Spring 2015

Assessing the Effectiveness of an Interval Estimation and a Visual-Spatial Secondary Task as Measures of Mental Workload During Laparoscopy

Levi Parker Warvel
Old Dominion University, leviwarvel@gmail.com

Recommended Citation

Warvel, Levi P.. "Assessing the Effectiveness of an Interval Estimation and a Visual-Spatial Secondary Task as Measures of Mental Workload During Laparoscopy" (2015). Master of Science (MS), Thesis, Psychology, Old Dominion University, DOI: 10.25777/7way-ec47
https://digitalcommons.odu.edu/psychology_etds/3

ASSESSING THE EFFECTIVENESS OF AN INTERVAL ESTIMATION AND A VISUAL-SPATIAL SECONDARY TASK AS MEASURES OF MENTAL WORKLOAD DURING LAPAROSCOPY

by

Levi P. Warvel
B.A. Psychology, December 2008, Gannon University

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

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2014

Approved by:

Mark W. Scerbo (Director)

Christopher Brill (Member)

Barbara Winstead (Member)

ABSTRACT

ASSESSING THE EFFECTIVENESS OF AN INTERVAL ESTIMATION AND A VISUAL-SPATIAL SECONDARY TASK AS MEASURES OF MENTAL WORKLOAD DURING LAPAROSCOPY

Levi P. Warvel
B.A. Psychology, December 2008, Gannon University
Old Dominion University, 2012
Director: Dr. Mark W. Scerbo

The goal of the present study was to compare two secondary workload tasks, specifically a time interval estimation and visual-spatial task, to determine which of these is most appropriate for the assessment of laparoscopic mental workload. Participants performed a peg transfer task in two conditions: a normal camera angle and a 90° camera angle intended to increase mental workload. Based on multiple resource theory, it was predicted the visual-spatial task would be more sensitive to the workload manipulation than the time estimation task because it draws upon the specific, as opposed to more general, attentional resources required by laparoscopy. Primary task results demonstrated that manipulation of camera angle did change workload levels. Secondary task results showed that the visual-spatial task possessed greater sensitivity and diagnosticity than the interval estimation task. However, interval estimation demonstrated a global sensitivity to workload changes. The findings suggest that a visual-spatial secondary task is an effective method to assess workload experienced during laparoscopy.

This thesis is dedicated to my father, Craig P. Warvel, without whom none of my achievements would be possible.

ACKNOWLEDGMENTS

To Dr. Mark Scerbo, I extend my most sincere thanks. If not for his patience and guidance during the thesis process, I would surely not be here today.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Laparoscopy, also known as minimally invasive surgery or endoscopic surgery, is an alternative to traditional open surgery. Laparoscopic surgery is performed with the use of long handled instruments that are inserted into a patient via a small fixed incision. The image of the operation site is obtained by a small camera inserted into one of these incisions and displayed on a monitor that is viewed by the surgeon.

The laparoscopic technique typically results in quicker recovery times and shorter hospital stays for patients as well as a reduced likelihood of postsurgical complications (Aziz et al., 2006; Braga et al., 2005; King et al., 2005; Milsom et al., 1998). Due to these advantages over traditional open surgery, laparoscopy can often be beneficial to the patient and may be preferred to other surgical options. However, these benefits come at a cost. Laparoscopic surgery is significantly more difficult for the surgeon (Berguer, Smith, & Chung, 2001). Because this increase in difficulty may also increase the risk of surgical error, it is important to develop methods to assess surgeons' readiness for the procedure. The present study sought to compare two secondary workload tasks to determine which of these is more appropriate for the assessment of laparoscopic mental workload which can then be used in assessment of laparoscopic surgical ability.

The first aspect of laparoscopic surgery that contributes to the difficulty of the procedure is the reduced tactile feedback experienced by the surgeon. Traditional open surgery allows the surgeon semi-direct contact with the internal structures of the patient. Although this is not direct contact as surgeons wear gloves, they can still determine the general structure, shape, and temperature of the tissues and organs (Westebring-van der Putten, Goossens, Jakimowicz, & Dankelman, 2008). Tactile feedback can help the surgeon determine if a sufficient amount of force is being applied to tissues to secure them for the procedure without causing damage. Laparoscopy can still be successfully performed without this feedback, but losing this source of information can be detrimental (Mohr et al., 2001).

