrelevant items (Coughlin & Patel, 1987; Muzzin et al., 1983; Patel et al., 1986). It was therefore anticipated that a mental model would indeed aid experts, resulting in higher recall of relevant items and lower recall of irrelevant items compared to novices. Further, this was a measure to help determine the type of processing each group was using for the clinical data. If active processing was being used it was anticipated that fewer irrelevant items would be recalled because they were being actively discarded from memory.

Regarding the relevant data, there was a significant effect for expertise. Bonferroni post hoc tests revealed that experts (239.29) and intermediates (211.89) recalled significantly more relevant items compared to novices (143.64). The total possible number of relevant items for recall was 468. Therefore, the experts recalled 51% of the relevant items while the novices only recalled 31%. The intermediates fell in between and recalled 45% of the relevant items. Further, there was a significant expertise effect for irrelevant data. Bonferroni post hoc tests also revealed that experts recalled significantly fewer irrelevant items (10.71) compared to intermediates (18.05) and novices (16.28). The total possible number of irrelevant items for recall was 108. Therefore, the experts only recalled 10% of the irrelevant items while the intermediates recalled 17% and the novices recalled 15%. Overall, these results are consistent with expectations and do suggest that the experts were actively processing the information they received by discarding more irrelevant items and recalling more relevant items compared to the novices. Regarding the intermediates recalling slightly more irrelevant items compared to the novices might be explained by the "intermediate effect" described by Schmidt and Boshuizen (1993). They noted that for clinical case recall intermediates tended to recall more information compared to other levels of expertise. It appears that for the current study higher recall was for irrelevant items, which might be attributed to the intermediates having trouble making determinations about what needs to be included/excluded because of only having partially developed mental models.

Recall that Schmidt et al. (1990) proposed a four-stage model of expertise development based on mental models. At stage one, novices are just learning basic pathophysiological processes. Because they are at stage one of clinical expertise development, their mental models are not developed and can result in poor patient handoffs due to struggling to differentiate critical from noncritical information. Indeed, consistent with the results of Patel et al. (1986), the novices in the current study did appear to have difficulty distinguishing between relevant and irrelevant information. On the other hand, experts are at stage four, allowing them to easily access stored patterns of disease from memory. Easily identifying common patient ailments allows the expert to distinguish relevant from irrelevant information for subsequent recall.

Clinical Total Handoff Score. Another goal of the present study was to examine how the different variables (expertise, list length, recall length, organization) interacted with each other and how the overall handoff scores were affected. Therefore all the data were combined and analyzed. There was a main effect for list length. Interestingly, and contrary to expectations, as list length increased the handoff scores increased as well. This will be discussed in more detail in the list length by expertise interaction discussion.

There was a main effect for recall length. Recall that participants were presented with lists in which either all the information was relevant and should be recalled (whole list) or only some of the information was relevant for recall (partial list). Participants performed better on the whole recall lists compared to the partial recall lists. This will be discussed in more detail in the recall length by expertise interaction.

List Length by Expertise Interaction. Three list lengths containing 8, 16, or 24 items, consistent with the running memory span task, were presented to the participants. It was anticipated that with increasing expertise the participants' handoff scores would also increase, which was indeed what occurred. Further, although it was anticipated that experts would excel at the clinical handoff task, it was still anticipated that increasing list length would hamper recall ability, thus resulting in lower performance with increasing list length. This was also predicted to occur with the intermediates and novices as well. Consistent with expectations and the running memory span task, there was a significant effect for list length. However, contrary to expectations, as list length increased, handoff scores actually significantly increased as well. Expert's handoff scores for the short, medium, and long lists were 11.05, 14.98, and 18.84, respectively. Novice's handoff scores for the short, medium, and long lists were 6.32, 7.82, and 9.48, respectively. Intermediate's handoff scores for the short, medium, and long lists were 9.38, 12.45, and 15.69, respectively.

Although the novices had no formal clinical training they were likely patients at some point in their lives. Further, the clinical scenarios used for the current study were often for common ailments such as asthma, strep throat, and allergies. Therefore, they might have had some knowledge that helped them recall the cases. However, their best mean recall was 9.48, which was still lower than the expert's lowest mean recall of 11.05. It appears that having a developed mental model significantly aided the experts to recall large amounts of information. As the list length increased for the clinical scenarios it allowed for more details to be provided about the "patient's" status. The additional details could have aided the experts in narrowing down a diagnosis in their minds, which could have helped them recall more details based on a determined diagnosis.

Examining the proportion data, however, did reveal that as list length increased the *proportion* of items recall decreased. Therefore, although the amount of items recalled did increase with list length the proportion recalled decreased in relation to how many details were presented.

Recall Length by Expertise Interaction. Regarding recall length it was anticipated that as expertise level increased handoff scores would also increase regardless of type of recall. Indeed, this was the case. Further, it was anticipated that experts would perform better on the partial lists compared to the whole lists because there were fewer items that actually needed to be recalled in the partial lists. It was also originally anticipated that there would be no difference between the whole and partial lists for novices and intermediates would fall between experts and novices.

Contrary to expectations, experts had higher handoff scores for the whole lists (19.98) compared to the partial lists (9.39). The most likely explanation for better performance on the whole lists may be related to the scoring scheme adopted for the partial lists. Recall that for each irrelevant item recalled a participant would lose one point. Only the partial lists contained irrelevant information; therefore, if irrelevant information was being recalled it would result in lower handoff scores. On the other hand, if irrelevant information were not recalled, the handoff scores would most likely be similar to or higher than the whole handoff scores. As is discussed later, the experts were not immune to recalling irrelevant data, which would have lowered their partial handoff scores.

Also contrary to expectations the novices performed better on the whole lists (10.57) compared to the partial lists (5.18); however, upon further reflection the results indicating that performance was better for whole lists seems logical. As was discussed regarding the expert performance, the inclusion of irrelevant information in the partial handoffs resulted in deductions of points. Like the experts, the novices did recall irrelevant information; therefore, it resulted in lower handoff scores compared to the whole recall scenarios.

Consistent with the experts and novices, intermediates had higher handoff scores for the whole lists (16.54) compared to the partial lists (8.48). Again, the scores fell in between the experts and novices.

Organization by Expertise Interaction. It was also anticipated that there would be a significant interaction between expertise and organization. Again, it was anticipated that as expertise increased handoff scores would also increase regardless of type of organization; however, based on previous research (Coughlin & Patel, 1987) it was anticipated that expert performance would decrease when the clinical cases were presented in an unorganized manor. Further, it was also anticipated that organization would have no bearing on the performance of the novices as they do not have a mental model of how clinical cases should be organize and the intermediates would fall between the experts and novices. Consistent with expectations, handoff scores were best for the experts, worst for the novices, and the intermediates fell in between.

Although the effect failed to reach significance, the trend was in the expected direction with experts did having slightly higher handoff scores for the organized list (15.62) compared to the unorganized list (14.29). Thus, receiving a handoff of information in an unorganized matter certainly does not help and can potentially make it more difficult to recall the information. Consistent with expectations, there was virtually no difference between the novices' organized handoff score (7.81) and unorganized score (7.94) or the intermediates' organized handoff score (12.46) and unorganized score (12.55). It appears that intermediates and novices do not have mental models that are as developed as an expert's and therefore show no real benefit from an organized presentation.

Recall that Muzzin et al. (1983) found that experts regrouped clinical data while novices relied more on sequential recall. Claessen and Boshuizen (1985) found similar results in that experts tended to cluster information. Therefore, it was also predicted that the experts would reorganize or group the data for the unorganized cases while the novices would be more likely to recall the cases in sequential order. Examination of the data does appear to support this hypothesis. Many of the experts recalled the clinical data in the same manner for each case, often following the SOAP (subjective, objective, assessment, plan) format. On the other hand, many of the novices recalled the information in a much more sequential-like order. However, some novice participants did attempt to organize the information at a basic level, such as starting with name, age, and gender when provided.

Memory researchers have shown that participants often attempt to organize incoming information into a pattern that they find easier to recall. For example, Tulving (1962) repeatedly presented a list of unrelated words to participants, in a different order each time, and found that they would reorganize the words into the same order each time they recalled them from memory. Recently, Polyn and colleagues (Polyn, Norman, &

Kahana, 2009a; 2009b) have explored memory organization and recall based on individuals' internal representations for the recall task. The internal representation is suggested to help organize the recall of information. For the current task the experts would have a more developed internal representation, their mental model for clinical cases. A possible explanation for the lack of a larger effect for organization might be that the experts and intermediates were already reorganizing the data in a personal manner and recalling it based on previous experience with similar cases.

Expertise by List Length by Recall Length by Organization Interaction. It was predicted that level of expertise would interact with list length, recall length, and organization with higher expertise levels resulting in better performance across variable combinations compared to the novices. Consistent with expectations, for every combination except the short partial organized and short partial unorganized lists the experts and intermediates had higher handoff scores compared to the novices. Further, for all but the medium partial unorganized list the intermediate's scores fell in between that of the experts and novices.

For the experts, it was anticipated that the combined effect of increasing list length coupled with the whole recall length and unorganized condition would result in poorer performance compared to shorter list lengths of partial recall length and organized data. Contrary to expectations it was the combination of long lists with whole recall and either organization type that resulted in the best performance for not only the experts but for each level of expertise. The only exception was the medium whole organized combination that resulted in better performance compared to the long lists. It is likely that the medium whole organized list was an anomaly with the higher handoff score compared to the long lists resulting from case(s) being easier to recall for the medium organized lists compared to the long lists. Verkoeijen, Rikers, Schmidt, van de Wiel, and Kooman (2004) presented six internal medicine cases to novices, intermediates, and experts and found a case effect. They concluded that the cases most likely differed in their degree of difficulty. The cases for the present study were not assessed for degree of difficulty, but the pattern observed does suggest some cases were more difficult to process compared to other cases.

For all the participants the combination of medium lists with whole recall and either organization type resulted in better performance compared to the short lists and the medium partial and long partial combinations. Further, in general the combination of short lists with whole recall and either organization type resulted in better performance compared to the short partial lists.

As previously discussed, the partial lists resulted in all participants recalling irrelevant information, which decreased their overall handoff scores. Further, overall organization does not have as large of an impact as originally anticipated for the experts. This result suggests that a developed mental model is a greater asset than originally anticipated, aiding experts in overcoming unorganized information to maintain the overall large picture for each clinical case. It appears that presenting more details rather than fewer details, as long as the information is relevant, aids the receiver in processing the information in the most effective manner.

To further examine how domain specific expertise and the resultant mental model might affect performance, participants performed a similar task for which they had no experience. The control handoff task was modeled on air traffic control and was used to investigate any potential performance differences between the three groups of participants. The same testing paradigm used for the clinical scenarios was used for the air traffic control scenarios.

Air Traffic Control Handoff Task

Air Traffic Control Expertise. Regarding the total handoff score for the air traffic control task it was anticipated performance would be equivalent across expertise levels. Equivalence testing was used to determine whether the performance of novices and intermediates was similar to that of the experts; however, equivalence testing only confirmed that the intermediate's handoff scores were comparable to the experts. The novice's confidence intervals fell outside the equivalency interval range indicating the novice's handoff scores were not equal to that of experts. Despite the lower handoff scores for novices compared to the experts, it must be pointed out that for all three groups overall performance was very poor. The overall handoff scores for novices, intermediates, and experts were 3.30 (9%), 5.83 (17%), and 6.16 (18%), respectively. The unanticipated differences in performance across expertise levels will be discussed in more detail below.

Air Traffic Control Relevant/Irrelevant Scores. It was hypothesized that performance would be similar for relevant information recalled across expertise level; however, equivalence testing only confirmed that the intermediates recalled a similar number of relevant items compared to the experts. The novice's confidence intervals fell outside the equivalency interval range indicating the novice recall of relevant items was not equal to that of experts. Despite the lower recall by novices compared to the experts, it again must be pointed out that all three groups performed quite poorly. The overall number of relevant items recalled was very low. Experts recalled an average of 150 relevant items, intermediates recalled an average of 137 items, and novices recalled an average of 93 items. The total possible number of relevant items for recall was 468. Therefore, the participants were only recalling about 20-30% of the relevant information needed to maintain the air space.

It was also hypothesized that performance would be comparable for irrelevant information recalled across levels of expertise. Consistent with expectations, the three groups of participants performed similarly for the ATC task. Further, the amount of irrelevant information recalled was quite high when compared to the amount of relevant information recalled. The experts and intermediates both recalled an average of 29 irrelevant items and the novices recalled an average of 24 irrelevant items. The total possible number of irrelevant items for recall was 108. Consequently, the participants recalled about 22-27% of the irrelevant items. Moreover, the percentage of irrelevant items recalled was comparable to the percentage of relevant items recalled. Collectively, participants in all of the groups failed to disregard the irrelevant information during recall. Thus, this pattern of results suggests that participants in all groups relied on passive processing.

Air Traffic Control Total Handoff Score. Consistent with expectations and the running memory and clinical handoff tasks, there was a significant effect for list length. As list length increased, handoff scores decreased significantly. There was also a main effect for recall length. Participants had higher handoff scores for the whole lists compared to the partial lists. As discussed above regarding the clinical data, a probable explanation for better performance on the whole lists may be related to the scoring scheme adopted for the partial lists.

