Retention and Retraining of Independent and Integrated Cognitive and Psychomotor Skills Related to Laparoscopic Surgery

A DISSERTATION

Submitted to the Faculty of the

Department of Psychology

School of Arts and Sciences

Of The Catholic University of America

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

© Copyright

All Rights Reserved

By

Anna D. Skinner

Washington, D.C.

2013
Retention and Retraining of Independent and Integrated Cognitive and Psychomotor Skills Related to Laparoscopic Surgery

Anna Skinner, M.A.
Director: Marc Sebrechts, Ph.D.

Maintenance of specialized skills during periods of nonuse presents a significant challenge across multiple domains, and is most relevant within applications in which skill degradation is common and has significant negative consequences, as is the case in the domain of military medicine. For example, laparoscopic surgery (LS) skills often are subject to decay during deployments in which military surgeons primarily practice open surgical procedures; however, empirical research is needed to characterize the nature of such specialized skill decay. Specifically, needs exist for 1) objective and sensitive metrics to support assessment of LS skill loss based on the underlying cognitive, psychomotor, and perceptual skill components, and 2) validated pedagogical approaches to retraining these perishable skills. The results of two experiments are presented here. Experiment 1 investigated the use of task decomposition- and instrumented glove-based metrics to objectively and reliably assess acquisition and retention of a psychomotor skill relevant to LS following short-term (<6 weeks) and long-term (7 months) retention periods. In addition to detecting significant skill acquisition, as well as significant decay for both retention periods, these metrics demonstrated significant differences in psychomotor skill acquisition for the dominant and non-dominant hands and a significant increase in bilateral dexterity over the course of training, which were not detectable by traditional time and error metrics. Specifically, for portions of the task requiring more precision, the task decomposition-based metrics detected faster performance by the dominant hand early in
training, with this difference decreasing in the later training blocks. In addition, the glove-based metrics indicated significantly faster velocity and significantly greater angular rotation for the non-dominant hand across all training blocks, with non-dominant hand velocity increasing during later training blocks. This finding may indicate that larger, gross movements were performed primarily by the non-dominant hand while smaller, more precise movements and manipulations were primarily performed with the dominant hand. Experiment 2 assessed training and retention of the same psychomotor task used in Experiment 1, but also introduced a novel cognitive task and novel integrated task to empirically assess the relative retention of psychomotor and cognitive skill components within an integrated/concurrent task relevant to LS, as well as in isolation. Relative skill decay across the three tasks, following a 3-week retention period demonstrated a significant increase in mean trial time for the cognitive task (80%) and for the integrated task (40%), but no significant change in psychomotor mean trial time. Significant decreases in accuracy were also observed following the retention period. Specifically, the integrated task demonstrated a greater decrease in accuracy (35%) than the cognitive task (23%) or psychomotor task, which demonstrated a nonsignificant accuracy increase (3%) when performed in isolation, suggesting greater decay overall for the constituent skills within an integrated context. Following retention testing, Experiment 2 also assessed the comparative effectiveness of video-based retraining to physical (hands on) retraining of the integrated skill. Results indicated significant skill recovery for both the video-based and hands-on retraining groups based on mean trial time and cognitive subtask component accuracy, with no between groups differences detected. These results are discussed within the context of training strategies to detect and reduce LS skill decay.
This dissertation by Anna D. Skinner fulfills the dissertation requirement for the doctoral degree in Applied Experimental Psychology, approved by Marc Sebrechts, PhD, as Director, by Deborah Clawson, PhD, as Reader, by James Howard, Jr., PhD, as Reader, and Corinna Lathan, PhD, as reader.

Marc Sebrechts, PhD, Director

Deborah Clawson, PhD, Reader

James Howard, Jr., PhD, Reader

Corinna Lathan, PhD, Reader
Table of Contents

Background ...................................................................................................................... 1
Skill Acquisition and Expertise .................................................................................... 1
Factors Impacting Skill Retention .............................................................................. 4
Assessment of Laparoscopic Surgical Skill Decay .................................................... 8
Retraining Specialized Skills ...................................................................................... 14
Research Questions ................................................................................................... 19
Methods ....................................................................................................................... 21
  General Methods ....................................................................................................... 21
  Experiment 1 Methods .............................................................................................. 23
  Experiment 2 Methods .............................................................................................. 35
Results ......................................................................................................................... 55
  Experiment 1 Results ............................................................................................... 55
  Experiment 1 Discussion ......................................................................................... 86
  Experiment 2 Results .............................................................................................. 96
  Experiment 2 Discussion ......................................................................................... 134
General Discussion .................................................................................................... 145
  Research Question 1: Task Decomposition-Based and Glove-Based Metrics ....... 146
  Research Question 2: Relative Skill Decay ............................................................ 149
  Research Question 3: Retraining Condition ............................................................ 152
  Future Directions .................................................................................................... 153
Appendix A ................................................................................................................ 156
Appendix B ................................................................................................................ 165
Bibliography .............................................................................................................. 180
Acknowledgements

I wish to acknowledge the Office of Naval Research for supporting this critical research effort, and to express sincere gratitude to my employer, AnthroTronix, Inc., for supporting both my education and doctoral research studies. I also wish to express deepest appreciation to my committee for providing insightful guidance and feedback throughout my graduate education and the completion of my doctoral research. In particular, I wish to thank my academic advisor for his extreme dedication to the success of this research, and to my success as a student and an individual. I will be forever grateful for his devotion of time and inspired insights. Finally, I will forever owe a debt of gratitude to my husband for his endless support and encouragement.
Background

Maintenance of specialized skills during periods of nonuse presents a significant challenge across multiple domains, and is most relevant within applications in which skill degradation is common and has significant negative consequences (Arthur, Bennett, Stanush, & McNelly, 1998), as is the case in the domain of military medicine. For example, specialized laparoscopic surgery (LS) skills often are subject to decay during deployments in which military surgeons primarily practice open surgical procedures (Wetter, 2011; Perez, et al., in press). Currently, needs exist for military medical sustainment training and assessment technologies to predict, prevent, and detect skill decay during periods of nonuse throughout deployment cycles. Perez et al. (in press) identify a critical need for research and development to support prevention of surgical skill attrition; and outline an approach for identifying critical skills, developing objective metrics of skill decay, developing skill decay models, and developing refresher training within the context of LS skills. In particular, sustainment training is needed to prevent loss of specialized clinical skills during overseas deployments and to prevent loss of combat casualty care skills between deployments (Deering, Rush, Lesperance, & Roth, 2011).

Skill Acquisition and Expertise

In order to examine skill retention, it is critical to understand the process by which the relevant knowledge and skills were initially acquired. While multiple theories and models of skill acquisition have been developed (e.g., Fitts & Posner, 1967; Dreyfus & Dreyfus, 1980; Anderson, 1982; Rasmussen, 1983), the prevailing paradigms maintain several commonalities, including the existence of multiple stages of learning in which knowledge and skills typically
become increasingly proceduralized, generalizable, and durable (Anderson, 2010). Fitts and Posner’s (1967) model of skill acquisition provided the basis for subsequent models, and thus is representative of the fundamental concepts common across models. This model is comprised of three phases of skill development: cognitive, associative, and autonomous. The cognitive stage consists of initial encoding, and is characterized by a requirement for high levels of attentional focus, strategies such as verbal mediation, and frequent and poorly self-diagnosed errors. The associative stage represents a higher level of skill development in which inputs are linked more directly to appropriate actions, strategies such as verbal mediation are implemented less frequently and errors are more readily diagnosed and eliminated by the learner. The autonomous stage involves increasingly automated and error-free performance, requiring less attention, effort, and conscious control (Proctor & Dutta, 1995).

The theoretical and research literature within the domain of expertise acquisition builds on such models, specifying criteria for skill learning that counteracts arrested development and performance improvement plateaus associated with automaticity (Ericsson, 2006). Ericsson (2006) describes the progression to expert performance as remaining within a cognitive or associative stage of performance while achieving higher levels of performance with increased experience. This contrasts with the progression of everyday skill, which leads to an autonomous stage with increased experience. This process, depicted in Figure 1, requires among other things, increasingly complex and flexible mental representations of the task.
Ericsson has specified that expertise is typically acquired through extensive “deliberate” practice that is “specifically designed to improve performance” (p. 693). Ericsson has further defined deliberate practice as involving four primary elements: motivation, pre-existing knowledge, immediate informative feedback, and repeated performance. The role of deliberate practice in expertise acquisition has been specifically demonstrated within the medical domain. Ericsson (2004) provided evidence suggesting that enhanced performance of medical diagnostic and treatment skills requires not only extensive experience via repeated exposures, but also feedback. Surgical skills are specifically cited as benefitting from real time feedback and knowledge of results during task performance, thereby accelerating skill acquisition. As highlighted by Ericsson (2004), performance of LS skills has been shown to increase
dramatically during the first 10 procedures, followed by continued skill improvement with additional experience (Rosser, Rosser, & Savalgi, 1997). Furthermore, several studies have demonstrated improved intra-operative performance within the initial clinical cases for surgical trainees who first completed simulation-based training (Ahlberg et al., 2007), including the use of video box trainers such as the Fundamentals of Laparoscopic Surgery (FLS) platform (McCluney et al., 2007; Sroka, et al., 2010), which provides the task environment for the current research. These studies suggest that simulation-based training can be used to provide effective deliberate practice, enabling novice surgeons to progress through the initial stages of skill acquisition in which errors are common within a context in which patients won’t be harmed (Schaverien, 2010). For a detailed review of the skill acquisition and expertise literature, see Appendix A.

**Factors Impacting Skill Retention**

Extensive past research has identified a variety of factors shown to impact the retention of knowledge and skills. In a meta-analysis of general skill decay literature, Arthur et al. (1998) categorized these factors in terms of task variables and methodological variables. Task variables are related to characteristics of the task itself, and include factors such as task dimensions, complexity, and demands. Conversely, methodological variables relate to the design of training and testing conditions by which the relevant skills are assessed. Examples of methodological variables include aspects of skill acquisition and retention such as spacing of training (i.e., massed versus distributed), degree of overlearning, part versus whole task training, length of retention interval, and conditions of retrieval (e.g., recognition versus recall). For a detailed review of these factors, see Appendix B.
Arthur et al. (1998) also demonstrated that interactions exist between task and methodological variables such that the effects of methodological conditions may be largely dependent on the specific task or skill, and may be moderated by underlying task characteristics. For example, Healy et al. (1993) reported that the effectiveness of part task training depends on the nature of the whole task and its subcomponents. In particular, Healy et al. concluded that part tasks should be maximally trainable, such that they are not excessively trivial, but also able to be mastered. Similarly, a primary component of deliberate practice theory, as specified by Ericsson (2006), is the assertion that optimal skill acquisition occurs when learners train on individual skill components that are challenging, but that can be mastered within a short period of time.

Constituent skill components of a complex task may involve a combination of task variables, as is the case for LS skills, which involve cognitive, perceptual, and psychomotor skill components. Of particular relevance to the current study, Arthur et al. (1998) concluded that in the absence of rehearsal, cognitive skills exhibit more decay than "physical" skills, with the difference in decay being close to half a standardized unit across various retention intervals. Additionally, a more recent meta-analysis conducted by Wang, Day, Kowollik, Schuelke, and Hughes (2013), demonstrated that tasks involving primarily cognitive demands exhibit higher decay rates than tasks involving primarily physical demands. Furthermore, in some cases retention may be related to underlying or inherent task factors that support implicit versus explicit learning. For example, Romano, Howard, & Howard (2010) demonstrated implicit learning and long-term retention of sequence-specific procedural skills using a modified serial reaction time task (SRTT); reaction times were faster for trial blocks in which a systematic
pattern was present despite participants’ reports that no regularities or patterns were detected. This effect was demonstrated during initial training and also following a 1-year retention period for young motor skill experts (piano and action video game players), older motor skill experts (piano players), and older controls. These results indicate reliance on procedural, rather than declarative, knowledge to support long-term skill retention without rehearsal; and are also consistent with procedural reinstatement theory (Romano et al., 2010). Procedural reinstatement requires that the underlying skill-based procedures present during training (i.e., cognitive, perceptual, psychomotor) are replicated or reinstated during subsequent testing to support enhanced retention (Healy, Wholdman, & Bourne, 2005). These principles are particularly relevant within the context of LS skill training and reinstatement as LS skills are largely procedural in nature and also consist of underlying perceptual, cognitive, and psychomotor skill components that can be trained in isolation, but ultimately must transfer to integrated task performance when tested in an operational context.

As stressed by Arthur et al. (2010) the majority of skill decay research has examined cognitive skills and psychomotor skills in isolation, and rarely within the context of integrated or complex tasks. As research has demonstrated, principles established within the context of simple skills do not necessarily generalize to complex skill domains (Wulf & Shea, 2002). Therefore, a need currently exists for empirical research to support the development of validated models to accurately characterize complex skill acquisition and decay; such models could be used to accelerate expertise acquisition, and support prediction and mitigation of skill loss via appropriately administered refresher training.
Kim, Ritter, and Koubek (2011) have proposed an integrated theory of skill acquisition and retention based on existing three-stage models of learning (e.g., Fitts, 1964; Anderson, 1982; Rasmussen, 1986; and VanLehn, 1996), which could be used to predict skill decay. The integrated model described by Kim et al. emphasizes theoretical constructs that are common across these models of skill acquisition, including the existence of an initial learning stage dedicated to the acquisition of declarative and procedural knowledge, an intermediate stage during which knowledge is consolidated, and a final stage of learning in which knowledge is tuned and proceduralized. Within this proposed model, the nature of skill decay is dependent on the stage of learning at which retention occurs. If a period of nonuse begins during the first (declarative) stage of learning a procedural skill, both the declarative and psychomotor components of the task are subject to decay, potentially resulting in "catastrophic" failures (i.e., inability to perform the task) due to declarative memory breakdowns. Conversely, periods of nonuse occurring during later stages of learning, at which time skills have become proceduralized or automated, are less likely to result in such catastrophic failures as reliance on declarative knowledge is greatly decreased. For example, as described by Kim et al. (2011), in the early stages of typing skill acquisition, declarative knowledge of key location on the keyboard is utilized in addition to coordinated motor movements. However, with practice, this skill becomes increasingly proceduralized and automated, eventually relying on declarative knowledge only for infrequently used keys or keystroke combinations. This model suggests that in order to effectively study complex skills, particularly those involving multiple skill components within an integrated task or procedure involving multiple steps and decision points, it is necessary to study their acquisition and retention within an integrated or concurrent context.
Furthermore, it may be necessary to examine retention of complex skills within the context of periods of nonuse beginning at various stages of learning. The current study examines relative skill retention for both procedural and declarative knowledge in the early stages of learning integrated and isolated skills tasks, at which time declarative knowledge may be especially susceptible to decay.

**Assessment of Laparoscopic Surgical Skill Decay**

A crucial first step in addressing the issue of retention for a given skill is determining what type and magnitude of skill decay, if any, has occurred. Deering et al. (2011) demonstrated significant perceived degradation in both the surgical and clinical skills of military physicians deploying for more than 6 months. Self-rated surgical skills based on a survey including specialty, length of deployment, perceived changes in skills, skill use while deployed, and time to get back to baseline clinically after deployment were significantly \((p < .001)\) lower following deployment as compared to pre-deployment self-rated skills. Deering et al. (2011) report that most deployed Army clinicians felt that the time needed to get back to pre-deployment skill levels was one to six months. Of the clinicians surveyed, more than 70% were deployed for more than six months, and 59% reported that they used their specialties less than 40% of the time while deployed. The majority of surgical specialists felt that it took them three to six months to return to their clinical and surgical pre-deployment performance baseline, and that six months was the most amount of time that they could be deployed without a significant decrement in skills. While this subjective data confirms that providers perceive that many of their skills decay while deployed, further empirical research involving objective metrics is
needed to reliably determine rates of skill decay and reacquisition, and validated training tools are necessary in order to accelerate refresher training across a wide variety of medical skills. However, currently very little empirical data exist to accurately quantify the nature and degree of decay for various clinical and surgical skills, and additional research is needed to identify optimal methods of sustainment and refresher training to prevent skill decay and support rapid reacquisition of lost skills (Wetter, 2011).

While self-report data and expert accounts have indicated that LS skills decay rapidly during periods of non-use (Deering et al., 2011), previous research examining the retention of relevant psychomotor skills has indicated durability up to 12 months following training to proficiency (Stefanidis, Korndorffer, Markley, Sierra, & Scott, 2006; Castellvi et al., 2009), and up to 24 months when refresher training to proficiency is provided every 6 months (Mashaud et al., 2010). This suggests that current metrics for assessing LS psychomotor skills, which are based on the Fundamentals of Laparoscopic Surgery™ (FLS) program (Derossis et al., 1998), may lack sensitivity, and highlights the need for sensitive and objective metrics of skill and performance over time in order to detect subtle indicators of LS skill degradation.

The FLS program is currently the “gold standard” for training, assessment, and certification of LS skills in the United States. This didactic training and assessment protocol consists of declarative knowledge training materials and an associated computer-based test, as well as training of five manual skills tasks to specified proficiency levels using a video trainer box. The video box training platform, shown in Figure 2, consists of a collapsible plastic box containing physical inanimate training materials, a standard video monitor that displays real
time video from a camera inside the box, and laparoscopic training instruments such as graspers for manipulating the materials.

*Figure 2. FLS manual skills training platform.*

The cognitive and psychomotor components of LS skills are currently trained and assessed independently from one another, and no tasks are performed during training that involve integration of the cognitive components (e.g., declarative knowledge, steps in a procedural task, decision-making) within the psychomotor skills tasks. Additionally, the specific mechanisms by which the FLS psychomotor skills are acquired and decay are poorly understood and measured, including the role of the dominant versus non-dominant hand. Ambidexterity has been identified as a primary skill required for achieving proficiency in laparoscopic surgery (Rosser, Rosser, & Savalgi, 1997; Derossis, et al., 1998); however, the current FLS scoring methods and metrics have been limited to time and overt errors, and do not account for performance with the left versus the right hands.
An increasingly large body of research has demonstrated the effectiveness of motion tracking metrics for LS skill assessment (Mackay, Datta, Mandalia, Bassett, & Darzi, 2002; Smith, Torkington, Brown, Taffinder, & Darzi, 2002; Datta, Chang, Mackay, & Darzi, 2002; Bann, Khan, & Darzi, 2003; Aggarwal, Moorthy, & Darzi, 2004; Dosis et al., 2005). In a recent meta-analysis, Mason et al. (2013) demonstrated evidence to validate motion analysis for use in assessing intra-operative laparoscopic skills, including path length and number of hand movements in addition to task completion time. Additional evidence has been provided for motion analysis within the context of LS training, including metrics such as smoothness of movements. While electromagnetic and camera-based motion tracking systems are preferred within the operating theatre, data-capture gloves have been shown to provide an effective measure of LS proficiency within simulation-based training. Kahol et al. (2006) concluded that hand movement tracking is representative of economy of motion and overall smoothness in execution. Furthermore, it has been demonstrated that data-capture gloves can be worn during simulation-based training, and can provide an effective measure of surgical proficiency, including expert analysis of surgical profiles in minimally invasive surgery (Kahol, Satava, Ferrara, & Smith, 2009). However, the use of data capture glove metrics has not been applied to skill decay.

To address these issues, Aim 1 of this research investigates the use of task decomposition to assess individual hand performance. In addition, motion capture data gloves will be used to develop metrics for comparing LS psychomotor skills with the dominant and non-dominant hands, especially in the context of LS psychomotor skill attrition. Extending psychomotor skill assessments beyond the traditional FLS manual skills metrics of total task
completion time and overt errors, to include decomposition of psychomotor skills and motion analysis, will enable assessment of subtasks, bimanual skills, and ambidexterity over time. It is anticipated that these more granular assessment metrics will provide greater sensitivity to detection of psychomotor skill acquisition, retention, and decay.

In addition to psychomotor skill decay assessments, perhaps an even greater need exists for objective assessments of relevant cognitive and perceptual skills within the LS domain as higher rates of decay have been indicated for cognitive skills over physical skills (Arthur et al., 1998; Wang et al., 2013). Furthermore, Perez et al. (in press) highlight the need for objective and reliable metrics to support assessment of LS skill loss based on the underlying cognitive, psychomotor, and perceptual skill components required for clinical and surgical competency, both independently and in conjunction with one another.

Complex skills have been shown to exhibit differential acquisition and decay characteristics from simple skills, and empirical research investigating appropriate training and retention paradigms related specifically to complex skills is lacking (Arthur et al., 2010). LS represents a complex task requiring integration of underlying psychomotor, cognitive, and perceptual skill components within procedures involving multiple steps and decision points. Psychomotor skills involved in the performance of laparoscopic surgical procedures include bimanual dexterity, eye hand coordinate, and fine motor control for precise manipulation and transfer of tissues and tools using laparoscopic instruments constrained by limited degrees of freedom (Cao, MacKenzie, & Payandeh, 1996; Rosser et al., 1997; Wetter, 2011). Requisite cognitive skills include declarative information recall, procedural knowledge, and decision-making, as well as skills and abilities relevant to communication and situation awareness such
as working memory capacity. Perceptual skills integral to LS are both visuo-spatial and perceptual-motor. Visuo-spatial skills include depth perception and discrimination, recognition, interpretation, and transformation of 2-dimensional visual information while working within a 3-dimensional physical space (Stefanidis et al, 2006). Relevant perceptual-motor skills within LS include haptic or tactile perception (Singapogu, 2012) and proprioception within a context in which a mismatch exists in visuomotor mapping between the visual image and the hands (Cao et al., 1996). Aim 2 of the current research is to assess the relative acquisition and retention of psychomotor and cognitive/perceptual skill components within an integrated task relevant to LS, as well as in isolation.

While previous studies have examined the retention of laparoscopic psychomotor skills in isolation (Stefanidis, Korndorffer, Markley, Sierra, & Scott, 2006; Castellvi et al., 2009; Mashaud et al., 2010), there is a paucity of research examining retention of these skills within the context of an integrated cognitive/perceptual and psychomotor task. Stephanidis, Scerbo, Montero, Acker, and Smith (2012) examined the long-term retention and transfer effects of training a laparoscopic psychomotor suturing skill to a specified level of automaticity by employing a secondary cognitive/perceptual skills task; however, this secondary task was performed in a concurrent, but not integrated, manner. Additionally, the secondary task consisted of monitoring a visual display indicating when a simple pattern emerged. Thus, the secondary task required constant monitoring of the simple visual stimuli and constant working memory load while performing the primary (suturing) task; however, the secondary task was not integrated within the primary task in such a way that delays in processing the secondary task would inherently impact performance on the primary task. For example, participants could
choose to favor one task at the expense of the other. Furthermore, the task used by Stefanidis et al. (2012) did not require participants to memorize more than one visual pattern, which would have required procedural knowledge and processing, and did not require participants to remember that pattern for retrieval during later retention assessments; this task purely served the purpose of increasing cognitive load and establishing when automaticity had been achieved on the primary task. Stefanidis et al. demonstrated increased retention and transfer to an operating room task following training to automaticity as compared to proficiency-based training. However, the time required to train to automaticity is substantial and may not be justified unless the secondary task also contributes to LS cognitive or perceptual skill learning as training time for these specialized skills is at a premium.

Retraining Specialized Skills

There is currently a dearth of both theoretical and empirical research to support scientifically-grounded retraining pedagogy within a domain related to LS, and as yet a validated methodology has not been developed for designing effective retraining of complex LS skills. In a report summarizing the outcomes of a gap analysis workshop for training for reintegration of surgical skills, Wetter (2011) emphasized that despite a recognized need, currently no formalized Department of Defense program exists to support reentry of military surgeons returning from deployments. The issue of retraining complex, high-risk skills following extensive periods of nonuse has been examined within the context of military aviator retraining, particularly with Individual Ready Reserve (IRR) pilots (Allnutt & Everhart, 1980; Wick, Millard, & Cross, 1986; Hendrickson, Goldsmith, & Johnson, 2006). However, these
retraining curricula focus primarily on identification of minimum flight hours required to bring aviators back up to contact checkride standards. This level of training has some parallels within the surgical domain, but does not sufficiently translate to the design of refresher training of LS skills.

Currently, refresher training of previously mastered LS skills typically consists of self-directed retraining, self-elected enrollment in hands on courses or “mini” fellowships, and self-sought opportunities for dual-attending, scrub/assist, or mentored surgical cases (Wetter, 2011). This method of retraining may be inefficient and ineffective since it relies on self-assessment of skill degradation and retraining assessment. A specific need exists for appropriate cognitive and perceptual skills sustainment training and assessment related to specialized surgical skills such as LS. The development of effective retraining must draw on both theoretical and empirical data. The procedural reinstatement principle may help to address the key components of retraining. The principle indicates that underlying cognitive, perceptual, and motor components must be the same between training and testing to maximize skill transfer (Healy et al., 1992). It has been applied and validated within the context of initial training and retention, but may also play a critical role in the design of refresher training, to prevent skill loss, and in retraining pedagogy following the loss of skills. Thus, an empirical question that warrants research concerns the extent to which each of the constituent components of an integrated or complex task must match the original learning context during subsequent refresher training. Furthermore, given the variable rates of decay for skills involving multiple task characteristics (e.g., physical versus cognitive) (Arthur et al., 1998; Wang et al., 2013), procedural reinstatement may be more critical for less durable skills. For example, cognitive skill
components may require closely-matched retraining elements, while psychomotor skill components may be reinforceable using more abstract means.

Simulation-based training has been shown to provide significant benefits for skill learning across a wide variety of domains, and in particular within domains in which the costs or risks associated with live training are exceedingly high (Skinner et al., 2010), as is the case within the surgical domain. Returning to Ericsson’s definition of deliberate practice, simulation provides all of the key elements necessary for providing deliberate practice: the game-like nature of simulations provides intrinsic motivation; embedded assessment and structured scaffolding can be used to tailor training to the individual learner’s pre-existing knowledge and skills; embedded feedback can be provided via a number of modalities; and in addition to providing repeated performance, simulation offers the opportunity to introduce a wide variety of training scenarios within a short period of time, including those that may be rarely encountered in live training.

Simulation-based training may also provide opportunities for various kinds of retraining, which may not require the specific matching suggested by procedural reinstatement. In particular interest within the current research is the question of whether integrated skills can be retrained via video-based refresher training, which has the potential to provide high fidelity cognitive skills content, but lacks physical interaction fidelity. Video-based training has been shown to be effective for a number of relevant skills, including LS perceptual/cognitive skills such as anatomical structure recognition (Guerlain, et al., 2004). Additionally, computer-based video instruction (CBVI) has been shown to enhance LS skill retention (Xeroulis, et al., 2007), but has not been studied within the context of retraining. If CBVI can be demonstrated to
provide an effective means of refresher training, supporting maintenance of relevant LS psychomotor, cognitive/perceptual, and integrated skills, the findings could be leveraged to provide low cost training on a wide variety of platforms, including mobile platforms suited for military surgical deployment settings. Aim 3 will examine the reacquisition of integrated psychomotor and cognitive/perceptual skills, and will include the use of video-based instruction, which can be delivered on a mobile platform such as a tablet.

The current research effort seeks to address several critical gaps in the skill decay and retraining research literature, particularly within the context of LS skills. It seeks to specifically explore the relationships between cognitive and psychomotor skill components within an integrated task, and in particular to explore the way in which the composite and constituent skills decay during a period of nonuse occurring within the early stages of learning. While the scope of the current research study is constrained to examination of this construct within novice learners in the general population, this research effort lays the groundwork for future, planned studies designed to extend the current findings to a more specialized population of surgical trainees as they progress to higher levels on the novice to expert skill acquisition continuum. Future work will also explore the extent to which the current study outcomes generalize to the retention and retraining of related skills by experts in the surgical domain following training to mastery.