Another aspect of the procedure that increases its difficulty is the loss of natural depth cues. In traditional open surgery, the surgeon is able to look directly at the operating site with all visual cues available. In laparoscopic surgery, the view of the operating site is projected onto a flat monitor, replacing the three-dimensional view with a two-dimensional view. This results in a loss of binocular vision, forcing the surgeon to rely on monocular depth cues to navigate the operating site. Monocular depth cues are generally useful in determining the distance and location of objects in the operating site but the use of a flat display often degrades these cues. These visual distortions may lead to misinterpretations of the anatomy of the patient, reduced surgical performance, longer operation times, and mental fatigue (Cuschieri, 1995, 2006; Tendick, Bhoyrul, & Way, 1997; Way et al., 2003), which is particularly important in light of research suggesting that an overwhelming number of injuries to patients during laparoscopic surgery were the result of degraded visual information. In an analysis of 252 cases of laparoscopic bile duct injuries, Way et al. (2003) found that 97% of injuries were the result of visual illusions caused by the degradation of depth cues with the remaining 3% being due to failures in technical skill. Further, these errors were attributable to experienced surgeons, not residents or novices.

A potential solution to the problems caused by degraded depth information is the use of stereoscopic displays that enable binocular depth cues. Unfortunately, these displays do not seem to be effective at enhancing laparoscopic performance. For example, several researchers have found that although surgeons may prefer stereoscopic displays to flat two-dimensional displays their performance does not differ between display types (Bittner, Hathaway, & Brown, 2008; Hanna & Cuschieri, 2000; Tendick, Bhoyrul, & Way, 1997).

The third factor that contributes to the increased difficulty of laparoscopic surgery is the misalignment of the surgeon's point of view. Open surgery preserves depth cues and also the relation between the visual axis and the forearm-instrument motor axis (Gallagher, Al-Akash, Seymour, & Satava, 2009; van Det, Meijerink, Hoff, Totte, & Pierie, 2009). The movements of the surgical instruments are natural as well, with forearm movements in any given direction resulting in instrument movements in the same direction. In contrast, laparoscopic surgery does not preserve the natural viewpoint found in open surgery. The angle in which the camera is inserted into the abdominal cavity is determined by an array of factors, such as the position of the patient, type of procedure, and arrangement of viewing monitors (Van Det et al, 2009). The camera's line of sight differs from the surgeon's line of sight. As the degree of separation between the surgeon's view and the camera's view increases, the procedure becomes more difficult (Conrad et al., 2006; Gallagher et al., 2009; Klein, Warm, Riley, Matthews, & Parsons, 2004).

Another related characteristic of laparoscopic surgery that can reduce performance is the fulcrum effect (Gallagher, McClure, McGuigan, Ritchie, & Sheehy, 1998). The phenomenon is a result of the long handled tools used to perform the procedure and the fixed points of entry into the operation site. Unlike open surgery, movement in laparoscopy is largely restricted to pivoting instruments about the incision point, which creates an inversion of motion compared to the more natural movements of open surgery. The visual-motor perceptual distortion created by the fulcrum effect can increase the surgeon's workload and reduce performance (Gallagher et al., 2009).

Since laparoscopic surgery is more difficult than traditional surgery, it is important to develop methods to train and evaluate performance (Tendick et al., 2000). Training and assessment tools, such as the Fundamentals of Laparoscopic Surgery (FLS) modules (www.flsprogram.org), offer methods to teach skills and measure performance through simulated surgical tasks. However, some surgical training programs do not utilize such methods, instead opting for a nonstandardized and subjective observation approach to assessment (Alkhayal et al., 2012). Training with simulation has been shown to have positive benefits in genuine operating settings (Dawe, Windsor, Broeders, Cregan, Hewitt, & Maddern, 2013) but some researchers have found that such training is not always superior to traditional training methods (Mansour et al., 2012). One possibility as to why simulated training may not successfully transfer may be the lack of standardized measures of the cognitive demands experienced by surgeons. Since the characteristics that result in increased difficulty in laparoscopy are perceptual and cognitive in nature, attention must also be paid to the mental workload experienced by surgeons if the demands of laparoscopy are to be fully understood. The purpose of the current research is to compare two methods of quantifying the cognitive demands of laparoscopy to determine which is a more accurate measurement of differences in mental workload. By doing so, assessment techniques may be developed to determine when residents are sufficiently qualified to move from simulation to supervised laparoscopy, allowing instructors to ensure patient safety and providing a more complete index of student progress. However, before such advanced training tools can be implemented, the nature of mental demands in laparoscopy must be studied further.