Further, there were a number of significant interactions including a 3-way interaction among list length, organization, and expertise. Overall, it was clear that list length had a significant impact on performance. The short lists resulted in higher handoff scores compared to the longer lists. Regarding expertise, some of the conditions resulted in experts having higher handoff scores compared to novices. Regarding organization, the results were mixed. For the short lists the scores were higher for the organized lists. However, at longer list lengths the opposite was observed. The likely explanation is that similar to the clinical cases, the information in some of the lists may have been easier to recall compared to other lists. However, a closer look at the medium unorganized list revealed that compared to the medium organized list the participants recalled about 10% fewer irrelevant items. In this instance this might have resulted in the lower handoff scores. However, if organization truly had an impact it would be expected to be consistent throughout the trials, resulting in better performance compared to unorganized lists across all list lengths, which did not happen.

A significant interaction for recall length, organization, and expertise was also observed. There were significant differences across expertise for the whole organized, whole unorganized, and partial organized lists. Novice handoff scores were significantly lower than those of the intermediates and experts in all conditions except compared to the intermediate’s partial organized score. There were also significant effects for recall length by organization within each level of expertise. For both novices, intermediates and experts organization again did not have a clear impact on performance. Regarding recall length the results support those for the main effect. Whole lists resulted in better handoff scores compared to partial lists.

Recall that both interactions revealed lower performance for the novices compared to the other levels of expertise. For the list length, organization, and expertise interaction the short lists, one medium list, and one long list resulted in poorer novice performance compared to intermediates and experts. Regarding the recall length, organization, and expertise interaction, novice performance was poorer for all combinations of recall and organization compared to the intermediates and experts except compared to the intermediate's partial organized score. One possible explanation for lower novice scores compared to the intermediates and experts may be tied to experience. Experts and intermediates in this study have received formal training in the medical field that requires assimilation of large amounts of complex information, similar to air traffic controllers. Stein et al. (2010) made the observation that experts from different disciplines displayed better use of memory compared to what might be expected from basic memory studies such as those by Miller (1956) and Yntema and colleagues (1960, 1962, 1963). Although this observation was made in relation to the expert's specific field, it stands to reason that the years of handling patient information in the medical field may have fostered some general skills to be able to synthesize, organize, and summarize large amounts of related information. This skill could have transferred to some extent to the ATC task giving the intermediates and experts a slight advantage over novices.

Recall that Schmidt et al. (1990) proposed a four-stage model of the development of *medical* expertise based on changes in mental models. However, other more general models have been proposed to help explain the development of expertise (O'Byrne, Clark, & Malakuti, 1997). One model proposed by Schumacher and Czerwinski (1992) consists of three basic stages. At the first stage a person is a considered a novice and is simply gathering knowledge about a certain domain mainly through personal experiences. At the second stage the learner is more advanced and can start detecting common patterns. Further, he or she can start to abstract relevant information rather than trying to retain all available information. By the final stage, the learner has developed expertise. At this stage, Schumacher and Cserwinski propose that individuals can detect system-wide patterns. Further, they suggest it might be possible to transfer knowledge from one system to another.

The general theory of expertise development proposed by Schumacher and Czerwinski (1992) can help explain the differences in the ATC data across expertise levels for the relevant items recalled and the interactions. Both fields of medicine and ATC require assimilation of large amounts of information. Patterns can be recognized to help organize the information. Although none of the participants had ATC experience, the need for the clinicians to recognize common features in clinical cases might have allowed them to identify similar patterns in the ATC scenarios, but to a much lesser extent.

Kimball and Holyoak (2000) point out that the concept of transfer in relation to expertise might seem counterintuitive as expertise is exhibited as optimal performance within a particular domain. However, the majority of expertise research to date has been performed predominantly within the expert's domain (Kimball & Holyoak, 2000), with a focus on what Hatano and Inagaki (1986) refer to as routine expertise. Hatano and Inagaki further suggested the concept of adaptive expertise which focuses on general reasoning skills and how development of these skills in one domain might transfer and aid an individual in performing a task in another domain.

Kimball and Holyoak (2000) suggested that adaptive expertise goes beyond knowing *how* and *what* to knowing *why* and that adaptive expertise might develop to a greater extent for tasks that are variable in nature. The present results are consistent with the notion of adaptive expertise: there was a positive relationship between the level of medical expertise and the degree of transfer to the ATC task, taken from another domain with similar characteristics.

Despite the recall differences observed for groups with different levels of medical expertise, it must be pointed out that all three groups performed quite poorly. Recall that the expert's overall handoff score was 6.16 (16%), the intermediate's score was 5.83 (15%), and the novice's score was 3.30 (8%). Therefore, although the experts and intermediates appeared to do better than the novices across conditions, the overall scores were so low that the differences can be considered minimal. Further, the amount of relevant items recalled for novices, intermediates, and experts were 92.94, 136.63, and 150.43, respectively, equaling 20-32% of recall of relevant items. In comparison the overall handoff scores and relevant items recalled for the clinical task were much higher. The overall handoff scores for novices, intermediates, and experts were 7.87 (20%), 12.51 (32%), and 14.95 (38%), respectively. It is important to note that the handoff performance of experts in the ATC task fell below that of novices in the clinical task. Further, the number of relevant items recalled by novices, intermediates, and experts in the clinical task was 143.64, 211.89, and 239.29, respectively, equaling 31% to 51% of recall of relevant items. Again the novice performance for clinical recall was almost the same for expert performance for the ATC task. It is clear that the ATC task was much more challenging and resulted in very poor performance across all three groups, minimizing the overall expertise differences that were observed.

Finally, a significant interaction for list length, recall length, and organization was observed. There were significant differences across list length for the whole organized, whole unorganized, and partial organized lists. There were also significant effects for recall length by organization within each level of list length. Regarding the recall length the results remain consistent with the previous findings. In general, the whole lists resulted in better handoff scores compared to the partial lists. Regarding organization, the effects were still mixed.

Limitations and Future Research

There are several limitations to the present study that should be noted. One limitation of the study was the total length of time it took to complete all three tasks. Although efforts were made to counterbalance the two handoff tasks, randomize the order of handoff scenarios, and provide breaks between tasks, there was still concern that participants may have become fatigued over the course of the study. The present results confirmed that all groups of participants had the same basic memory capacity; therefore, it might be possible to present an abbreviated version of this task in future research to help reduce fatigue effects. Further, other modifications could be made to all cases, such as only presenting two list lengths rather than three.

A second limitation of the current study was that the handoff cases were not assessed for potential differences in difficulty to minimize case-specific effects. Indeed, some of the results were the opposite of what was predicted regarding list length. It appears these effects were due to some cases being more difficult to recall compared to others, although this could not be confirmed definitively. The same is true for the ATC cases. For the ATC cases guidance was sought from multiple sources including an ATC professor from a local university, military air traffic controllers, academic publications, and FAA guidelines. Despite these efforts to ensure realistic cases an air traffic control expert did not examine the final ATC scenarios. In the future, it would be advisable to have one or two experts review the case scenarios and then run a pilot test to minimize potential differences among the cases.

Further, no ATC personnel participated in the study. Future research should include ATC experts to determine whether the effects observed for different levels of medical expertise would also be observed different levels of ATC experience. Indeed, Stein and Garland (2009) have noted that there is little literature on mental models in relation to ATC; however, the literature that does exist appears to support the concept that air traffic controllers use mental models to facilitate retention and that more advanced controllers have better mental models. In addition, it would be of interest to explore further the concept of adaptive expertise between the medical and ATC domains.

Another limitation of the current study was that the clinical scenarios were developed based on a number of subdisciplines from medicine such as emergency medicine, pediatrics, and primary care. Further, the physicians who participated represented a variety of specialties. Therefore, some physicians may have been less familiar with the content of some of the clinical scenarios used in this study, even though a physician vetted the cases for realism and to ensure the cases would be recognizable across specialties. However, future research should focus on examining handoff performance within specialties of medicine using cases only from that discipline.

Another limitation concerns the scoring of the cases. Although an extensive effort was made to follow a standard scheme for scoring, only one rater was used to score all the data who was not blind to which data belonged to which conditions. However, multiple measures were employed to remain consistent and unbiased. First, a blind, unbiased rater analyzed a subset of the data and overall IRR was high. Further, as discussed previously, very specific scoring guidelines were developed and closely followed to ensure all data were scored in an objective and consistent manner. Recall also that the data were scored using both lenient and strict criteria. A strict criterion was used adopted because it greatly reduced the subjectivity surrounding scoring the scenarios by either not assigning any points or only assigning half points for information that was not 100% correct. As noted above, the pattern of results was similar with both criteria, although a few effects in the clinical data failed to reach significance with the lenient criterion. Thus, using the strict criterion helped minimize the potential sources of bias; however, it is recommended that at least two blind raters be used to score the cases in the future.

Study Implications

One of the main purposes of the current study was to examine characteristics that might affect communication during transitions of care. As previously discussed, the restriction in resident work hours by the ACGME (2003) has resulted in an increase in transitions of patient care (Chang, et al., 2010; Nemeth et al., 2006; Patterson & Wears, 2010), which have been associated with errors related to patient care (Chu et al., 2010;

Riesenberg, et al., 2009; Weinger, et al., 2010). Therefore, the current study sought to assess how memory, mental models, and expertise might affect an individual's ability to receive and process information.

The results revealed that all three groups have the same basic memory capacity for disyllabic word lists, indicating that clinical experts with advanced education and highly specialized training do not possess unique running memory capabilities over medical students and novices. Further, all three groups passively processed the disyllabic word lists, again indicating that experts do not process this type of basic information differently from novices. These results also confirm the findings of Broadway and Engle (2010) and Elosúa and Ruiz (2008) that when presented with abstract information people rely on passive processing over active processing. It should be noted, however, that under genuine transition in care conditions the information that one receives during a handoff is not abstract, but relevant to performing a task.

Many occupations require effective handoffs of information (Arora & Johnson, 2008; Patterson, et al., 2004). Indeed, a critical component of a patient handoff is determining what information needs to be passed along to the receiver. Further, it is important for the receiver to be able to comprehend and thus prioritize the information they receive to perform the task at hand. Unfortunately, as previously discussed, patient handoffs are susceptible to errors and the source of those errors is still unclear. Therefore, the current study sought to examine how providing only relevant information affected recall as compared to providing both relevant and irrelevant information. Further, the study sought to examine how mental models affected information recall.

As previously discussed, a mental model is a person's own internal cognitive representation for domains that are mechanical-causal in nature (Brewer, 2006). Moray (1997/2004) further classified mental models into three different types. Regarding this study, a Type 2 mental model is an operator's model of a device or system, such as a person's model of a patient's condition. As Schmidt et al. (1990) described, mental models vary based on level of expertise associated with the domain in question. Therefore, the mental models of clinical experts were expected to be better developed than those of novices and thus the experts should be able to recall more relevant than irrelevant items from the clinical scenarios.

Indeed, the experts did recall more relevant items and fewer irrelevant items compared to novices, they still recalled irrelevant items, which ultimately resulted in lower handoff scores for the partial recall scenarios compared to the whole recall scenarios. The same pattern was also observed with intermediates and novices (i.e., lower handoff scores for the partial as compared to the whole scenarios). Moreover, this same pattern of poorer handoff scores for the partial scenarios was observed for all groups in the ATC task. These results indicate that passing along irrelevant information during a handoff results in the recall of irrelevant items to the detriment of relevant items.

A finding such as this suggests that it may be critical to minimize or eliminate irrelevant information from a handoff to reduce the potential for interference with the recall of important relevant information. For any occupation where critical information needs to be handed off, training should be developed to help less experienced personnel learn to identify what information should be included in a handoff and what should not. This is not an easy task because what is important can vary according to each case. For medical personnel developing practice cases, such as the ones used in the current study, it might be beneficial to help students understand what is truly relevant, what is completely irrelevant, and then what might be relevant or irrelevant. Using a structured format might also prove beneficial.

In addition to the relevance of information, the organization of information was also examined in the current study. For the clinical task organization had an effect on the experts' performance, with better performance for the organized as compared to unorganized scenarios. Regarding the ATC task organization did not affect performance for any of the groups. Thus, it appears that with lower levels of expertise or unfamiliar domains receivers fail to benefit from organization, perhaps because they cannot recognize the structure embedded in the material. However, with expertise when one has a developed mental model and can anticipate organizational structure it appears that an unorganized presentation format has a negative effect on recall. Therefore, teaching a structured format for handoffs may be critical to ensure proper transfer of information.

Regarding the observed differences in expertise, for the clinical task, the experts recalled more relevant items and fewer irrelevant items compared to the novices. The experts also had higher handoff scores for both the partial and whole scenarios and the organized and unorganized scenarios compared to the novices. The intermediates' performance fell in the middle. These results support the notion that different levels of expertise result in different levels of developed mental models. For the ATC task the experts did recall more relevant items compared to novices, but the overall level of recall was quite low. Moreover, all groups recalled equal amounts of irrelevant items and had similar handoff scores. Therefore, for the clinical task, although experts were negatively affected by irrelevant data and poor organization, they were still able to rely on their mental model for clinical cases resulting in superior performance compared to the other groups.

Finally, the current study sought to determine whether active or passive processing was used for more realistic tasks and whether the benefit of experience and a developed mental model would improve outcomes. Overall, the ATC results indicate that when individuals do not have expertise and thus a mental model for the task, they are forced to rely on passive processing of the information, as they do not know what is important to keep in working memory and what can be discarded. On the other hand, as the clinical results indicate when individuals do have experience and a mental model for a task it is more likely they can actively process the information they receive by retaining more relevant information and discarding irrelevant information.