In addition, the results of the current research are applicable within the context of medical procedural skills for which practitioners may not be trained or experienced at the expert level, and further, may have limited opportunities for skill rehearsal following initial training. An example of a skill with such limited rehearsal is cardiopulmonary resuscitation (CPR) and
basic life support skills. Extensive study of these skills has demonstrated extraordinarily poor retention rates by medical and lay trainees in the absence of rehearsal or employment of these skills (Hamilton, 2005; Wayne et al., 2006). CPR and basic life support skills comprise both psychomotor (e.g., administering chest compressions) and cognitive skills; including declarative, procedural, and decision-making components. Therefore, a deeper understanding of the way in which integrated and constituent medically-relevant skills decay and are subsequently reacquired by non-experts can have important real-world consequences.
Research Questions

Three Specific Aims have been identified, targeting three distinct research questions within the framework of a laboratory-based laparoscopic surgery (LS) relevant task. Aim 1 of this research is to investigate the use of task decomposition-based and instrumented glove-based metrics to assess individual hand performance (time and errors) for a standardized LS manual skills training task using an inanimate box training platform. The specified metrics will be used to compare LS psychomotor skills with the dominant and non-dominant hands, in order to assess LS psychomotor skill acquisition and attrition. Aim 2 is to develop a better understanding of the retention of integrated skill components relevant to LS. Using a modified FLS manual skill task that includes a cognitive skill component, Aim 2 will test the hypothesis that the cognitive skill component will demonstrate significantly more decay than the psychomotor skill component following a retention period, and that significantly more relative skill decay will be observed for the integrated task as compared to the psychomotor task or cognitive tasks in isolation. Finally, Aim 3 will assess the comparative effectiveness of video-based instruction to physical retraining of the integrated LS skill; it is hypothesized that both physical retraining and video-based retraining will lead to significant skill recovery, with greater recovery of skills in the physical retaining group.

Thus, the following research questions will be addressed: 1) Do task decomposition-based and hand motion-based metrics provide more sensitive means than traditional metrics of total task completion time and overt errors for assessing skill acquisition and decay within the context of LS psychomotor task performance? 2) For the specified task paradigm, does the cognitive skill component of the integrated task demonstrate significantly more decay than the
psychomotor skill component following a brief (3 week) retention period, and does significantly more relative skill decay occur for the integrated task as compared to the psychomotor or cognitive subtasks in isolation? 3) Do both hands-on retraining and video-based retraining lead to significant skill recovery, and does hands-on retraining result in greater recovery of skills?
Methods

General Methods

Experiments 1 and 2 involved Institutional Review Board (IRB)-approved human subjects protocols. IRB approval was obtained from The Catholic University of America (CUA) IRB following full reviews of each protocol. Participants for each experiment were recruited through university-wide flyers, a Psychology Department bulletin board, and in-class and web announcements. Participants were compensated with cash, psychology research points, or a combination of the two. All participants were assigned a subject identification number, and all questionnaires and data files were stored using this number; individual identification was used to match codes across sessions, but was not otherwise stored with the data.

Given that the current research sought to address basic research questions regarding skill acquisition, retention, and reacquisition, but within the context of LS skills; the tasks selected are relevant to LS skills, but do not require medical or surgical knowledge. Specifically, it has been demonstrated in the research literature that the selected psychomotor tasks translate to LS performance in the operating room, but that they can also be learned, and even mastered, by undergraduate university students (Stefanidis et al., 2012). Both experiments used the Fundamentals of Laparoscopic Surgery™ (FLS) standardized training protocol (Derossis et al., 1998) and associated video box training platform, targeting fundamental manual skills relevant to LS. Specifically, a peg transfer task, which is the first of five psychomotor tasks that make up the manual skills portion of the FLS standardized training protocol, was used in both Experiments (see Figure 3). The Experiment 1 training and retention sessions also included data
collection for a suturing task (following peg transfer); however, this task was not included in Experiment 2, and is not included within the scope of the current research study.

Figure 3. FLS peg transfer task

The peg transfer peg board consists of 12 black wooden pegs arranged on a white plastic board, which is affixed to the bottom of the inside of the FLS box via Velcro to prevent it from moving. As shown in Figure 3, the wooden pegs are arranged with six on one side in an oval shape, and six on the other side in a rectangular shape; the peg board can be placed in the FLS box with either wooden peg configuration on the left or right side. While these shapes bear no relevance to task performance within the standardized FLS training and credentialing protocols, for consistency, in both Experiments each trial began with the rubber triangles on the rectangular side of the board (which could be positioned on the right or on the left) as shown in Figure 3.

Both experiments also included hand motion analysis. Participants wore a commercial off-the-shelf (COTS) instrumented glove on each hand, the Acceleglove™, shown in Figure 4, which automatically captures hand movement data based on embedded inertial sensors sewn into
the glove. The sensors, located on top of each finger and the back of the hand, are small and unobtrusive. Additionally, the gloves do not cover the tips of the fingers, and thus do not impede tactile sensation in the fingertips during task performance. A previous assessment of the Acceleglove™ by an experienced laparoscopic surgeon (based on number of cases performed) indicated that the form factor does not interfere with performance of the FLS manual skills tasks (J. Buller, M.D., F.A.C.O.G., personal communication, November 5, 2010).

Figure 4. Acceleglove instrumented data capture gloves.

Experiment 1 Methods

Experiment 1 began to address the first research question by assessing FLS psychomotor skill acquisition and retention following short-term (<6 weeks) and long-term (~7 months) retention periods, including assessment of the role of ambidexterity in FLS manual skill acquisition and retention based on task decomposition and instrumented glove-based metrics for the left and right hands, independently.

Participants in Experiment 1 included a total of 27 undergraduate students (9 male, 18 female). Of these, 24 were right-handed, 2 were left-handed, and 1 was ambidextrous with a
left-hand preference. This distribution of handedness is representative of the general population, of which approximately 10% of individuals have been reported to be left-handed and 1% are ambidextrous (Hardyck & Petrinovich, 1977). Handedness, per se, was not a variable of interest within the current study. The assessment and potential acquisition of ambidextrous performance of the task across all participants was the primary variable of interest; furthermore, while statistical analyses were conducted to ensure that the left hand dominant and ambidextrous participants did not introduce systematic variability within the data, the primary analysis of interest revolved around comparison of performance with the dominant versus non-dominant hands across all participants.

All participants completed a total of 5 hours of initial training over the course of two training sessions, on separate days. The second training session was performed within 6 weeks of the original training session for each participant; the mean time between sessions across all subjects was 23 days (range = 5 – 42 days) with a normal, unskewed distribution across participants ($p>.05$ for Shapiro Wilkes test of normality). Figure 5 provides a histogram to demonstrate this distribution.
Figure 5. Histogram representing frequency of retention period between Session 1 and Session 2.

The large range and lack of retraining sessions occurring at or near the mean number of days was due to a number of scheduling issues encountered, including failure of participants to respond to follow-up scheduling requests, participants missing appointments, and constraints of the academic calendar. Despite these scheduling issues, which impacted the length of retention period, all participants who completed Session 1 (N = 27) also completed Session 2 within the specified 6-week timeframe.

While these issues resulted in an undesirably large range of inter-session intervals across participants, no participants were excluded from the data analysis, and additional analyses were conducted (see results section), which indicated that the length of retention period between Sessions 1 and 2 had no systematic effect on performance during Session 2. Specifically, a 2-
tailed bivariate correlation was conducted, which indicated that the length of retention period between Sessions 1 and 2 did not significantly correlate to changes in task completion time between the last trial of Session 1 and the Session 2 retention test $r(25) = -.01, p = .951$.

Of the 27 participants who completed Sessions 1 and 2, a total of 15 participants also completed a follow-on skill assessment and retraining session (Session 3), following a retention period of approximately 7 months. This length of retention period was selected in order to emulate a typical military medical deployment, accounting for the typical 180-day deployment and 30-day pre-deployment family leave, training and travel time to assigned location (Wetter, 2011). This retention period did not account for mandatory post-deployment transition, reintegration, family leave time, and time prior to starting clinics and booking elective cases, which typically results in total time away from clinical cases of 250 days (Wetter, 2011). The ultimate goal of the current research effort and planned subsequent research is to develop skill decay assessment and retraining to be administered during or immediately following deployment, prior to surgeons’ return to clinical practice. Therefore, post-deployment time was not included in the designed long-term retention period, and the target date for the Session 3 retention test was set for 210 days following Session 2 for all returning participants. The mean time between Sessions 2 and 3 across all returning participants ($n = 15$) was 223 days (range = 182 – 260 days) with a normal, unskewed distribution across participants ($p>.05$ for Shapiro Wilkes test of normality). Figure 6 provides a histogram to demonstrate this distribution.
Figure 6. Histogram representing frequency of retention period between Session 2 and Session 3.

The large range was due to a number of scheduling issues encountered, including failure of participants to respond to follow-up scheduling requests, participants missing appointments, and constraints of the academic calendar. Despite these scheduling issues, the resulting distribution provided data points representing a typical military surgical deployment of 180 days, as well as typical pre- and post-deployment time, resulting in total time away from clinical practice of up to 280 days. Additionally, to account for the large range of inter-session intervals across participants, a 2-tailed bivariate correlation was conducted, which indicated that the length of retention period between Sessions 2 and 3 did not significantly correlate to changes in task completion time between the last trial of Session 2 and the Session 3 retention test, $r(13) = .14, p = .623$. Furthermore, correlations were conducted comparing these participants to those who did
not return for a third session based on demographic and task performance factors in order to determine whether the Session 3 participants systematically differed from participants having only completed Sessions 1 and 2. The only significant correlation detected was for handedness, \( r(27) = -0.40, p = .041 \), due to the fact that the three participants having a left-handed preference did not return for Session 3. Given that these participants represented a small subset of the overall sample and given the high rate of overall attrition, this was not surprising. Additionally, attempts to schedule Session 3 appointments revealed that the left-handed participants had left the university; the ambidextrous participant with a left-handed preference was simply unresponsive to follow-up emails. The following sections provide detailed descriptions of the experimental methods for each session.

**Session 1.** Participants first completed an informed consent form, acknowledging the risks and benefits of participation prior to participating, and the experimental procedures were explained. Each participant completed a demographic questionnaire, which included questions regarding frequency and proficiency of video game play and instrument playing skill, as well as self-reported handedness. Analyses were conducted to assess any potential correlations between task performance and gender, handedness, video game skill, instrument-playing skill. A detailed summary of these analyses is provided in the Session 1 Results section; however, no significant correlations were found. Participants also responded to a series of questions from the Annett Hand Preference Questionnaire (Annett, 1970) and completed a simple assessment task to confirm self-reported hand dominance.
Participants were given detailed instructions for completing the FLS peg transfer task, including a scripted description of the task and a video demonstration in order to ensure consistency in instruction. Participants were encouraged to ask questions if further clarification was required. Participants then stood in front of the FLS video box training platform and rehearsed the peg transfer task for a total of 60 minutes, divided into three 20-minute training blocks (Block 1, Block 2, and Block 3), with brief breaks between blocks to prevent boredom or fatigue effects. The number of trials completed in each block varied based on the average time to complete each trial. Participants with slower trial times completed fewer trials within each 20-minute block, and participants generally completed fewer trials in their first training block than in subsequent blocks in which their individual trial times improved. On average, participants completed 5-10 trials within the first training block, and 10-15 trials within subsequent blocks.

The FLS peg transfer task requires trainees to lift six rubber triangles, one at a time, from wooden pegs on one side of a peg board using a laparoscopic grasper instrument in one hand, transfer the rubber triangle in midair to a grasper instrument held by the other hand, and place the rubber triangle on a peg on the opposite side of the peg board. Once all six rubber triangles have been transferred to the opposite side of the board, they are then moved back, one at a time, to pegs on the original side of the board. As shown in Figure 3, the rubber triangles consist of three orange and three green triangles; however, no importance is assigned to the colors of the triangles, which is made clear within the instructional video. Participants were randomly assigned to begin with the triangles in a rectangular array on the side of either their dominant or non-dominant hand, with half of the participants beginning each trial by transferring triangles from their dominant side to their non-dominant side, and half of the participants beginning each
trial by transferring from their non-dominant side to their dominant side. All participants completed transfers in both directions on each trial and completed an equivalent number of transfers in each direction with each hand within each training block; the only difference between the randomly assigned groups was the direction in which each participant began each individual trial. The two left-handed participants and one ambidextrous subject with a left-hand preference were randomly assigned to either the dominant or non-dominant starting direction groups as well. Two were randomly assigned to the group beginning each trial with the triangles on their non-dominant (or non-preferred) side, and thus began each trial with the triangles on the right side of the peg board, and one was randomly assigned to the group beginning each trial with the triangles on their dominant side, and thus began each trial with the triangles on the left side of the peg board. Analyses were conducted to assess any potential correlations between task performance and randomly-assigned starting side (dominant or non-dominant side). A detailed summary of this analysis is provided in the Session 1 Results section; however, no significant correlations were found. For consistency, all trials began with the triangles in a rectangular, rather than circular, array formation. The two possible starting positions can be seen in Figure 7.
Figure 7. Sample left starting position and sample right starting position.

This counterbalanced design controlled for potential learning effects resulting from completion of the first segment of the first trial within the first training block during each data collection session. Participants were instructed to complete the task as quickly as possible without dropping the rubber triangles. A trained researcher recorded time and errors (i.e., dropped triangles) according to the FLS standardized scoring and assessment protocol. In addition, the researcher recorded individual times and errors for the first segment of the task (i.e., moving the triangles to one side of the board) and the second segment of the task (i.e., moving the triangles back to the original side of the board). These task decomposition-based metrics were developed for the purposes of this study in order to enable assessment of comparative task performance with the dominant and non-dominant hands over time. The first trial of Block 1 served as a baseline test for task performance in both directions (moving the triangles to the opposite side of the board and then back to the original side).
Instrumented gloves were worn for all trials in order to provide additional metrics of task performance for each hand. The existing data acquisition software for the gloves required approximately one minute to start and stop between each trial. Thus, in order to prevent the participants from having to pause for 1 minute between trials, potentially interrupting task flow and skill acquisition and reducing the total training time and number of training trials, a method was devised in which participants were required to raise both hands in the air for 1-2 seconds after completing each trial. This motion, when performed properly created a distinct signature in the data files, which could be used to separate the files during post-processing.

Following peg transfer training, all participants in Experiment 1 also completed training on an FLS suturing task; however, this task was not included in Experiment 2, and is not included within the scope of the current research study.

Session 2. Upon returning for Session 2, all participants were first required to watch the same instructional video viewed in Session 1 as a reminder of task procedures, and all participants were invited to ask questions if further clarification was needed. Participants then stood in front of the FLS video box training platform and performed one task trial in order to assess skill retention between sessions for performance in both directions (moving the triangles to the opposite side of the board and then back to the original side). Again, a trained researcher recorded time and errors (i.e., dropped triangles), including individual times and errors for the first segment of each trial (i.e., moving the triangles to one side of the board) and the second segment of each trial (i.e., moving the triangles back to the original side of the board). As in Session 1, participants were instructed to complete the task as quickly as possible in both
directions, without dropping triangles. All participants also completed a single trial of the FLS suturing task; again, this task was not included in Experiment 2, and is not included within the scope of the current research study.

Participants then rehearsed the peg transfer task for a total of 40 minutes, divided into two 20-minute training blocks (Block 4 and Block 5), with brief breaks between blocks to prevent boredom or fatigue effects. Each participant began each trial by moving the triangles from the same side of the peg board to which they were assigned to begin all trials in Session 1 (dominant or non-dominant side). As in Session 1, this counterbalanced design was used to control for potential learning effects resulting from completion of the first segment of the first trial within the first training block. Also, as in Session 1, instrumented gloves were worn for all trials, and participants were required to raise both hands in the air for 1-2 seconds after completing each trial in order to create a distinct signature in the data files that could be used to separate the files during post-processing. Following peg transfer training, all participants again completed an FLS suturing task training; however, this task was not included in Experiment 2, and is not included within the scope of the current research study.

Session 3. Upon returning for Session 3, all participants were first required to watch the same peg transfer instructional video viewed in the previous sessions as a reminder of task procedures, and all participants were invited to ask questions if further clarification was needed. Participants then stood in front of the FLS video box training platform and performed one task trial in order to assess skill retention between sessions for performance in both directions (moving the triangles to the opposite side of the board and then back to the original side). Again,
a trained researcher recorded time and errors (i.e., dropped triangles), including individual times and errors for the first segment of each trial (i.e., moving the triangles to one side of the board) and the second segment of each trial (i.e., moving the triangles back to the original side of the board). As in the previous sessions, participants were instructed to complete the task as quickly as possible in both directions, without dropping triangles. All participants also completed a single trial of the FLS suturing task; again, this task was not included in Experiment 2, and is not included within the scope of the current research study.

Participants then stood in front of the FLS video box training platform and rehearsed the peg transfer task for a total of 40 minutes, divided into two 20-minute training blocks (Block 6 and Block 7), with brief breaks between blocks to prevent boredom or fatigue effects. Each participant began each trial by moving the triangles from the same side of the peg board to which they were assigned to begin all trials in Sessions 1 and 2 (dominant or non-dominant side). As in the previous sessions, this counterbalanced design was used to control for potential learning effects resulting from completion of the first segment of the first trial within the first training block.

Also, as in Sessions 1 and 2, instrumented gloves were worn for all trials. New glove data acquisition software was available and was employed for Session 3 data collection, enabling individual glove data trials to be separated without requiring participants to raise their hands between trials and also enabling individual trial time data to be recorded in the glove data files. However, due to a miscommunication, a research assistant did not properly indicate the beginning and end of individual retraining trials. Therefore, while time and error data were recorded for all 15 retention tests, and the primary researcher recorded retraining trial time, error,
and glove data for 7 participants, the assistant did not properly record retraining trial time, error, and glove data for the remaining 8 participants.

While Session 3 participants retrained on the peg transfer task following the initial retention test to assess potential skill decay, the primary trial of interest for analysis was the retention test trial.

**Experiment 2 Methods**

The objective of Experiment 2 was to build on Experiment 1 in order to address all three research questions. This experiment used a combination of task decomposition-based and glove-based metrics to assess acquisition and retention of the same psychomotor task (FLS peg transfer), an added cognitive/perceptual skills task, and an integrated task. It also assessed the comparative effectiveness of video-based retraining to physical retraining of the integrated skill.

Experiment 2 specifically sought to empirically assess the relative retention of psychomotor and cognitive/perceptual skill components within an integrated/concurrent task relevant to LS, as well as in isolation. Unlike a dual task paradigm in which a primary and secondary task are performed simultaneously but with task switching, the integrated task designed for Experiment 2 consisted of a psychomotor skill component and a cognitive/perceptual skill component that can be performed in an interdependent concurrent manner as well as in isolation. Additionally, the cognitive/perceptual skill component requires long-term retention of procedural knowledge in addition to adding cognitive load and a working memory component to the integrated task during skill acquisition and retention testing.
**Pilot testing.** Pilot testing was conducted with five participants, recruited under the Experiment 2 IRB-approved protocol. Pilot testing was completed with the primary goal of assessing the appropriateness of the level of difficulty for the Cognitive and Integrated tasks in order to determine whether these tasks could be learned and performed to criterion specifications within the time allotted for Session 1. A secondary goal was to determine whether the specified retention period of three weeks was sufficient to induce partial, but not complete, skill loss on the Cognitive and Integrated tasks as these tasks were not included in Experiment 1; since they were novel tasks, previous studies could not be drawn on to establish prediction of skill decay rates. All five pilot participants completed the experiment according to the same protocol specified below for the primary experiment cohort, with two exceptions: 1) one pilot participant was assigned a 4-week retention period, rather than 3 weeks in order to assess whether a longer retention interval might be optimal, and 2) one pilot participant did not complete Session 2 as she was a research assistant (RA) assigned to assist with data collection following her Session 1 data collection. Data for pilot participants was used only to confirm the experimental protocol; these data were not included in experimental results analyses.

**Experiment enrollment and design overview.** A new cohort of participants was recruited and enrolled in Experiment 2. Based on a conservative power analysis (set at .95), which indicated a required sample size of 36 participants, a total of 40 undergraduate students (20 male, 20 female) were recruited and enrolled in the study in order to account for attrition rates common within studies involving multiple data collection sessions. Of these, two were left-handed, with the remaining 38 being right hand dominant. As in Experiment 1, the Psychomotor
task (peg transfer) was performed by transferring the six rubber triangles (three orange, three green) from one side of a peg board to the other and then back again using laparoscopic instruments (graspers); each triangle is picked up with a grasper in one hand, transferred to the grasper in the other hand, and then placed on a peg with the second hand. Participants were trained on this task according to the FLS standard protocol, in which no importance is given to the color of the triangles.

Participants were then trained to create a series of six patterns of virtual orange and green triangles using a simple touchscreen tablet app (Figure 8), allowing rehearsal of pattern creation without using the graspers or requiring transfer of the objects between hands (Cognitive Task).

![Figure 8. Tablet-based Cognitive (pattern formation) task.](image)

Participants were shown a series of six patterns formed from the colored triangles on a virtual peg board via a PowerPoint slideshow, and were instructed to memorize the individual patterns as well as the order in which they were to be formed. Training procedures and performance criteria are detailed below. Finally, participants completed the integrated task, in which they were required to arrange the colored triangles in the six color patterns learned on the tablet while performing the FLS peg transfer task using the laparoscopic graspers.
Figure 9 provides a graphical representation of the Session 1 and Session 2 procedures, including indicators of when criterion-level performance was required for each task, as well as when glove-based metrics were recorded and when retraining was completed with half of the participants (n=20) randomly assigned to each retraining condition, video or hands on.

![Graphical representation of experiment design](image)

**Figure 9.** Experiment 2 design.

Criterion levels were established during the experiment design phase, and the number of required trials to reach criteria was then confirmed by pilot testing. As shown in Figure 9, multiple criterion-level trials were required for inclusion in the data analysis to ensure that sufficient training of the individual and integrated skills was achieved prior to the retention
period. The following provides a detailed description of experimental procedures for each session.

**Session 1.** Upon arriving for the initial data collection session, participants first completed an informed consent form and were permitted to ask questions about the experimental procedure. Participants completed a brief color vision test to rule out color blindness, verbally responded to a series of questions from the Annett Hand Preference Questionnaire (Annett, 1970) and completed a simple assessment task to assess hand dominance. No participants were found to have color vision impairment.

**Task 1: Psychomotor task.** Participants were given detailed instructions for completing the FLS peg transfer task, including a scripted description of the task and a video demonstration in order to ensure consistency in instruction. Participants were encouraged to ask questions if further clarification was required. Participants then stood in front of the FLS video box training platform and rehearsed the peg transfer task for a total of 40 minutes, divided into two 20-minute training blocks (Block A and Block B), with a brief break between blocks to prevent boredom or fatigue effects. As in Experiment 1, the instructional video indicated that no importance is assigned to the colors of the triangles for the peg transfer task. In Experiment 2 participants were again randomly assigned to begin with the triangles in a rectangular array on the side of either their dominant or non-dominant hand, with half of the participants beginning each trial by transferring triangles from their dominant side to their non-dominant side, and half of the participants beginning each trial by transferring from their non-dominant side to their dominant
side in order to control for learning effects resulting from completion of the first segment of the first trial within the first training block. All participants completed transfers in both directions on each trial and completed an equivalent number of transfers in each direction with each hand within each training block; the only difference between the randomly assigned groups was the direction in which each participant began each individual trial. The two left-handed participants were randomly assigned to either the dominant or non-dominant starting direction groups as well. In this case, both were randomly assigned to the group beginning each trial with the triangles on their dominant side, and thus began each trial with the triangles on the left side of the peg board. Figure 9 provides a sample image of triangles placed in the left starting position, and an image of triangles placed in the right starting position. Again, the ordering of the colored triangles on each side of the board was irrelevant for this task; however, each trial began with the triangles in a rectangular formation, rather than a circular formation, for consistency.

A trained researcher recorded time and errors (i.e., dropped triangles) for both training blocks according to the FLS standardized scoring and assessment protocol. The researcher also recorded individual times and errors for the first segment of the task (i.e., moving the triangles to one side of the board) and the second segment of the task (i.e., moving the triangles back to the original side of the board). Block A was conducted in the same manner as Blocks 1-5 of Experiment 1, without feedback from the researcher regarding performance. Prior to beginning Block B, all participants were told the following:

“You will complete the same task for another 20 minutes and then we’ll switch to the next task. Again, your goal is to complete the task as quickly as possible without dropping any triangles. This time, after each trial I will tell you your time as I write it down. To
give you a frame of reference, when surgeons complete their training and certification on this task, they are required to transfer the pegs from one side of the board to the other and back in 48 seconds."

The researcher then said aloud the total time for each trial upon its completion throughout Block B. This was done to provide motivation to the participants to continue to improve their speed and accuracy.

As indicated in Figure 8, instrumented gloves were worn for all trials, and a minimum of two of the last five Block B trials were required to be performed at criterion-level for inclusion in the final data set used for statistical analysis of results. The criterion level set for this task was double the time required for a perfect FLS test score by surgeons seeking FLS credentialing (this time is different from the 48-second criterion used during FLS training, and the exact time was provided to the researchers for the purposes of this study, but must be kept confidential and therefore is not reported here) with no triangles dropped. Additionally, the same time criterion was required for the Psychomotor task PostTest. However, no accuracy criterion (i.e., dropped triangles) was specified for PostTest performance as this type of error within a single trial may not be indicative of overall skill level, and the accuracy assessment for the last five trials of Block B served the purpose of demonstrating that the participants were capable of error-free task performance. The first trial of Block A served as a baseline test for task performance in both directions (moving the triangles to the opposite side of the board and then back to the original side).
Task 2: Cognitive task. Following a brief break, participants were seated at a table with a computer monitor and shown a series of PowerPoint slides, which were advanced by the researcher. The first slide, shown in Figure 10, provided a brief description of the task to be completed and provided an indication of expectations for criterion-level performance.

![Pattern Formation Task](image)

*Pattern Formation Task*
- You will memorize 6 orange/green patterns
- You will practice creating them on the tablet

*You must complete all 6 in order from memory at least 2 times*

Figure 10. Cognitive task introduction slide.

A tablet mobile computing device (Motorolla Xoom) with the pattern formation custom application (app) was then placed on the table in front of the participant. Two distinct versions of the app were developed. In both versions, the orange and green triangles were arranged in a starting position on the rectangular-shaped portion of the virtual peg board, as was the case for the starting positions used for the physical (rubber) triangles in the peg transfer task on the FLS box trainer. Participants who were randomly assigned to begin the Psychomotor Task with the triangles on the left side (as shown in Figure 7), also began the Cognitive Task with the virtual triangles on the left side, as shown in Figure 11. Participants who were randomly assigned to
begin the Psychomotor Task with the triangles on the right side (as shown in Figure 7), also began the Cognitive Task with the virtual triangles on the right side, as shown in Figure 11.