CHAPTER II

MENTAL WORKLOAD

Historical Background

Workload refers to the relationship between the amount of one’s mental processing capacity and the demands required for a task (Hart & Staveland, 1988). Workload can be differentiated by the cognitive and physical components of a task. The cognitive component of workload, or mental workload, describes the processing of information and formulation of plans and responses. The concept of mental workload as a finite resource began with Moray (1967) who proposed that the human operator was a “limited capacity processor.” Until this point, it was generally believed that any task requiring attention was processed via a single attentional channel. Additional tasks would need to be completed sequentially before any other attentionally demanding tasks could be initiated (Broadbent, 1958). However, Cherry (1953) argued that some tasks could be processed in parallel to a limited degree, such as the tendency for people to be able to differentiate two numbers presented auditorily at the same time. Subsequent modification to the attentional channel model described this limit as a single pool of mental resources (Kahneman, 1973). Processing of information is limited by the amount of resources available to the operator at any given time and difficulty affects the proportion of the resources needed to perform a task (Kahneman, 1973; Norman & Bobrow, 1975). However, subsequent evidence was inconsistent with the Kahneman (1973) model of mental resources and suggested that there was not one single pool of attentional resources but many (Navon & Gopher, 1979; Wickens, 1980, 2002, 2008), though they were still considered limited.

Of these models, one of the most influential in the domain of workload research is Multiple Resource Theory (MRT). Wickens (1980) performed a meta-analysis on multitask experiments in an attempt to account for variance in time-sharing performance. He considered the processing structure for the tasks and the degree to which tasks using separate processing structures failed to affect one another. The results indicated the existence of three orthogonal dimensions, each comprised of limited resources.

The first dimension of this model is the information processing stage dichotomy. These resources are differentiated by the perceptual/cognitive processing stage and response processing stage. Perceptual/cognitive resources are consumed through thought processes and the organization of information. Response resources are consumed in executing actions. The second dimension is the processing code dichotomy, distinguished by spatial and verbal resources. Spatial resources are consumed by processes needed for object location or distance judgments. Verbal resources are consumed by language processes. The third dimension is processing modality. These resources are separated into auditory and visual channels. Additionally, there is a differentiation between focal and ambient visual processing within the visual processing modality pool (Wickens, 2002). Focal visual processing is typically foveal and related to the pattern recognition and discrimination of details, while ambient visual processing is largely peripheral and related to movement detection and environmental changes outside the fovea.

Although MRT is a widely accepted theory of workload, there are others that should be noted. Boles and Law (1998) have proposed that the degree to which processing structures affect one another may be due to both structural and attentional resources. In this expanded multiple resource theory, resources are considered from a more general view of association with processes instead of structural or attentional models. The modeling of the resources is not through dichotomous dimensions as specified in MRT but rather through independent process-specific resources (Boles, 2010). For example, visual-spatial resources are consumed during the spatial processing of visual stimuli whereas visual-verbal resources are consumed during the verbal processing or visual stimuli. These are considered separate resource channels instead of two processing coding channels sharing input from the same processing modality channel. Although a promising as an extension of MRT, there is a limited amount of research using this model at present.

Workload assessment methodology

Multiple resource theory is useful in assessing mental workload because it provides a model from which predictions of an operator’s ability to multitask can be made. Specifically, the degree of similarity between the demands imposed by two tasks should determine the ability to complete both tasks. If the task demands are similar, they may compete for resources from the same pool and increase the level of mental workload. If the tasks are dissimilar, they may draw upon different pools of resources and therefore may have minimal or no effect on mental workload. Understanding this relationship allows researchers to assess the workload associated with different tasks.

There are a variety of methods that can be used to measure workload (O’Donnell & Eggemeier, 1986). Workload measurement techniques can vary greatly and should be selected based on five main criteria (Carswell, Clarke, & Seales, 2005; O’Donnell &