Conclusion

The main goal of the current study was to examine how individuals with different levels of expertise in medicine process disyllabic word lists, air traffic control handoffs, and clinical handoffs. List length and recall length were manipulated for all three tasks and organization was manipulated in the handoff tasks. Consistent with expectations, all three groups had the same basic memory capacity for disyllabic words and they exhibited passive processing for the running memory task. Regarding the clinical handoff task, as anticipated there were differences among the three groups. The experts exhibited superior performance indicated by higher overall handoff scores, higher recall of relevant items, and lower recall of irrelevant items compared to the novices. Years of specialized training in medicine led to developed mental models of disease for the experts. These mental models aided the experts in determining what information needed to be recalled and what was not important to recall. In addition, as they received the handoffs of information the items received would trigger a pattern of disease that allowed the experts to recall more details over the novices. Further, as anticipated, the intermediate's scores fell in between that of novices and experts. Regarding the ATC data there were some differences in performance for the novices compared to the intermediates and experts, but in general all three groups performed similarly. The participants exhibited passive processing of the data and overall performance was very poor. None of the participants had knowledge about air traffic control; thus, they did not know what was or was not important to recall for maintaining control of airspace. Further, the ATC language was unfamiliar to the participants, again making it difficult to receive and process the information in anything but a passive capacity.

REFERENCES

Accreditation Council for Graduate Medical Education (2003). Report of the ACGME work group on resident duty hours. Chicago. June 11. http://www.acgme.org/DutyHours/wkgroupreport611.pdf (January 23, 2012)

Accreditation Council for Graduate Medical Education. (2011). Common Program Requirements.http://www.acgme2010standards.org/pdf/Common_Program_Requirements_07012011.pdf

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135,* 298-313. doi:10.1037/0096-3445.135.2.298

Arora, V., & Johnson, J. (2006). A model for building a standardized hand-off protocol. *The Joint Commission Journal on Quality and Patient Safety, 32,* 646-655.

Arora, V., & Johnson, J. (2008). Handoffs. In S. C. McKean, A. L. Bennett, & L. K. Halasyamani (Eds.), *Hospital medicine: Just the facts.* New York: McGraw-Hill.

Arora, V., Johnson, J., Lovinger, D., Humphrey, H. J., & Meltzer, D. O. (2005). Communication failures in patient sign-out and suggestions for improvement: A critical incident analysis. *Quality and Safety in Healthcare, 14,* 401-407. doi: 10.1136/qshc.2005.015107

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press.

Baddeley, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition, 10,* 17-23. doi: 10.1016/0010-0277(81)90020-2

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 11,* 417-423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56,* 851-864. doi: 10.1037/0003-066X.56.11.851

Baddeley, A. D. (2004). The psychology of memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The essential handbook of memory disorders for clinicians* (pp. 1-13). London: John Wiley & Sons, Ltd.

Baddeley, A. D., Allen, R.J., Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49,* 1393-1400. doi:10.1016/j.neuropsychologia.2010.12.041

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). 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. doi:10.1016/j.jml.2009.05.004

Baddeley, A. D., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia, 40,* 1737-1743. doi:10.1016/S0028-3932(01)00146-4

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

Bernbach, H. A. (1970). A multiple copy model for postperceptual memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 103-116). New York: Academic Press.

Brewer, W. F. (1987). Schemas versus mental models in human memory. In P. Morris (Ed.), *Modelling Cognition* (pp. 187-197). Oxford, England: John Wiley & Sons.

Brewer, W. F. (2006). Mental models. *Encyclopedia of Cognitive Science.* DOI: 10.1002/0470018860.s00576

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42,* 563-570.

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

Bower, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 3-32). New York: Oxford University Press.

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59,* 1691-1700. DOI:10.1080/17470210600848402

Burnham, W. H. (1888). Memory, historically and experimentally considered. *The American Journal of Psychology, 2,* 39-90. doi: 10.2307/1411406

Chabner, D. E. (2001). *The language of medicine* (6th ed.). Philadelphia, PA: W. B. Saunders Company.

Chang, V. Y., Arora, V. M., Lev-Ari, S., D'Arcy, M., & Keysar, B. (2010). Interns overestimate the effectiveness of their hand-off communication. *Pediatrics, 125,* 491-496. doi: 10.1542/peds.2009-0351

Chi, M. T. H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167-184). New York, NY: Cambridge University Press.

Chu, E. S., Reid, M., Burden, M., Mancini, D., Schulz, T., Keniston, A., … Albert, R. K. (2010). Effectiveness of a course designed to teach handoffs to medical students. *Journal of Hospital Medicine, 5,* 344-348. DOI: 10.1002/jhm.633

Chu, E. S., Reid, M, Schulz, T., Burden, M., Mancini, D., Ambardekar, A. V., … Albert, R. K. (2009). A structured handoff program for interns. *Academic Medicine, 84,* 347-352. doi: 10.1097/ACM.0b013e3181970829

Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education, 19,* 61-67. DOI: 10.1111/j.1365-2923.1985.tb01140.x

Clair, L. L., & Trussell, P. M. (1969). The change of shift report: Study shows weaknesses, how it can be improved. *Hospitals, 43,* 91-95.

Cleland, J. A., Ross, S., Miller, S. C., & Patey, R. (2009). "There is a chain of Chinese whispers…": Empirical data support the call to formally teach handover to prequalification doctors. *Quality and Safety in Healthcare, 18,* 267-271. doi:10.1136/qshc.2008.029983

Cofer, C. N. (1976). An historical perspective. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 1-14). New York: W. H. Freeman and Company.

Cohen, M. D., & Hilligoss, P. B. (2008). Handoffs in hospitals: A review of the literature on information exchange while transferring patient responsibility or control. http://hdl.handle.net/2027.42/61498.

Cohen, M. D., & Hilligoss, P. B. (2010). The published literature on handoffs in hospitals: Deficiencies identified in an extensive review. *Quality and Safety in Healthcare, 19,* 493-497. doi:10.1136/qshc.2009.033480

Collins, R. D. (2008). *Differential diagnosis in primary care* (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Journal of Medical Education, 62,* 818-828.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-185. doi: 10.1017/S0140525X01003922

Craik, K. J. W. (1943). *The nature of explanation.* Cambridge, UK: Cambridge University Press.

Davis, S. F., & Palladino, J. J. (1997). *Psychology* (2nd ed.). New Jersey: Prentice-Hall Inc.

de Groot, A. D. (1965). *Thought and choice in chess.* The Hague: Mouton.

Dunn, S. A. (1998). *Mosby's primary care consultant.* St. Louis, MO: Mosby, Inc.

Ebbinghaus, H. (1885). On memory. (H. A. Ruger and C. E. Bussenius, Trans.). New York: Teacher's College, 1913. Paperback edition, New York: Dover, 1964.

Elosúa, M. R., & Ruiz, R. M. (2008). Absence of hardly pursued updating in a running memory task. *Psychological Research, 72,* 451-460. DOI: 10.1007/s00426-007-0124-4

Ericsson, K. A. (1998). The scientific study of expert levels of performance: General implications for optimal learning and creativity. *High Ability Studies, 9,* 75-100. DOI:10.1080/1359813980090106

Ericsson, K. A. (2006). An introduction to the Cambridge handbook of expertise and expert performance: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 3-19). New York, NY: Cambridge University Press.

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

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review, 85,* 114-121.

Eva, K. W., Norman, G. R., Neville, A. J., Wood, T. J., & Brooks, L. R. (2002). Expert-novice differences in memory: A reformulation. *Teaching and Learning in Medicine, 14,* 257-263. DOI:10.1207/S15328015TLM1404_10

Fay, N., Page, A. C., & Serfaty, C. (2010). Listeners influence speakers' perceived communication effectiveness. *The Journal of Experimental Psychology, 46,* 689-692. doi:10.1016/j.jesp.2010.02.012

Fay, N., Page, A. C., Serfaty, C., Tai, V., & Winkler, C. (2008). Speaker overestimation of communication effectiveness and fear of negative evaluation: Being realistic is unrealistic. *Psychonomic Bulletin and Review, 15,* 1160-1165. doi:10.3758/PBR.15.6.1160

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publications Inc.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-72). New York, NY: WH Freeman/Times Books/Henry Holt & Co.

Hess, S. M., Detweiler, M. C., & Ellis, R. D. (1994). The effects of display layout on monitoring and updating system states. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1336-1340). Santa Monica, CA: Human Factors and Ergonomics Society. DOI: 10.1177/154193129403801919

Hess, S. M., Detweiler, M. C., & Ellis, R. D. (1999). The utility of display space in keeping track of rapidly changing information. *Human Factors, 41,* 257-281. DOI: 10.1518/001872099779591187

Hill, J. C., & Bliss, J. C. (1968). Modeling a tactile sensory register. *Perception and Psychophysiscs, 4,* 91-101. DOI: 10.3758/BF03209516

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology, 25,* 104-111. DOI:10.1080/14640747308400328

Horwitz, L. I., & Detsky, A. S. (2011). Physician communication in the 21st century: To talk or to text? *The Journal of the American Medical Association, 305,* 1128-1129. doi: 10.1001/jama.2011.324

Horwitz, L. I., Moin, T., Krumholz, H. M., Wang, L., & Bradley, E. H. (2008). Consequences of inadequate sign-out for patient care. *Archives of Internal Medicine, 168,* 1755- 1760. DOI: 10.1001/archinte.168.16.1755

Horwitz, L. I., Moin, T., Krumholz, H. M., Wang, L., & Bradley, E. H. (2009). What are covering doctors told about their patients? Analysis of sign-out among internal medicine house staff. *Quality and Safety in Healthcare, 18,* 248-255. doi:10.1136/qshc.2008.028654

Hunt, R. R., & Ellis, H. C. (2004). *Fundamentals of cognitive psychology* (7th ed.). New York: McGraw-Hill Higher Education.

James, W. (1890). *The principles of psychology.* New York: Henry Holt.

Jarvis, C. (2000). *Physical examination and health assessment* (3rd ed.). Philadelphia, PA: W. B. Saunders Company.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society, 16,* 1-13.

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13,* 207-212. doi: 10.1111/1467-9280.00439

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.) *The Oxford handbook of memory* (pp. 109-122). New York: Oxford University Press.

Lawrence, R. H., Tomolo, A. M., Garlisis, A. P., & Aron, D. C. (2008). Conceptualizing handover strategies at change of shift in the emergency department: A grounded theory study. *BMC Health Services Research, 8,* 256. doi:10.1186/1472-6963-8-256

Li, J. Z., Kohrt, H. E., Caughey, A. B. (2007). *Blueprints clinical cases* (2nd ed.). Baltimore, MD: Lippincott Williams and Wilkins.

Lužar-Stiffler, V., & Stiffler, C. (2002). Equivalence testing the easy way. *Journal of Computing and Information Technology, 3,* 233-239.

Mackworth, J .F. (1959). Paced memorizing in a continuous task. *Journal of experimental psychology, 58,* 206-211. doi:10.1037/h0049090

Manser, T., Foster, S., Gisin, S., Jaeckel, D., & Ummenhofer, W. (2010). Assessing the quality of patient handoffs at care transitions. *Quality and Safety in Healthcare, 19,* 1-5. doi:10.1136/qshc.2009.038430

Matlin, M. W. (2002). *Cognition* (5th ed.). Fort Worth, TX: Harcourt College Publishers.

Matthews, A. L., Harvey, C. M., Schuster, R. J., & Durso, F. T. (2002). Emergency physician to admitting physician handovers: An exploratory study. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 1511-1515). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120204601622

Mayo Clinic health information. (n.d.). Retrieved from http://www.mayoclinic.com/health-information/

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. doi:10.1037/h0043158

Moray, N. (2004). Models of models of… mental models. In N. Moray (Ed.), *Ergonomics: Major writings* (pp. 506-526). London, UK: Taylor and Francis. (Original work published in 1997).

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81,* 111-121. DOI: 10.1111/j.2044-8295.1990.tb02349.x

Murdock Jr., B. B. (1970). Short-term memory for associations. In D. A. Norman (Ed.), *Models of human memory* (pp. 285-304). New York: Academic Press.

Muzzin, L. J., Norman, G. R., Feightner, J. W., Tugwell, P., & Guyatt, G. (1983). Expertise in recall of clinical protocols in two specialty areas. In *Proceedings of the Annual Conference on Research in Medical Education* (pp. 122-127).

Nemeth, C. P., Kowalsky, J., Brandwijk, M., Kahana, M., Klock, P. A., & Cook, R. I. (2006). Before I forget: How clinicians cope with uncertainty through ICU sign-outs. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 939-943). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120605001030

Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133-153). Cambridge, UK: Cambridge University Press.

Nolan, M. S. (2011). *Fundamentals of air traffic control* (5th ed.). Clifton Park, NY: Delmar.

Norman, D. A. (1970). Introduction: Models of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 1-15). New York: Academic Press.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 7-14). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education, 39,* 418-427. DOI: 10.1111/j.1365-2929.2005.02127.x

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1166-1174. doi: 10.1037/0278-7393.15.6.1166

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 339-353). New York, NY: Cambridge University Press.

Norman, G. R., Jacoby, L. L., Feightner, J. W., & Campbell, E. J. M. (1979). Clinical experience and the structure of memory. In *Proceedings of the Annual Conference on Research in Medical Education* (pp. 214-218).

Oberauer, K., & Vockenberg, K. (2009). Updating of working memory: Lingering bindings. *The Quarterly Journal of Experimental Psychology, 62,* 967-987. DOI:10.1080/17470210802372912

O'Byrne, K., Clark, R. E., & Malakuti, R. (1997). Expert and novice performance: Implications for clinical training. *Educational Psychology Review, 9,* 321-332.