![Diagram of starting patterns](image)

**Figure 11.** Cognitive task left starting pattern (Left) and Cognitive task right starting pattern (Right).

As indicated in Figure 11, participants were instructed that the first pattern, is the starting pattern, which was provided for them and did not require memorization. The remaining six patterns were presented systematically and in a consistent order a total of 12 times, and participants were required to practice creating the patterns with gradually reduced cues for support. Slides providing an overview of the task procedures (Figure 12) and instructions for using the tablet interface to complete the task (Figure 13) were also presented.
Participants then performed three trials of each portion of the Cognitive task (Part A, Part B, Part C, and Part D) according to the procedures laid out in the task overview slide (Figure 12). For Part A, each pattern to be formed was visible on the computer monitor in front of the participant while the participant formed the pattern on the tablet. All six patterns were presented in order a total of three times for Part A (Figure 14). In Part B, the pattern to be formed was visible on the computer monitor for just one second, and then automatically advanced to a blank screen using PowerPoint automated slideshow timing settings. The blank screen remained on the computer monitor until the participant completed forming the target pattern. After each pattern formation, participants were shown the correct (target) pattern again and were given time to correct any mistakes in the pattern formed. Only times to create the first attempt at each pattern were recorded; pattern accuracy, but not time to correct mistakes was recorded for scoring purposes. Part B consisted of three complete trials (all six patterns) performed in this manner (Figure 15).
Part C consisted of three complete trials in which all six patterns were formed one at a time from memory, with the correct pattern displayed after each attempt (Figure 16). Part D required participants to form all six patterns, one after another, from memory, only pausing between patterns to view the correct pattern if an incorrect pattern was formed (Figure 17).

Part A:
The pattern will be visible on this screen while you create the pattern on the tablet
Try to learn the patterns and memorize the order of the patterns as you go
You will do this 3 times
You will have to tap the timer, but don’t worry about time
Say “Next” when ready

Part B:
The pattern will be visible on this screen for a second and then will disappear while you create the pattern on the tablet
You will do this 3 times
You will have to tap the timer and I will write down your time, but focus more on learning the patterns for now
Say “Next” when ready

Part C:
The pattern will NOT be visible on this screen at all.
You will create the patterns one at a time from memory
You will do this 3 times
You will have to tap the timer after completing each pattern
You will pause briefly so that I can write down the time before starting the next pattern

Part D:
The pattern will NOT be visible on this screen at all.
You will create all six patterns from memory
You will do this 3 times
You will have to tap the timer after completing each pattern
You will pause briefly so that I can write down the time before starting the next pattern

Figure 14. Part A instructions slide.  
Figure 15. Part B instructions slide.

Figure 16. Part C instructions slide.  
Figure 17. Part D instructions slide.
Criterion performance for inclusion in the final data set used for statistical analysis of results for this task required at least two of the three Part D trials to be performed without any errors in pattern formation. Error-free performance was also required for the PostTest. No time criterion was set for this task as the emphasis was on correct pattern formation. Time to complete each pattern and the number of incorrect color assignments (0, 2, 4, or 6) were recorded for each trial.

Identical but opposite (mirror image) patterns were presented for each group of participants based on their random assignment to starting side. The six patterns assigned to participants starting with the triangles on the left side are provided in Figure 18. The six patterns assigned to participants starting with the triangles on the right side are provided in Figure 19.
Figure 18. Patterns 1-6 for left starting position.
Figure 19. Patterns 1-6 for right starting position.

Upon completion of the last trial of Part D of the Cognitive task, the slide shown in Figure 20 was displayed on the computer monitor, indicating that following a brief break the participants would next be required to form the patterns that they had just learned on the tablet.
using the colored rubber FLS peg task triangles while completing the peg transfer task on the FLS box training platform.

**Integrated Skill Task**

- You will now take a quick break
- You will then create the patterns you've just learned while you complete the Peg Transfer Task on the FLS box

*Figure 20.* Integrated task introduction slide.

**Task 3: Integrated task.** For the Integrated task, participants again stood at the FLS box and were given the laparoscopic grasping instruments used to complete the Psychomotor Task while wearing the instrumented gloves. The researcher arranged the colored rubber triangles on the peg board in the same starting position used for the Cognitive task for each participant based on random assignment. Participants then proceeded to form all 6 trained patterns, in order, from memory by transferring the rubber triangles from one side of the board to the other; this required using the grasper held by one hand to pick up each triangle, transferring each triangle to the grasper held by the other hand, and then using the second grasper to place each triangle on the other side of the board in the same manner employed for the Psychomotor task. After each pattern was formed, the correct pattern was displayed on the tablet, which was visible to the participant. Incorrect patterns were rearranged by participants prior to proceeding to subsequent
pattern formation; however, only times to create the first attempt at each pattern were recorded; pattern accuracy, but not time to correct mistakes, was recorded for scoring purposes.

Criterion performance for inclusion in the final data set used for statistical analysis of results for this task required all 3 trials to be performed without any errors in pattern formation. No time criterion was set for this task as the emphasis was on correct pattern formation during initial skill acquisition. Participants were not informed of criterion specifications. Time to complete each pattern and the number of incorrect color assignments (0, 2, 4, or 6) were recorded for each trial.

*Psychomotor and cognitive task post-tests.* Lastly, a post-test was completed, consisting of a single trial of the Psychomotor task (peg transfer with no importance assigned to colors or patterns) and a single trial of the Cognitive task on the tablet (all 6 patterns). Criterion-level total trial time was again required for the Psychomotor task for inclusion in the final data set for analysis. However, no accuracy criterion (i.e., dropped triangles) was specified for Psychomotor task PostTest performance as this type of error within a single trial may not be indicative of overall skill level, and the accuracy assessment for the last five trials of Block B served the purpose of demonstrating that the participants were capable of error-free task performance. Criterion-level (error-free) task performance was again required for the Cognitive task PostTest in order to demonstrate that the Cognitive task had been sufficiently learned prior to the retention period. This post-training trial was conducted as a test to be used for statistical analyses comparing performance at various time points, within and across subjects; however, the researcher referred to this post-test as simply a “final trial” of the tasks to mitigate the potential
for anxiety-induced performance variations resulting from participants’ awareness of being tested.

**Session 2.** Upon arriving for the second data collection session, participants first completed a demographic questionnaire. Questions included those related to video game and musical instrument expertise as these skills have been shown to correlate to enhanced perceptual-motor skills within other domains involving visual perception and attention, as summarized in a review by Green and Bavelier (2008). Participants also completed two spatial abilities tests: matching to sample (Figure 21) and mental rotation (Figure 22). The matching to sample task consists of presentation of a single 4x4 checkerboard pattern for a brief study period, followed by presentation of a blank display for 5 seconds, and then two patterns presented side by side; participants indicate which of the two patterns matches the first as quickly as possible. This test has been purported to assess spatial memory and pattern recognition (Englund, Reeves, Shingledecker, Thorne, & Wilson, 1987). In a factor analytic study assessing the construct validity of the matching to sample test with a population of healthy controls, Bleiberg, Kane, Reeves, Garmoe, and Halpern (2000) found that the test loaded on a factor with a test of verbal learning and memory (i.e., the Hopkins Verbal Learning Test). A second factor analytic study using a clinical population demonstrated that matching to sample loaded on a factor with other tests presumed to measure processing speed and cognitive efficiency, suggesting that this test may be more closely related to measures of complex attention than memory/retention (Kabat et al., 2001).
The mental rotation task assesses the ability to mentally rotate a line drawing (Shepard & Metzler, 1971), and presents pairs of perspective line drawings in which the second drawing is either different from the first or is the same shape but rotated 20-180 degrees in either the picture-plane or depth dimensions. Participants indicate whether the second drawing is the same or different from the first.

Figure 21. Matching to sample test.  
Figure 22. Mental rotation test.

**Integrated, Psychomotor, and Cognitive task retention tests.** The retention test consisted of a single complete trial of each task, beginning with the Integrated task, in order to assess retention of the composite and constituent skill components. The Integrated task was assessed first as the emphasis of the current research is to develop a deeper understanding of the manner in which complex and integrated skills, such as laparoscopic surgical procedure skills, decay with nonuse. While retention testing of the integrated task inherently provided some practice and feedback on the constituent psychomotor and pattern formation tasks, this was determined to be preferable to providing practice on the individual tasks prior to assessing
retention of the integrated task skills. The Integrated task retention test consisted of formation of each of the six previously memorized patterns on the FLS box. After each pattern was formed, the correct pattern was displayed on the tablet, which was visible to the participant. Incorrect patterns were rearranged by participants prior to proceeding to subsequent pattern formation; however, only times to create the first attempt at each pattern were recorded; pattern accuracy, but not time to correct mistakes, was recorded for scoring purposes.

The Psychomotor task retention test consisted of a single trial in which participants transferred the 6 rubber triangles on the FLS platform to the opposite side of the peg board and then back again with no importance given to the color of the triangles or their arrangement on either side of the board.

The Cognitive task retention test consisted of formation of each of the six patterns, in order, from memory using the tablet. Incorrect patterns were rearranged by participants prior to proceeding to subsequent pattern formation; however, only times to create the first attempt at each pattern were recorded; pattern accuracy, but not time to correct mistakes, was recorded for scoring purposes.

**Integrated task retraining.** All participants then retrained on the Integrated task in either the same manner employed during the initial training session (n=20) for three trials, or by watching a video demonstrating correct formation of each pattern on the FLS box (n=20) a total of three times. The videos were embedded in a PowerPoint™ slide show, which the participants in the video retraining group viewed by clicking through the slides in a self-paced manner. Figure 23 provides a sample screenshot of the video-based retraining slide show for participants randomly assigned to begin trials with the triangles on the right side of the board.
Figure 23. Video based retraining slide show screenshot for participants randomly assigned to begin trials with the triangles on the right side of the board.

Integrated final test. Lastly, all participants then completed one final trial of the Integrated Task using the same method employed for the retention test. This post-retraining trial was conducted as a test to assess relative training improvement within and across the groups; however, the researcher referred to this post-test as simply a “final trial” of the integrated task to mitigate the potential for anxiety-induced performance variations resulting from participants’ awareness of being tested. Participants again wore the Acceleglove™ instrumented gloves (Figure 4) throughout performance of all Session 2 Psychomotor and Integrated task trials.
Results

Experiment 1 Results

The primary objective of the analyses within Experiment 1 was to assess skill acquisition and retention for the selected psychomotor task (FLS peg transfer) for both the dominant and non-dominant hands using task decomposition-based metrics and instrumented glove-based metrics. The task decomposition-based metrics initially identified included comparative analyses of task time and errors (dropped triangles) to determine the relationship between these task performance metrics, as well as decomposition of task trials into segments led with the dominant versus non-dominant hands. While the peg transfer task represents a bimanual psychomotor skill requiring coordinated dexterity between the two hands, particularly when transferring triangles from one grasper to the other, each trial consists of a portion that is led by each hand, meaning that the triangles are picked up using the grasper in that hand. The analyses conducted sought to determine whether differences in task performance exist for the portion of the task that involves picking up the triangles from one side of the board as compared to the trial segments that involve placing the triangles on the other side of the board. Additionally, these analyses sought to determine whether handedness plays a role in performance of the constituent trial segments. The objective was not to compare right-handed to left-handed individuals, but rather to compare performance with the dominant and non-dominant hands within and across individuals. This research objective was inspired by previous reports of the role of ambidextrous task performance by experts (Rosser et al., 1997; Derossis, et al., 1998), as well as unpublished pilot study research indicating a potential trend for
increased ambidexterity correlated to higher overall time-based task performance within the context of FLS manual skills.

The use of instrumented glove-based metrics sought to build on the task decomposition-based analyses by providing additional, objective measures of relative task performance between the dominant and non-dominant hands. The glove-based metrics selected for the current study include two existing metrics: velocity and angular rotation. The Acceleglove™ collects raw x, y, and z-axis acceleration data from a small, inertial measurement sensor sewn into the gloves on the back of the hand. The raw values are converted to measures of velocity and angular rotation via existing algorithms. Velocity is calculated by taking the integral of the raw acceleration values, providing raw rate of position change data for each axis in meters per second (m/s). Given the small space in which the Psychomotor and Integrated tasks were performed within the FLS box, the resulting mean velocity values have been converted to centimeters per second (cm/s) to provide a more relevant metric. Angular rotation calculates the overall change in x, y, and, z angles from the original (starting) hand position in degrees, providing a measure of overall flexion/extension, radial/ulnar deviation, and pronation/supination of the hands during task performance. Figure 24 provides an illustration of these ranges of motion.
Sessions 1 and 2. Analysis of both the task decomposition-based metrics and glove-based metrics for Sessions 1 and 2 included assessment across all five of the 20-minute training blocks (three in Session 1 and two in Session 2) for the peg transfer task. The following provides a summary of analyses across both sessions for the task decomposition-based metrics, followed by a summary of analyses across both sessions for the glove-based metrics for all 27 participants for Experiment 1.

Task decomposition-based metrics. Data for all trials were decomposed into 2 segments for each trial: transfer of the triangles from the dominant hand side of the board and from the non-dominant hand side of the board. For each trial segment, both time and accuracy
(number of triangles dropped) were assessed. It is worth noting here that both performance data and subjective report data from a previous (unpublished) pilot study indicated that removing triangles from wooden pegs is easier (i.e., requires less precision) than placing triangles on wooden pegs, and also that holding and releasing triangles during the mid-air transfer is easier (i.e., requires less precision) than grasping triangles during the transfer portion of the task. Thus, moving triangles from the dominant hand side of the board represents the trial segment in which the non-dominant hand is performing the more difficult portions of the task. Figure 25 provides sample images demonstrating a trial segment in which a triangle is transferred from the left side of the peg board to the right side of the board. For the majority of participants (i.e., right-handed participants), the example shown in Figure 25 represents a trial in which the non-dominant hand is performing the easier portions of the task (removing a triangle from a wooden peg and holding/releasing the triangle during the transfer), with the dominant hand performing the more difficult portions of the task (grasping the triangle during the transfer and placing it on a wooden peg).
Figure 25. Sample peg transfer task trial in which the left grasper is used to remove a triangle from a wooden peg on the left side of the board (left image), the triangle is transferred to the right grasper (center image), and the right grasper is used to place the triangle on the right side of the board (right image).

Furthermore, while each hand is required to remove triangles from pegs and place triangles on pegs independently using only the grasper in that hand (without assistance from the other grasper), the portion of the task involving transfer of the triangles from one grasper to the other represents a bimanual skill. The bimanual portion of the task requires coordination of the dominant and non-dominant hands and represents the portion of the task in which performance with one hand could be used to compensate for performance with the other hand. The task decomposition-based metrics were designed such that timing for the first trial segment begins when the first grasper comes in contact with the first triangle; this is consistent with the FLS official certification testing protocol and enables the trainee to begin each trial when ready. Following transfer of all six triangles, timing for the first trial segment ends when the sixth (last) triangle is released by the second grasper after being placed on a wooden peg on the opposite side of the board. Timing for the second trial segment begins immediately in order to
maintain the continuous task flow required within FLS certification standards and to ensure that total trial time is not increased by the use of decomposition-based metrics. Following transfer of all six triangles back to the original side of the board, timing for the second trial segment ends when the sixth triangle is released by the first grasper after being placed on a wooden peg on the original side of the board. Therefore, the decomposition-based timing metrics differentiate between trial segments in which each hand performs the more difficult portions of the task.

The official FLS scoring metrics for the peg transfer task consist of normalized calculations based on a set cutoff time (300 seconds) from which time to perform the task and any penalties (16.7 seconds per triangle dropped outside the camera view) are subtracted (Fraser et al., 2003). Therefore, only if a triangle is dropped outside the view of the camera is the official FLS score impacted by anything other than time. Previous research has demonstrated that the official FLS scoring metrics are no more effective than total trial time alone for assessing LS skill level, and that the two measures are nearly perfectly correlated (Kowalewski, 2012). Within the standardized FLS training and certification protocol, triangles dropped within view of the camera can be retrieved; however, retrieval can be time consuming, adding to total trial time, and can interrupt task flow. While surgeons performing this task for purposes of certification are not explicitly penalized for triangles dropped within view of the camera during the manual skills portion of the certification exam, the time and expense of certification dictates that surgeons seeking certification train to high levels of proficiency prior to the exam, likely reducing the incidence of such errors.
Within the context of the early stages of skill acquisition for this task, triangles are dropped more frequently as trainees acquire proper instrument manipulation techniques. It stands to reason that such manipulation errors are indicative of poor instrument handling performance and are associated with poor LS skill. A primary objective of this research study was to explore the use of more granular metrics based on decomposition of the task. Therefore, while the FLS scoring methodology only penalizes trainees for triangles dropped outside of view of the camera, the current study explored the assessment of triangles dropped overall was examined as a potential means for providing additional indicators of skill acquisition and subsequent decay. In order to first determine the relationship between time and errors (triangles dropped) for the peg transfer task, bivariate correlations were conducted for a sampling of individual trial segments, composite trials (including performance of both the dominant and non-dominant segments for a given trial), and for means derived from entire blocks of training trials. In particular, the sampling of individual trials selected included the first trial from Block 1 and last trial from Block 3 (last trial of Session 1). The sampling of blocks selected included the mean for all Block 1 trials and the mean for all Block 2 trials. Table 1 provides a summary of individual trial correlations assessed, and Table 2 provides a summary of blocks assessed. Given the consistent and highly correlated nature of these factors for both blocks and individual trials during the early stages of skill acquisition, no further trials and blocks were assessed for correlation within later training sessions. As additional analyses reported within this research effort will demonstrate, the incidence of dropped triangles decreases over the course of training, and thus is even less of a concern within later training sessions.
### Table 1

**Sample Correlations for Peg Transfer Mean Trial Times and Mean Triangles Dropped Per Trial**

<table>
<thead>
<tr>
<th>Block/Trial</th>
<th>Hand</th>
<th>$r(25)$</th>
<th>$p^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1, Trial 1</td>
<td>Dominant</td>
<td>0.64</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 1, Trial 1</td>
<td>Non-dominant</td>
<td>0.79</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 1, Trial 1</td>
<td>Combined</td>
<td>0.70</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 3, Last Trial</td>
<td>Dominant</td>
<td>0.54</td>
<td>0.004</td>
</tr>
<tr>
<td>Block 3, Last Trial</td>
<td>Non-dominant</td>
<td>0.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 3, Last Trial</td>
<td>Combined</td>
<td>0.55</td>
<td>0.003</td>
</tr>
</tbody>
</table>

*Note: *All significant at the $p<0.01$ level.

### Table 2

**Sample Correlations for Peg Transfer Mean Block Times and Mean Triangles Dropped Per Block**

<table>
<thead>
<tr>
<th>Block/Trial</th>
<th>Hand</th>
<th>$r(25)$</th>
<th>$p^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Dominant</td>
<td>0.62</td>
<td>0.001</td>
</tr>
<tr>
<td>Block 1</td>
<td>Non-dominant</td>
<td>0.79</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 1</td>
<td>Combined</td>
<td>0.70</td>
<td>0.000</td>
</tr>
<tr>
<td>Block 2</td>
<td>Dominant</td>
<td>0.61</td>
<td>0.001</td>
</tr>
<tr>
<td>Block 2</td>
<td>Non-dominant</td>
<td>0.56</td>
<td>0.002</td>
</tr>
<tr>
<td>Block 2</td>
<td>Combined</td>
<td>0.63</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Note: *All significant at the $p<0.01$ level.

The results of these analyses indicate that task time is highly correlated to number of triangles dropped. This was as expected, since recovering dropped triangles is time-consuming and should contribute to overall task time. However, previous reports of such a correlation were
not found in the existing research literature, warranting investigation of this relationship. Based on these results, it was determined that triangle drops do not provide a significantly more sensitive metric than time alone. Therefore, the remaining performance analyses for Experiment 1 focused only on task time for the decomposed trial segments.

In order to rule out unintended systematic effects for demographic and methodological variables within the task performance data, 2-tailed bivariate correlations were conducted to assess relationships between task performance and the following demographic and methodological variables: gender, hand preference, frequency of video game play, instrument playing skill, time between Session 1 and Session 2, number of sessions completed (two or three), and hand used to start each trial (based on random assignment). These variables were correlated to individual trial segment mean times and total trial mean times for Session 1 and Session 2, as well as to the mean difference between Session 1 and Session 2 trial segments and total trial times. Gender correlated significantly to frequency of video game play, \( r(25) = -.64, p < .001 \), with male participants reporting greater frequency of video game play. Gender was also significantly correlated to time between Session 1 and Session 2, \( r(25) = -.44, p = .021 \), with overall shorter retention periods for female participants. Gender did not significantly correlate to any task performance indicies. Self-reported instrument playing skill was also significantly correlated to time between Session 1 and Session 2, \( r(25) = .43, p = .027 \), such that participants reporting greater instrument playing ability had overall longer retention periods. As previously reported, handedness was significantly correlated to number of sessions completed due to the fact that the three left-hand preferred participants did not return for Session 3;
handedness (right-versus left-hand preference) was not significantly correlated to any other factors.

Additionally, in order to further assess any potential effects of time between sessions, a 2-tailed bivariate correlation was conducted comparing the relationship between the number of days from Session 1 to Session 2 and the difference in total trial completion time for the first trial of Session 2 and the last trial of Session 1 for each participant. The total trial time for the Session 1 last trial was subtracted from the total trial time for the Session 2 first trial time for each participant. These values are plotted (y-axis) against the value for the number of days between sessions (S1 and S2) for each participant (x-axis) in Figure 26. This correlation did not approach significance, $r(25) = -.01, p = .951$.

**Figure 26.** Scatterplot representing the relationship between number of days from Session 1 training to Session 2 training (x-axis) and the mean total trial time difference between the last Session 1 trial and the first Session 2 trial (y-axis).
As shown in Figure 26, there was no significant trend present for trial time increase or decrease with respect to increased inter-session interval. Some participants from both ends of the retention period spectrum demonstrated slower times following the retention period (i.e., positive y-axis values) and some demonstrated faster times following the retention period (i.e., negative y-axis values). Therefore, no participants were excluded from further statistical analyses based on length of retention between Sessions 1 and 2.

Lastly, correlations for hand used to start each trial (based on random assignment) were conducted in order to rule out the possibility of a systematic effect for first trial segment versus second trial segment across participants. Starting timing for the first trial segment once the first grasper is already in contact with the first triangle could potentially provide a slight time advantage for the first trial segment over the second, although the impact of this potential effect was controlled for in the data by the fact that half of the participants were randomly assigned to begin each trial on their dominant side while the other half began each trial on their non-dominant side. Additionally, bivariate correlations were conducted in order to identify any significant relationships between starting hand and task performance. Starting hand did not significantly correlate to individual trial segment mean times or total trial mean times for Session 1 or Session 2; additionally, starting hand did not significantly correlate to the mean difference between Session 1 and Session 2 trial segment or total trial times. This result confirmed that further analyses could be conducted without differentiating between participants based on starting hand.
In order to assess performance across all five training blocks (Sessions 1 and 2), a 5 (Block) x 2 (Hand) repeated measures analysis of variance (RM ANOVA) was conducted for mean trial time. Levels for Hand included dominant and non-dominant, indicating the trial segment based on which hand led each trial. For example, the dominant level for Hand indicates trial segments in which triangles were moved from the dominant hand side of the board to the non-dominant hand side of the board. Again, this represents the trial segment in which the non-dominant hand is required to perform the subtasks requiring more precision (grasping triangles during transfers and placing triangles on pegs). Mean times for both the dominant and non-dominant hands across all 5 training blocks are presented in Figure 27.

![Figure 27. Mean peg transfer times, in seconds, for trial segments led with the dominant and non-dominant hands across training blocks for Session 1 (Blocks 1, 2, and 3) and Session 2 (Blocks 4 and 5).]
Mauchly’s test of sphericity was significant for both Block and Block x Hand; therefore, the Greenhouse-Geisser correction is reported for the F value, p value, and degrees of freedom for these effects. A significant main effect was found for Block, \( F(2, 50) = 30.60, p < .001 \), with the slowest mean trial segment time for Block 1 (\( M = 102.24 \)), followed by Block 2 (\( M = 60.68 \)), Block 4 (\( M = 60.62 \)), and Block 3 (\( M = 54.57 \)), with the fastest mean time for Block 5 (\( M = 49.50 \)). A simple contrasts post hoc analysis was conducted, comparing Block 1 to each subsequent block, in order to assess performance changes over time relative to the beginning of training. The Bonferroni adjustment was used to account for familywise error. This analysis resulted in significant contrasts between Block 1 and all four subsequent blocks, with a significance level of \( p < .001 \) for all contrasts. Additionally, a planned post hoc repeated contrasts analysis was conducted, comparing each block to the subsequent block, in order to determine whether significant differences were present between all individual training blocks within sessions and also to determine whether a significant difference was present between the last block of Session 1 and the first block of Session 2 during which the retention period took place. This analysis was again conducted using the Bonferroni adjustment for familywise error. In addition to the significant contrast reported between Block 1 and Block 2, this analysis also revealed a significant contrast between Block 2 and Block 3, \( F(1,26) = 9.94, p = .004 \). However, significant contrasts were not present between Block 3 and Block 4, which reflects a mean performance comparison between the last block of Session 1 and the first block of Session 2, or between Block 4 and Block 5, representing the Session 2 training blocks.
A significant main effect was also found for Hand, $F(1, 26) = 7.57, p = .011$, with the average time for trial segments moving the triangles from the non-dominant side of the board being significantly faster ($M = 69.37$) than the average time for trial segments moving the triangles from the dominant side of the board ($M = 73.69$). This was anticipated based on previous reports and the task decomposition-based conclusion that more precision is required to grasp the triangles during transfer and to place the triangles on a peg than to hold and release the triangles during transfer and to remove the triangles from a peg. No significant interaction effect (Block x Hand) was present, indicating that changes in mean trial segment times across blocks were statistically similar for the dominant and non-dominant hands.

It is important to note that the repeated contrast for the main effect of Block between Block 3 and Block 4 reflects the lack of a statistically significant difference for the mean trial segment times across the entire blocks, and thus does not indicate that no skill loss was present between the end of Block 3 (Session 1) and beginning of Block 4 (Session 2). Additionally, the comparisons of training blocks did not provide insight into performance changes occurring within each block, which is of interest from a skill acquisition perspective. In order to assess changes in performance occurring within each training block, as well as between training sessions, during which the trained skills were subject to decay, a 4 (Trial) x 2 (Hand) RM ANOVA analysis was conducted comparing the first and last trials for each Session (1 and 2). The 4 levels for Trial were Session 1 First Trial, Session 1 Last Trial, Session 2 First Trial, and Session 2 Last Trial. The 2 levels for hand were dominant and non-dominant. Means
for trial segments led by the dominant and non-dominant hands for each trial are presented in
graph form with standard error indicated in Figure 28.

![Figure 28](image)

**Figure 28.** Mean time in seconds for trial segments led with the dominant and non-
dominant hands for the first Session 1 trial (S1First), last Session 1 trial (S1Last), first
Session 2 trial (S2First), and last Session 2 trial (S2Last).

A significant main effect was found for Trial, $F(3, 78) = 56.32, p < .001$, with the mean
time for Session 1 First Trial being slowest ($M = 148.77$), followed by Session 2 First Trial ($M$
$= 67.39$), and Session 2 Last Trial ($M = 60.31$), with the mean time for Session 1 Last Trial
being fastest ($M = 58.46$). Tests of post hoc repeated contrasts for Trial were conducted using
the Bonferroni adjustment in order to assess changes in performance from the first to last trial of
each training session, as well as between the last trial of Session 1 and the first trial of Session
2, during which a retention period occurred. This analysis indicated significant contrasts for
Session 1 First Trial versus Session 1 Last Trial, $F(1, 78) = 104.35, p < .001$, as well as for Session 1 Last Trial versus Session 2 First Trial, $F(1, 26) = 6.63, p < .001$, but not for Session 2 First Trial versus Session 2 Last Trial. While no significant main effect was found for Hand, nor was a significant interaction effect found for Trial x Hand. These results indicate a significant decrease in trial time at the end of Session 1 training compared to the beginning of Session 1 training, indicating performance improvement over the course of the session. These results also indicate a significant increase in trial time at the beginning of Session 2 training compared to the end of Session 1 training, indicating performance degradation over the course of the retention period. Additionally, the lack of a significant contrast between the first and last trials of Session 2 indicate that performance at the end of Session 2 was not significantly better than at the beginning of Session 2. Finally, these results indicate that while mean times for trial segments moving the triangles from the dominant hand side of the board to the non-dominant hand side of the board across all trials in a Block were slower overall than times for trial segments moving the triangles from the non-dominant hand side of the board to the dominant hand side of the board, individual trials did not reflect this difference in every instance.

In an effort to further explore the role of ambidexterity in the acquisition of fundamental laparoscopic psychomotor skills, and in particular the relationship between increased equivalence in dominant and non-dominant hand performance and overall task performance, ambidexterity scores were calculated for each participant by individual training session and compared to the mean total trial time for each participant by session. Ambidexterity scores were calculated based on the absolute time difference between the dominant and non-dominant
hands. A 2-tailed bivariate correlation was then conducted for the initial training block in Session 1 (Block 1), as well as for the final block in Session 2 (Block 5); these data are plotted in Figure 29 with a linear trend line. The mean total trial times and mean ambidexterity scores across all subjects were significantly correlated for both Block 1, $r(25) = .395$, $p = .042$, and Block 5 $r(25) = .590$, $p = .001$.

![Figure 29](image-url)

Figure 29. Correlations between mean calculated ambidexterity score and mean total trial time in seconds for Block 1 (left) and Block 5 (right).

The results of this analysis indicate that overall trial time tended to be faster for trials in which the time to perform the trial segment moving triangles to the non-dominant hand side of the board was similar to the time to move the triangles to the dominant hand side of the board. This has to be the case to some extent due to the fact that trials with shorter overall trial times inherently have a shorter maximum potential time difference, as well as the fact that there is a limit to how quickly the task can be completed with either hand, and thus a limit to the potential
time difference for shorter overall trials. However, it is also possible for the task to be performed equally slowly with both hands, in which case the total trial time would be large, while the ambidexterity score would indicate equivalent performance for the two trial segments. Contrary to this potential explanation, the results indicate that equivalent performance for the two trial segments is associated with better overall task performance.

Finally, median trial time values were also examined to control for outliers and non-normal distributions typical of timed tasks in which speed accuracy tradeoffs may be present. Overall, the trends of significance were the same for both the mean and median analyses.

**Glove-based metrics.** All participants wore instrumented gloves for all trials. As detailed in the Experiment 1 methodology, participants were required to raise both hands in the air for 1 – 2 seconds following completion of each trial in order to generate a distinct signature in the data files that could be used to separate the files during post-processing. This was done in order to prevent the participants from having to pause for 1 minute between trials to allow the glove data acquisition software to be stopped and restarted, potentially interrupting task flow and skill acquisition, as well as reducing the total training time and number of training trials. However, many participants failed to perform the required movement correctly or consistently across all trials, despite reminders to continue to do so. This resulted in data files that were extremely noisy and difficult to separate during post-processing, even when compared side by side to recorded trial time data. Therefore, complete and clean glove data sets for all trials of Session 1 and Session 2 were only obtained for a total of 10 participants. Despite this small
Of the 10 participants included in this study, one participant was left-handed. One-sample t-tests were used to compare the mean dominant and non-dominant hand velocity and angular rotation data for the left-handed participant to the means for the 9 right-handed participants. For each t-test, a single data point (i.e., the mean for the left handed participant) was compared to the data set of means for the right-handed participants. These analyses indicated significant differences for the left-handed participant, and thus the data for that participant were excluded from further analyses.

**Velocity.** A 5 (Block) x 2 (Hand) RM ANOVA for glove-based hand velocity was conducted for the 9 remaining right-handed data sets. Levels for Hand included dominant and non-dominant hand, indicating the hand from which the sensor data samples were acquired, rather than trial segment. Data for each trial were separated using the hand-raising method. This was performed after each complete trial (i.e., transferring the triangles from one side of the board to the other and then back to the original side of the board), rather than after individual trial segments simply moving the triangles from one side of the board to the other. This was done to maintain consistency with the standardized FLS method of task completion and to enable comparison of trial segment analyses to traditional (total trial) analyses. Figure 30 presents mean velocity in centimeters per second (cm/s) for the dominant and non-dominant hands across all trials of all five training blocks for 9 participants.
Figure 30. Mean glove velocity in centimeters per second for the dominant and non-dominant hands across 5 training blocks for 9 participants.

Despite the small sample size, a significant main effect was found for Block, $F(4, 32) = 5.13$, $p = .003$, with the slowest mean velocity for Block 2 ($M = 8.79$), followed by Block 1 ($M = 8.86$), Block 3 ($M = 8.95$), and Block 5 ($M = 9.68$), with the fastest mean velocity for Block 4 ($M = 10.04$). A simple contrasts post hoc analysis was conducted, comparing Block 1 to each subsequent block in order to compare velocity for each block to velocity at the beginning of training. The Bonferroni adjustment was used to account for familywise error. This analysis resulted in a significant contrast between Block 1 and Block 4, $F(1, 8) = 8.79$, $p = .018$. Additionally, a repeated contrasts analysis was conducted in order to identify significant changes in velocity from one training block to the next, and in particular to compare the last block of Session 1 (Block 3) to the first block of Session 2 (Block 4). This analysis resulted in a significant post hoc contrast for Block 3 versus Block 4, $F(1, 8) = 9.48$, $p = .015$, indicating a
significant increase in mean hand velocity, across both hands, from the last block of Session 1 to the first block of Session 2. No other repeated contrasts for Block were significant.

A significant main effect was also found for Hand, $F(1, 9) = 19.79$, $p = .002$, with mean velocity being faster for the non-dominant hand ($M = 10.64$) than for the dominant hand ($M = 7.89$). Additionally, a significant interaction was detected for Block x Hand $F(4, 32) = 6.15$, $p = .001$. Post hoc simple contrasts for Block x Hand were conducted, comparing Block 1 to each subsequent block for each hand, in order to assess relevant changes in performance with the dominant and non-dominant hands later in training as compared to early in training. The Bonferroni adjustment was again used to account for familywise error. This analysis resulted in a significant contrast for the dominant versus non-dominant hands between Block 1 and Block 4, $F(1, 8) = 10.38$, $p = .012$, as well as between Block 1 and Block 5, $F(1, 8) = 8.43$, $p = .020$. Thus, while the velocity of the dominant hand did not change significantly across sessions, the velocity of the non-dominant hand increased significantly in Blocks 4 and 5 (i.e., the Session 2 training blocks). Additionally, a repeated contrasts analysis was conducted in order to identify significant changes in dominant versus non-dominant hand velocity from one training block to the next, and in particular to compare the last block of Session 1 (Block 3) to the first block of Session 2 (Block 4). This analysis indicated that the post hoc contrast for Block 3 versus Block 4 neared, but did not reach, significance, ($p = .051$); no other repeated contrasts for Block were significant.

The main effect for hand indicates an overall higher velocity movement profile for the non-dominant hand based on complete trials, rather than trial segments. This result may
indicate that overall, non-dominant hand movements are characterized by larger, gross movements; with dominant hand movements being characterized by smaller, fine movements and manipulations consistent with higher precision. This effect was detected for means across entire trials, representing both the higher and lower precision subtasks for each hand. The main effect for Block indicates that overall velocity, across both hands increased from Session 1 to Session 2. As demonstrated by the significant interaction, this increase in mean velocity was primarily driven by a significant increase in non-dominant hand velocity while the dominant hand velocity remained relatively consistent across training blocks and sessions. This result may indicate that in order to perform the task more quickly and efficiently over the course of training, larger, gross movements are performed with increasing speed using the non-dominant hand; while smaller, fine movements and manipulations continue to be performed with the dominant hand in order to maintain consistent precision, particularly during the triangle transfer portions of the task. This finding is particularly interesting given that equivalent time-based performance for the dominant-led and non-dominant-led trial segments was associated with better overall task performance, as indicated by the correlation of ambidexterity score to overall trial time. These results are explored more deeply, and within the context of related findings within the research literature, in the discussion section.

**Angular Rotation.** A 5 (Block) x 2 (Hand) RM ANOVA was also conducted for glove-based angular rotation. Levels for Hand again included dominant and non-dominant hand, indicating the hand from which the sensor data samples were acquired. Figure 31 presents
mean angular rotation in degrees for the dominant and non-dominant hands across all trials of all five training blocks for the 9 right-handed participants.

*Figure 31.* Mean glove angular rotation, in degrees, for the dominant and non-dominant hands across five training blocks in Session 1 (Blocks 1, 2, and 3) and Session 2 (Blocks 4 and 5) for 9 right-handed participants.

Despite the small sample size, a significant main effect was again found for Hand, $F(1, 8) = 26.55, p = .001$, with mean angular rotation being greater for the non-dominant hand ($M = 24.53$) than for the dominant hand ($M = 15.73$). No significant main effect was found for Block; nor was a significant interaction for Hand x Block present. As with the velocity results, these results may indicate that the non-dominant hand is performing larger, gross movements, while the dominant hand is performing smaller, more precise and efficient movements, particularly during the triangle transfer portions of the task.
**Session 3.** Due to a variety of factors involving responsiveness of participants to follow-up emails, availability of participants during the subsequent school semester, and scheduling issues, only 15 of the original 27 participants completed the final session (Session 3); thus, independent analyses were conducted including only the data for those 15 participants across trials for all three sessions. A 2-tailed bivariate correlation was conducted in order to assess the relationship between the number of days from Session 2 to Session 3 and the difference in total trial completion time for the Session 3 retention test and the last trial of Session 2 for each participant. The total trial time for the Session 2 last trial was subtracted from the total trial time for the Session 3 retention test trial time for each participant. These values are plotted (x-axis) against the value for the number of days between sessions (S2 and S3) for each participant (y-axis). This correlation did not approach significance, $r(13) = .14, p = .623$.

![Figure 32](image)

*Figure 32.* Scatterplot representing the relationship between number of days from Session 2 training to Session 3 retention test (x-axis) and the mean total trial time difference between the last Session 2 trial and the Session 3 retention test (y-axis).
As shown in Figure 32, no trend was present for trial time increase or decrease with respect to increased inter-session interval. Some participants from both ends of the retention period spectrum demonstrated slower times following the retention period (i.e., positive y-axis values) and some demonstrated faster times following the retention period (i.e., negative y-axis values). Based on this analysis, it was determined that data for all Session 3 participants could be included in statistical analyses.

As previously noted, time and error data were recorded for all 15 retention tests, which was the primary trial of interest within Session 3; however, retraining trial time, error, and glove data were only recorded for 7 participants. Therefore, task decomposition-based metrics were analyzed for the retention test trial for all 15 returning participants; however, analyses of decomposition-based metrics were not conducted for the retraining trials and are not presented here. Additionally, glove-based metrics were successfully captured for 6 right-handed participants. Of the 6 successfully recorded glove data sets for the Session 3 retention test, one was matched to glove data sets for the same participant in Sessions 1 and 2 (overlapping). The remaining five Session 3 retention test glove data sets were for participants not represented in the Session 1 and Session 2 glove data set (non-overlapping). Univariate ANOVAs for glove velocity and angular rotation were only conducted for the non-overlapping data sets to avoid violating the assumption of independence across sample means.

Task decomposition-based metrics. A 5 (Trial) x 2 (Hand) RM ANOVA was conducted. The 5 conditions for Trial were Session 1 First Trial, Session 1 Last Trial, Session 2
First Trial, Session 2 Last Trial, and Session 3 First Trial (retention test). The 2 conditions for hand were dominant and non-dominant, indicating the side of the board from which the individual trial segments were started. Figure 33 provides mean trial times in seconds for trial segments led by the dominant and non-dominant hands across 5 trials for the participants that completed all three sessions (n=15).

![Figure 33](image)

**Figure 33.** Mean trial time in seconds for trial segments beginning on the dominant and non-dominant side of the board for the first Session 1 trial (S1First), last Session 1 trial (S1Last), first Session 2 trial (S2First), last Session 2 trial (S2Last), and Session 3 retention test (S3First) for all Session 3 participants (n=15).

A significant main effect was found for Trial, \(F(4, 56) = 50.17, p < .001\), with the mean time for Session 1 First Trial being the slowest (\(M = 142.27\)), followed by Session 2 First Trial
(M = 66.12), Session 3 First Trial (M = 62.31), and Session 2 Last Trial (M = 51.91), with the mean time for Session 1 Last Trial being fastest (M = 51.74). A significant main effect was not found for Hand, nor was a significant interaction effect found for Trial x Hand. Tests of within-subjects contrasts for Trial using the repeated method (comparing each trial to the subsequent trial) indicated significant contrasts for Session 1 First Trial versus Session 1 Last Trial, F(1, 14) = 96.44, p < .001; Session 1 Last Trial versus Session 2 First Trial, F(1, 14) = 10.85, p = .005; and Session 2 Last Trial versus Session 3 First Trial, F(1, 14) = 7.73, p = .015. These results indicate a statistically significant difference for the comparison of the last trial of Session 2 to the Session 3 retention test trial 7 months later. As observed in the previous individual trial analysis, this analysis also indicates that while mean times for trial segments moving the triangles from the dominant hand side of the board to the non-dominant hand side of the board across all trials in a Block were slower overall than times for trial segments moving the triangles from the non-dominant hand side of the board to the dominant hand side of the board, individual trials do not reflect this difference in every instance.

**Glove-based metrics.** Univariate ANOVAs were conducted for the non-overlapping data sets for glove velocity and angular rotation. These analyses included glove data for the nine participants included in the Session 1 and Session 2 glove data set, as well as Session 3 retention test glove data for five different participants.
**Velocity.** Despite a small sample size, a 2 (Hand) x 2 (Session) univariate ANOVA analysis was conducted comparing glove velocity with each hand for the Session 3 Retention Test to Session 1 means by hand across Blocks 1, 2, and 3. Session 1 and Session 3 Retention Test mean velocity data are presented in Figure 34 in centimeters per second (cm/s) for the dominant and non-dominant hands. Error bars indicate standard error of the mean.

![Figure 34](image)

**Figure 34.** Mean velocity for the dominant and non-dominant hands for different groups of participants for Session 1 (n=15) and the Session 3 retention test (n=5).

A significant main effect was detected for Hand, $F(1, 26) = 17.50, p < .001$, with the mean velocity for the non-dominant hand being faster ($M = 10.70$) than the mean velocity for the dominant hand ($M = 7.38$) across Session 1 and the Session 3 Retention Test; however, the main effect for Session was not significant and no significant interaction was detected.
Additionally, a 2 (Hand) x 2 (Session) univariate ANOVA analysis was conducted comparing glove velocity with each hand for the Session 3 Retention Test to Session 2 means by hand across Blocks 4 and 5. Session 2 and Session 3 Retention Test mean velocity data are presented in Figure 35 in centimeters per second (cm/s) for the dominant and non-dominant hands. Error bars indicate standard error of the mean.

![Figure 35](image.png)

*Figure 35.* Mean velocity for the dominant and non-dominant hands for different groups of participants for Session 2 (n=15) and the Session 3 retention test (n=5).

A significant effect was detected for Hand, $F(1, 26) = 17.95$, $p < .001$, with the mean velocity for the non-dominant hand being faster ($M = 11.40$) than the mean velocity for the dominant hand ($M = 7.37$) across Session 1 and the Session 3 Retention Test; however, the main effect for Session was not significant and no significant interaction was detected.
Angular rotation. Again, despite a small sample size, a 2 (Hand) x 2 (Session) univariate ANOVA analysis was conducted comparing glove angular rotation with each hand for the Session 3 Retention Test to Session 1 means by hand across Blocks 1, 2, and 3. Session 1 and Session 3 Retention Test mean angular rotation data are presented in Figure 36 in degrees for the dominant and non-dominant hands. Error bars indicate standard error of the mean.

Figure 36. Mean angular rotation for the dominant and non-dominant hands for different groups of participants for Session 1 (n=15) and the Session 3 retention test (n=5).

A significant main effect was detected for Hand, $F(1, 26) = 9.00$, $p = .006$, with the mean angular rotation for the non-dominant hand being greater ($M = 33.80$) than the mean angular rotation for the dominant hand ($M = 10.50$) across Session 1 and the Session 3
Retention Test; however, the main effect for Session was not significant and no significant interaction was detected.

Additionally, a 2 (Hand) x 2 (Session) univariate ANOVA analysis was conducted comparing glove angular rotation with each hand for the Session 3 Retention Test to Session 2 means by hand across Blocks 4 and 5. Session 2 and Session 3 Retention Test mean angular rotation data are presented in Figure 37 in degrees for the dominant and non-dominant hands. Error bars indicate standard error of the mean.

![Figure 37. Mean velocity for the dominant and non-dominant hands for different groups of participants for Session 2 (n=15) and the Session 3 retention test (n=5).](image)

A significant effect was detected for Hand, $F(1, 26) = 7.85$, $p = .009$, with the mean angular rotation for the non-dominant hand being greater ($M = 32.61$) than the mean velocity for
the dominant hand ($M = 10.01$) across Session 1 and the Session 3 Retention Test; however, the main effect for Session was not significant and no significant interaction was detected.

**Experiment 1 Discussion**

Experiment 1 began to address identified gaps in the skill decay and retraining research literature within the context of laparoscopic surgery (LS) skills. Specifically, the aim of this experiment was to investigate the use of task decomposition-based and instrumented glove-based metrics to assess individual hand performance (i.e. dominant versus non-dominant) for a standardized LS manual skills training task (peg transfer) using the Fundamentals of Laparoscopic Surgery (FLS) inanimate box training platform. As reported by Ritter and Scott (2007):

“This task was designed to develop depth perception and visual-spatial perception in a monocular viewing system and the coordinated use of both the dominant and nondominant hands. It also replicates the important action of transferring and positioning an object laparoscopically, as required to adjust a needle between needle holders when suturing p. 108.”

While this task purports to develop bimanual dexterity, the standardized FLS scoring metrics do not differentiate between performance with the dominant and non-dominant hands. Rather, the normalized metrics used for FLS scoring, and by extension for LS certification and credentialing, are derived from time to complete the entire task, with time-based penalties
assessed for irrecoverable errors (i.e., triangles dropped outside the view of the camera) (Fraser et al., 2003). Previous research has demonstrated that the official FLS scoring metrics, are no more effective than time trial time alone for assessing LS skill level, and that the two measures are nearly perfectly correlated (Kowalewski, 2012). Additionally, previously reported durability of the skills involved in this task (e.g., Stefanidis et al., 2005; Stefanidis et al., 2006; Castellvi et al., 2009; Mashaud et al., 2010) despite subjective reports of noticeable LS skill decay within shorter time periods (Deering et al., 2011) suggest that these metrics may lack sensitivity to decay characteristics. Previous studies have established the construct validity of motion analysis and demonstrated that motion-based metrics provide a reliable means for assessment of simple surgical dexterity (Bann, Kahn, & Darzi, 2003). However, motion-based metrics have not been reported within the context of skill decay assessment in the research literature to date. Additionally, few studies have reported on performance with the dominant versus non-dominant hands, and a specific gap was identified for research assessing variable performance within subtasks that rely more heavily on one hand.

In particular, Experiment 1 sought to address Research Question 1: Do task decomposition-based and hand motion-based metrics provide more sensitive means than traditional metrics of total task completion time and irrecoverable errors for assessing skill acquisition and decay within the context of LS psychomotor task performance? This Experiment assessed FLS psychomotor (peg transfer) skill acquisition and retention following short-term (<6 weeks) and long-term (~7 months) retention periods, applying the specified metrics to assess performance with the dominant versus non-dominant hands. The task
decomposition-based metrics specifically compared two trial segments: moving triangles from the dominant hand side of the board to the non-dominant side versus moving triangles from the non-dominant hand side to the dominant hand side. Each segment involves three primary subtasks: removing triangles from pegs, transferring triangles in the air, and placing triangles on pegs. Removing triangles from pegs was determined to be easier (i.e., requires less precision) than placing triangles on pegs based on task decomposition and the results of an unpublished pilot study. Additionally, transferring triangles involves holding/releasing each triangle with one instrument and grasping each triangle with the other instrument; of these subtask components, task decomposition indicated that holding/releasing triangles is easier (i.e., requires less precision) than grasping triangles. Trial segments led by the dominant hand involve performance of the easier subtask components (removing triangles from pegs and holding/releasing triangles during transfers) by the dominant hand, with the non-dominant hand performing the subtask components requiring more precision (grasping triangles during transfers and placing triangles on pegs). Conversely, the non-dominant hand performs the easier subtask components for trials led by the non-dominant hand.

In addition to decomposing the task itself, potential performance metrics (time and errors) were evaluated in order to identify sensitive and objective metrics. Assessment of the relationship between task time and recoverable errors (i.e., triangles dropped within view of the camera) for individual trial segments and complete trials indicated that task time is highly correlated to number of triangles dropped. This was as expected, since recovering dropped triangles is time-consuming, contributing to overall task time, and interrupts task flow.
However, this correlation had not previously been reported in the research literature, warranting investigation and demonstration of this relationship. Based on this assessment, it was determined that triangle drops do not provide a more sensitive metric than time alone. Therefore, further performance analyses for Experiment 1 were limited to assessments of time to complete the decomposed trial segments and total trial times. Additionally, instrumented glove-based metrics were selected to support comparisons of dominant and non-dominant hand motion profiles during task performance. The specific glove-based metrics selected included velocity and angular rotation. Velocity measures speed of hand movements in centimeters per second, while angular rotation provides a measure of the overall range of hand rotation about three axes (i.e., flexion/extension, radial/ulnar deviation, and pronation/supination) in degrees.

The results of Experiment 1 demonstrated significant psychomotor skill acquisition over the course of five training blocks within two training sessions (Session 1 and Session 2), spaced less than 6 weeks apart. While this improvement in task performance was evident based on traditional metrics of total trial time and irrecoverable errors (i.e., triangles dropped outside the view of the camera), the task decomposition-based metrics and instrumented glove-based metrics also detected significant differences in task performance between the dominant and non-dominant hands. Additionally, the task decomposition-based metrics detected significant performance degradation across both hands following both short-term (<6 weeks) and long-term (~7 months) retention periods.

Specifically, across all Session 1 and Session 2 training blocks, task decomposition-based analyses indicated significantly faster times for trial segments led by the non-dominant
hand, for which the dominant hand performs the more difficult task subcomponents. Furthermore, a non-significant trend was demonstrated for increased ambidexterity in task performance over the course of training. While mean times for both trial segments improved over time, greater improvement was observed for trial segments led by the dominant hand, resulting in equivalent performance for the dominant hand-led and non-dominant hand-led trial segments later in training. Additionally, increased ambidexterity indexes, based on absolute differences in time to complete trial segments led with the dominant and non-dominant hands, were significantly correlated to faster overall trial times. While this was expected and necessarily must be the case to some extent due to limitations in how quickly the task can be completed with either hand and the fact that trials with shorter overall trial times inherently have a shorter maximum potential time difference, the results ruled out the possibility that trials exhibiting higher ambidexterity indexes were simply performed equally slowly with both hands. This result was consistent with Mettler, Zuberi, Rastogi, and Schollmeyer’s (2006) observations of improved performance for both the dominant and non-dominant hands individually, as well as together, for a task similar to FLS peg transfer performed on a video box training platform. Additionally, Dosis et al. (2005) demonstrated, using motion analysis, that expert surgeons displayed significantly greater ambidexterity than novices.

Examination of individual trials (first/last of each training session), demonstrated a significant effect for faster times at the end of Session 1 compared to the beginning of Session 1, indicating improved performance over the course of Session 1 training, as well as significantly slower times at beginning of Session 2 as compared to the end of Session 1,
indicating significant decay over the course of the short-term (<6 weeks) inter-session interval. The detection of significant skill decay within less than 6 weeks is relevant within the context of the broader research literature with respect to LS skills specifically and more generally across skilled psychomotor task domains. This finding does not necessarily contradict previous reports of LS skill durability over longer periods of time as the previously cited studies tested retention following training to FLS certification proficiency standards (Stefanidis, et al., 2006; Castellvi et al., 2009), and in one case also included refresher training to proficiency every 6 months (Mashaud et al., 2010). However, this result serves to highlight the important role of initial training in the subsequent retention of complex psychomotor skills. Additionally, the overall mean performance for Block 4 indicated rapid reacquisition of the lost skill and continued improvement with further training. Furthermore, while significant skill loss was also demonstrated for the long-term (~7 months) retention period based on the comparison of the Session 3 retention test trial to the last Session 2 trial, the degree of skill loss for the 7-month retention period was not significantly greater than the degree of skill loss for the short-term retention period. There are two interesting points to be made here. First, the additional training that occurred over the course of Session 2 (Blocks 4 and 5) resulted in trainees achieving a higher level of task proficiency at the end of Session 2 than Session 1, preventing a direct comparison of retention rates for the subsequent retention periods following the respective training sessions. A second interesting point to note is that previous research has demonstrated that in the case of complex skills optimal distribution of practice is inter-related to retention interval (Arthur et al., 2010; Wang et al., 2013). While specific recommendations are limited
within the domain of psychomotor skills, Pashler, Rohrer, Cepeda, and Carpenter (2007) suggested that an optimal inter-session interval (ISI) for verbal recall task training sessions is 10-20% of the intended retention interval, indicating for example, that for a retention interval of 10 days, a 1-day ISI is optimal. Cepeda et al., (2006) specifically demonstrated that for a verbal recall task, 28 days was the optimal ISI for a 6-month retention interval. The findings of the current study may serve to contribute to the development of similar recommendations for psychomotor skills.

In addition, the analyses for individual trials at the start and end of each training session (1 and 2) also tested for a main effect for hand, indicating no significant effect for performance trials led with the dominant versus non-dominant hand for this analysis. Similarly, analyses including performance of the retention test following a 7-month retention period (Session 3) demonstrated no significant effect for subtask performance with the dominant versus non-dominant hand in terms of a main effect or interaction with time of assessment, but demonstrated significant overall changes in performance over time. The lack of consistency in findings for relative performance with the dominant and non-dominant hands may be an artifact of high performance variability for the selected psychomotor task within a novice population, particularly within the early stages of skill acquisition, which has been observed in previous work (Kowalewski. 2012). More importantly, these results confirm that this psychomotor skill is extremely durable following initial skill acquisition, with significant but not total skill loss occurring even within long-term retention periods. However, it is important to note that any skill loss may be unacceptable within the context of surgical skill performance; thus,
understanding the precise nature of the skill loss is critical in order to reduce skill attrition and accelerate skill reacquisition via refresher training.

The glove-based metrics were employed in an effort to provide additional insight into the nature of the skill itself with this goal in mind. Glove-based hand velocity indicated significantly faster velocities for the non-dominant hand across all training blocks for Sessions 1 and 2, as well as the Session 3 retention test. The velocity of the dominant hand did not change significantly across sessions, but the velocity of the non-dominant hand increased significantly in Blocks 4 and 5 (i.e., the Session 2 training blocks). These results may indicate that in order to perform the task more quickly and efficiently over the course of training, larger, gross movements are performed with increasing speed using the non-dominant hand; while smaller, fine movements and manipulations continue to be performed with the dominant hand in order to maintain consistent precision. This finding is particularly interesting given that equivalent time-based performance for the dominant-led and non-dominant-led trial segments was associated with better overall task performance, as indicated by the correlation of ambidexterity score to overall trial time. As previously noted, the transfer of triangles from one grasper to the other represents the only bi-manual portion of the task, in which the dominant and non-dominant hands must coordinate, and in which it is possible for one hand to compensate for the other by dominating the subtask or facilitating difficult aspects of the subtask. Specifically, the main effect for velocity may indicate that the non-dominant hand was performing larger, gross movements, with the dominant hand performing smaller, fine movements during triangle transfer. Additionally, this result may indicate that the dominant hand exhibited more efficient
movements, which would have involved some high velocity movements, but over shorter distances (path lengths). Such efficient movements with the dominant hand may have been coupled with longer time periods in which the dominant hand wasn’t moving, as would be the case during the trial segments in which the non-dominant hand was removing triangles from pegs and placing triangles on pegs. This would have resulted in lower mean velocity for the dominant hand, driven by longer periods of non-movement despite the presence of high velocity movements during active trial segments. In addition to longer periods of non-movement while “waiting” for the non-dominant hand to complete the independent portions of the task (removing and placing triangles on pegs), the dominant hand may have also exhibited longer periods of non-movement during triangle transfers in the direction of the non-dominant side of the board as this involves holding each triangle still with the first grasper while the second grasper attempts to grasp it. Figure 38 provides an example of the right instrument (i.e., grasper) holding a triangle still while the left instrument attempts to grasp it.