Parush, A., Simoneau, Y., Foster-Hunt, T., Thomas, M., & Rashotte, J. (2010). The use of resources during shift hand-offs in a pediatric intensive care unit. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 962-966). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193121005401234

Parker, J., & Coiera, E. (2000). Improving clinical communication: A view from psychology. *Journal of the American Informatics Association, 7,* 453-461. DOI: 10.1111/j.1467-0658.2001.0106h.pp.x

Patel, V. L., Groen, G. J., & Frederiksen, C. H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education, 20,* 3-9. DOI: 10.1111/j.1365-2923.1986.tb01033.x

Patterson, E. S. (2008). Structuring flexibility: The potential good, bad and ugly in standardization of handovers. *Quality and Safety in Healthcare, 17,* 4-5. doi:10.1136/qshc.2007.022772

Patterson, E. S., Roth, E. M., & Render, M. L. (2005). Handoffs during nursing shift changes in acute care. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1057-1061). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120504901112

Patterson, E. S., Roth, E. M., Woods, D. D., Chow, R., & Gomes, J. O. (2004). Handoff strategies in settings with high consequences for failure: Lessons for health care operations. *International Journal for Quality in Health Care, 16,* 125-132. doi: 10.1093/intqhc/mzh026

Patterson, E. S., & Wears, R. L. (2010). Patient handoffs: Standardized and reliable measurement tools remain elusive. *The Joint Commission Journal on Quality and Patient Safety, 36,* 52- 61.

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

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57,* 137-146. doi: 10.1037/h0046137

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009A). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129-156. doi:10.1037/a0014420

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009B). Task context and organization in free recall. *Neuropsychologia, 47,* 2158-2163. doi:10.1016/j.neuropsychologia.2009.02.013

Postman, L., Turnage, T. W., & Silverstein, A. (1964). The running memory span for words. *The Quarterly Journal of Experimental Psychology, 16,* 81- 89. DOI:10.1080/17470216408416353

Poulose, B. K., Ray, W. A., Arbogast, P. G., Needleman, J., Buerhaus, P. I., Griffin, M. R., … Holzman, M. D. (2005). Resident work hour limits and patient safety. *Annals of Surgery, 241,* 847-860. doi: 10.1097/01.sla.0000164075.18748.38

PubMed Health diseases and conditions. (n.d.). Retrieved from http://www.ncbi.nlm.nih.gov/pubmedhealth/s/diseases_and_conditions/a/

Raduma-Tomàs, M. A., Flin, R., Yule, S., & Close, S. (2010). Doctors' handovers in an acute medical assessment unit: A hierarchical task analysis. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 967-971). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193121005401235

Raduma-Tomàs, M. A., Flin, R., Yule, S., & Williams, D. (2011). Doctors' handovers in hospitals: A literature review. *BMJ Quality and Safety, 20,* 128-133. doi:10.1136/bmjqs.2009.034389

Rapid differential diagnosis. (2002). Philadelphia: Lippincott Williams & Wilkins.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Motivation, 16,* 163-178.

Riesenberg, L. A., Leitzsch, J., & Cunningham, J. M. (2010). Nursing handoffs: A systematic review of the literature. *American Journal of Nursing, 110,* 24-34. doi: 10.1097/01.NAJ.0000370154.79857.09

Riesenberg, L. A., Leitzsch, J., & Little, B. W. (2009). Systematic review of handoff mnemonic literature. *American Journal of Medical Quality, 24,* 196-204. doi: 10.1177/1062860609332512

Riesenberg, L. A., Leitzsch, J., Massucci, J. L., Jaeger, J., Rosenfeld, J. C., Patow, C., … Karpovich, K. P. (2009). Residents' and attending physicians' handoffs: A systematic review of the literature. *Academic Medicine, 84,* 1775-1787. doi: 10.1097/ACM.0b013e3181bf51a6

Robinson, E. S., & Darrow, C. W. (1924). Effect of length of list upon memory for numbers. *The American Journal of Psychology, 35,* 235-243.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100,* 349-363. doi: 10.1037/0033-2909.100.3.349

Rutherford, A., & Wilson, J. R. (2004). Models of mental models: An ergonomist-psychologist dialogue. In N. Moray (Ed.), *Ergonomics: Major writings* (pp. 309-323). London, UK: Taylor and Francis. (Original work published in 1991).

Schmidt, H. G., & Boshuizen, P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5,* 205-221. DOI: 10.1007/BF01323044

Schmidt, H. G., Boshuizen, H. P. A., and Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: The "intermediate effect" in clinical case representation studies. In *Proceedings of the Cognitive Science Society Meeting,* Lawrence Erlbaum, Hillsdale, NJ.

Schmidt, H. G., Norman, G. R., & Boshuizen, P.A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65,* 611-621.

Schumacher, R. M., & Czerwinski, M. P. (1992). Mental models and the acquisition of expert knowledge. In R. R. Hoffman (Ed.), *The psychology of expertise: Cognitive research and empirical AI.* Mahwah, NJ: Lawrence Erbaum Assosiate.

Sharit, J., McCane, L., Thevenin, D. M., & Barach, P. (2005). Examining issues in communicating patient care information across shifts in a critical care setting. In

Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (pp. 1062-1066). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120504901113

Simon, H. A., & Chase, W. G. (1973). Skill in chess: Experiments with chess-playing tasks and computer simulation of skilled performance throw light on some human perceptual and memory processes. *American Scientist, 61,* 394-403.

Simonton, D. K. (2006). Historiometric methods. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 319-335). New York, NY: Cambridge University Press.

Snow, M. P., Reising, J. M., Barry, T. P., & Hartstock, D. C. (1999). Comparing new designs with baselines. *Ergonomics in Design: The Quarterly of Human Factors Applications, 7,* 28-33.

Solet, D. J., Norvell, J. M., Rutan, G. H., & Frankel, R. M. (2005). Lost in translation: Challenges and opportunities in physician-to-physician communication during patient handoffs. *Academic Medicine, 80,* 1094-1099.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1-29. doi: 10.1037/h0093759

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19-31. DOI: 10.1177/001872086300500103

Stein, E. S., Garland, D. J., & Muller, J. K. (2010). Air-traffic controller memory. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Handbook of aviation human factors* (pp. 21.1-21.39). Boca Raton, FL: CRC Press.

Strachan, M. W. J., Sharma, S. K., & Hunter, J. A. A. (Eds.). (2012). *Davidson's 100 clinical cases.* Churchill Livingstone Elsevier.

Tabachnick, B. G. & Fidell, L. S. (2001). *Computer-Assisted Research Design and Analysis.* Boston, MA: Allyn and Bacon.

Tulving, E. (1979). Memory research: What kind of progress? In L. G. Nilsson (Ed.), *Perspectives in memory research* (p. 19-34). Hillsdale, NJ: Erlbaum.

Venturino, M. (1997). Interference and information organization in keeping track of continually changing information. *Human Factors, 39,* 532-539. doi: 10.1518/001872097778667942

Venturino, M., Romano, N. J., Miller, S. L., Murphy, M., & Coffey, T. M. (1994). Dynamic memory: Keeping track of continually changing information. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1317-1321). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193129403801915

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., Schmidt, H. G., van de Wiel, M. W. J., Kooman, J. P. (2004). Case representations by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Medical Education, 38,* 617-627.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89-104. doi: 10.1037/h0021797

Wayne, J. D., Tyagi, R., Reinhardt, G., Rooney, D., Makoul, G., Chopra, S., & DaRosa, D. A. (2008). Simple standardized patient handoff system that increases accuracy and completeness. *Journal of Surgical Education, 65,* 476-485. doi:10.1016/j.jsurg.2008.06.011

Wears, R. L., Perry, S. J., Eisenberg, E., Murphy, L., Shapiro, M., Beach, C., … Behara, R. (2004). Transitions in care: Signovers in the emergency department. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 1625-1628). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120404801414

Weinger, M. B., Slagle, J. M., Kuntz, A., Banerjee, A., Schilderout, J. S., Mercaldo, N. D., … Wallston, K. (2010). Improving actual handover behavior with a simulation-based training intervention. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 957-961). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193121005401233

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* (2nd ed.). Boca Raton, FL: Chapman& Hall/CRC Press.

Wickelgren, W. A. (1970). Multitrace strength theory. In D. A. Norman (Ed.), *Models of human memory* (pp. 67-102). New York: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). New Jersey: Prentice-Hall Inc.

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors, 31,* 617-634. doi: 10.1177/001872088903100601

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors, 5,* 7-17. doi: 10.1177/001872086300500102

Yntema, D. B., & Mueser, G. E. (1960). Remembering the present states of a number of variables. *Journal of Experimental Psychology, 60,* 18-22. doi: 10.1037/h0040055

Yntema, D. B., & Mueser, G. E. (1962). Keeping track of variables that have few or many states. *Journal of Experimental Psychology, 63,* 391-395. doi: 10.1037/h0045706

APPENDIX A

INFORMED CONSENT DOCUMENT
OLD DOMINION UNIVERSITY

PROJECT TITLE: The Effects of Mental Models and Expertise on Running Memory and Clinical Handoff Effectiveness

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 effect of mental models and expertise on running memory and clinical handoff effectiveness (located in the Mills Godwin Jr. Life Sciences Building, Room # 132 F) is completely voluntary. It is your right and responsibility to inform the researcher if you wish to cease participation at any time.

RESEARCHERS

Brittany L. Anderson-Montoya, Graduate Student, College of Science, Psychology Department
Mark W. Scerbo, Ph.D., Associate Professor, College of Sciences, Psychology Department

DESCRIPTION OF RESEARCH STUDY

The purpose of this study is to assess how an individual's level of expertise and their mental model for different handoff scenarios affect their ability to process information presented via a running memory task.

If you decide to participate, then you will join a study involving research of undergraduate Old Dominion University students, medical students, and physicians. You will participate in three tasks. One task is a running memory task where you will be asked to listen to lists of words and you will be asked to recall words from the lists. You will also participate in air traffic control and clinical handoff tasks where you will receive information that you will need to determine if it is important to recall.

If you say YES, then your participation will last for approximately 2.5 hours in the Mills Godwin Jr. Life Sciences Building, Room # 132 F. Approximately 40 undergraduates of Old Dominion University will be participating in this study.

EXCLUSIONARY CRITERIA

To the best of your knowledge, you should not have vision or hearing deficiencies that would keep you from participating in this study. You must at least be 18 years old to participate. You must have no aviation or air traffic control experience and no clinical experience.

RISKS AND BENEFITS

RISKS: If you decide to participate in this study, then you may face a risk of hand fatigue from writing down your responses. The researcher tried to reduce these risks by providing breaks between the different tasks. 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 that you will gain knowledge about how research is performed.

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.

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 written answers 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 Old Dominion University, or otherwise cause a loss of benefits to which you might otherwise be entitled. 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, neither Old Dominion University nor the researchers are 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 W. Scerbo at 757-683-4217 or Dr. George Maihafer the current IRB chair at 757-683-4520 at Old Dominion University, or the Old Dominion University Office of Research at 757-683-3460 who will be glad to review the matter with you.

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 W. Scerbo: 757-683-4217
Brittany L. Anderson-Montoya

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

SUBJECT CONSENT FORM
EASTERN VIRGINIA MEDICAL SCHOOL

Eastern Virginia Medical School (EVMS) Institutional Review Board

STUDY TITLE

The Effects of Mental Models and Expertise on Running Memory and Clinical Handoff Effectiveness

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

Brittany L. Anderson-Montoya, MS
College of Sciences
Department of Psychology
Old Dominion University

WHY IS THIS STUDY BEING DONE?

The decrease in resident work hours to 80 per week has resulted in an increase in the number of patient handoffs. While several research papers have discussed the deficiencies associated with clinical handoffs there has been little empirical research addressing what factors might affect communication during transitions of care. The present study is examining how an individual's memory and expertise for clinical information affect how handoffs of information are processed.

The purpose of this study is to determine how individuals of varying clinical expertise process information handed off to them. 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 have clinical experience and are 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 80 people will take part in this study. In addition, 40 individuals will be recruited to participate from the Old Dominion University campus. We will need you to be in the study for no more than 2.5 hours.

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
- You have aviation or air traffic control knowledge or experience
- You have uncorrected hearing or vision deficiencies

What Is Involved in the Study?

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

You will be asked to perform three types of tasks. You will perform a memory task, a clinical handoff task, and an air traffic control handoff task. The memory task will be used to establish baseline memory performance. The clinical handoff task will be used to assess how individuals with varying levels of expertise process information handed off to them. The air traffic control task will be used to assess how individuals with no experience process information handed off to them that is critical to perform air traffic control. For each task you will listen to recorded lists of information. At the end of the list for the memory task you will be asked to recall a certain number of words from the list and you will record them in an answer packet. For the clinical handoff tasks and the air traffic control handoffs at the end of each list you will be asked to recall and record only the information you feel is relevant.

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

We are studying how an individual's expertise and knowledge for a subject area affects his or her ability to process information passed to him or her.

What Are the Risks of the Study?

There are very few known risks to you, although there is some risk that you may become fatigued.

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.

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

ARE THERE BENEFITS TO TAKING PART IN THE STUDY?

If you agree to take part in this study, there may or may not be direct benefit to you. There is no guarantee that you will personally benefit from taking part in this study. We hope the information learned from this study will benefit healthcare providers in their ability to perform better clinical handoffs in the future.

WHAT ABOUT CONFIDENTIALITY?

All protected health information will be maintained in strict confidence as required by law. However, your protected health information may be disclosed if required by law. Once your protected health information is disclosed for research, such as to the sponsor, federal privacy laws may no longer protect the information.

- You also have the right to review your research records, or someone you designate may review your research records on your behalf, once the study has ended unless prohibited by law.

Your study records may be reviewed and/or copied in order to meet state and/or federal regulations. Reviewers may include, for example, an Eastern Virginia Medical School Institutional Review Board and an Old Dominion University Institutional Review Board.

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.

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. Mark Scerbo, at (757) 683-4217, Dr. Thomas Hubbard at (757) 446-7093, or Brittany Anderson-Montoya at (757) 636-1815.

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. Thomas Hubbard at (757) 446-7093 or 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.

Approved by EVMS IRB. IRB # 13-01-EX-0007

APPENDIX C

EMPLOYEE/STUDENT ADDENDUM CONSENT FORM

Eastern Virginia Medical School (EVMS) Institutional Review Board

Study Title:	The effects of mental models and expertise on running memory and clinical handoff effectiveness
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, 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.

Approved by EVMS IRB. IRB #: 13-01-EX-0007

APPENDIX D

PARTICIPANT BACKGROUND INFORMATION FORM

Participant #:______ Date:______ Time:______

The purpose of this questionnaire is to obtain background information that will be used for research purposes only.

General Information

1. Age______

2. Gender______ (0 = Female; 1 = Male)

3. Do you have normal or corrected-to-normal vision?______ (0 = Yes; 1 = No)

4. Do you have normal or corrected-to-normal hearing?______ (0 = Yes; 1 = No)

5. Please briefly list any occupations you have held that required you to handoff information at the end of your shift to the incoming personnel:

___________________________ ___________________________
___________________________ ___________________________
___________________________ ___________________________

Aviation Experience

5. Do you have any formal training in air traffic control?______

6. Do you have any flight experience?______

7. Have you taken any classes significantly covering aviation/air traffic control?______

8. Do you play any air traffic control games?______

9. Do you play any flight simulation games?______

10. Do you have a family member or close friend who is a pilot or air traffic controller? ______

Clinical Experience

Undergraduate University Students

10. Do you have any formal clinical training?______

11: If yes, please describe:____
____

12. Have you or do you currently work in a clinical setting (including dental, veterinary, etc.)?______

13. If yes, please describe your position: ____
____

14. Are you a nursing or pre-med student?______

15. Do you have a family member or close friend who works in the healthcare field? ______

Medical Students/Physicians

15. If you are a medical student please indicate what year you are:______

16. If you are an experienced physician please indicate the number of years you have been in practice:________
What is your specialty area? :____

17. Have you ever received any formal training for handoffs?______

18. If yes, please briefly describe:____
____

19. Please briefly describe your handoff experience with patients. If you have never handed off patient care please indicate with NA:_____________________________________
____
____
____
____

APPENDIX E

GENERAL PARTICIPANT INSTRUCTIONS

Today you will participate in three types of tasks: a running memory span task, an air traffic control handoff task, and a clinical handoff task. It will take approximately 2.5 hours to complete the three tasks. You will be given the opportunity to take a break between each task. You will be provided with specific instructions prior to performing each task. Further, you will be given a short practice to familiarize yourself with the task. Do you have any questions at this point? (Once all questions are addressed each participant will be provided with the task specific instructions for the first task they are assigned).

APPENDIX F

INSTRUCTION SCRIPT RUNNING MEMORY SPAN TASK

For this task you will listen to lists of words. The start of each list will be indicated by the announcement of the list number you are about to start. For example, you will hear “list 1” for the first list. The first word after the list number will be the first word belonging to the list of words. When the list is complete you will hear a tone like this (tone sounded). Then you will be instructed to recall a certain number of the words from the list in serial order. For example, if you were given the following list “red, bat, nap, one, road, cake, sit” and you were asked to recall the last three words you would write your answer in the following order “road, cake, sit”. If you do not recall a word please mark that space with an “X”. For example, if from the list above you do not recall “cake” you would write down your answer like this: “road, X, sit”.

You will listen to a total of 18 lists and have an answer packet to write down your answers for each list. Your answer for each list of words will be written in the corresponding slot based on what list you are instructed you are starting. For example, when you are told you are starting “list 1”, record your answers on the “list 1” page in that location. The list lengths will vary and you will not be informed of the length of each list. The number of words that you will be asked to recall will also vary. At the end of each list you will be informed of the number of words you need to recall.

You will now perform two practice trials to familiarize yourself with the task. Do you have any questions?

APPENDIX G

EXAMPLE OF ANSWER PACKET RUNNING MEMORY SPAN TASK

List 1: ____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
List 2: ____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
List 3: ____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________
____________	____________	____________	____________

APPENDIX H

INSTRUCTION SCRIPT AIR TRAFFIC CONTROL TASK

At the start of their shift, air traffic controllers receive a handoff of information about the current state of the location where they are working and details about the airspace they are monitoring. We will refer to this as a “big brief”. They then take place at their assigned position. There are four major positions that an air traffic controller performs: ground control, local control, flight data, and clearance delivery. At specified time periods during their shift, they rotate positions with each other and handoff their old position to the controller taking over their spot.

For this task you will listen to air traffic control handoffs. The start of each handoff will be indicated by an announcement of the handoff number. For example, you will hear “handoff 1” for the first air traffic control handoff. Each handoff will begin with the position being handed off (big brief, ground control, local control, flight data, and clearance delivery). When the handoff is complete you will hear a tone like this (tone sounded). At this point you will record the information you think you would need if you were to assume responsibility for the airspace. If you feel you do not need all the information, simply write down the information you do need. Further, you may record the information in any order you feel is appropriate. You will be given scratch paper to write down your notes. You may use abbreviations if you wish. You will then be asked to write your final response in an answer packet. Do not use abbreviations in your final answer and please write as clearly as possible. The researcher will scan your answer to check legibility and make any clarifications if necessary.

You will listen to a total of 36 air traffic control handoffs. Your final answer for each handoff will be written in your answer packet in the corresponding slot based on the handoff you have listened to. For example, when you are told you are starting "ATC handoff 1" you will find "handoff 1" in your answer packet under the heading of "Air Traffic Control Handoff" and record your answers for "ATC handoff 1" in that location. The handoff length will vary from scenario to scenario. You will not be informed of the length of each handoff, but again the end will be marked by the tone you heard previously.

You will now perform two practice trials to familiarize yourself with the task. Do you have any questions?

APPENDIX I

EXAMPLE OF ANSWER PACKET FOR AIR TRAFFIC CONTROL HANDOFF

Handoff 1: ______________________________

____

____

____

____

Handoff 2: ______________________________

____

____

____

____

Handoff 3: ______________________________

____

____

____

____

Handoff 4: ______________________________

____

____

____

____

APPENDIX J

INSTRUCTION SCRIPT CLINICAL TASK

At the start of their shift, clinicians receive a handoff of information about the patients they will care for during their shift.

For this task you will listen to clinical handoffs of patients who have not been diagnosed. The start of each handoff will be indicated by an announcement of the handoff number. For example, you will hear “handoff 1” for the first clinical handoff. Each handoff will begin with the patient’s name. When the handoff is complete you will hear a tone like this (tone sounded). At this point you will record the information you think you would need if you were to assume responsibility for the patient. If you feel you do not need all the information, simply write down the information you do need. Further, you may record the information in any order you feel is appropriate. You will be given scratch paper to write down your notes. You may use abbreviations if you wish. You will then be asked to write your final response in an answer packet. Do not use abbreviations in your final answer and please write as clearly as possible. The researcher will scan your answer to check legibility and make any clarifications if necessary.

You will listen to a total of 36 clinical handoffs. Your final answer for each handoff will be written in your answer packet in the corresponding slot based on the handoff you just listened to. For example, when you are told you are starting “handoff 1” you will find “handoff 1” in your answer packet under the heading of “Clinical Handoff” and record your answers for “clinical handoff 1” in that location. The handoff length will vary from scenario to scenario. You will not be informed of the length of each handoff, but again the end will be marked by the tone you heard previously.

You will now perform two practice trials to familiarize yourself with the task. Do you have any questions?

APPENDIX K

EXAMPLE OF ANSWER PACKET FOR CLINICAL HANDOFF

Handoff 1: ____

____

____

____

____

Handoff 2: ____

____

____

____

____

Handoff 3: ____

____

____

____

____

Handoff 4: ____

____

____

____

____

APPENDIX L

WORDS USED TO GENERATE RUNNING MEMORY SPAN TASK

Absent
Accent
Acid
Actor
Address
Advice
Album
Almond
Angel
Apple
Arrow
Artist
Aspen
August
Austin
Author
Awful
Baby
Badger
Bali
Basket
Beaver
Blossom
Bobcat
Bonsai
Boston
Brazil
Bronco
Bubbly
Buffer
Building
Bunny
Burden
Butter
Candle
Camper
Candy
Carrot
Cattle
Chapel
Chapter
Cherry
Chicken
Chilly
Clover
Coffee
Comet
Complete
Concrete
Contact
Copy
Cuba
Dallas
Define
Dentist
Deny
Dessert
Detail
Digit
Dinner
Direct
Display
Dizzy
Doctor
Dolphin
Doodle
Double
Dragon
Durham
Eager
Eagle
Early
Earring
Easter
Edit
Eggnog
Encode
Enjoy
Entry
Error
Escape
Estate
Ethics
Europe
Exclude
Fasten
Facial
Falcon
Famous
Fancy
Farmer
Father
Feather
Feline
Ferret
Ferry
Fifty
Finish
Flavor
Folder
Forest
Forty
Freezer
Gecko
Georgia
Geyser
Giant
Goalie
Goblet
Goblin
Golden
Gourmet
Graphic
Grateful
Guppy
Gurgle
Gutter
Gymnast
Gypsy
Hammer
Habit
Hamster
Handle
Happy
Harbor
Harvest
Highest
Hippo
Hollow
Honey
Hornet
Human
Humid
Hunger
Hurdle
Iceland
Igloo
Image
Immune
Imply
Import
Improve
Index
Inkwell
Inlay
Insect
Inside
Intent
Invite
Iowa
Island
Jacket
Jagged
Jaguar
Jailer
Jasmine
Jasper
Jelly
Jester
Jingle
Jockey
Journal
Joyful
July
Jungle
Junior
Justice
Kansas

Kernel	Never	Question	Trample
Keyhole	Ninja	Quiet	Travel
Kidney	Noble	Quiver	Trumpet
Kindle	Noodle	Quota	Tumble
Kingdom	Normal	Rabbit	Tuna
Kitten	Norway	Radar	Turkey
Kiwi	Notice	Raffle	Turtle
Label	Novel	Rainbow	Tutor
Labor	Novice	Rattle	Twilight
Lady	Number	Repair	Ulcer
Laser	Nutmeg	River	Ultra
Lawyer	Object	Robber	Umpire
Lazy	Observe	Robot	Under
Lion	Obtain	Rocket	Until
Liquid	Occur	Roller	Unzip
Little	Octave	Romance	Upper
Liver	Offend	Royal	Upset
Lizard	Offer	Rubber	Urchin
Llama	Office	Rugby	Useful
Lobster	Olive	Rupture	Utah
Lumber	Only	Rustic	Vaccine
Luster	Optic	Scooter	Valet
Lyric	Orchid	Salmon	Valley
Marble	Order	Sheriff	Vampire
Magic	Orphan	Shower	Vantage
Marine	Otter	Silver	Velvet
Market	Outfit	Sixty	Venom
Medal	Oyster	Sluggish	Verbal
Mistake	Panther	Snail	Vermont
Mister	Parrot	Sorbet	Violet
Money	Patio	Standard	Violet
Monkey	Peacock	Story	Viper
Monster	Peanut	Student	Visit
Mother	Peru	Sturdy	Vital
Motion	Phony	Sugar	Vocal
Mountain	Pillow	Sullen	Voltage
Muscle	Pilot	Surgeon	Vulture
Music	Pirate	Surprise	Whistle
Mustang	Plaster	Target	Waffle
Mustard	Pony	Teacher	Wagon
Napkin	Powder	Temper	Waitress
Narrow	Prepare	Thirsty	Waiver
Nature	Primate	Thursday	Wallet
Navy	Puppet	Tiger	Walnut
Needle	Purple	Traffic	Walrus
Neon	Quarter	Trailer	Wander

Weather	Widow	Yankee	Zebra
Wedding	Window	Yearling	Zigzag
Whisker	Wonder	Yellow	Zombie
Whisper	Worry	Youthful	

APPENDIX M

RUNNING MEMORY SPAN TASK: SHORT

Whole Recall Recall all 8 Items	Partial Recall Recall Last 6 Items
Bonsai	Traffic
Display	Harvest
Georgia	Bobcat
Waffle	Junior
Fancy	Purple
Jockey	July
Fasten	Address
Silver	Useful
Laser	Novel
Vocal	Beaver
Dizzy	Temper
Ethics	Building
Neon	Whisker
Complete	Trumpet
Utah	Apple
Freezer	Navy
Kitten	Goalie
Early	Wallet
Honey	Acid
Rugby	Gutter
Earring	Market
Burden	Chapter
Only	Zombie
Austin	Humid