Figure 38. Example of the right instrument holding a triangle still while the left instrument grasps it.
Higher velocity movement with the non-dominant hand later in training could have been achieved via 1) increased distance traveled in the same amount of time 2) the same distance traveled but in less time, or 3) increased distances traveled in less time. In order to determine the most likely cause of this increase, the research literature was consulted.

In a task analysis, Cao et al. (1996) concluded that laparoscopic psychomotor skills tasks requiring more precision (having more precision constraints) add to the perceived difficulty of the task. Cao et al. (1996) also demonstrated that a comparison of the frequencies of motions in each subtask revealed inconsistencies between an expert and novice surgeons. In some instances, the novice surgeons performed more movements in one motion category, while in other instances, the expert surgeon performed more movements within one motion category. Despite this, the duration of each subtask for the expert surgeon remained shorter than for the novices.

Bann et al. (2003) specifically demonstrated that the relationship between time and motion was variable depending on the task across three laparoscopic suturing tasks. When the number of movements in a minute (standardized movements per minute) were considered, both groups were found to work at a similar rate (i.e., velocity), depending on the task, implying that the more experienced surgeon is more economical, performing the same exercise with fewer moves rather than with higher speed. While the current study did not assess number of movements, the significant interaction for hand velocity across Sessions 1 and 2 indicated significantly higher velocities for the non-dominant hand from training Block 1 to training Block 4, with mean velocity remaining high in Block 5, while velocities for the dominant hand
remained consistent across blocks. This may indicate implementation of a strategy in which the non-dominant hand not only performs the majority of gross, speeded movements, but also speeds up over time in an effort to perform the overall task more quickly. Meanwhile, despite increasingly faster trial times over the course of training, the dominant hand continues to perform consistently slower, and likely more precise, movements. This is particularly relevant during the transfers in both directions.

Similarly, glove-based angular rotation data demonstrated that angular rotation was significantly greater for the non-dominant hand across all training and retention assessment trials. While the sample size was small and the variability of this measure was large, this statistically significant main effect indicates a prominent role of dominant versus non-dominant hand performance with respect to angular rotation within the current task. Angular rotation measures the amount of rotation of the hands about three axes during task performance; thus, the nondominant hand is demonstrating greater overall movement in these degrees of freedom, again indicating that the non-dominant hand may be performing more gross motor movements with less efficiency, while the dominant hand performs more of the precise, fine motor movements requisite of this psychomotor task. Further elucidation of these effects are presented in the general discussion within the broader context of both experiments.

**Experiment 2 Results**

Experiment 2 sought to build on the results of Experiment 1, providing additional data relevant to research question 1, which concerns assessment of skill acquisition and retention for
the selected psychomotor task (FLS peg transfer) for both the dominant and non-dominant hands using task decomposition-based metrics and instrumented glove-based metrics. Additionally, Experiment 2 sought to address research questions 2 and 3, assessing relative skill decay for the Psychomotor, Cognitive, and Integrated tasks following a 3-week retention period; and to assess the comparative effectiveness of video-based instruction to physical retraining of the integrated LS skill, respectively.

The primary objectives of the analyses within the current study were to 1) assess skill acquisition and retention for the selected Psychomotor task (FLS peg transfer) for both the dominant and non-dominant hands using task decomposition-based metrics and instrumented glove-based metrics 2) assess relative skill decay for the constituent cognitive and psychomotor skills within the context of the Integrated tasks, as well as in isolation, and 3) assess the comparative effectiveness of video-based versus hands-on Integrated task retraining. In order to effectively assess retention, it was necessary to demonstrate that the individual and integrated skills had been sufficiently acquired prior to the retention period. For this reason inclusion criterion were established and confirmed via pilot testing.

These inclusion criteria were designed to establish a level of difficulty sufficient to require some training to achieve, but within a range that is achievable by most individuals. Specifically, for the Psychomotor and Cognitive tasks, an initial learning curve was desirable in terms of both time and accuracy in order to have a point of comparison for subsequent skill decay, and also to ensure that a floor effect was not present in which case decay would not be detectable. The requirement that the skills be sufficiently acquired during initial training was
equally important. While a comparison of the effects of various levels of initial integrated and isolated skill acquisition on subsequent retention is relevant to the general skill decay literature, the objective of the current research was to lay the groundwork for research specifically pertaining to retention of previously mastered skills, as is the case for military surgeons. The current research emphasizes the early stages of skill acquisition, rather than training to mastery; however, it was not within the scope of the current effort to examine retention across a continuum of initial training below proficiency, proficiency-level training, and overlearning.

The established criteria included time and accuracy-based requirements for the Psychomotor task and accuracy-based requirements for the Cognitive and Integrated tasks. Specifically, criterion performance for the Psychomotor task excluded participants with mean trial times at the end of Session 1 training exceeding double the time required for a perfect FLS test score by surgeons seeking FLS credentialing; additionally participants were required to have performed two errorless Psychomotor task trials (no triangles dropped) within the last five Session 1 training trials. Cognitive task criterion specifications required a minimum of two errorless trials from memory at the end of Session 1 training, prior to completing the Integrated task, and an additional errorless Session 1 PostTest trial following Integrated task training. Integrated task criterion specifications required all three Session 1 trials to be performed without errors in the cognitive task component (pattern formation accuracy). Despite the fact that all five pilot participants achieved the specified criterion levels, a large subset of the official study cohort did not meet criterion requirements for one or more tasks. All 40 participants, including those that did not meet criterion standards in Session 1, completed Session 2 following a retention period
of 3 weeks. However, data analyses were only conducted for participants who achieved criterion level performance on all three tasks in Session 1 (n=24). While the reduced sample size does not meet the original, conservative, sample size estimation based on .95 power, it does meet the required sample size for .80 power, and thus was considered to be sufficient for the planned statistical analyses.

**Demographic data.** Demographic factors assessed included gender, video game play frequency, dominant hand, and starting hand. Correlations were assessed for comparisons of these factors to each of the time and accuracy measures included in the Experiment 2 analyses for each task. Gender correlated only to video games play frequency $r(22) = .59, p=.002$, indicating that male participants play video games more frequently. Video game play frequency did not significantly correlate to any other variables. No significant correlations were found for hand dominance (left versus right) or for randomly assigned starting hand. Significant correlations for all other variables are presented in Table 3.
Table 3
Significant Correlations for Demographic Factors, Methodological Variables, and Matching to Sample Test to Dominant (D) and Non-dominant (ND) Task Trial Segments for All Participants (N=24)

<table>
<thead>
<tr>
<th>Demographic/ Methodological Variable</th>
<th>Significant Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument Playing Skill</td>
<td>Psychomotor PostTest Accuracy (ND)*</td>
</tr>
<tr>
<td>Time Between Sessions</td>
<td>Psychomotor PostTest Time (D)**</td>
</tr>
<tr>
<td>Retraining Condition (Video vs Hands on)</td>
<td>Psychomotor First Trial Time (ND)*</td>
</tr>
<tr>
<td></td>
<td>Integrated First Trial Time*</td>
</tr>
<tr>
<td></td>
<td>Integrated Final Test Time*</td>
</tr>
<tr>
<td>Matching to Sample Mean Correct</td>
<td>Psychomotor PostTest Accuracy (ND)*</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>Integrated Retention Test Time*</td>
</tr>
<tr>
<td></td>
<td>Integrated Time Difference (S2-S1)*</td>
</tr>
<tr>
<td></td>
<td>Instrument Playing*</td>
</tr>
<tr>
<td></td>
<td>Retraining Condition*</td>
</tr>
<tr>
<td>Matching to Sample Throughput</td>
<td>Integrated First Trial Time*</td>
</tr>
<tr>
<td>Matching to Sample % Correct</td>
<td>Integrated Retention Test Time*</td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01

As indicated in Table 3, self-reported instrument playing skill was only significantly correlated to the number of pegs dropped within the Psychomotor task Session 1 PostTest for the trial segment led by the non-dominant hand. Time between Session 1 and Session 2 was only significantly correlated to a task performance factor that occurred prior to the retention period and therefore cannot be caused by the length of retention period. Retraining condition (video versus hands on based on random assignment) was significantly correlated to three task performance factors, indicating that participants randomly assigned to the hands on retraining condition demonstrated significantly faster mean trial segment times for the Psychomotor task Session 1 first trial, Integrated task Session 1 first trial, and the Integrated task Session 2 final
test. Multiple significant correlations were detected for the matching to sample spatial ability test variables. However, the matching to sample measures were not significantly correlated to Cognitive task performance in either session, which was expected given that the Cognitive task in isolation (performed on the tablet) seems more similar, at least on the surface, to the tablet-based matching to sample test. Particularly relevant significant correlations were found for the relationship of both matching to sample percent correct and reaction time for correct trials with Integrated task Session 2 retention test mean trial time. Specifically, higher percent correct and faster reaction times for correct trials on the matching to sample test correlated to faster Integrated retention test trial times. Additionally, slower matching to sample correct trial reaction times correlated to larger increases in trial times between the end of Session 1 to the Session 2 retention test for the Integrated task. This finding indicates that the matching to sample test may provide a useful measure for predicting skill decay for complex and integrated skill tasks. Matching to sample is intended as a measure of visual-spatial processing, working memory, and visual recognition memory. This test relies on working or “immediate” memory, given the relatively short delay period of five seconds between stimulus presentation and response choices. Additionally, visual recognition memory, which differs from free recall memory, is tested.

As demonstrated by the lack of significant correlations to Session 2 performance, variance in the retention period did not impact task performance across participants. The second training session was performed approximately 3 weeks (21 days) after the original training session for each participant; the mean time between sessions across all subjects
included in the final data set for analysis (n=24) was 20 days (range = 15 – 22 days), with the majority of participants completing their second session exactly 21 days after their first. However, a small minority of participants completed their second session several days early due to scheduling issues, resulting in a statistically significant skewed distribution ($p > .05$ for Shapiro Wilkes test of normality). Figure 39 provides a histogram to demonstrate this distribution.

![Histogram](image)

Figure 39. Histogram presenting frequency of retention period in days between Session 1 and Session 2 for all participants included in the final data set for analysis (n=24).

**Task decomposition-based metrics.** Analyses were conducted for both time and accuracy for each task at the beginning and end of Session 1 training and following the 3-week
retention period. A final test was also conducted for the Integrated task following Session 2 retraining.

*Psychomotor task time and accuracy.* Initial analysis of the Psychomotor task performance metrics included a 4 (Trial) x 2 (Hand) RM ANOVA for task completion time. The four levels for Trial were Session 1 First Trial, Session 1 Mean Best (a calculated mean of the fastest 3 of the last 5 Session 1 trials, without triangles dropped), Session 1 PostTest, and Session 2 Retention Test. These trials were selected in order to provide a baseline of task performance at the beginning of Session 1 training, an assessment of maximum performance achieved at the end of Psychomotor task training, a final assessment at the end of Session 1 in order to detect any further improvements in psychomotor skill resulting from rehearsal of the Integrated task, and retention of the Psychomotor skill following the 3-week retention period. The two levels for Hand were dominant and non-dominant, indicating as in Experiment 1 trial segments led with the dominant and non-dominant hands. Means for trial segments led by the dominant and non-dominant hand for each assessment are presented in graph form with standard error indicated in Figure 40.
Figure 40. Mean Psychomotor task time in seconds for trial segments led by the dominant and non-dominant hands for the first Session 1 trial (S1 First Trial), mean of the best error-free trials at the end of Session 1 training (S1 Mean Best), the Session 1 PostTest trial (S1 PostTest), and the Session 2 retention test trial (S2 Retention Test).

A significant main effect was found for Trial, $F(3, 69) = 87.15, p < .001$, with the mean time for Trial 1 being the slowest ($M = 142.27$), and with mean times for Trials 2, 3, and 4 being nearly identical ($M = 38.32$, $M = 38.61$, and $M = 38.57$, respectively). Not surprisingly, tests of within-subjects contrasts for Trial using the repeated method (comparing each trial to the subsequent trial) indicated a significant contrast only for Trial 1 versus Trial 2, $F(1, 23) = 99.28$, $p < .001$. A significant main effect was not found for Hand, nor was a significant interaction effect found for Trial x Hand.

All individual task analyses were also conducted for accuracy. The accuracy analysis for the Psychomotor task included a 3 (Trial) x 2 (Hand) RM ANOVA for triangles dropped.
The two levels for Hand were dominant and non-dominant, indicating the hand used to lead the trial segment. The three levels for Trial analysis were First Trial Session 1, Session 1 PostTest, and Session 2 Retention Test. Accuracy means for Session 1 Mean Best trials were not included in the analysis since, as shown in Figure 41, mean accuracy for those trials was constrained to 100% across all participants. This was by design since criterion-level performance was required at the end of Session 1 training in order for participants to be included in the final data set for analysis. This requirement was set in order to ensure that the requisite psychomotor skill had been adequately acquired within the individual (Psychomotor) task prior to participants attempting the Integrated task, and also to ensure that retention could be assessed, requiring adequate proficiency to be established prior to the retention period. The Session 1 Mean Best trials were not included in the statistical analysis because the absence of variability would artificially reduce the error variance. Figure 40 presents mean Psychomotor task accuracy, expressed as percentage of triangles dropped out of six, for trial segments led by the dominant and non-dominant hands across the 4 specified trials in graphical form. Figure 41 includes error bars indicating standard error for all trials plotted; however, given that all accuracy scores for Session 1 Mean Best Trials were 100%, the standard error for those trials is 0, and therefore no error bars are visible.
Figure 41. Mean Psychomotor task accuracy expressed as percentage of triangles dropped out of 6 for trial segments led by the dominant and non-dominant hands for the first Session 1 trial (S1 First Trial), mean of the best error-free trials at the end of Session 1 training (S1 Mean Best), the Session 1 PostTest trial (S1 PostTest), and the Session 2 retention test trial (S2 Retention Test).

Mauchly’s test of sphericity was significant ($p < 0.05$) for Trial; test statistics are reported using the Greenhouse-Geisser adjustment. A significant main effect was found for Trial, $F(2, 46) = 6.83$, $p = .003$, with the mean accuracy for Session 1 First Trial being the lowest ($M = 86.5\%$), followed by Session 1 PostTest ($M = 94.1\%$), and with Session 2 Retention Test accuracy being the highest ($M = 96.5\%$). Repeated post hoc contrasts indicated a significant comparison for Session 1 First Trial to Session 1 PostTest $F(1, 23) = 5.96$, $p = .023$, but not for Session 1 PostTest versus Session 2 Retention Test. Also, it is important to note that the Integrated retention test was conducted prior to the Psychomotor retention test in Session 2,
providing the equivalent of three Psychomotor task trials. The RM ANOVA for Psychomotor task accuracy did not reveal a significant main effect for Hand or a significant interaction effect for Trial x Hand.

*Cognitive task time and accuracy.* Task completion time analysis of the Cognitive task included a 4 (Trial) x 6 (Pattern) RM ANOVA for mean pattern formation time using the tablet app. The four conditions for Trial were Session 1 First Training Trial From Memory, Session 1 Last Training Trial From Memory, Session 1 PostTest, and Session 2 Retention Test. As for the Psychomotor task analyses, these trials were selected in order to provide a baseline of task performance at the beginning of Session 1 training, an assessment of maximum performance achieved at the end of Cognitive task training, a final assessment at the end of Session 1 in order to detect any further improvements in cognitive skill resulting from rehearsal of the Integrated task, and retention of the Cognitive skill following the 3-week retention period. Figure 42 presents the mean cognitive task times in seconds for all six patterns across four trials presented in graphical form with standard error indicated.
Figure 42. Mean Cognitive task time, in seconds, by pattern for the first Session 1 training trial from memory (S1 First), the Session 1 last trial from memory (S1 Last), the Session 1 PostTest trial (S1 PostTest), and the Session 2 retention test (S2 Retention Test).

Mauchly’s test of sphericity was significant \( p < .05 \) for Trial and Trial x Pattern; test statistics are reported using the Greenhouse-Geisser adjustment. A significant main effect was found for Trial, \( F(2, 43) = 55.71, p < .001 \), with the mean time for the PostTest Session 1 being the fastest \( (M = 8.73) \), followed by Session 1 Last Training Trial From Memory \( (M = 9.14) \), Session 1 First Training Trial From Memory \( (M = 13.86) \), and with the Session 2 Retention Trial being slowest \( (M = 15.16) \). Tests of within-subjects contrasts for Trial using the repeated method (comparing each trial to the subsequent trial) indicated significant contrasts for Trial 1 (First Training) versus Trial 2 (Last Training), \( F(1, 23) = 85.45, p < .001 \), as well as for Trial 3
(Session 1 PostTest) versus Trial 4 (Retention Test), $F(1, 23) = 77.56, p < .001$, indicating a significant change in accuracy for the time period between Session 1 and Session 2.

A significant main effect was also found for Pattern, $F(5, 115) = 35.72, p < .001$, with the fastest time being for Pattern 1 ($M = 6.38$), followed by Pattern 5 ($M = 12.02$), Pattern 3 ($M = 12.44$), Pattern 2 ($M = 12.88$), Pattern 6 ($M = 13.05$), and with the mean time to form Pattern 4 being the slowest ($M = 13.56$). Significant repeated contrasts were found for comparisons of Pattern 1 versus Pattern 2, $F(1, 23) = 143.05, p < .001$, and Pattern 4 versus Pattern 5, $F(1, 23) = 6.93, p = .015$.

Finally, a significant interaction effect was found for Trial by Pattern, $F(5, 183) = 3.53, p = .005$, with significant contrasts for Pattern 1 versus Pattern 2 for the comparison of Trial 1 (First Training) to Trial 2 (Last Training) and the comparison of Trial 3 (PostTest) to Trial 4 (Retention Test). Mean pattern formation times provided in Table 4, indicate decreased times from Trial 1 to Trial 2 and increased times from Trial 3 to Trial 4; however, these changes were much larger for Pattern 2 than for Pattern 1.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Pattern 1</th>
<th>Pattern 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>5.99</td>
<td>13.97</td>
</tr>
<tr>
<td>Trial 2</td>
<td>5.71</td>
<td>9.61</td>
</tr>
<tr>
<td>Trial 3</td>
<td>6.37</td>
<td>9.60</td>
</tr>
<tr>
<td>Trial 4</td>
<td>7.45</td>
<td>18.35</td>
</tr>
</tbody>
</table>
The accuracy analysis for the Cognitive task included a 2 (Trial) x 6 (Pattern) RM ANOVA for mean pattern formation accuracy using the tablet app. Accuracy was calculated as a percentage out of six total patterns per trial, and partial points were awarded by trial if the target pattern design was formed but the triangle colors were reversed or if another one of the trained patterns was formed (other than the target pattern for that trial). Participants were not informed of the scoring criteria. As indicated in the methodology, they were simply instructed to create all patterns in order, with an emphasis on learning the patterns and performing them accurately. For the accuracy RM ANOVA, the two levels for Trial were Session 1 First Training Trial From Memory and Session 2 Retention Test. These trials were selected in order to provide a baseline of performance at the beginning of training and an assessment of potential performance degradation following the retention period. In addition to these trials, accuracy means are plotted in graph form for Session 1 Last Training Trial From Memory and Session 1 PostTest in Figure 43. As shown in the graph, mean accuracy for these trials was 100% for all 6 patterns across all participants. This was by design since criterion-level performance was required for the Session 1 last training trial from memory and for the Session 1 post-test in order for participants to be included in the final data set for analysis. This requirement was set in order to ensure that the patterns were learned adequately within the individual (Cognitive) task prior to participants attempting the Integrated task, and also to ensure that retention could be assessed, requiring proficiency to be established prior to the retention period. These trials were not included in the statistical analysis because the constrained lack of variability would artificially reduce the error variance, creating a positive bias for significant outcomes. Figure 43
includes error bars indicating standard error for the Session 1 First Trial means and the Retention Test means; however, given that all accuracy scores for Last Trial Session 1 and Post-Test Session 1 were 100%, the standard error for those trials is 0, and therefore no error bars are visible.

![Graph showing mean cognitive task accuracy percentage by pattern for different trials.](image)

**Figure 43.** Mean Cognitive task accuracy percentage by pattern for the first Session 1 training trial from memory (S1 First), the Session 1 last trial from memory (S1 Last), the Session 1 PostTest trial (S1 PostTest), and the Session 2 retention test (S2 Retention Test).

Mauchly’s test of sphericity was significant ($p < .05$) for Trial and Trial x Pattern; test statics are reported using the Greenhouse-Geisser adjustment. A significant main effect was found for Trial, $F(1, 23) = 16.39, p < .001$, with the mean accuracy for S1 First Training Trial From Memory being higher ($M = 86\%$), than the S2 Retention Trial ($M = 67\%$). No significant effect was found for Pattern or for the interaction of Trial x Pattern.
As with the Psychomotor retention test, the Cognitive retention test was administered after the participants completed the Integrated retention test in Session 2, providing cognitive skill rehearsal and feedback during the formation of the six patterns within the context of the Integrated task, providing the equivalent of a single trial of the cognitive task with feedback. Given that participants were shown the correct pattern for any incorrect patterns formed during the Integrated retention test just prior to completing the Cognitive retention test, the low accuracy scores are even more striking.

*Integrated task time and accuracy.* Initial analysis of the Integrated task performance metrics included a 4 (Trial) x 6 (Pattern) RM ANOVA for mean pattern formation time using the FLS box. The 4 levels for Trial were Session 1 First Training Trial From Memory, Session 1 Last Training Trial From Memory, Session 2 Retention Test, and Session 2 Final Test. Figure 44 presents the mean integrated task time in seconds for all six patterns across four trials in graphical form with error bars indicating the standard error of the mean for each trial.
Figure 44. Mean Integrated task time in seconds for all six patterns across the first Session 1 training trial (S1 First), the Session 1 last trial (S1 Last), the Session 2 retention test (S2 Retention Test), and the Session 2 final test (S2 Final Test).

A significant main effect was found for Trial, $F(3, 69) = 30.19$, $p < .001$, with the mean time for the Session 2 Final Test being fastest ($M = 40.90$), followed by Session 1 Last Training Trial From Memory ($M = 41.67$), Session 1 First Training Trial From Memory ($M = 45.76$), and with the Session 2 Retention Trial being slowest ($M = 57.67$). Tests of within-subjects contrasts for the main effect of Trial, using the repeated method (comparing each trial to the subsequent trial), indicated significant contrasts for all trial comparisons, as shown in Table 5. Also presented in Table 5, multiple repeated post hoc contrasts for the interaction between Trial and Pattern were detected. In summary, these significant interactions revealed a trend in which Patterns 1, 2, and 4 demonstrated greater improvement in mean pattern formation time during Session 1 training, greater decay during the retention period, and then greater improvement
during retraining than Patterns 3, 5, and 6. While Patterns 3, 5, and 6 also showed performance degradation during the retention period and improvement during Session 2 retraining, these changes were significantly smaller than the changes observed for Patterns 1, 2, and 4. Mean times for all significant contrasts are presented in Table 6.

Table 5

Integrated Task 4 (Trial) x 6 (Pattern) RM ANOVA for Mean Time

<table>
<thead>
<tr>
<th>Main Effect</th>
<th>Significant Contrasts*</th>
<th>df</th>
<th>F</th>
<th>η</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial**</td>
<td></td>
<td>2, 43</td>
<td>30.19</td>
<td>0.57</td>
<td>0.000</td>
</tr>
<tr>
<td>1 vs. 2</td>
<td></td>
<td>1, 23</td>
<td>7.32</td>
<td>0.24</td>
<td>0.013</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td></td>
<td>1, 23</td>
<td>34.68</td>
<td>0.60</td>
<td>0.000</td>
</tr>
<tr>
<td>3 vs. 4</td>
<td></td>
<td>1, 23</td>
<td>103.16</td>
<td>0.82</td>
<td>0.000</td>
</tr>
<tr>
<td>Interaction Effect</td>
<td></td>
<td>8, 175</td>
<td>3.14</td>
<td>0.12</td>
<td>0.003</td>
</tr>
<tr>
<td>Trial x Pattern**</td>
<td></td>
<td>1 vs. 2</td>
<td>2 vs. 3</td>
<td>1, 23</td>
<td>11.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 vs. 4</td>
<td>4 vs. 5</td>
<td>1, 23</td>
<td>7.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 vs. 3</td>
<td>2 vs. 3</td>
<td>1, 23</td>
<td>12.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 vs. 5</td>
<td>1, 23</td>
<td>5.19</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 vs. 4</td>
<td>2 vs. 3</td>
<td>1, 23</td>
<td>10.89</td>
</tr>
</tbody>
</table>

Note: *Repeated post hoc contrasts with Bonferroni adjustment. **Greenhouse-Geisser adjustment used due to violation of Mauchly’s test of sphericity.
The accuracy analysis for the Integrated task included a 2 (Trial) x 6 (Pattern) RM ANOVA for mean pattern formation accuracy using the FLS box. As with the Cognitive task in isolation, accuracy for the cognitive component (i.e., pattern formation) within the Integrated task was calculated as a percentage out of six total patterns per trial, and partial points were awarded by trial if the target pattern design was formed but the triangle colors were reversed or if another one of the trained patterns was formed (other than the target pattern for that trial). The 2 levels for Trial were Session 2 Retention Test and Session 2 Final Test. As in the previous analyses, trials for which 100% performance was required (i.e., all three Session 1 training trials for the Integrated task) were not included in the statistical analyses. Figure 45 presents mean Integrated task pattern accuracy, expressed as a percentage, for all six patterns across four trials, including Session 1 First Training Trial From Memory and Session 1 Last Training Trial From Memory as a frame of reference for performance at the beginning and end of Session 1 training. Error bars indicate the standard error of the mean for each trial. Given
that all accuracy scores for First Trial Session 1 and Last Trial Session 1 were 100%, the standard error for those trials is 0, and therefore no error bars are visible.

**Figure 45.** Mean Integrated task cognitive component (pattern formation) accuracy expressed as a percentage for all six patterns across the first Session 1 training trial (S1 First), the Session 1 last trial (S1 Last), the Session 2 retention test (S2 Retention Test), and the Session 2 final test (S2 Final Test).

A significant main effect was found for Trial, $F(1, 23) = 87.01, p < .001$, with the mean accuracy for the Session 2 Final Test being significantly greater ($M = 94.1\%$) than mean accuracy for the Session 2 Retention Test ($M = 51.4\%$). The main effect for Pattern was not significant, and no significant interaction effect was detected. These results indicate significant recovery of the cognitive skill component during retraining.