APPENDIX N

RUNNING MEMORY SPAN TASK: MEDIUM

Whole Recall Recall all 16 Items	Partial Recall Recall Last 10 Items
Violet	Durham
Highest	Justice
Define	Kindle
Camper	Magic
Exclude	Gypsy
Zebra	Hammer
Window	Blossom
Accent	Sluggish
Copy	Worry
Jasper	Igloo
Estate	Famous
Tuna	Kansas
Lyric	Nutmeg
Intent	Quiet
Yellow	Phony
Error	Rupture
Vulture	Romance
Folder	Tiger
Offend	Cuba
Bubbly	Arrow
Norway	Rattle
Hamster	Almond
Label	Tumble
Under	Basket
Lady	Robot
August	Mistake
Sixty	Chicken
Trample	Thursday
Peru	Kiwi
Absent	Entry
Lizard	Jacket
Orchid	Turtle
Travel	Yearling
Ferret	Pirate
Dinner	Vital
Verbal	Marine
Mister	Image
Contact	Marble
Awful	Lazy
Cattle	Kernel

Inside	Hurdle
Royal	Dessert
Narrow	Buffer
Immune	Panther
Sorbet	Order
Little	Mother
Double	Plaster
Geyser	Joyful

APPENDIX O

RUNNING MEMORY SPAN TASK: LONG

Whole Recall Recall all 24 Items	**Partial Recall Recall Last 14 Items**
Wonder	Weather
Jungle	Rubber
Candle	Nature
Butter	Widow
Digit	Quarter
Handle	Llama
Vantage	Wagon
Rustic	Aspen
Comet	Office
Repair	Unzip
Money	Monster
Whisper	Rainbow
Tutor	Deny
Optic	Jelly
Umpire	Peacock
Valet	Yankee
Octave	Motion
Doodle	Clover
Grateful	Orphan
Detail	Coffee
Feline	Inkwell
Index	Bronco
Boston	Jester
Pillow	Standard
Ferry	Imply
Waiver	Venom
Noodle	Keyhole
Author	Dentist
Wedding	Edit
Muscle	Shower
Island	Candy
Hollow	Valley
Waitress	Father
Forest	Inlay
Upset	Pony
Liquid	Concrete
Raffle	Sullen
Encode	Roller
Viper	Bali
Trailer	Eagle

Facial	Until
Building	Fifty
Eager	Direct
Obtain	Salmon
Prepare	Whistle
Lion	Iceland
Flavor	Urchin
Twilight	Lumber
Number	Sugar
Dolphin	Dallas
Jingle	Ninja
Cherry	Liver
Youthful	Album
Quota	Chilly
Ulcer	Goblin
Import	Pancake
Actor	Baby
Hippo	Ultra
Radar	Falcon
Student	Occur
Rabbit	Human
Target	Voltage
Upper	Gourmet
Offer	Brazil
River	Rocket
Forty	Walrus
Visit	Outfit
Feather	Scooter
Lawyer	Farmer
Velvet	Surprise
Olive	Mustard
Medal	Wander

APPENDIX P

AIR TRAFFIC CONTROL: SHORT TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 8/Organized
Ground control[1]: Southwest[2] Four Six Four[3] departing Gate niner[4], taxiing[5] to runway five[6], awaiting IFR departure clearance[7]. Runway seven inactive[8].

Local control[1]: United[2] five seven five[3] needs clearance[4] to land runway six[5]. AeroMexico[6] three niner seven[7] on runway eight[8].

Flight data[1]: Temp: negative eight degrees[2]. runway one zero[3] icy[4]. Pass IFR departure clearance[5] to clearance delivery controller[6]. Plane Captain training[7] occurring on Taxiway seven[8].

Whole: 8/Unorganized
Clearance delivery[1]: Cleared to the Dallas/Fort Worth Airport[2]. American[3] eight three eight[4]. Informed to cross Wasco[5] at one five thousand[6]. Maintaining three one thousand[7]. Cleared as filed[8].

Ground Control[1]: ready to taxi[2]. Request local controller[3] if clear to cross[4] runway three one[5]. Southwest[6] six two five[7]. Use scenic taxiway[8].

Local Control[1]: Holding[2]. traffic landing[3]. Taxiing into position[4]. Virgin[5] two three three[6]. Runway one niner[7] right[8].

Partial: 6/Organized

Clearance Delivery[1]: Southwest[2] eight three seven[3], cleared to Orlando International Airport[4], <u>no flight plan problems</u>[5]. Heading two zero five[6], <u>no NOTAMS</u>[7], maintain three six thousand[8].

Ground Control[1]: <u>Nothing in holding</u>[2]. Delta[3] seven four niner[4] just crossed[5] runway three four[6]. Inform local control[7]. <u>No vehicles on runway</u>[8].

Flight Data[1]: <u>Processing equipment working fine</u>[2]. Jet Blue[3] three niner two[4] awaiting departure clearance[5]. <u>No SNOWTAM</u>[6]. Heavy cloud coverage[7]. Lightning visible[8].

Partial: 6/Unorganized

Flight data[1]: Newark airport[2] experiencing delays[3]. Need to update daily records[4]. Birdtam in effect[5]. <u>No clouds[6].</u> Flight Data Processing Equipment malfunctioning[7]. <u>Sunny[8].</u>

Clearance Delivery[1]: Assign transponder code[2]. <u>Terminal two open</u>[3]. Report reaching eight thousand[4]. <u>No holding pattern issued</u>[5]. Air Canada[6] five one seven[7]. Cleared to Chicago O'Hare International[8].

Local Control[1]: Frontier[2] two four one[3]. <u>No reports for ground control[4].</u> In holding pattern[5]. Lighting equipment damaged[6] runway five[7]. <u>No departing aircraft[8].</u>

APPENDIX Q

AIR TRAFFIC CONTROL: MEDIUM TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 16/Organized

Clearance Delivery[1]: American[2] eight three eight[3] cleared to Houston Bush International Airport[4]. Maintain three two thousand[5]. Update automatic terminal information system[6]. Raleigh Durham International Airport[7]. Information delta one five two zulu[8], weather winds one zero five[9] at two[10]. Visibility three.[11] Light rain[12]. Temperature niner[13]. Dew-point one seven[14]. Departing runways three[15] and eight right[16].

Ground Control[1]: United[2] seven five three[3] on taxiway five[4]. Called back[5] to gate three[6]. Radio malfunctioned[7]. Runway incursion[8] on runway seven[9]. Category Charlie[10]. No collision[11]. All airplane ground movement[12] halted[13]. Coordinating with local control[14]. All inbound flights[15] holding[16].

Local Control[1]: category 1 aircraft[2] runway three[3]. Crossed departure line[4]. Cessna[5] one five two[6] cleared to takeoff[7] same runway[8]. Notify departure control[9]. Traffic landing[10] runway seven[11]. Ground control request[12] crossing runway one[13]. Delta[14] three two four[15] was told go around[16].

Whole 16/Unorganized

Flight Data[1]: Malfunctioning lights[9] at Uniform tango sierra building[10]. United[3] three four four[4] to Orlando International airport[5]. Light snow[13]. Tabulate[15] daily records[16]. Level 4[11] BIRDTAM[12]. Collect[14] records first. Relay departure clearance[2] to clearance delivery[6]. Caution men working[7] in safety area[8].

Big Brief[1]: Reroute all flights[2] going to Baltimore-Washington International Airport[3]. Notice to airmen[4] taxiway foxtrot closed[5]. Snowstorm[6] in Baltimore[7]. Visibility one mile[8]. Short one air traffic controller[9]. Wind one seven knots[10]. Slush[11] on runway three right[12]. Tree knocked down taxiway foxtrot[13]. Delta[14] one five four[15] being de-iced[16].

Local Control[1]: Alert Cessna[13] two tango niner[14] of BIRDTAM[15]. Issue landing clearance[6] for runway one two[7]. Taxiway two whiskey[11] closed[12]. Air Wisconsin[2] niner four two[3] delayed arrival[4]. Two local controllers[8]. Put in holding pattern[5]. Field condition[9] moderate rain[10]. BIRDTAM level three[16].

Partial: 10/Organized

Big Brief[1]: No wind[2]. Clear skies[3]. Southwest[4] five seven two[5] on taxiway seven[6]. Returning[7] to Gate four[8]. Disorderly person onboard[9]. Automated radar terminal system working[10]. No India Lima Sierra critical areas[11]. Runway eight right[12] closed[13]. Gas spill[14]. Boston Logan International[15] running on time[16].

Flight Data[1]: Training occurred yesterday[2] on runway seven[3] and taxiway four[4]. Relay departure clearance[5] Continental[6] one two seven[7] to clearance delivery[8]. Automated flight data device broken[9]. Clearances being relayed through phone[10]. Clear skies[11]. Ground light outage[12] runway two four[13]. No BIRDTAM[14]. New controller working[15]. No restricted airspace[16].

Ground Control[1]: Parking lot delta[2] under construction[3]. Request Frontier[4] six seven four[5] position[6]. Skylane[7] November two four[8] runway three two[9] via taxiway hotel[10]. No cautionary advisory[11]. Gate two bravo[12] closed[13]. Local control[14] operating well[15]. No requests pending[16].

Partial 10/Unorganized

Flight Data[1]: Forest fires[2] Great Dismal Swamp[3]. Light outage[4] runway two one[5]. Flight progress strip[6] up to date[7]. Training exercises[8] on taxiway Romeo[9]. Lights working[10] all other runways[11]. Flight restricted[12] immediately[13] over swamp[14]. Departure clearance[15] already sent[16].

Clearance Delivery[1]: Cessna[2] four five three mike[3]. Cleared to Denver International Airport[4]. Clear skies[5] on route[6]. Filed via victor eighty-three[7]. Bird threat[8] zero[9]. To climb[10]. No clouds[11]. Maintain five thousand five hundred[12]. Sunny[13]. Expect one two thousand[14] one five minutes[15] after departure[16].

Ground Control[1]: No incursions[2]. Terminal Juliett[3] closed[4]. Visibility good[5]. Speedbird[6] taxi[7] to runway one five[8]. No aircraft[9] in holding areas[10]. Unauthorized vehicle[11] on taxiway four[12]. Runway three[13] inactive[14]. Do not need local control permission[15] to cross[16].

APPENDIX R

AIR TRAFFIC CONTROL: LONG TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 24/Organized

Big Brief[1]: heavy traffic volume[2], taxiway three[3] being used for training[4], runway two four[5] closed[6] due to power line falling[7]. Also knocked tree down[8] and took out power to lights[9]. Reopen once fixed[10]. BIRDTAM in effect[11], intensity level 7[12], low altitude one thousand[13], high altitude two thousand[14]. Reroute all flights[15] going to Philadelphia International Airport[16]. Severe winds[17], icy runways[18], heavy sleet[19], poor ground visibility[20]. Military exercises ongoing[21] so restricted airspace[22] Norfolk Naval Base[23] until zero niner three zero[24].

Clearance Delivery[1]: Norfolk International airport[2] information Foxtrot[3] one one three niner zulu[4] weather, winds two eight zero[5] at five[6], gust one zero[7], visibility one two[8]. Few clouds[9] at five hundred[10], one thousand one hundred[11] scattered[12], ceiling three thousand[13] overcast[14]. Temperature two two[15], dew point three[16]. Altimeter two niner seven three[17]. Runways six right[18] and zero one in use[19]. Expect visual approach[20]. Simultaneous approaches in use[21]. Departures four left[22]. Low level wind[23] advisory is in effect[24].

Flight Data[1]: Relay departure clearance[2] to Virgin[3] three niner five[4]. Cleared to Minneapolis/St. Paul International airport[5]. Climb to one two thousand[6], expect flight level[7] one five minutes after departure[8]. Departure frequency[9] one three two point eight five[10]. Update automatic terminal information system[11] to the following: Information Kilo[12]. One eight three zero zulu[13]. Winds two one five[14] at one three[15]. Visibility seven[16]. Eight hundred[17] few[18], one thousand five hundred[19] scattered[20], measured ceiling five thousand[21] overcast[22]. Temperature three one[23]. Dewpoint one eight[24].

Whole: 24/Unorganized

Big Brief[1]: Delta[2] four eight seven[3] departing[4] on runway one[5]. Chicago O'Hare airport[6] has 4-hour delay[7] due to severe weather conditions[8]. BIRDTAM[9] moderate intensity[10], wind speed[11] moderate[12], runway one zero[13] closed[14], BIRDTAM high altitude 4000[15], light fog[16], Piper Cherokee[17] approaching for landing[18], gate seven[19] closed[20], light traffic volume[21], Piper Cherokee speed one five zero[22], storm approaching[23], Gate seven inoperable[24].

Clearance Delivery[1]: Ceiling three seven thousand[13] broken[14]. Crane[23] near taxiway Xray[24]. Temperature three one[15]. Two eight thousand[11] scattered[12]. India Lima Sierra[20] runway one eight approach[21] in use. Miami International airport[2] information quebec[3]. Dew point seven eight[16]. Wind two niner five[5] at one zero[6]. Runways one eight[18] and two three left[19] in use. two two three zero zulu[4] observation. Few clouds[9] at one seven thousand[10]. Departing runway two three left[22]. Altimeter two niner six eight[17]. Visibility eight[8]. Gust one five[7].

Big Brief[1]: Ground control[2] use frequency one two one point eight zero[3]. Caution for construction equipment[21] near terminal foxtrot[22]. Parallel approaches in use[12] between John F Kennedy International airport[13] and La Guardia airport[14]. Relay departure clearance[4]. Visibility one five[18]. Hold short[8] of runway one two right[9], traffic departing[10] runway one two right[11]. Aircraft taxi[23] with transponder[24]. Cleared to San Francisco International airport[5]. Runway one seven left[6] via taxiway papa[7]. Land and hold short operations[19] are in effect[20]. Runways one two right[15], one seven left[16] and one niner in use[17].

Partial: 14/Organized

Big Brief[1]: All lighting systems operable[2]. All flight radio aids operating[3]. All flights delayed[4] terminal alfa[5] due to crashed luggage cart[6]. Airport shuttle[7] running on schedule[8]. Aeromexico[9] five four one six[10] unscheduled landing[11]. In hold position[12] at altitude niner thousand[13]. Medical emergency landing[14], ambulance on standby[15]. No military training[16] occurring at Norfolk Naval base[17]. No air restrictions[18]. Men working[19] and equipment[20] on Taxiway hotel[21]. Runway five[22] free of debris[23]. No ASHTAM[24].

Ground control[1]: Use frequency one two one point seven zero[2] for south ground[3]. Taxiway whiskey[4] clear of standing water[5], no visual impediments[6]. Runway two left[7] inactive[8]. Training finished[9] scenic taxiway[10] this morning[11], is active[12]. No runway incursions[13]. Issue taxi clearance[14] to Frontier[15] four three seven[16] from Terminal Charlie[17] Gate two eight[18]. No glide slope[19] critical area[20]. Aircraft and vehicles[21] on taxiways Alfa[22] and bravo[23] and use north ground.[24]

Big Brief[1]: Clear skies[2], sunny[3], no rain[4]. No airshows[5] or military training[6]. Issue departure clearance[7]. Cactus[8] five four two[9] runway three two[10] via taxiway Quebec[11], November[12]. Altimeter two niner niner one[13]. No BIRDTAM reported[14]. Caution for men[15] and fallen trees[16] and wires[17] adjacent to taxiway Charlie[18]. Runway seven clear[19], no obstacles[20]. Reroute flights[21] to Portland International Airport[22]. Denver International Airport[23] has no delays[24].

Partial: 14/Unorganized

Clearance Delivery[1]: Winds one niner five[2] at eight[3]. Issued clearance[4] to Midex[5] six seven three[6] cleared to Kansas City International[7]. Update ATIS[8]: San Francisco International airport[9] information golf[10]. Midex already departed[11] one hour ago[12].

NOTAM heavy fog[13]. Midex altitude[14] one five thousand[15]. Runways eight right[16] and one zero in use[17]. Low level wind[18] advisory in effect[19]. Midex no departure amendments[20] or delays[21]. Wind gust one five[22]. Altimeter two eight eight seven[23]. Visibility one[24].

Big Brief[1]: New restaurants[2] open in Terminal Golf[3] near Gate two eight[4]. Parallel approaches[5] not in use[6]. NOTAM air show[7]. Runways eight left[8] and one two closed[9]. Runway two two[10] in use[11]. Moderate BIRDTAM[23]. All flights arriving[12] from Philadelphia International Airport[13] are delayed[14]. No delays[15] from Miami International Airport[16]. Cactus[17] seven eight three[18] awaiting clearance[19] to land[20]. No cranes[21] near airfield[22]. Radios operating[24].

Big Brief[1]: South[2] and North ground[3] operating[4]. Runway two eight[5] flooded[6]. Kite flying festival ongoing[7]. No ice[8] on taxiway Lima[9]. Runway seven[10] clear of water[11]. Gate two seven alfa[12] operating[13] in terminal Charlie[14]. Dog loose[15] near Terminal bravo[16] gate one eight[17]. Crane[18] and workers[19] near taxiway Yankee[20]. Lights operating[21]. BIRDTAM in effect[22]. No lightning[23] or severe wind[24].

APPENDIX S

CLINICAL HANDOFF: SHORT TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 8/Organized

Hendrik Johansson[1], 70 years old[2], severe upper right abdominal pain[3], yellowing of skin[4]. Had bariatric surgery 2 months ago[5]. Receiving morphine[6]. Abdominal ultrasound ordered[7]. Bloodwork ordered[8]. (dx: gallstones)

Paul Wells[1], has headache[2], chills[3], achy muscles[4], temp of 103[5]. Sick contacts at work[6]. Flu test ordered[7]. IV fluids started[8]. (dx: flu)

Lee Sanderson[1], 28 year old[2], unexplained fatigue[3], last menstrual period 4 days ago[4], heavy bleeding during menstruation[5], CBC[6] and HCG[7] ordered, pelvic ultrasound ordered[8] (dx: menorrhagia)

Whole: 8/Unorganized

Hannah Fields[1], X-ray ordered[2]. Tender to palpation over later malleolus[3]. Splint needs to be ordered[4]. Ankle swollen[5]. Non-weight bearing[6]. Lortab administered[7]. Heard "pop" while playing soccer[8]. (dx: ankle sprain vs. fracture)

Tim Romero[1], administered Ketoralac[2]. Has had similar past episodes[3]. Severe pain right side of head[4]. Check CT results[5]. Blurred vision[6]. Nauseous[7]. Allergic to sulfa[8]. (dx: migraine)

Samantha Burgess[1], complains of sternal chest pain[2] and shortness of breath[3]. Echocardiogram has been ordered[4]. Is febrile[5]. Has arrhythmia[6]. Has rheumatoid arthritis[7]. T-wave inversions on EKG[8]. (dx: Myocarditis)

Partial: 6/Organized

Tiffany Harris[1], presenting with chest pain[2]. Anxious appearing[3]. Temp: 98.8[4]. Increased respiratory rate[5]. No nasal congestion[6]. Orapharynx clear[7]. Just started grad school[8]. (dx: anxiety)

Kendra Ross[1], abdominal cramps[2], diarrhea for 4 days[3]. Ate at new restaurant tonight[4]. Pregnant[5]. Having a boy[6]. No vaginal bleeding[7]. Does not smoke[8]. (dx: food poisoning)

Ben Hicks[1], presenting with fever[2] and joint pain[3]. Has had similar episodes in past[4]. Gets occasional nasal congestion[5]. Ears are clear[6]. Started on prednisone[7], wife wants hourly checks performed[8] (dx: rheumatoid arthritis)

Partial: 6/Unorganized

Chad Kensington[1], upper GI series ordered[2]. Difficulty swallowing[3]. 63 years old[4]. No skin lesions[5]. Recently had a stroke[6]. Lost weight in past month[7]. No recent travel out of country[8]. (dx: dysphagia)

Katelyn McEvoy[1], Renal ultrasound ordered[2]. Normal lung sounds[3]. Being treated for urinary tract infection[4]. Respiratory rate 18[5]. Experiencing right flank pain[6]. Latex allergy[7]. Noticed blood in urine[8]. (dx: pylenephritis/kidney stones)

William Becker[1]. Rash spreading on right forearm[2]. Works as a bank teller[3]. Pain in arm[4]. No past surgies[5]. Cleaning out back of garage[6]. Redness and swelling[7] with central punctate lesion on right forearm[8]. (dx: brown recluse bite)

APPENDIX T

CLINICAL HANDOFF: MEDIUM TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 16/Organized

Zack Dunn[1], 20 years old[2]. Abdominal pain[3]. Pain correlated with eating[4]. Bloated feeling[5]. Mucous in stool[6]. Nausea[7]. Started 10 weeks ago[8]. Tender to palpation in lower abdomen[9]. Stool Guaiac ordered[10]. Awaiting CBC results[11]. Need to obtain stool culture[12]. Forgot to ask if having diarrhea[13]. Made NPO[14]. Heavy smoker[15]. Need to provide counseling regarding quitting smoking[16]. (dx: IBS)

Taneka Greene[1]. 5 months old[2]. African American[3]. Swollen hands/feet[4]. Is not up to date on vaccines[5]. Has had no visits with primary care physician[6]. Appears pale[7]. Tachycardic[8]. Febrile[9]. Hemoglobin: 8[10]. Platelets: 450,000[11]. Blood culture sent[12]. Mom has history of sickle cell disease[13]. Needs pain medication[14]. Needs antibiotics[15]. Social worker consult needed[16]. (dx: sickle cell anemia)

Sasha Green[1], 5 year old[2], female[3], sore throat[4], fever[5], headache[6], fatigue[7], itching[8], red facial rash[9], has Type I diabetes[10], uses Humulin R to manage[11], throat culture sent[12], blood sugar 95[13], acetaminophen[14] and diphenhydramine given[15], here with grandmother[16]. (dx: viral illness)

Whole 16/Unorganized

Whitney Huffman[1], Uric acid: 7.3[2]. Appears pale[3]. UA shows proteinuria[4]. Mild headache[5]. Ultrasound ordered[6]. Awake and alert[7]. Creatinine: 2.1[8]. Given Ondansetron[9]. BUN: 27[10]. Fatigued[11]. Calcium: 11.3[12]. BP: 140/95[13]. Vomiting[14]. Potassium: 5.4[15]. Diabetic[16]. (dx: renal failure)

Jose Gutierrez[1], skeletal survey pending[2]. Social work called[3]. Fussy past two days[4]. Swelling on right scalp[5]. 9 months old[6]. Interactive on exam[7]. Bruises noted on child[8]. Parents don't know how he obtained bruises[9]. Tension between parents[10]. Heart rate 126[11]. Consider head CT[12]. Mom reports lethargy[13]. No vomiting[14]. Had a femur fracture at 6 months of age[15]. Allergic to amoxicillin[16]. (dx: suspected physical child abuse)

Paul Burton[1], blood cultures ordered[2], 19 years old[3], tachycardic[4], lives in college dorm[5], had splenectomy[6], I.V. fluids started[7], temp: 104[8], CT ordered[9], stiff neck[10], severe headache[11], WBC slightly elevated[12], vomiting[13], roommate showing similar symptoms[14], no history of migraines[15], symptoms started 3 days ago[16] (dx: meningitis)

Partial: 10/Organized

Cooper Lawrence[1], 8 years old[2]. Pain with swallowing[3], decreased oral intake[4]. Penicillin allergy[5]. Febrile[6]. Oxygen saturation 99%[7]. Tonsils red and enlarged[8]. Pulse: 90[9]. Respiratory Rate: 19[10]. Tylenol[11] and amoxicillin started[12]. No pain with urination[13]. Does not want to go to school today[14]. Throat culture obtained[15]. No x-rays needed[16]. (dx: strep throat)

Josh Warner[1], 11 years old[2]. Nighttime rash on trunk[3], spiking fevers[4], fatigue[5]. Did not eat breakfast[6]. Swelling of knee joints past 8 weeks[7]. Normal vision[8]. Enlarged spleen on exam[9]. No nasal congestion[10]. White Blood Count: 12[11]. Red Blood Count: 4.7 million[12]. Hemoglobin: 12[13]. Glucose: 100[14]. Chloride: 112[15]. Platelet count high[16]. (dx: juvenile rheumatoid arthritis)

Jennifer Coleman[1], joint pain[2], fatigue[3]. Rash on face[4]. Eyes clear[5]. Irregular heartbeat on exam[6]. No bowel movement today[7]. Friction rub heard on cardiac exam[8]. No tinnitus[9]. Chest x-ray ordered[10]. UA shows increased protein[11]. Chloride: 100[12]. Magnesium: 1.7[13]. CBC pending[14]. ANA test ordered[15]. Patient has no questions[16]. (dx: lupus)

Partial 10/Unorganized

Ryan Jenkins[1], Family man[2]. UA results pending[3]. Normal appetite[4]. CBC ordered[5]. Hand tremors[6]. No abdominal guarding[7]. Occasional drooling[8]. Voice becoming monotone[9]. History of normal cholesterol[10]. Chemistry profile ordered.[11] No current medications[12]. Retired[13]. Slow to move limbs[14]. 67 years old[15]. No rash[16]. (dx: Parkinson's disease)

Natasha Cooke[1], Respiratory Rate: 16[2]. Not using new skin care products[3]. Heart Rate: 70[4]. Vesicles on hands[5]. Last menstrual period 2 weeks ago[6]. No history of allergies[7]. No other rash noted[8]. Social drinker[9]. No difficulty breathing[10]. No headache[11]. Corticosteroid cream applied[12]. Florist[13]. Started yesterday[14]. No visual changes[15]. Severe itching of hands[16]. (dx: contact dermatitis from plant)

Leslie Carr[1], Eyes clear[2]. Last food digested was ice cream[3]. 8 years old[4]. Has nausea[5]. Blood pressure 100/80[6]. Has diarrhea[7]. No influenza vaccination[8]. Throat clear[9]. Abdominal pain occurs after ingestion of dairy products[10]. Temp: 98.9[11]. Has abdominal cramps[12]. Has had same symptoms in the past[13]. Bubbly personality[14]. Allergic to penicillin[15]. Nose clear[16]. (dx: lactose intolerance)

APPENDIX U

CLINICAL HANDOFF: LONG TRIALS

The underlined items in the partial cases represent the items considered irrelevant to recall.