Additionally, in order to determine whether accuracy performance differed significantly from Session 1, for which all trials were performed with 100% accuracy, and the S2 Retention
Test, one-sample t-tests were used to compare retention test accuracy means for each pattern to 100. Additionally, in order to assess whether accuracy for each pattern following Session 2 retraining (S2 Final Test) significantly differed from baseline performance (100%), one-sample t-tests were also used to compare final test accuracy means for each pattern to 100. The results of these analyses are presented in Table 7. These results demonstrate significant performance degradation for the cognitive accuracy component of the Integrated task during the 3-week retention period. These results also demonstrate that following Session 2 retraining, accuracy for the cognitive component of the Integrated was only significantly different from Session 1 baseline performance for Pattern 5 \( (p = .050) \); performance for all other patterns returned to a level equivalent to the 100% accuracy performance demonstrated at baseline.
Table 7

One Sample t-tests comparing mean Integrated task retention test and final test pattern accuracy to 100% across all participants (n=24)

<table>
<thead>
<tr>
<th>Trial</th>
<th>t</th>
<th>df</th>
<th>p*</th>
<th>Mean Difference</th>
<th>95% Confidence Interval (Lower/Upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention Pattern 1</td>
<td>-5.87</td>
<td>23</td>
<td>&lt; .001</td>
<td>-50.00%</td>
<td>-67.60% -32.40%</td>
</tr>
<tr>
<td>Retention Pattern 2</td>
<td>-4.70</td>
<td>23</td>
<td>&lt; .001</td>
<td>-41.67%</td>
<td>-60.00% -23.30%</td>
</tr>
<tr>
<td>Retention Pattern 3</td>
<td>-7.23</td>
<td>23</td>
<td>&lt; .001</td>
<td>-62.50%</td>
<td>-80.40% -44.60%</td>
</tr>
<tr>
<td>Retention Pattern 4</td>
<td>-5.94</td>
<td>23</td>
<td>&lt; .001</td>
<td>-52.08%</td>
<td>-70.20% -34.00%</td>
</tr>
<tr>
<td>Retention Pattern 5</td>
<td>-3.76</td>
<td>23</td>
<td>0.001</td>
<td>-33.33%</td>
<td>-51.70% -15.00%</td>
</tr>
<tr>
<td>Retention Pattern 6</td>
<td>-5.35</td>
<td>23</td>
<td>&lt; .001</td>
<td>-52.08%</td>
<td>-72.20% -31.90%</td>
</tr>
<tr>
<td>Final Pattern 1</td>
<td>N/A</td>
<td>23</td>
<td>N/A</td>
<td>0.00%</td>
<td>N/A - N/A</td>
</tr>
<tr>
<td>Final Pattern 2</td>
<td>-1.00</td>
<td>23</td>
<td>0.328</td>
<td>-2.08%</td>
<td>-6.40% 2.20%</td>
</tr>
<tr>
<td>Final Pattern 3</td>
<td>-1.37</td>
<td>23</td>
<td>0.185</td>
<td>-6.25%</td>
<td>-15.70% 3.20%</td>
</tr>
<tr>
<td>Final Pattern 4</td>
<td>-1.81</td>
<td>23</td>
<td>0.083</td>
<td>-6.25%</td>
<td>-13.40% 0.90%</td>
</tr>
<tr>
<td>Final Pattern 5</td>
<td>-2.07</td>
<td>23</td>
<td>0.050</td>
<td>-14.58%</td>
<td>-29.20% 0.00%</td>
</tr>
<tr>
<td>Final Pattern 6</td>
<td>-1.37</td>
<td>23</td>
<td>0.185</td>
<td>-6.25%</td>
<td>-15.70% 3.20%</td>
</tr>
</tbody>
</table>

Note: *2-tailed significance.

The accuracy analysis for the Integrated task also included a 4 (Trial) x 6 (Pattern) RM ANOVA for mean motor accuracy (triangles dropped) using the FLS box. The four levels for Trial were Session 1 First Training Trial From Memory, Session 1 Last Training Trial From Memory, Session 2 Retention Test, and Session 2 Final Test. Figure 46 presents mean Integrated task motor accuracy, expressed as a percentage (out of 6), for all six patterns across four trials in graphical form with error bars indicating the standard error of the mean for each trial.
Figure 46. Mean Integrated task motor component accuracy expressed as a percentage of triangles dropped (out of 6) for all six patterns across the first Session 1 training trial (S1 First), the Session 1 last training trial (S1 Last), the Session 2 retention test (S2 Retention Test), and the Session 2 final test (S2 Final Test).

Mauchly’s test of sphericity was significant ($p < 0.05$) for Trial and Trial x Pattern; therefore, statistical significance was assessed based on the Greenhouse-Geisser adjustment. No significant main effect was present for Trial or Pattern, and no significant interaction was present. As demonstrated in Figure 46, mean motor accuracy within the Integrated task was near ceiling at all time points assessed and across all patterns ($M = 95.75\%$). This result indicates that significant variations in Integrated task performance are not related to triangles dropped.
**Comparative task analysis.** In order to specifically explore relative skill acquisition during the initial training session and decay for the individual and integrated task skills during the 3-week retention period, the overall mean trial times and relative changes in trial times were compared across the three tasks. While the previously reported statistical analyses for each task demonstrated significant changes in task trial times within tasks, a direct comparison across tasks is provided here in order to specifically address research question 2: For the specified task paradigm, does the cognitive skill component of the integrated task demonstrate significantly more decay than the psychomotor skill component following a brief (3 week) retention period, and does significantly more relative skill decay occur for the integrated task as compared to the psychomotor or cognitive subtasks in isolation?

In order to account for the variation in time scales across tasks, a 3 (Task) x 2 (Time Period) RM ANOVA was conducted, comparing the percent change in mean trial segment times for each task between S1 First and S1 Last, as well as between S1 Last and S2 Retention. To determine time-based performance improvement over the course of training, the percent decrease in mean trial time was calculated by subtracting mean trial time for the last Session 1 trial (S1 Last) from the first Session 1 trial (S1 First); that number was then divided by the S1 First mean trial time and multiplied by 100 to express the change as a percentage. To determine time-based performance degradation over the retention period, the percent increase in mean trial time was calculated by subtracting mean trial time for the last Session 1 trial (S1 Last) from the Session 2 retention test trial (S2 Retention); that number was then divided by the S1 Last mean trial time and multiplied by 100 to express the change as a percentage. Mauchly’s test of
sphericity was significant ($p < 0.05$) for Task and Trial x Time Period; test statics are reported using the Greenhouse-Geisser adjustment. A significant main effect was found for Task, $F(1, 32) = 17.68, p < .001$, with the greatest overall change in mean trial times for the Cognitive task ($M = 57.3\%$), followed by the Psychomotor task ($M = 36.7\%$), and with the smallest overall change in mean trial time being for the Integrated task ($M = 23.6\%$) across both time periods assessed. Simple contrasts with Bonferroni adjustment, comparing each of the constituent tasks to the Integrated task, indicated significant differences in percent change between the Integrated task and the Psychomotor task, $F(1, 23) = 11.61, p = .002$, and well as between the Integrated task and the Cognitive task, $F(1, 23) = 36.52, p < .001$. A significant main effect was not found for Time Period; however, a significant interaction was present for Task x Time Period, $F(2, 36) = 88.81, p < .001$. Post hoc contrasts revealed a significant interaction between the Psychomotor and Integrated tasks for the Training time period compared to the Retention time period. As shown in Figure 47, during the Training time period, the percent change in the Psychomotor task was largest ($M = 66.1\%$), followed by the Cognitive task ($M = 35.3\%$), and with the smallest change occurring for the Integrated task ($M = 7.3\%$) from the first trial of Session 1 to the last trial of Session 1. Conversely, for the Retention time period, the percent change in the Integrated task trial mean times was significantly greater ($M = 40.0\%$) than the percent change in the Psychomotor task mean trial times ($M = 7.3\%$), and was significantly smaller than the percent change in the Cognitive task mean trial times ($M = 79.2\%$) from the last trial of Session 1 to the Session 2 Retention test.
Figure 47. Mean change in task trial segment times expressed as a percentage for all three tasks across two trials occurring before and after the 3-week retention period.

These results directly address research question 2, indicating that time-based skill decay for the 3-week retention period was greatest for the Cognitive task, followed by the Integrated task, and with minimal skill decay occurring for the Psychomotor task, based on mean trial times. Based on the previous analyses, the time-based skill decay for both the Cognitive and Integrated tasks was significant; this analysis served to simply compare these significant changes as a percent-based change across tasks. However, it is again important to note that some rehearsal of the psychomotor and cognitive skills took place during the Integrated task retention test.

In terms of task accuracy, as shown in Figure 48, the Cognitive task demonstrated the greatest improvement (26%) for Session 1 training trials completed from memory. The Psychomotor task demonstrated a smaller but significant improvement in accuracy (13%), as
measured by triangles dropped per trial subtask. The cognitive component of Integrated task accuracy (pattern formation accuracy) was required to be at criterion (100%) for all Session1 training trials; therefore, no change was present during the skill acquisition phase; additionally, no significant effect was present for the motor component of Integrated task accuracy (triangles dropped), which also demonstrated a ceiling effect during Session 1 training.

Figure 48. Mean task trial segment accuracy (%) for all three tasks across three trials occurring at the beginning of Session 1 training, at the end of Session 1 training, and at the beginning of Session 2 training, following the 3-week retention period.

Figure 48 also demonstrates relative task accuracy decay based on a comparison of Session 1 Last Trial to Session 2 Retention Test. The Integrated task demonstrated the greatest decline (35%) for the cognitive accuracy component of the task between S1 Last Trial and S2 Retention Test. As previously reported no significant change was detected for the psychomotor
component of the Integrated task across trials. The Cognitive task also exhibited a significant but smaller decline (23%) in accuracy. No significant decay was detected for Psychomotor task accuracy; in fact, a non-significant increase (7%) in Psychomotor task accuracy was observed. Relative accuracy attrition was significantly greater for the Integrated task than for the Cognitive task in isolation, suggesting greater decay overall for the cognitive skill within an integrated context, although the integrated retention test may have boosted performance on the cognitive retention test.

Next, a direct comparison across tasks was conducted to provide some insight into the processes underlying performance of the individual versus integrated tasks, and thus is worth consideration on a cursory level. Despite the fact that the overall time scales for the various tasks differ, the factor of interest in this comparison was the interactive effects between tasks (Psychomotor, Cognitive, and Integrated). Mean trial segment times for each task across the 3 trials of interest are displayed in Figure 49, with error bars indicating standard error. A segment of the Psychomotor task was defined as the time to transfer 6 triangles from one side of the peg board to the other; given that no significant differences were detected for dominant versus non-dominant hand for the Psychomotor task, the mean time was taken across both hands. A segment of the Cognitive task was defined as the mean time to form a single pattern of six virtual triangles using the tablet app. A segment of the Integrated task was defined as the mean time to form a single pattern by transferring six triangles from one side of the peg board to the other.
Figure 49. Mean task trial segment times in seconds for all three tasks across three trials occurring at the beginning of Session 1 training, at the end of Session 1 training, and at Session 2 retention testing following the 3-week retention period.

The mean trial times presented in Figure 49 indicate mean trial segment times by task at critical points before and after integrated training trials in an effort to better understand the effects of task integration across training and retention periods. The first set of trials (Last Individual Training Trials & First Integrated Trial) represent that last training trials for the individual (Psychomotor and Cognitive) tasks, as well as the first Integrated task trial, which immediately followed the last trial of the Cognitive task training. If the mean times for the Psychomotor and Cognitive task trial segments are added together, the total (45.9 seconds) is nearly identical to the mean trial segment time for the first Integrated task trial (45.8 seconds).
This simple analysis suggests that the integration of the two individual tasks within the Integrated task likely has a simple additive effect on task performance time.

The second set of trials presented in Figure 49 (Last Integrated Trial & Individual PostTest Trials) represents mean trial segment time for the third and final Integrated task training trial in Session 1, as well as the mean trial segment times for the individual (Psychomotor and Cognitive) task PostTests, which were performed immediately following the final Integrated task training trial in order to assess any performance changes in the individual tasks resulting from rehearsal of the Integrated training trials. In this case, if the individual task mean trial times are added together, the total (47.7 seconds) is less than the mean Integrated task trial segment time for the immediately preceding integrated skill trial. This may suggest that on the final Integrated task training trial, the individual task components no longer had a strictly additive effect, and had perhaps become increasingly integrated and proceduralized as a single task, requiring less time to perform overall.

The third set of trials presented in Figure 49 (S2 Integrated and Individual Retention Test Trials) represent mean trial segment times for the Session 2 Integrated task retention test and individual (Psychomotor and Cognitive) task retention tests, which immediately followed. In this case, if the individual mean trial segment times are added together, the total (53.8 seconds) is less than the mean trial segment time for the Integrated task (57.4 seconds). This indicates that the integration of the two constituent tasks following the 3-week retention period may have had more than a simple additive effect. It is also possible that the rehearsal provided by the single trial of the Integrated task during performance of the Integrated retention test
resulted in some skill recovery, which may have been reflected in faster trial times for the subsequent individual task trials. While this particular issue was not identified as an objective of the current research effort, these findings provide an interesting perspective warranting further investigation.

Retraining condition. The following analyses sought to directly address research question 3: Do both hands-on retraining and video-based retraining lead to significant skill recovery, and does hands-on retraining result in greater recovery of skills? A 2 (Retraining Condition) x 2 (Trial) mixed factors ANOVA was conducted for mean time for the Integrated task. The levels for condition were video-based retraining (n=12) and hands-on retraining (n=12). The levels for Trial included S2 Retention Test and S2 Final Test. Mean times for each retraining condition group across the three trials of interest are displayed in Figure 50 with error bars indicating standard error.

![Figure 50. Mean Integrated task trial times in seconds for video and hands on retraining groups across two trials: after retention and after retraining.](image-url)
A significant within-subjects effect was found for Trial $F(1, 22) = 95.51, p < .001$, with S2 Final Test having faster mean trial time ($M = 40.91$) than S2 Retention Test ($M = 57.38$). A significant between-subjects effect was also present, with participants randomly assigned to the hands-on retraining condition exhibiting faster trial mean trial times ($M = 44.78$) than participants randomly assigned to the video retraining condition ($M = 53.51$) overall. No significant within-subjects interaction was present for Trial x Condition.

In order to assess the effect of retraining condition directly, controlling for the between groups difference, which existed prior to retraining, the relative change in mean integrated times was calculated for inter-trial time periods prior to retraining, representing changes in mean times occurring over the course of the retention period (S2 Retention Test- S1 Last), as well as the retraining inter-trial period (S2 Final Test - S2 Retention Test). These mean time differences, displayed in Figure 51 with error bars indicating standard error, were then used to conduct a 2 (Inter-Trial Time Period) x 2 (Training Condition) RM ANOVA.
Figure 51. Mean time differences in Integrated task trial times in seconds for video and hands on retraining groups across two inter-trials intervals: over the retention period and during Session 2 (S2) retraining.

No significant main effect was found for Inter-Trial Time Period. The test for main effect of Training Condition did not approach significance ($F<1$), nor did the test for an interaction between Time Period and Training Condition ($F<1$). Furthermore, the 95% confidence intervals for video-based training condition ($CI=5.53$ to $12.45$) and for hands on training condition ($CI=6.28$ to $13.20$) were almost completely overlapping, indicating that it is highly unlikely that the true population means for these two groups differ.

These results support the hypothesis that both groups would improve significantly following retraining. The additional hypothesis that the hands on retraining group would
demonstrate even greater recovery of skill than the video-based retraining group is not supported by these results.

**Glove-based metrics.** While all participants wore instrumented gloves for all trials, in several cases the glove software did not record properly. Additionally, repeated donning and doffing of the gloves by multiple participants caused wires within the gloves to become loose over time, resulting in incomplete or faulty data (missing trials or accelerometer axes). Thus, complete glove data sets for all trials of Session 1 and Session 2 were only obtained for a total of 13 participants. Despite low power, velocity and angular rotation were computed for all Psychomotor task trials for complete data sets, and these data sets and statistical analyses were conducted for all trials of these data sets. The gloves were also worn during completion of the Integrated task trials across both sessions; however analysis of these data was not within the scope of the current research effort.

A 2 (Hand) x 4 (Trial) RM ANOVA for glove-based hand velocity was conducted. Levels for Hand included dominant and non-dominant hand, indicating the hand from which the sensor data samples were acquired. Levels for Trial included S1 First Trial, S1 Mean Last (mean velocity for the last 5 trials of Session 1), S1 Post-Test, and S2 Retention. Figure 52 presents mean velocity in centimeters per second for the dominant and non-dominant hands across the four trials of interest for the Psychomotor task.
Figure 52. Mean velocity for the dominant and non-dominant hands for the first Session 1 trial (S1 First), mean of the last 5 Session 1 trials (S1 Mean Last), Session 1 post-test trial (S1 PostTest), and the Session 2 retention test (S2 Retention) for the Psychomotor task.

Mauchly’s test of sphericity was significant ($p < 0.05$) for Hand x Trial; therefore, statistical significance was assessed based on the Greenhouse-Geisser adjustment. A significant main effect was found for Trial, $F(3, 33) = 9.83, p < .001$, with mean velocity being fastest for S1 PostTest ($M = 9.27$), followed by S1 Mean Last ($M = 9.23$), and S2 Retention ($M = 8.77$), and with the slowest velocities observed for S1 First Trial ($M = 7.14$). A planned post hoc simple contrasts analysis was conducted, comparing S1 First Trial to each subsequent trial in order to assess changes in velocity over time, compared to baseline velocity early in training. The Bonferroni adjustment was used to account for familywise error. This analysis resulted in significant contrasts comparing S1 First to S1 Mean Last, $F(1, 11) = 15.69, p = .002$; as well as
to S1 PostTest, $F(1, 11) = 22.53, p = .001$; and to S2 Retention Test, $F(1, 11) = 10.54, p = .008$.

This is to be expected given that trial times were becoming faster over the course of training. Additionally, a repeated post hoc contrasts analysis was conducted in order to determine whether changes in mean velocity from one trial to the next were significant. Of particular interest was the change between the end of Session 1 (S1 PostTest) and the beginning of Session 2 (S2 Retention) as this would indicate any significant change in hand velocity related to skill decay. However, only the repeated contrast for S1 First versus S1 Mean Last (reported above) was significant.

More interestingly, a significant main effect was found for Hand, $F(1, 11) = 45.82, p < .001$, with velocities for the non-dominant hand being faster ($M = 10.56$) than for the dominant hand ($M = 6.64$) across all trials. This finding is particularly pertinent given that the performance data indicate no significant differences for task performance between the dominant and non-dominant hands. No significant interaction was present for Trial x Hand; however, a trend is evident for increased velocity for the non-dominant hand at the end of Session 1 and Session 2 retention.

A 2 (Hand) x 4 (Trial) RM ANOVA for glove-based angular rotation was also conducted. Levels for Hand included dominant and non-dominant hand, indicating the hand from which the sensor data samples were acquired. Levels for Trial included S1 First Trial, S1 Mean Last (mean angular rotation for the last 5 trials of Session 1), S1 Post-Test, and S2 Retention. Figure 52 presents mean angular rotation in degrees for the dominant and non-dominant hands across the four trials of interest.
Mauchly’s test of sphericity was significant ($p < 0.05$) for Trial and Hand x Trial; therefore, statistical significance was assessed based on the Greenhouse-Geisser adjustment. No significant main effect was found for Trial; however, a significant main effect was again found for Hand, $F(1, 11) = 13.70, p = .003$, with mean angular rotation for the non-dominant hand being greater ($M = 16.42$) than for the dominant hand ($M = 6.01$). As with the result for velocity, this finding is particularly pertinent given that the performance data indicate no significant differences for task performance between the dominant and non-dominant hands. No significant interaction was present for Trial x Hand; however a trend is apparent for
decreased angular rotation for the non-dominant hand for trials at the send of Session 1 and at Session 2 retention testing compared to the first trial of Session 1.

**Experiment 2 Discussion**

Experiment 2 expanded on the outcomes of Experiment 1, extending the area of study beyond strictly psychomotor skill, to include assessments of acquisition and attrition for cognitive and integrated skills. This Experiment again included LS psychomotor skill assessments for the FLS peg transfer task using task decomposition-based and instrumented glove-based metrics, providing continuity and a basis for comparison with the previous experiment. In addition, Experiment 2 introduced a novel Cognitive task and novel Integrated skills task in order to empirically assess the relative retention of psychomotor and cognitive skill components within an integrated/concurrent task relevant to LS, as well as in isolation. Unlike a dual task paradigm in which a primary and secondary task are performed simultaneously but with task switching, the Integrated task designed for Experiment 2 consisted of a psychomotor skill component and a cognitive skill component that can be performed in an interdependent concurrent manner as well as in isolation. Additionally, the cognitive skill component requires long-term retention of declarative knowledge (i.e., individual patterns) and procedural knowledge (i.e., order of patterns), in addition to adding cognitive load and a working memory component to the integrated task during skill acquisition and retention testing.

The novel Integrated skill task was designed to emulate the cognitive task demands of recalling steps in a surgical procedure while concurrently performing a psychomotor skills task.
relevant to LS. As detailed in Appendix B, the skill decay literature indicates that within the context of procedural skills, the number of steps represents a primary indicator of task complexity, and by extension, skill retention (Hurlock & Montague, 1982; Hagman & Rose, 1983; Druckman & Bjork, 1991; Wisher et al., 1999). Task complexity, and thereby retention, has also been shown to be related to the level of integration of steps (Annett, 1979); whether steps must be performed in a specific sequence and whether built-in feedback is provided regarding correct performance of task steps (Wisher et al., 1999); and the extent to which cueing is provided to indicate next steps (Hagman & Rose, 1983; Druckman & Bjork, 1991). Both the Cognitive and Integrated tasks required recall of six steps (i.e., pattern) in order, from memory. In both tasks one step consistently preceded the next (e.g., when forming Pattern 3, the triangles started in the Pattern 2 formation); however, individual steps/patterns did not inherently provide cueing for the subsequent step/pattern. Within a surgical task, the one step may inherently cue the subsequent step in some cases. For example, placing a clamp on an artery may intuitively be followed by cutting that artery based on training and general surgical knowledge. While such cueing was not inherently present within the Integrated task, some participants may have developed individual strategies for cueing transitions from one pattern to the next. Additionally, given that the patterns were numbered and always performed in the same order, some participants may have developed mnemonics to associate each formation with its assigned pattern number. In this case, participants would rely on recall of the associations between individual patterns and their assigned numbers, rather than on relationships or transitions between one pattern and the next. The Integrated task was specifically used to test the
hypothesis that the cognitive skill component would demonstrate significantly more decay than the psychomotor skill component following a 3-week retention period, and that significantly more relative skill decay would be observed for the Integrated task as compared to the Psychomotor task or Cognitive task in isolation.

While Experiment 1 indicated that dropped triangles were highly correlated to total task time, scoring of dropped triangles was included in the Experiment 2 analyses in order to support diagnosis whether increased times with each hand were due to overt errors or simply slower movement. Additionally, the inclusion of an accuracy component for the Psychomotor task was included to differentiate between changes in Integrated performance related to underlying cognitive and psychomotor components. The Psychomotor task, performed independently as in Experiment 1, demonstrated significant improvement in both time and accuracy (dropped triangles) over the course of Session 1 training; and as anticipated, no significant decay was observed for time or accuracy following the brief (3-week) retention period. Additionally, no effect was detected for comparative subtask performance with the dominant versus non-dominant hand for time or accuracy across all 4 time points, which included the first Session 1 training trial, the mean of the fastest error-free trials at the end of Session 1, the Session1 PostTest trial (following training on the Integrated task, which inherently involved rehearsal of the Psychomotor skill), and the Session 2 retention test trial, which again occurred following testing of the Integrated skill task, providing some psychomotor skill rehearsal. The lack of a consistent effect for dominant versus non-dominant hand performance across individual trials may again be representative of variability within a novice population (university undergraduate
students) during the early stages of skill acquisition, which has been shown in previous research (Kowalewski, 2012).

Despite the lack of a significant effect for subtask time and error-based differential performance between the dominant and non-dominant hands, and despite a smaller sample size for glove-based metrics, statistically significant effects were detected for left versus right hand velocity and angular rotation across all assessment trials. As in Experiment 1, both hand velocity and angular rotation were significantly higher for the non-dominant hand across all time points. This finding is particularly pertinent given that the performance data indicate no significant differences for task performance between the dominant and non-dominant hands. Furthermore, a significant main effect was detected for the time of assessment for velocity, with the first training trial of Session 1 demonstrating significantly lower overall velocity than the subsequent trials assessed across both hands. These results are consistent with the Experiment 1 results and may indicate that the dominant hand motion profile was characterized by more efficient and precise movements (i.e., fewer and more direct), with the non-dominant hand motion profile reflecting larger, speeded movements that were less precise. As in Experiment 1, the increased velocity observed for the non-dominant hand later in training may indicate that attempts to increase task performance with the non-dominant resulted in movements that were faster, but not as precise or efficient as the dominant hand. These differences in dominant and non-dominant hand performance may have been most pronounced during the bimanual portion of the task (i.e., triangle transfers); however, this level of task decomposition was not conducted. As previously noted within the context of Experiment 1, these findings are
consistent with previous research indicating that skilled LS psychomotor task performance is associated with more efficient, rather than higher velocity, hand movements (e.g., Cao et al., 1996; Bann et al., 2003). Unlike Experiment 1, a significant interaction effect was not detected, indicating equivalent increases in velocity for both hands over time for the current data set. This is likely due to the fact that Experiment 1 included more training overall on this task (a total of 60 minutes in Session 1 and 40 minutes in Session 2) than was provided in Experiment 2 (only 40 minutes total within Session 1).

Like the Psychomotor task, the Cognitive task also demonstrated significant improvement in time and accuracy over the course of Session 1 training; however, performance for this task (both time and accuracy) declined significantly over the course of the 3-week retention period, returning to baseline or worse across all six trained patterns for both time and accuracy. Additionally, a significant effect was observed for individual pattern performance in terms of time, but not accuracy, across all time points. Thus, the participants did not have increased difficulty learning or recalling any particular patterns over others, but were able to recall and form some patterns faster than others. Specifically, the time to form Pattern 5 was significantly faster than Pattern 4; and Pattern 1 was significantly faster than all other patterns, demonstrating a primacy effect. Pattern 1 could also be considered a simpler pattern.

In terms of time to complete trials, the Integrated task demonstrated significant improvement over the course of Session 1 training, significant decline during the retention period, and significant improvement following retraining, with mean times returning to baseline or better for the final Session 2 trial across all participants. These significant performance
improvements within Session 1 training are particularly striking given the fact that only three complete training trials (each consisting of formation of all six patterns) were performed. Unlike the Cognitive task, no significant effect was found for individual pattern formation time; however, a significant interaction was demonstrated in which Patterns 1, 2, and 4 demonstrated greater improvement in mean pattern formation time during Session 1 training, greater decay during the retention period, and then greater improvement during retraining than Patterns 3, 5, and 6.

Accuracy for the Integrated task was assessed in terms of both cognitive subtask accuracy (pattern formation accuracy) and psychomotor subtask accuracy (triangles dropped). No effect was found for psychomotor accuracy across all time points assessed; however, a significant effect was present for cognitive subtask accuracy. Perfect pattern formation accuracy was required at the first two time points for inclusion in the final data set for analysis; this was done to ensure that the patterns had been learned sufficiently prior to performance of the Integrated task and prior to the retention period. Significant decline was observed in accuracy for the cognitive component of the Integrated task at retention, followed by significant improvement after retraining. As with the Cognitive task in isolation, no significant accuracy effect was detected for specific patterns, and no interaction was present within the context of the Integrated task.

Analyses of relative skill acquisition and decay across the three tasks revealed significant variation across the tasks in terms of both time and accuracy. Over the course of Session 1 training trials assessed, the Psychomotor task demonstrated the largest change in
mean trial time (66%) during initial skill acquisition, followed by the Cognitive task (35%), with the Integrated task demonstrating the smallest change in mean trial time (7%) between the first and last Session 1 training trials. In terms of task accuracy, the Cognitive task demonstrated the greatest improvement (26%) for Session 1 training trials completed from memory. The Psychomotor task demonstrated a smaller but significant improvement in accuracy (13%), as measured by triangles dropped per trial subtask. The cognitive component of Integrated task accuracy (pattern formation accuracy) was required to be at criterion (100%) for all three Session 1 training trials; therefore no change was present during the skill acquisition phase; additionally, no significant effect was present for the motor component of Integrated task accuracy (triangles dropped), which also demonstrated a ceiling effect during Session 1 training.

It is important to note that Session 1 training for the Psychomotor task consisted of two 20-minute training blocks, for a total of 40 minutes of training time, in which participants completed 20 trials, on average. The total Cognitive task training time was shorter (approximately 20 minutes), and the period of training assessed in these results, which began with the first trial performed completely from memory (Part C of the task) took place over the course of less than 10 minutes, on average. Similarly, Integrated task training in Session 1 consisted of only three complete trials over the course of approximately 20 minutes of training time.

Relative skill decay across the three tasks following the 3-week retention period was assessed based on percent change in mean trial time and accuracy from the last trial performed
at the end of Session 1 to the retention test performed at the beginning of Session 2. The
greatest relative change in trial time was observed for the Cognitive task, with a 79% increase in
mean trial time at retention. The Integrated task also demonstrated significant time-based skill
decay, with an increase in mean trial times of 40%. A significant difference was not detected in
Psychomotor mean trial time following the retention period; however, a 7% increase in mean trial time was observed. In terms of relative task accuracy decay, the Integrated task
demonstrated the greatest decline (35%) for the cognitive accuracy component of the task, and
no significant decline for the motor accuracy component of the task. The Cognitive task also
exhibited a significant but smaller decline (23%) in accuracy. No significant decay was
detected for Psychomotor task accuracy, which demonstrated a 3% increase in accuracy
following the retention period, based on mean number of triangles dropped. These results
indicate that relative accuracy attrition was significantly greater for the Integrated task than for
the Cognitive task in isolation, suggesting greater decay overall for the cognitive skill within an
integrated context. However, it is important to note that the Integrated retention test was
conducted prior to the Psychomotor and Cognitive retention tests in Session 2, providing
psychomotor and cognitive skill rehearsal and feedback during the formation of the six triangle
patterns within the context of the Integrated task. The Integrated task provided the equivalent of
three Psychomotor task trials and the equivalent of one Cognitive task trial, with feedback
indicating all six correct patterns. Given that participants were shown the correct pattern for
any incorrect patterns formed during the Integrated retention test just prior to completing the
Cognitive retention test, the low accuracy scores for Cognitive task retention are even more
striking. This finding indicates that performance of the Integrated (whole) task may have introduced an interference effect by which participants were unable to focus on relearning the patterns they had forgotten while also completing the psychomotor subtask.

A potential interference effect was further explored, on a cursory level, by comparing mean trial segment times for the individual (Psychomotor and Cognitive) and Integrated tasks for trials performed in succession (back to back) for the first trial of the Integrated task in Session 1, the last trial of the task in Session 1, and for the Session 2 retention test. While this examination could not be tested statistically, trends were observed suggesting a strictly additive effect for the first Integrated task trial, increased task component integration for the last Session 1 task trial, and a compounded effect for reintegration of the task components within the Session 2 retention test in which the added time to complete the Integrated task was more than a simply additive factor of the constituent task mean trial times. While the current experiment tested the Integrated skill retention, it is also noted that completing a single trial of the constituent (part) tasks prior to attempting the Integrated task may have produced greater skill recovery of the cognitive task component within the Integrated task at retention; however, this was not tested.

While part versus whole task retraining was not assessed within the current study, the comparative effects of two forms of whole task retraining were studied. In particular, a primary objective of Experiment 2 was to assess the comparative effectiveness of video-based instruction to physical (hands on) retraining of the integrated skill; it was specifically hypothesized that both hands on retraining and video-based retraining would lead to significant
skill recovery. Results indicated significant skill recovery for both the video-based and hands-on retraining groups in terms of both mean trial time and cognitive subtask component accuracy (i.e., pattern formation). It was further hypothesized that the hands on retraining group would demonstrate even greater recovery of skill than the video-based retraining group, which is not supported by the results obtained in the current study. The randomly assigned hands on retraining group demonstrated overall faster mean trial times and accuracy than the video-based retraining group across all time points assessed, both before and after retraining. However, the relative improvement in performance following three Integrated task retraining trials in Session 2, lasting less than 20 minutes on average, was equivalent for the two groups.

Demographic data analyses indicated that self-reported amount of video game play was not significantly correlated to measures of task performance within Session 1 or Session 2. However, self-reported instrument playing skill was significantly correlated to the number of pegs dropped within the Psychomotor task Session 1 PostTest for the trial segment led by the non-dominant hand. The lack of a significant correlation to Psychomotor performance on the first trial of Session 1 is consistent with findings by Madan, Harper, Frantzides, and Tichansky (2008), which demonstrated that nonsurgical skills did not predict baseline scores in inanimate box or virtual reality trainers. However, Madan et al. (2008) only assessed baseline scores; whereas the current assessment demonstrated a significant correlation for post-training skills. Additionally, Madan et al. (2008) performed a purely categorical assessment, comparing for example performance for participants who did and did not play video games or musical instruments. It is possible that frequency and level of skill in non-surgical skills may be more
predictive of baseline scores in inanimate and virtual reality trainers. Additionally, Madan et al. (2005) demonstrated a significant correlation between chopstick use, based on a survey of self-reported nonsurgical dexterity skills and performance of a specific LS task. Future planned studies will inquire about, and perhaps test, chopstick use, as a potentially predictive measure. Multiple significant correlations were also detected between the matching to sample spatial ability test variables and performance on aspects of both the Psychomotor and Integrated tasks, but not the Cognitive task. Particularly relevant significant correlations were found for the relationship of both matching to sample percent correct and reaction time for correct trials with Integrated task Session 2 retention test mean trial time. Specifically, higher percent correct and faster reaction times for correct trials on the matching to sample test correlated to faster Integrated retention test trial times. Additionally, slower matching to sample correct trial reaction times correlated to larger increases in trial times between the end of Session 1 to the Session 2 retention test for the Integrated task. This finding indicates that the matching to sample test may provide a useful measure for predicting skill decay for complex and integrated skill tasks.
General Discussion

Overall, the objective of this research was to begin to address several critical gaps in the skill decay and retraining research literature, particularly within the context of laparoscopic surgery (LS) skills. Specifically, needs were identified for 1) objective and sensitive metrics to support assessment of LS skill loss based on the underlying cognitive, psychomotor, and perceptual skill components, and 2) validated pedagogical approaches to retraining these perishable skills. As emphasized by Smith et al. (2002), “the ability to make an objective evaluation of a surgeon’s operative ability remains an elusive goal (p. 640).” Smith et al., (2002) demonstrated that motion analysis metrics, including “speed of movement” provide potential objective measures for effectively assessing laparoscopic dexterity, specifically within the context of detecting novice skill acquisition. Therefore, assessment of the effectiveness of motion-based metrics compared to traditional metrics, within the context of not only skill acquisition but also skill decay was a primary objective of the current research.

Three Specific Aims were identified for the current research effort, targeting three distinct research questions within the framework of a laboratory-based LS relevant task. Aim 1 of this research was to investigate the use of task decomposition-based and instrumented glove-based metrics to assess individual hand performance (time and errors) for a standardized LS manual skills training task using an inanimate box training platform. The specified metrics were used to compare LS psychomotor skills for the dominant and non-dominant hands in order to assess LS psychomotor skill acquisition and attrition. Aim 2 was to develop a better understanding of the retention of integrated skill components relevant to LS. Using a modified FLS manual skill task that includes a cognitive skill component, Aim 2 specifically sought to
test the hypothesis that the cognitive skill component would demonstrate significantly more decay than the psychomotor skill component following a retention period, and that significantly more relative skill decay would be observed for the Integrated task as compared to the Psychomotor task or Cognitive tasks in isolation. Finally, Aim 3 was to assess the comparative effectiveness of video-based instruction to physical retraining of the integrated LS skill; it was hypothesized that both physical retraining and video-based retraining would lead to significant skill recovery, with greater recovery of skills in the physical retaining group. These Aims and the study results are addressed within the context of the stated research questions.

**Research Question 1: Task Decomposition-Based and Glove-Based Metrics**

The first research question was 1) Do task decomposition-based and hand motion-based metrics provide more sensitive means than traditional metrics of total task completion time and overt errors for assessing skill acquisition and decay within the context of LS psychomotor task performance? Cao et al. (1996) demonstrated an effective approach to assessing performance of LS skills using task analysis and decomposition-based methods and motion analysis. Cao et al. specifically demonstrated that task decomposition and motion analyses of subtasks illuminated key differences in task performance between an expert and a sample of novice laparoscopic surgeons for four fundamental surgical tasks. Key differentiating factors, based on this analysis, included precision and safety constraints, variations in tool manipulation, tool positioning and orientation, number of motions, and serial versus parallel execution of motions. While the current study did not include an expert for comparison, the intent was to leverage this approach, comparing novice performance across various time points in training in order to identify similar
factors for comparisons of novice trainees progressing towards proficiency. A subsequent objective of the study was to compare performance by those same novices following retention periods of varying lengths in order to determine whether regression to performance characteristics observed early in training were evident. Of particular relevance from Cao et al.’s task analysis was the decomposition of psychomotor tasks into constituent subtasks, and specifically the identification of subtasks requiring “precision constraints” and “safety constraints”. A similar methodology was applied within the current study for the selected psychomotor task (FLS peg transfer), resulting in the differentiation of subtasks based on precision constraints (i.e., removing triangles from pegs and holding/releasing triangles during transfers versus placing triangles on pegs and grasping pegs during transfers) and potential safety constraints (i.e., dropped triangles). Additionally, the current study leveraged the motion analysis approach presented by Cao et al. by identifying motion-based indices of performance likely to vary based on level of training for the specified task: velocity and angular rotation. The current study further specified analysis of the identified subtasks and motion-based indices for the dominant versus non-dominant hands in order to examine the role of ambidexterity in LS psychomotor skill acquisition and retention. Ambidexterity, or bilateral dexterity, has been identified as a critical component of LS proficiency and expertise (Rosser et al., 1997; Derossis, et al., 1998). Application of these metrics in Experiment 1 highlighted significant differences in performance for the dominant and non-dominant hands, including a trend for increased bilateral dexterity over the course of training, which were not detectable by traditional time and error metrics. Specifically, the task decomposition-based metrics detected faster performance for the dominant hand during completion of trial segments containing subtasks requiring increased
precision (placing triangles on pegs and grasping triangles during transfers) across all training blocks. In addition, the glove-based metrics indicated significantly higher velocity and significantly greater angular rotation for the non-dominant hand across all training blocks. These results taken together may indicate that the dominant hand motion profile was characterized by more efficient and precise movements (i.e., fewer and more direct), with the non-dominant hand motion profile reflecting larger, speeded movements that were less precise. In both experiments increased velocity was observed for the non-dominant hand later in training, which may indicate that attempts to increase task performance with the non-dominant resulted in movements that were faster, but not as precise or efficient as the dominant hand. These differences in dominant and non-dominant hand performance may have been most pronounced during the bimanual portion of the task (i.e., triangle transfers); however, this level of task decomposition was not conducted.

Bann et al. (2003) specifically demonstrated that the relationship between time and motion was variable depending on the task across three laparoscopic suturing tasks. When the number of movements in a minute (standardized movements per minute) were considered, both groups were found to work at a similar rate (i.e., velocity), depending on the task, implying that more experienced surgeons are more efficient with their movements, performing the same tasks with fewer movements rather than with higher speed. It is possible that a similar distinction exists between the more skilled (i.e., dominant) and less skilled (i.e., non-dominant) hands during acquisition of these skills. Related to this finding, Mackay et al. (2002) demonstrated differential task performance for speeded versus precise LS psychomotor tasks. Mackay et al. (2002) specifically demonstrated that overall time and number of movements were lower for
trials in which participants were instructed to perform the task as quickly as possible, favoring speed over precision (speeded). While these outcomes are generally in agreement with the results of the present study, without decomposing the task further to determine dominant and non-dominant hand performance for the speeded versus precise trials, these results cannot be directly correlated to the present study results. Additionally, while an overall decrease in trial time and decrease in number of movements necessarily indicates lower mean velocity over the entire task trial, it is also possible that high velocity movements were used, with pauses or slower velocity movements intermixed within the trials. Therefore, again a direct comparison cannot be drawn; however, these results are in agreement on the surface, indicating overall that speeded task performance is associated with faster mean velocity and fewer movements, while precise task performance is associated with slower mean velocity and an increased number of movements. Taken together with previous reports indicating that skilled LS psychomotor task performance is associated with more efficient, rather than higher velocity, hand movements (e.g., Cao et al., 1996), the glove data indicate more skilled performance with the dominant hand across all training periods, despite observations of equivalent time-based task performance later in training. This may have been particularly pronounced during the bi-manual portions of the task (i.e., triangle transfers); however, further task decomposition within the glove-based data would be required to assess this.

**Research Question 2: Relative Skill Decay**

Research Question 2 was: For the specified task paradigm, does the cognitive skill component of the integrated task demonstrate significantly more decay than the psychomotor
skill component following a brief (3 weeks) retention period, and does significantly more relative skill decay occur for the Integrated task as compared to the Psychomotor or Cognitive tasks in isolation? Experiment 2 assessed training and retention of the same Psychomotor task used in Experiment 1, but also introduced a novel Cognitive task and novel Integrated task to empirically assess the relative retention of psychomotor and cognitive skill components within an integrated/concurrent task relevant to LS, as well as in isolation. Relative skill decay across the three tasks, following a 3-week retention period demonstrated a significant increase in mean trial time for the Cognitive task and for the Integrated task, but no significant change in Psychomotor mean trial time. Significant decreases in accuracy were also observed following the retention period. Specifically, the Integrated task demonstrated a greater decrease in accuracy than the Cognitive or Psychomotor tasks in isolation, with the Psychomotor task exhibiting no significant change. This result suggests greater decay overall for the constituent skills within an integrated context. A potential interference effect was further explored in which trends were observed suggesting a strictly additive effect for the first Integrated task trial, increased task component integration for the last Session 1 task trial, and a compounded effect for reintegration of the task components within the Session 2 retention test in which the added time to complete the Integrated task was more than a simply additive factor of the constituent task mean trial times. It has been suggested that completing a single trial of the constituent (part) tasks prior to attempting the Integrated task may have produced greater skill recovery of the cognitive task component within the Integrated task at retention; although, this was not tested. These results are consistent with previous reports indicating greater decay for cognitive over physical skills (e.g., Arthur et al., 1998). More specifically, Wang et al. (2013) reported
that tasks consisting of primarily cognitive skill characteristics are subject to greater decay than tasks consisting of primarily psychomotor skill characteristics. The Integrated task designed and implemented within Experiment 2 was considered to have equally represented cognitive and psychomotor skill components, particularly given the precision requirements of the psychomotor task and previous reports by Cao et al. (1996) indicating that precision constraints increase perceived task difficulty. However, more training was provided for the Psychomotor task than the Cognitive and Integrated skills tasks, and thus the psychomotor skill may have been more proceduralized and more automated both before and after the retention period. The additional training for the Psychomotor task was necessary in order to develop the skill sufficiently prior to requiring performance of the psychomotor skill within the context of the Integrated task. Thus, the emphasis was on training both tasks to minimum proficiency levels while also providing equivalent training time on each task across participants, rather than providing equivalent training time on each task. It could also be argued that 40 minutes of training on the Cognitive task would have led to extensive overlearning of that task, while 40 minutes of training on the psychomotor task was simply enough to establish criterion levels of performance for most participants. Therefore, overall, the Integrated task seems to have consisted of equally represented cognitive and psychomotor components. Thus, the increased decay rates for the Integrated task indicate decay that is more closely associated with cognitive skills tasks than psychomotor skills tasks. This suggests that decay rates for complex, integrated tasks may be based on the “lowest common denominator” skill component, or the constituent task characteristic associated with the highest rate of decay. This is in agreement with the integrated theory of skill retention proposed by Kim et al. (2011), which also suggested
that empirical data, such as the data presented here, could be used in conjunction with cognitive architectures such as ACT (Anderson, 1982) to develop predictive models of skill decay.

**Research Question 3: Retraining Condition**

Research question 3 posed the following: do both hands-on retraining and video-based retraining lead to significant skill recovery, and does hands-on retraining result in greater recovery of skills? Following retention testing, Experiment 2 also assessed the comparative effectiveness of video-based retraining to physical (hands on) retraining of the integrated skill. It was specifically hypothesized that both hands on retraining and video-based retraining would lead to significant skill recovery. Results indicated significant skill recovery for both the video-based and hands-on retraining groups in terms of both mean trial time and cognitive subtask component accuracy (i.e., pattern formation). It was further hypothesized that the hands on retraining group would demonstrate even greater recovery of skill than the video-based retraining group, which is not supported by the results obtained in the current study. While the randomly assigned hands on retraining group demonstrated overall faster mean trial times and accuracy than the video-based retraining group across all time points assessed, both before and after retraining, the relative improvement in performance following three Integrated task retraining trials in Session 2, lasting less than 20 minutes on average, was equivalent for the two groups. Despite a small sample size, the confidence intervals for this result indicate that the true population means for these representative samples would also be equivalent. While replication with a larger sample is warranted, this finding provides a significant contribution to the research literature, building on previous work indicating the effectiveness of video-based
training within LS skills (Guerlain, et al., 2004; Xeroulis, et al., 2007), but demonstrating equivalent effectiveness within the context of LS-relevant skills retraining, which has not previously been demonstrated. This finding is particularly promising within the intended domain of military medicine as video-based retraining can be readily implemented on mobile computing systems for use within deployed environments where it is needed most (Perez et al., in press). Additionally, an anecdotal observation was made in which some participants assigned to the video-based retraining group adopted a strategy that was used within each of the video retraining trials. This strategy involved transferring the three orange triangles first and then filling in the green triangles to form each pattern. No audio instruction was provided in the videos; therefore all changes in performance, including adoption of such a strategy were based only on visual observation of performance of the task within the videos. Formal exploration of this phenomenon within future studies may provide valuable insights into effective video-based training strategies.

**Future Directions**

Stefanidis and Heniford (2009) concluded that, “a successful laparoscopic skills curriculum should encompass goal-oriented training, sensitive performance metrics, appropriate methods of instruction and feedback, deliberate, distributed, and variable practice, an amount of overtraining, maintenance training, and a cognitive component (p. 81)”. The present research began to address several aspects of this future vision, including exploration of more sensitive metrics, maintenance training, and the inclusion of a cognitive skills component within training and assessment. Future planned research involving expert surgeons and surgical residents
Completing FLS training to mastery will be conducted in a subsequent study to further explore the relevance of task decomposition-based metrics within LS skill acquisition and retention. Furthermore, in addition to assessment within the context of video box training such as the FLS platform, application within the context of virtual reality training systems may be beneficial as well. Aggarwal et al. (2004) assert that video box training platforms such as the FLS platform lack objective assessment measures of skill acquisition, and suggest that virtual reality simulators may provide objective assessment of LS psychomotor skills. The metrics employed within the current study could be readily integrated into virtual reality training systems across a variety of platforms ranging from immersive systems to mobile computing platforms. In particular, unlike may electromagnetic sensor and camera-based motion tracking systems, glove-based motion analysis metrics provide the possibility of incorporating motion analysis within low fidelity, mobile systems. In order to further explore the specific glove-based assessment findings, future work analyzing the decomposition-based subtasks by individual hand may further elucidate these findings. Also, while velocity and angular rotation provided relevant and insightful measures of skill acquisition within the psychomotor task, additional metrics from the research literature may be explored in future work. For example, number of movements has been shown to correlate to laparoscopic skill within the context of both simulation and video-based training platforms, as well as in the operating room. It is also relevant to note that some training tasks translate more readily from video box training to the operating theater. The peg transfer task is intended as a basic skills task to teach instrument handling; however it does not directly map to a specific intraoperative task. On the other hand, suturing tasks can be rehearsed using video box trainers and are also replicated within surgical
procedures. Future planned studies will leverage the findings of the current study, expanding the task analysis and motion-based measures to additional tasks including suturing.

Future research should also explore the application of observational learning theory by addressing questions regarding optimal video content for retraining. For example, during initial skill acquisition optimal learning may result from watching videos of tasks performed correctly by experts; however, in later stages of learning and within the context of retraining previously mastered skills, watching videos of one’s self performing procedures may be more effective for reactivating prior knowledge and skills. Furthermore, at this stage of learning, watching videos containing examples of improper techniques and suboptimal performance may be more beneficial than detrimental. Future research is needed to assess the effects of these retraining techniques within a population of expert surgeons; as well as assessments of surgical residents over the course of training to mastery, retention, and retraining in order to determine how well the current results generalize to the target population, and to continue to identify subtleties associated with relevant skill proficiency. Self-report from a panel of expert surgeons at the 2013 Medicine Meets Virtual Reality Conference indicated that these experts notice or “feel” a change in their psychomotor skills after even a few days of nonuse. Such subtle changes in perceived skill are likely undetectable by traditional metrics; however, more granular metrics such as hand motion-based metrics may support more sensitive assessment of expert level skills.
Appendix A

Skill is defined by Proctor and Dutta (1995) as “goal-directed, well-organized behavior that is acquired through practice and performed with economy of effort (p. 18).” Skill is further characterized by these authors as not being innate, and, rather, developing in response to some demand. While multiple taxonomies of skills exist, skills are typically classified, based on their underlying mechanisms, as being primarily perceptual, cognitive, or motor in nature (Carlson & Yaure, 1990; Rosenbaum, 1987). Fleishman (1984) developed a human performance taxonomy that classifies skills as primarily cognitive (e.g., verbal, mathematical, reasoning), physical (e.g., strength, flexibility, gross body coordination), or psychomotor (e.g., reaction time, hand and finger dexterity, control precision, and arm-hand steadiness).

Formal, empirical study of the acquisition of various types of skills has been conducted and documented for over a century. Some of the earliest published skill acquisition research was conducted by Bryan and Harter (1897, 1899) and involves individual differences and the development of skill improvement curves within the context of telegraphic writing tasks. This work demonstrated that skill acquisition involves plateaus in performance improvement. Later work by Snoddy (1926) first characterized this function mathematically, demonstrating that skill acquisition follows a power function in which a linear relationship exists between the logarithms of performance and trials performed for a given task. This theory was later popularized and promoted to a law, known as the Power Law of Practice, by Newell and Rosenbloom (1981). While much of the early skill research focused on primarily psychomotor skills, Bartlett (1958) introduced the concept of experimentally studying cognitive skills, or “thinking”, as he termed it, experimentally. Bartlett argued that thinking is a form of skill that can be studied within the context of closed systems. He specifically differentiated thinking in
closed systems from what he termed “adventurous thinking”, which he further categorized as everyday thinking, experimental thinking, and artists’ thinking.

Fitts (1962, 1964; Fitts & Posner, 1967) expanded on the early work by Bryan and Harter, developing a three-phase model of skill acquisition, consisting of a cognitive, associative, and autonomous phase. Newell and Simon (1972) further developed these concepts by modeling human problem solving within an information processing system paradigm, which was instantiated in the Soar cognitive architecture (Laird, Rosenbloom, & Newell, 1987). Formerly known as State, Operator And Result, Soar is a production model based on a symbolic system of information processing and intelligence representation. While the Soar architecture was developed as a model of general human intelligence within the domain of Artificial Intelligence (AI) in order to model human behavior within intelligent agents, the underlying principles of cognitive and psychological theory are highly relevant to the study of human cognition in general. The Soar architecture is based on a Unified Theory of Cognition (UTC), developed by Newell (1990), which encompasses concepts related to goal-directed behavior, stimulus response, learning, and knowledge representation. Dreyfus and Dreyfus (1980) developed a 5-stage model of mental activities involved in directed skill acquisition using piloting tasks as a primary example. This model specified five developmental stages through which students pass when acquiring a skill by means of instruction and experience: novice, competence, proficiency, expertise, and mastery. Rasmussen (1983) proposed a model representing human performance at the skill, rule, and knowledge-based levels. This model, developed within the context of human computer interface design further defines information perception at each level in terms of signals, signs, and symbols. Anderson (1982, 1983) revised Fitts and Posner’s (1967) model and developed a framework for cognitive skill acquisition
consisting of just two phases or “stages”: declarative and procedural, which map to Fitts’ first and third phases, respectively. Anderson’s model also includes a process of knowledge compilation by which skills transition from the declarative stage to the procedural stage. Anderson’s model has been formally implemented within the Adaptive Control of Thought (ACT) (Anderson, 1987) and ACT-Rational (ACT-R) (Anderson, 1996) cognitive architectures. Finally, VanLehn (1996) proposed three phases of cognitive skill acquisition, which included early, intermediate, and late learning. Thus, while a single, unified theory of skill acquisition has not been accepted across domains, the existing theories are generally complimentary to one another and consist of similar underlying principles of cognition and learning processes.

Furthermore, these various models all incorporate the Power Law of Practice, which can be applied to predict relative acquisition curves for skills with various underlying components (Kim et al., 2011). Anderson (1982) states that, “It requires at least 100 hours of learning and practice to acquire any significant cognitive skill to a reasonable degree for proficiency (p. 369).” This claim is exceedingly broad and oversimplified, but it provides a point of departure and context from which to examine the application of established theoretical models and existing empirical data to quantify the acquisition of proficiency across various skill types. Furthermore, the question of what constitutes proficiency is somewhat subjective and may vary across tasks and domains. A common method for demonstrating consolidation of a skill, particularly motor skills, is the ability to concurrently perform a secondary task, demonstrating automaticity of the primary task or skill (Proctor & Dutta, 1995). However, on the continuum of skill acquisition, from expert to novice, automaticity does not represent expert performance. In fact, as described by Ericsson (1998, 2006), the autonomous stage of learning is characteristic of everyday skills, and represents a plateau in performance similar to those
described in Bryan and Harter’s early work (1897, 1899). By contrast, Ericsson asserts that expertise counteracts automaticity and progresses beyond plateaus of arrested development and thereby remains in the cognitive or associative stages of learning, but at higher levels of performance with increased experience.

Theoretical and empirical research has proposed several mechanisms by which this process occurs. Simon and Chase (1973) proposed that expertise development involves gradual acquisition of patterns and knowledge regarding appropriate reactions to stimuli based on previous exposures to similar situations, which are stored in memory. Ericsson (1996, 2002, 2004) proposed an expert-performance approach that describes both general skill acquisition and expertise acquisition as an extended series of gradual changes in physiological and cognitive mechanisms mediating performance. The primary mechanism by which this is said to occur is by straining the physiological systems, including the brain, inducing adaptations over time. Thus, tasks that are initially physically and/or cognitively challenging become easier with practice. In particular, Ericsson specifies that designed, or deliberate, practice is required in order to provide learners with appropriate, typically guided, training tasks. These tasks are ideally beyond the trainee’s existing skill level and capabilities initially, but are not beyond a level that can be attained sequentially and within a short time period (typically within a matter of hours) via practice with appropriate feedback. In addition, Ericsson describes concentration as a critical aspect of the deliberate practice paradigm such that mindless, rote task repetitions and playful engagement do not constitute the form of practice required to achieve expertise. Ericsson further highlights the fact that the concentration required during true deliberate practice inherently limits the amount of practice time that can be engaged in within a given training session. Ericsson (2006) also suggests that expertise acquisition involves the
acquisition of increasingly complex mental representations in order to support continued learning and performance improvement, and further specifies that such mental representations must also become increasingly flexible.