Whole: 24/Organized

Lauren Miller[1], 45 years old[2], female[3]. Increased sweating[4], feeling nervous[5], intolerance to heat[6], recent sudden weight loss[7], increased appetite[8]. Prior hysterectomy[9]. Allergic to penicillin[10] and iodine[11]. Asthma[12]. Uses albuterol as needed[13]. On exam clammy skin[14]. Noted hand tremor[15]. Hair is brittle[16]. Heart Rate: 115[17]. Irregular heart[18]. BP: 140/80[19]. Eyes slightly bulging[20]. Hyperactive reflexes[21]. Thyroid panel ordered[22]. CBC results pending[23]. Need to gather family history[24]. (dx: hyperthyroidism)

Ava Moss[1], 65 years old[2], female[3]. Upper abdominal pain radiating to back[4]. Pain started 4 days ago[5]. Has been losing weight[6]. History of gallstones[7]. Prior hysterectomy[8]. Allergic to penicillin[9]. Does not consume alcohol[10]. Mid-epigastric tenderness on abdominal palpation[11]. Abdominal distension[12]. Diminished bowel sounds[13]. Blood pressure: 90/60[14]. Temp: 100.8[15]. White Blood Count: elevated[16]. Normal electrolytes[17]. IV fluids started[18]. No pain medication given[19]. Abdominal CT results pending[20]. Husband is en route from out of town[21]. Patient refuses to make any health decisions without him here[22]. Consider surgery consult[23]. Asking what can be done for pain[24]. (dx: pancreatitis)

Melanie Carter[1], 3 month old[2], female[3]. Coughing[4]. Has runny nose[5]. Labored breathing[6]. Extremely fussy[7]. Premature birth[8]. Attends daycare[9] and is often watched by grandparents[10]. Grandfather smokes[11]. On exam febrile[12]. Tachypneic[13]. Hypoxic[14]. Appears to be in moderate respiratory distress[15]. Wheezing[16]. Dehydrated[17]. Pale skin[18]. IV fluids started[19]. On supplemental oxygen[20]. Chest X-ray ordered[21]. Pending RSV test[22]. UA pending[23]. Will need admission[24]. (dx: Respiratory syncytial virus)

Whole: 24/Unorganized

Tyler Wright[1], No sick contacts[2]. Swollen lymph nodes[3]. Decreased red blood cells[4]. Oxygen being administered[5]. 5 years old[6]. Gets nosebleeds[7]. Pale skin on exam[8]. Parents very agitated/upset[9]. Increased white blood cells[10]. Male[11]. Temp: 100.2[12]. No family history available[13]. Gets dizzy spells[14]. Fevers at home[15]. Trouble breathing during exam[16]. Mom states bruises easily[17]. On exam enlarged liver[18]. Increased fatigue last two weeks[19]. Splenomegaly[20]. Adopted[21]. Enlarged axillary and inguinal lymph nodes[22]. Decreased platelet count[23]. Coughing[24]. (dx: Acute lymphocytic leukemia)

Jasmine Bennett[1], On oxygen[2]. Intubated twice in past[3]. Female[4]. Complaining of shortness of breath[5]. IV bolus started[6]. Started 2 days ago[7]. Parents want to discuss new asthma treatment[8]. Ibuprofen given[9]. Tachycardic[10]. Father is heavy smoker[11]. 11 years old[12]. Crackles at base of left lung[13]. Febrile[14]. Chest x-ray ordered[15]. Wheezing[16]. A lot of kids sick at school[17]. Need to educate about smoking and asthma[18]. Flu test pending[19]. Three prior hospitalizations for asthma attacks[20]. Check on vaccine status[21]. Subcostal retractions noted[22]. Receiving Albuterol treatment[23]. Dehydrated[14]. (dx: asthma exacerbated by pnemonia)

María Garcia[1], 16 year old[2]. Slight cough[3], Increased white blood cell count[4]. Spanish speaking only[4]. Swollen lymph nodes[5], prednisone started[6], headache[7], allergic to penicillin[8], father demanding to speak to attending[9], swollen tonsils[10], whitish-yellow membrane on tonsils[11], ibuprofen administered[12], temp: 102.5[13], fatigued[14], brother with recent history of mono[15], mono spot titer ordered[16], symptoms for 5 days[17], waiting on liver function tests[18], no tonsillar asymmetry[19] or palatal swelling,[20] does not appear dehydrated[21], no difficulty breathing[22], parents very agitated[23], loss of appetite[24] (dx: mononucleosis)

Partial: 14/Organized

Lillian Morgan[1], 20 years old[2]. Anthropology student[3]. Attends William and Mary[4]. Recurring high fever[5] and chills[6]. Fatigue.[7] Profuse sweating[8]. Headache[9]. Diarrhea[10]. Last menstrual period: 1 week ago[11]. Period occurs in regular intervals[12]. Spent summer in Africa[13]. Returned to US 1 week ago[14]. No tobacco use[15]. Used Chloroquine but ran out[16]. Normal hearing[17]. Oxygen saturation: 98%[18]. Splenomegaly[19]. Heart rate: 72[20]. No pain on abdominal palpation[21]. CBC pending[22]. Blood smear obtained[23]. Has steady boyfriend[24]. (dx: malaria)

Chase Hudson[1], 17 years old[2]. Male[3]. Loss of appetite[4]. Dark urine[5]. Plays varsity soccer[6]. Penicillin allergy[7]. Doesn't eat a lot of red meat[8]. Jaundice on exam[9]. Normal lung sounds[10]. No shortness of breath[11]. Slight fever[12]. No vision changes[13]. Tender to abdominal palpation in the right upper quadrant[14]. Blood pressure: 110/70[15]. BUN: 28[16]. Potassium: 4.1[17], Sodium: 138[18], Magnesium: 1.7[19]. Liver function test pending[20]. Sexually active[21]. Admits experimenting with illegal drugs[22]. Parents not aware he is here[23]. Starting college in the fall[24]. (dx: hepatitis)

Jason Mullins[1], 7 years old[2], good student[3], Presenting with chest tightness[4] and increasing shortness of breath[5]. No vomiting[6]. Was playing outside when started having difficulty breathing[7]. Has ADHD[8]. Uses Albuterol inhaler at home[9]. Good oral intake[10]. Reports normal bowel movements[11]. Heart rate 118 bpm[12], No heart murmur on exam[13], Respiratory rate 30[14]. Ears clear[15], eyes clear[16]. Wheezing[17]. No abdominal tenderness[18]. Started on prednisone[19]. Received albuterol nebulizer treatment[20]. Electrolytes normal[21]. Put on oxygen[22]. Both parents are smokers[23]. Parents very friendly[24]. (dx: asthma)

Partial: 14/Unorganized

Sasha Logan[1], Most likely need surgery consult[2]. UA results pending[3]. Reflexes normal[4]. Abdominal pain increasing over few days[5]. Temp: 101[6]. Lungs sound clear[7].

Abdominal/pelvic CT ordered[8]. No prescription medication[9]. Tender to palpation right lower quadrant[10]. College student[11]. Heart Rate: 110[12]. Sexually active[13]. Lymph nodes not swollen[14]. Does not always use protection[15]. Not complaining of myalgias[16]. No rashes[17]. CBC ordered[18]. Has abdominal rebound on exam[19]. Urine pregnancy test negative[20]. Respiratory rate: 14[21]. No wheezes[22]. 25 years old[23]. Korean-American[24]. (dx: appendicitis)

Luis Santiago[1], No weight change[2]. Sinuses not tender[3]. Tried Tylenol cold with no improvement[4]. Heart rate: 85[5]. 20 years old[6]. Circles under eyes[7]. Normal reflexes[8]. Mexican American[9]. New to the area[10]. Normal bowel movements[11]. Runny nose[12]. CBC pending[13]. No abdominal pain[14]. Clear drainage from eyes[15]. No known sick contacts[16]. Coughing[17]. In law school[18]. Enjoys occasional glass of wine[19]. Throat itches[20]. Sleep pattern normal[21]. Complaining of red itchy eyes[22]. Has a sister in Seattle[23]. Sneezing[24]. (dx: seasonal allergies)

Hunter Ross[1], No change in diet[2]. Head CT ordered[3]. 68 years old[4]. No skin lesions on exam[5]. Wife very concerned[6]. LDH: 155[7]. Patient is combative[8]. Height: 5'9"[9]. Wife states his personality is changing[10]. Only drinks socially[11]. Weight: 158[12]. Forgetting names of familiar objects[13]. No change in diet[14]. Uncooperative during exam[15]. Retired mechanic[16]. Wants to go home[17]. Hearing normal[18]. History of high blood pressure[19]. BUN: 15[20]. Wife notes increasingly forgetful[21]. No abdominal pain[22]. Getting lost on familiar routes[23]. Need to administer mental status exam[24]. (dx: change in mental status)

APPENDIX V

PRACTICE CASES

Running Memory

Whole
Journal
Hornet
Vampire
Enjoy
Observe

Partial (Recall Last 3)
Mustang
Walnut
Escape
Gurgle
Story

Clinical Handoff

Whole (unorganized)
Chandler Ellis, EKG ordered. Chest pain. 56 years old. Past history of heart attack.

Partial (organized)
Mallory Walters, cute child, ankle swollen. Eyes clear. Xray ordered.

ATC Handoff

Whole (organized)
Ground control: American seven two eight taxiing to runway niner.

Partial (unorganized)
Flight Data: No SNOWTAM. Strong winds. No ground delays. Fog approaching.

APPENDIX W

CLINICAL PROPORTION ANOVA

Results of the Analysis of Variance for Clinical Proportion Handoff Scores

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	1.69	2	.85	20.79	.00***	.36
Error	2.97	73	.04			
Within Subjects						
List Length (L)	4.33	1.80	2.40	255.64	.00***	.78
L x E	.12	3.61	.03	3 .39	.01*	.09
Error (L)	1.24	131.67	.01			
Recall Length (R)	1.02	1	1.02	84.14	.00***	.54
R x E	.15	2	.08	6.13	.00**	.14
Error (R)	.89	73	.01			
Organize (O)	.02	1	.02	3.29	.07	.04
O x E	.04	2	.00	3.45	.04*	.09
Error (O)	.45	73	.01			
L x R	.19	1.80	.10	11.77	.00***	.14
L x R x E	.14	3.60	.04	4.47	.00**	.11
Error (L x R)	1.15	131.47	.01			
L x O	.06	1.99	.03	5.26	.00**	.07
L x O x E	.04	3.99	.01	1.90	.12	.05
Error (L x O)	.82	145.66	.01			
R x O	.16	1	.16	23.20	.00***	.24
R x O x E	.02	2	.01	1.14	.33	.03
Error (R x O)	.51	73	.01			

L x R x O	.26	1.83	.14	17.99	.00***	.20
L x R x O x E	.06	3.66	.02	2.22	.08	.06
Error (L x R x O)	1.05	133.58	.01			

Note. *$p < .05$; ***$p < .001$

APPENDIX X

AIR TRAFFIC CONTROL PROPORTION ANOVA

Results of the Analysis of Variance for Air Traffic Control Handoff Scores

Source	*SS*	*df*	*MS*	*F*	*p*	partial η^2
Between subjects						
Expertise (E)	1.69	2	.85	20.79	.00***	.36
Error	2.97	73	.04			
Within Subjects						
List Length (L)	5.94	2	2.97	392.15	.00***	.84
L x E	.56	4	.14	18.46	.00***	.34
Error (L)	1.11	146	.01			
Recall Length (R)	4.12	1	4.12	350.03	.00***	.83
R x E	.00	2	.00	.15	.86	.00
Error (R)	.86	73	.01			
Organize (O)	.00	1	.00	.12	.73	.00
O x E	.00	2	.00	.02	.98	.00
Error (O)	.44	73	.01			
L x R	.30	1.73	.07	25.45	.00***	.26
L x R x E	.03	3.46	.01	1.28	.28	.03
Error (L x R)	.86	126.25	.01			
L x O	.35	2	.17	37.32	.00***	.34
L x O x E	.10	4	.03	5.59	.00***	.13
Error (L x O)	.68	146	.01			
R x O	.18	1	.18	28.77	.00***	.28
R x O x E	.04	2	.02	3.13	.05*	.08
Error (R x O)	.45	73	.01			

L x R x O	.13	1.82	.07	10.98	.00***	.13
L x R x O x E	.03	3.64	.01	1.09	.36	.03
Error (L x R x O)	.89	132.88	.01			

Note. *$p < .05$; ***$p < .001$

VITA

Brittany L. Anderson-Montoya
Department of Psychology
250 Mills Godwin Building
Norfolk, VA 23529

Education

2009-2015	**PhD, Human Factors Psychology**, Old Dominion University, VA
2006-2008	**MS, Human Factors Psychology**, Old Dominion University, VA
2002-2005	**BS, Biology and Psychology,** Old Dominion University, VA
2001-2002	**Associate in Science,** Tidewater Community College, VA

Work Experience

2014-Current	**Human Factors Coach, Quality,** Carolinas Healthcare System
2006-2014	**Graduate Research Assistant**, Old Dominion University
2006	**Graduate Teaching Assistant**, Old Dominion University
2004-2006	**Research assistant**, Old Dominion University
2003-2005	**Research assistant**, Portsmouth Naval Hospital

Selected Publications

Anderson, B. L., Scerbo, M. W., Belfore II, L. A., & Abuhamad, A. Z. (2011). Time and number of displays impact critical signal detection in fetal heart rate tracings. *American Journal of Perinatology, 28,* 435-441.

Anderson, B. L., Scerbo, M. W., & Belfore II, L. A., Abuhamad, A. Z. (2009). Detecting critical patterns in maternal-fetal heart rate tracings over time. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 1121-1125). Santa Monica, CA: Human Factors and Ergonomics Society.

Anderson, B. L., Scerbo, M. W., Belfore, L. A., & Abuhamad, A. Z. (2009). When is a deceleration perceived as a late deceleration*? Simulation in Healthcare, 4,* 311.

Bustamante, E. A., Bliss, J. P., & **Anderson, B. L.** (2007). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics, 50*, 1127-1147.

Scerbo, M. W., & **Anderson, B. L**. (2012). Medical simulation. In P. Carayon (Ed.), *Handbook of human factors and ergonomics in health care and patient safety* (2nd ed.). Boca Raton, FL: Taylor and Francis Group.