Ericsson and Smith (1991) conducted a series of lab-based studies assessing chess players, typists, and musicians, which provided evidence for deliberate practice as the primary mechanism by which expertise is achieved across domains. Ericsson’s (2004) proposed framework is of particular relevance to the current research study. He proposes a framework by which to examine expertise within the medical domain based on the assumption that acquisition of expert performance is not based on innate capabilities but, instead, requires deliberate practice. This controversial theory directly contradicts the notion proposed by Sir Francis Galton (1869/1979) that innate, heritable characteristics determine an individual’s performance limitations, and that these limitations cannot be overcome through training. Meinz and Hambrick (2010) demonstrated that within the context of piano sight-reading deliberate practice was highly correlated to expertise, with the exception of working memory capacity, which was significantly correlated over and above the effects of deliberate practice. Within the context of memory specifically, Ericsson and Chase (1982) demonstrated that an individual could be trained to expand digit span memory from 7 to 81 digits within approximately 200 hours of practice. However, as Sebrechts et al. (2003) highlight, the resulting increase in memory span did not transfer from digits to letters.

Perhaps the most critical component of learning theory as it relates to operational skills training is the issue of training transfer. As cited by Singley and Anderson (1989), the Doctrine of Formal Discipline, which is credited to John Locke and was a generally accepted theory throughout the eighteenth and nineteenth centuries, held that the mind comprises various
general faculties such as attention, observation, discrimination, and reasoning. It was believed that exercising or developing these broad faculties of cognition would support transfer and application of intellectual skill across a wide variety of contexts. As cited by Sebrechts et al. (2003), early work by Thorndike and Woodworth (1901) conversely pointed to a high degree of specificity in skill learning, which resulted in the development of the Theory of Identical Elements (Thorndike, 1906). This theory attributes transfer from one task to another to shared or identical elements across the tasks, rather than to strength or weakness of general cognitive faculties. More recent theories have expanded upon Thorndike’s theory, such as Tulving and Thompson’s (1973) encoding specificity principle, which posits that retrieval increases when an overlap exists between cues stored at the time of encoding and at the time of retrieval. Similarly, multiple studies have demonstrated effects for transfer-appropriate processing (Craik & Tulving, 1975; Morris, Bransford, & Franks, 1977), indicating that retention is maximized for matching modes of encoding retrieval. Singley and Anderson (1989) demonstrated the instantiation of a similar principle, the use specificity of procedural knowledge, within the ACT framework such that productions or declarative precursors of transferable skills overlap. Finally, the procedural reinstatement principle (Healy et al. 2005), also grounded in Thorndike’s theory of identical elements, maintains that tradeoffs exist between procedural and declarative information, whereby procedural information leads to strong retention but limited transfer. Conversely, declarative information leads to limited retention but increased transfer. Healy, Wohldmann, Parker, and Bourne (2005) demonstrated this phenomenon within the context of a dual task paradigm in a series of experiments in which an easy or difficult secondary task was conducted during the performance of a primary (duration production) task under initial training conditions. Follow-up testing was assessed for the same secondary task, a different secondary
task, or without a secondary task following a retention period of one week. The results of this study demonstrated that during initial training, performance on the primary task was poorer for participants completing the more difficult task as compared to participants completing the easier secondary task. Additionally, participants in both initial training groups demonstrated perfect retention for the task they had trained on after one week, but showed poorer performance when they switched to a different secondary task. This result for poorer performance was present even when switching to an easier task (although to a lesser extent). Lastly, removing the secondary task altogether at retest reduced performance on the primary task. Healy et al. report that these results suggest that transfer performance was disrupted because the cognitive operations used during training in the difficult condition could not be applied under the retest conditions. While these results are compelling, further research is needed in order to replicate these findings within more complex tasks. In a review of previous research, Wulf and Shea (2002) compared studies of motor tasks in an effort to determine whether principles demonstrated to be effective for improved learning of simple motor skills could be applied to the learning of more complex motor tasks. In particular, this investigation focused on the practice variables of attentional focus and observational learning. The results of this review provided evidence that manipulations shown to enhance simple skill learning are in fact detrimental to complex skill learning in some cases, and the authors recommend further study of complex skills in order to identify optimal learning strategies across skill types.

A promising approach to addressing some of the gaps in the skill acquisition research literature is the development of computer-based models of complex skills. It has been demonstrated that computer-based cognitive architecture programs such as ACT-R, EPIC (Executive-Process/Interactive Control), Soar, and EASE (for Elements of ACT-R, Soar, and
EPIC), which combines capabilities of the previous three, can be used to facilitate the development of models of human behavior and learning across a variety of tasks and domains (Chong, 2004). These models, however, must be grounded in not only theoretical, but also empirical research, particularly when applied to new tasks and domains with the goal of designing optimized training. For example, Reznick and MacRae (2006) suggest that surgical skills ought to be examined from the perspective of the established stages of skill acquisition, and they further assert that fundamental skills such as suturing should be trained to automaticity prior to introducing more complex and cognitive skills within the setting of the operating room.

Kim et al. (2011) proposed an integrated theory of learning and forgetting based on the existing theories developed by Fitts (1964), Anderson (1982), Rasmussen (1986) and VanLehn (1996) that model skill decay based on the stage of learning during which a retention period begins. Kim et al. highlight the commonalities existing across the selected pre-existing models, including the existence of distinct three stages, phases, or processes of learning. As previously noted in the case of Anderson’s model, the third component is the process of knowledge compilation by which learners progress from one stage to another. Kim et al., further highlight the similarities in the various tri-phasic models, indicating that each consists of a stage in which declarative and procedural knowledge are acquired. The acquired knowledge is consolidated, and that knowledge is tuned to produce overlearning. These authors suggest that a cognitive architecture such as ACT-R, which currently models decay of declarative knowledge, could be extended to model degradation of procedural skills, as well as skills involving both declarative and procedural components. Such a model, as described by Kim et al., would differentially characterize forgetting curves of declarative, procedural, and mixed memory items. Specifically, increased response times, decreased retention, and decreased accuracy over time
could be modeled based on a variety of parameters and programmed critical thresholds. Appendix B provides a detailed review of relevant skill decay research literature, indicating the current knowledge base from which such models could be developed.
Appendix B

As emphasized by Schmidt and Bjork (1992), training and retention are inseparable from one another; retention is largely dependent on a variety of training factors, and training must be designed with optimized retention in mind. Ebbinghaus proposed the first formal decay curve, which he called the forgetting curve (1885, 1913) based on a series of experiments involving memorization and recall of series of nonsense syllables. This model indicated that the rapidity of learning series of syllables was a function of their length, and that the rate of forgetting was a function of multiple factors, including number of repetitions, time (retention period), repeated learning, and the order of succession of the members of a series. This model specifically indicated that typically single trial learning results in exponential skill decay, with each additional learning trial resulting in decreased rate of decay.

Over the past few decades multiple review articles and meta-analyses have been conducted in an effort to consolidate what are often seemingly contradictory findings related to factors influencing skill decay, and conversely retention. For example, in an investigation of skill retention and its implications for navy tasks, Hurlock and Montague (1982) concluded that the primary factors associated with skill retention are the amount of learning prior to a period of nonutilization, the length of the nonutilization period, previous experience, ability level, the type of skill in question, the quantity of practice, and the quality of feedback during training. In a review of 13 experiments examining retention of military skills tasks, Hagman and Rose (1983) reported that enhanced retention is achieved by 1) increasing the amount of task repetition 2) incorporating testing into the training process 3) providing distributed (spaced) repetitions, and 4) tailoring training for the specific targeted training transfer environment improved retention. This review also indicated that the single best predictor of forgetting within
the context of procedural tasks is the number of steps required, and that the use of mnemonic techniques during training was no more effective than rote memorization for promoting retention.

In another review, including over 360 articles and spanning 100 years of research, Adams (1987) examined critical factors impacting the learning, retention, and transfer of motor skills. This review specifically examined the impact of knowledge of results, the distribution of practice, transfer of training, retention intervals, and individual differences. Wells and Hagman (1989) reviewed over 220 articles, identifying specific training procedures for enhancing learning, retention, and transfer of both verbal and perceptual-motor task skills. The effects of factors such as distribution of practice, testing during training, and elaboration on information and skills to be learned were included across a variety of task and skill types. Wisher, Sabol, Sukenik, and Kern (1991) investigated decay of skills and knowledge with 20,000 military reservists. Additionally, Wisher, Sabol & Ellis (1999) conducted an extensive review of general skill acquisition and retention/decay literature, and, based on this review, categorized military tasks into three components: knowledge, decision and execution. Wisher et al. (1999) also identified specific task factors impacting skill acquisition and decay such as task complexity and task demands. Other factors identified as affecting decay included task time pressure, whether or not job aids were used, and the quality of job aids used. Arthur et al.’s (1998) meta-analysis examined a variety of methodological and task variables relevant to skill decay, including task characteristics (i.e. physical versus cognitive, open versus closed loop, natural versus artificial, and speed versus accuracy), overlearning, method of testing for original learning and retention, conditions of retrieval, evaluation criteria (i.e., learning versus behavioral criteria), as well as instructional strategies and training method. Arthur et al. also
examined the relative effects of moderators on skill decay. Finally, in a recent meta-analysis, which sought to build on the previous review by Arthur et al., Wang et al. (2013) assessed a similar set of task and methodological variables. The results of each of these reviews and meta-analyses are organized in the following summary by task and training variables assessed.

**Task Variables**

**Open versus closed.** Arthur et al. (1998) defined closed loop tasks as consisting of a series of discrete components having definable beginning and ending states, which are typically completed in a fixed sequence. An example of a closed loop task is a fixed-sequence task such as a preflight check. Conversely, open loop tasks such as tracking and problem solving are characterized as consisting of continuous response components, which lack definitive beginning and ending states and are often repetitive. Open loop tasks have been shown to be highly resistant to skill decay over long periods of time and have demonstrated greater retention rates than closed loop tasks (Hurlock & Montague, 1982; Farr, 1987). Arthur et al., asserted that the closed- versus open-loop nature of a task acts as a moderator, and demonstrated that overall retention was actually higher for closed-loop tasks in their meta-analysis. However, Wang et al., (2013), in a book edited by Winfred Arthur and Winston Bennett, indicated that Arthur et al.’s moderator analyses were confounded by the fact that several of the variables examined covaried. Wang et al. expanded on Arthur et al.’s examination of the open-closed loop task distinction, demonstrating less decay for open loop tasks overall. Wang et al. also examined the moderation effects, demonstrating covariance of retention interval with the closed/open loop distinction. Specifically, Wang et al. demonstrated that open-looped tasks tended to have longer
retention intervals than closed-loop tasks within the studies assessed in the meta-analysis. This meta-analysis also demonstrated that open-loop tasks were significantly correlated with three measures of increased task complexity.

**Physical versus cognitive.** Physical tasks are characterized as requiring physical strength, endurance, and coordination, while cognitive tasks are those involving mental operations, problem solving, perceptual processing, and decision-making (Arthur et al., 1998). Cognitive tasks include both declarative and procedural knowledge. Declarative knowledge typically refers to factual information, while procedural knowledge indicates knowledge representing behaviors (Anderson & Lebiere, 1998), also known as “how-to-do-it” knowledge (Kieras, 1997). Wells and Hagman (1989) indicated that repetition facilitates proficiency on both verbal and perceptual-motor tasks, but that repetition did not impact extremely simple tasks in either category of learning. Wisher et al. (1991) found that motor skills such as lifting decayed after about 10 months, while cognitive skills assessed by written tests decayed within about 6 months. The Naval Education and Training (NAVEDTRA) Command developed a categorization of naval tasks on a scale of proneness to decay; tasks most prone to decay included recalling procedures and voice communications tasks, while those least prone to decay included gross motor skills and attitude learning. Konoske and Ellis (1991) asserted that the more cognitive factors involved in a task, the more likely it is to break down. Similarly, Wang et al. (2013) concluded that skill decay is generally greater for skills and tasks involving greater cognitive demands. Specifically, the most decay occurred for tasks combining moderate cognitive and low physical characteristics, and the highest retention was found for tasks involving low cognitive and high physical characteristics. Furthermore, Wang et al. assert that
decay is most likely to occur for simple cognitive tasks, and that less decay will be exhibited for tasks that are more cognitively complex, as well as tasks that are primarily physical in nature. Thus, tasks that require information integration or engagement in multiple cognitive processes are more robust.

**Natural versus artificial.** Natural tasks are those that are typically performed in operational settings and are distinguished from artificial tasks based on their complexity and motivation. Naturalistic tasks tend to be more elaborate and integrated, and the operational nature of such tasks makes them more goal-directed and motivating, whereas artificial tasks, such as those used in lab-based experiments lack integration and intrinsic motivation (Arthur et al., 1998; Annett, 1979).

**Task complexity.** Annett (1979) indicated that, among other factors, the level of task integration impacts retention. Integrated tasks consist of inter-related steps and processes, while non-integrated tasks consist of discrete sub-components. Additional task characteristics include difficulty, complexity, and level of integration required. Difficulty may be based on physical, psychomotor, or cognitive/perceptual demands, and is often determined based on subjective rating scales provided by trainees. Complexity is based on the number of discrete steps or subtasks involved in completion of a task (Hurlock & Montague, 1982); as well as the degree to which steps are cued by the equipment or previous steps in the task itself (Hagman & Rose, 1983). Druckman and Bjork (1991) found that retention of procedural tasks is affected by task organization, number of steps involved in the task, the amount of cueing, and the amount of elaboration possible. Wisher et al. (1999) also identified specific task factors impacting skill
acquisition and decay, including task complexity and task demands. Task complexity was defined based on the number of steps required to complete a task, whether the steps must be performed in a set sequence, and whether built-in feedback is provided that indicates correct performance of task steps. Task demands include knowledge demands (i.e., number and complexity of facts), cognitive skill demands (i.e., processing of large amounts of technical information or rapid decision making), and execution demands (i.e., degree of motor control required and whether motor control is discreet or continuous).

Donovan and Rodosevich (1999) conducted a meta-analysis that classified tasks based on the dimensions of overall task complexity (number of distinctive behaviors required, number of choices needed, and degree of uncertainty involved to complete a task), mental requirements of the task, and physical requirements of the task. This review suggested that the advantage of spaced practice over massed practice diminished as task complexity increased across all three dimensions. A more recent meta-analysis (Wang et al., 2013) categorized tasks based on 4 dimensions of complexity: closed versus open loop, discretion, dynamic complexity, and component complexity. This taxonomy was based on a framework of task complexity developed by Wood (1986), consisting of component complexity (number of distinct acts, information cues, or elements required for task performance), coordinative complexity (representing the form and strength of task component inter-relationships, as well as the chronology of components), and dynamic complexity (requirement to adapt to changes during task performance). Wang et al. demonstrated mixed results across complexity components, indicating less decay for tasks with high versus low, but not high versus moderate dynamic complexity, as well as less decay for higher component complexity; however more decay was found for tasks involving higher discretion, as well as for closed-loop tasks. Additionally,
optimal lengths of inter-study intervals for spaced training may be largely dependent on the targeted skill complexity and components. Within the context of cognitive skills, Arthur et al. (2010) suggest that longer interstudy intervals may enhance retention of cognitively complex tasks.

**Methodological Variables**

**Retention interval.** Arthur et al. (1998) demonstrated a correlation between length of retention interval and increased rate of skill decay for 8 data points ranging from less than 1 day to greater than 365, but noted that these effects were impacted by moderators and a full analysis was not possible given the available data points. While it is generally accepted that retention interval is directly correlated to skill decay (e.g., Hurlock & Montague, 1982), Wang et al. (2013) demonstrated that this is not always a linear relationship. It depends on task factors, specifically the level of cognitive and physical demands of a task. In fact, in some cases, longer periods of nonuse were associated with less decay or even performance improvement. This conclusion was based primarily on findings from Cepeda et al. (2006, 2008), providing evidence for a joint function of inter-study interval (ISI) and retention interval (RI). Based on a review of previous research, Pashler, Rohrer, Cepeda, and Carpenter (2007) have suggested that an optimal ISI for verbal recall tasks is between 10% and 20%. Cepeda et al., (2006) specifically demonstrated that a 1-day ISI was best for a 10-day retention period (compared to various time intervals ranging from 5 minutes to 14 days), and that 28 days was the optimal ISI for a 6-month RI (compared to various time intervals ranging from 20 minutes to 168 days).
Spacing of training (massed versus distributed practice). Overall, research indicates that spaced or distributed practice is superior to massed practice for complex tasks (e.g., Hagman and Rose, 1983). However, it is important to note here that Cepada (2006) define spaced or distributed practice as a time lag of 1 second or more between study episodes, and that massed practice is characterized by continuous practice, with no interruptions. The beneficial effects of distributed practice schedules have been demonstrated within a variety of complex skill domains, including psychomotor surgical skills such as vascular anastomosis; although complex tasks such as this will always include a lag of at least 1 second between trials (Moulton et al., 2006; Mitchell et al., 2011). Additionally, this effect has been shown to be largely dependent on the task (Arthur et al., 1998) and is generally based on effects observed within simple motor and cognitive tasks (Arthur et al., 2010). Wells and Hagman (1989) concluded that distributed practice, as opposed to massed practice, results in higher retention rates for verbal tasks, but that perceptual-motor skills retention rates are not impacted by the type of practice (distributed versus massed). Arthur et al. (2010) demonstrated that for 8-week retention of a cognitively complex task (a PC-based naval wargaming platform), 10 hours of practice over 2 weeks resulted in higher immediate post-test performance and retention than 10 hours over 1 week. Donovan and Radosevich’s (1999) meta-analysis also demonstrated that longer ISIs were beneficial for simple tasks but detrimental for complex tasks.

Overlearning. Fleishman and Parker (1962) demonstrated that performance at the end of practice (EOP) was the biggest factor in the retention of a complex piloting/tracking task. Training to specified levels of proficiency, and in some cases over-training certain skills, has been shown to promote long-term retention, particularly for difficult tasks (e.g., Farr, 1987).
Similarly, Hagman and Rose (1983) reported that enhanced retention is achieved by increasing the amount of task repetition. For example, overtraining was explored by Schendel and Hagman (1980) within the context of an M60 machine gun disassembly/assembly task. Following training either up to or beyond proficiency, soldiers were retested after eight weeks; soldiers that had overtrained made significantly fewer errors and reacquired a proficient level of performance in fewer trials than those that had simply trained to proficiency. Likewise, Schendel & Hagman (1991) concluded that overlearning was the biggest predictor of the retention of motor skills tasks, and may be cost effective. In this case, for example, the military benefited in terms of a reduced need for retraining.

In a meta-analysis of overlearning, Driskell Willis, and Copper (1992) demonstrated that the overall effect of overlearning on retention is moderated by degree of overlearning, type of task, and retention period length. The effect of overlearning was shown to be more beneficial for cognitive over physical skills. Despite no overall relationship between length of retention interval and effect of overlearning on retention, a significant interaction between retention interval and task type was detected whereby longer retention intervals reduced the effect of overlearning on retention for cognitive tasks, but the study actually showed an increase in the effect of overlearning for physical tasks. Driskell et al. attribute this increase to the fact that participants may have intentionally or unintentionally practiced the physical skills between testing intervals; however, this finding is directly in line with the previously reported findings that longer retention periods tend to result in better retention of physical skills. Driskell et al. also identify what they refer to as the half-life of the overlearning effect, the point at which the effect of overlearning (i.e., the increase in retention) is reduced by half; for cognitive skills, the half-life was 19 days, and the point at which the positive effect of overlearning on retention
reduced to zero was 38 days. Arthur et al (1998) also attempted to assess this effect via a meta-analytic approach; however, results were inconclusive due to a limited number of studies included and a limited range of degree of overlearning within those studies.

The amount of training or improvement that takes place during original learning has been shown to be directly related to the degree of overlearning. For example, Goldberg and O’Rourke (1989) demonstrated that early performance predicted later performance on a video game task. Also within the context of a video game-based task, Jones (1989) presents evidence that when end-of-practice skill level is statistically controlled for, the more an individual is overpracticed, the better skill retention tends to be. Similarly, Jones demonstrated that the more slowly an individual is improving late in practice, indicating that rate of improvement late in practice (RILP) is reaching asymptote, the greater the resulting retention. Furthermore, Jones presents evidence that rate of improvement early in practice (RIEP) has no correlation to retention that is not mediated by end-of-practice skill level, combined with either overpractice or RILP. Therefore, more overpractice or shallower RILP leads to greatest retention. However, this method of assessment controls only for the rate of improvement early in practice, not the absolute skill or performance level early in practice, which was assessed by Goldberg and O’Rourke (1989). Additionally, the measures assessed by Jones were correlated to retention, presumably as a function of performance level at the end of practice, not absolute performance level at the time of retention testing. Thus, Goldberg and O’Rourke’s findings remain relevant and are consistent with the conclusion by Schendel & Hagman (1991) that individuals starting at a higher level of performance achieve higher levels with training and subsequently maintain higher levels over time, including during periods of nonuse. Similarly, Hurlock & Montague
(1982) concluded that, in addition to amount of learning prior to a period of nonutilization and ability level, previous experience is a determinant of retention.

**Individual differences.** Vineberg (1975) demonstrated that while ability levels of individuals predicted original learning, skills decayed at the same rate across individuals, regardless of ability level. Similarly, as cited by Arthur et al. (1998), Shendel et al. (1978) and Farr (1978) have argued that individuals having higher abilities demonstrate higher performance overall and acquire knowledge and skills more quickly during initial training. Thus, Driskell et al. (1992) suggest that overlearning can be used as a means to even the playing field by getting all trainees, including lower ability individuals to equivalent performance levels prior to the onset of retention periods. Additionally, Arthur et al. suggest that individual differences could be used to determine optimal training schedules.

**Testing effect.** Testing during training has been shown to increase the retention of both verbal and perceptual-motor skills (Hagman & Rose, 1983; Wells & Hagman, 1989). For example, in a series of experiments involving a naval motor skill, Hagman (1980, 1981) demonstrated significantly higher retention for trainees that received increased testing trials during training as compared to those that received increased presentation trials during initial training following a retention period of only twenty-four hours. More recent studies have further explored the effects of various study and test schedules for increased retention within a variety of tasks. Karpicke and Roediger (2008) demonstrated that repeated testing on both correctly and incorrectly tested items within word pair lists resulted in greater overall retention than repeated testing including only previously incorrectly tested items during each subsequent
Of particular relevance to the current research study, Kromann, Jensen, & Ringsted (2009) demonstrated benefits of the test effect within the context of a medical procedural skill (resuscitation) following a two-week retention period. Also of relevance, Rohrer, Taylor and Sholar (2010) demonstrated that the beneficial test effect was even larger for a transfer task, in this case a task requiring both declarative knowledge recall and reasoning skills. Finally, Karpicke and Bauernschmidt (2011) assessed the effect of following “study and test to correct retrieval of all trained items” with additional test sessions that were also spaced. Compared to immediate repeated testing following testing to correct retrieval of all items, the spaced testing approach demonstrated 200% improvement in retention.

**Method of testing and conditions of retrieval.** Arthur et al. (1998) also included in their review a meta-analytic assessment for the effect of method of testing on retention; this included an assessment of the method of testing of both original learning and retention. The results demonstrated, as hypothesized, that recognition elicited higher retention than recall. Additionally, the effect of conditions of retrieval was assessed, demonstrating that the similarity of original learning and retention contexts had a large effect on retention; however, data points for only 4 studies were included in this analysis. Arthur et al. also demonstrated that less decay was associated with behavioral evaluation criteria than for traditional learning evaluation criteria; this result was attributed to the fact that behavioral criteria are typically assessed in a more naturalistic context, whereas learning criteria such as tests in a classroom tend to be more artificial in nature. Finally, Arthur et al. examined the relative effects of various moderators on skill decay. Similarity in conditions of learning and retrieval was shown to be the most important moderator. The second most influential moderator was whether the dependent
variable assessed was speed-based or accuracy-based. Interestingly, skill decay for accuracy-based task assessments was over three times higher than for speed-based task assessments. The least influential moderator was the natural versus artificial task distinction. Wang et al., (2013) further classified task dependent variables as cognitive or “skill-based” within their meta-analytic assessment of the methodological factor of type of evaluation criteria based on a previous meta-analysis by Kraiger, Ford, and Salas (1993). Within this framework, Wang et al. further categorized cognitive dependent variables based on declarative knowledge, procedural knowledge, strategic knowledge, knowledge organization, and cognitive strategies criteria. Skill-based criteria were then identified as either proceduralized or adaptive. Wang et al. indicated less decay for declarative knowledge than for tasks involving both declarative and procedural knowledge; furthermore, skills assessed based on adaptive criteria were shown to decay more than those assessed based on proceduralized skill criteria.

**Refresher training.** The effects of refresher training, or retraining at strategic points in time, on skill retention have been shown to vary based on factors such as length of the retention period, nature and characteristics of the task, frequency and timing of refresher training sessions, and initial skill levels of trainees (Shendel & Hagman, 1991). Additionally, Shendel and Hagman (1991) demonstrated that motor skills retraining generally requires less than 50% of time required for original training in order to return to pre-retention performance levels. Retraining to proficiency at various intervals following initial training has been shown to be highly effective in sustaining skills over long periods of time. For example, O’Hara (1991) demonstrated that periodic retraining on Marine watch-standing tasks improved retention at 9 months. Of particular relevance to the current research study, Castellvi et al. (2009) reported
that training to proficiency for 42 surgical residents on all 5 FLS manual skills resulted in skill maintenance of approximately 86% for Tasks 4 and 5 (suturing tasks) after 6.5 months. Retraining was provided at that time for any subjects not at criterion. All participants were then tested again at 12.5 months, at which time maintenance of 96% was demonstrated for both Tasks 4 and 5. Mashaud et al. (2010) continued the study reported by Castellvi et al. (2009) with the same subjects, administering retention testing on the FLS suturing tasks approximately every 6 months, as well as retraining as needed following testing at 6, 12, and 18 months if proficiency had not been maintained at 100%. The results of this study indicated that for the 18-month retention test, mean skill retention was 98% of original skill level for Task 4 and 96% for Task 5. At 18 months, retraining was required for 47% of the participants on Task 4 and 44% on Task 5. At 24 months, skill retention was reduced to 91% for Task 4, but was measured as 100% of original skill level for Task 5. Thus, a significant ($p<.001$) performance decrease was detected between initial training post-test and 6 months for both tasks, and a significant ($p<.001$) performance increase was detected between 6 months and 24 months for Task 5, indicating the effectiveness of the retraining for long-term skill maintenance and even improvement. These results suggest that basic psychomotor skills, as trained and assessed by the FLS protocol, are extremely durable following training to proficiency, and specifically highlight the value of periodic retraining to further enhance retention of these skills during periods of nonuse. However, as previously noted, current FLS assessment metrics may lack the necessary sensitivity to detect subtle changes in psychomotor skill related to skill degradation.

**Mental rehearsal.** Finally, mental rehearsal of skills and tasks has been shown to decrease skill decay and, in some cases, even result in greater transfer on a related task than
physical practice (Wohldmann, Healy, & Bourne Jr., 2008). These outcomes provide evidence that mental rehearsal may involve abstract representations of task elements (Healy & Wohldmann, 2012). Ryan and Simons (1981) argue that cognitive skills are more conducive to mental rehearsal and ought to be retainable for longer time periods than physical tasks; however, in the absence of mental rehearsal, physical skills are in fact more robust over periods of retention. Adams (1987) suggests that mental imagery research can also be used to help to identify shifts from controlled to automatic processing of motor behavior.
Bibliography


Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic...


Gronlund & Kimball (2012). Remembering and Forgetting: From the laboratory looking out.


selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures.


ANZ Journal of Surgery, 72(9), 632-634.


Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to account for individual differences in piano sight-reading: The role of working memory


Rosser, J.C., Rosser, L.E., & Savalgi, R.S. (1997). Skill Acquisition and Assessment for


W. C. (2006). A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Academic Medicine, 81*(10), S9-S12.