Eggmeier, 1986). The first criterion is sensitivity and refers to the ability of a measure to reflect differences in operator workload. For a measurement technique to be effective, it must be able to distinguish between variations in workload imposed by tasks and be resistant to the effects of extraneous influences. The degree of task sensitivity should be matched to the objectives of research being performed. For example, if one is interested in singling out occurrences of extreme workload in a task, a measure with less sensitivity may be sufficient. On the other hand, if the goal is to detect more subtle changes in workload, a greater amount of sensitivity is necessary. A second characteristic, diagnosticity, is the degree to which a measure reflects the demands imposed on a particular resource. Further, there is a differentiation between global and specific diagnostic measures (Carswell et al., 2005). Some measures are intended to be more sensitive to a specific type of resource that affects workload (e.g., visual-spatial processing versus auditory processing). Others may be less sensitive to an individual resource and instead target workload in general. Similar to sensitivity, the degree of diagnosticity of a measure should be determined by the purpose of the research. If the intent is to assess workload changes in a task that imposes on several resources, a global measure is satisfactory. If a task places high demands on a particular resource or set of resources, the measure should be sensitive to these dimensions alone. A third characteristic is intrusiveness, which refers to how a measure interferes with the task of interest. Implementation requirements generate practical concerns. For example, the need to implement a measure, the cost of a measure, and potential training requirements for an operator to use a measure are all factors related to implementation requirements. The last criterion is operator acceptance and refers to the willingness of an individual to use the measure as instructed and comply with the requests of the researcher (Carswell et al., 2005; O'Donnell & Eggemeier, 1986; Wierwille & Eggemeier, 1993).

Although each of the selection criteria for a workload assessment technique is important, the level of each should be determined by the domain or system being researched. In the example of laparoscopy, the perception of the operating area is primarily visual, the coding of this information is spatial, and the response modality is motor (Cuschieri, 1995, 2006; Klein et al., 2004). As such, an ideal measure for laparoscopy would be one with a high amount of sensitivity and diagnosticity in the visual, spatial, and motor dimensions. Additionally, it should have a low level of intrusion as well as a sufficient level of compliance to ensure that the performance measured is representative of the actual abilities of the operator. Once the criteria for the optimal measure of a task are known, an appropriate workload measure should to be selected to ensure that all criteria are satisfied.

Workload measures fall into three categories: subjective, physiological, and performance. Subjective measures require operators to report their perceived experience of workload, typically through a survey or rating scales. Subjective measures are categorized by the dimensions of workload that they assess (O'Donnell & Eggemeier, 1986). Unidimensional measures require the respondent to give a single, global judgment of workload. An example of a unidimensional measure would be an index that asked the respondent to rate the perceived demand associated with a task on a single Likert scale. Alternatively, multidimensional measures require the respondent to provide ratings on a number of subscales. In essence, multidimensional measures are comprised of a number of unidimensional scales, each assessing workload in relation to a different task characteristic. Unidimensional measures often allow for simpler data analyses but possess less diagnosticity than multidimensional measures. In contrast, multidimensional measures have a higher degree of diagnosticity but can be more time-consuming and difficult to analyze (Young & Stanton, 2004). Generally, subjective measures are advantageous due to being inexpensive and quick to use, non-intrusive, flexible, and largely generalizable across different tasks (Stanton, Salmon, Walker, Baber, & Jenkins, 2005). However, there are some disadvantages to subjective measures. The ability of an operator to self-assess is limited and most data are collected after the fact, leading to potential decay of memory during periods of high workload (Carswell et al., 2005). Subjective measures can also correlate with performance, resulting in high levels of workload reported by individuals who perform poorly and low levels reported by better performers (Stanton et al., 2005).

Many subjective measures of mental workload are available but two of the more frequently used instruments are the Subjective Workload Assessment Technique (SWAT; Reid & Nygren, 1988) and the National Aeronautics and Space Administration Task Load Index (NASA-TLX; Hart & Staveland, 1988). The SWAT procedure requires respondents to rank 27 possible combinations of three workload dimensions (time load, mental load, and stress load) on a scale of 1 to 3. The responses to these combinations determine how each dimension is weighted. The NASA-TLX also requires respondents complete a weighing procedure by presenting 15 pairwise comparisons of its six subscales (mental demand, physical demand, temporal demand, effort, performance, and frustration level). Each subscale is selected based on how much it contributed to the workload of the task. The number of times each subscale is selected is tallied and summed, leading to values ranging from 0 to 5. Following this, ratings are provided on an interval scale from 1 to 20. Finally, the interval ratings are multiplied by their respective weights and divided by the sum of the weights. Both of these measures require little time to apply and require minimal training to use compared to other multidimensional measures (Stanton et al., 2005).

In contrast to subjective measures which are dependent on the respondent’s ability to report perceived mental workload, physiological measures reflect an operator’s autonomic activity in response to workload. Options for physiological measures are numerous and varied but some cardiac measures, heart rate variability in particular, have shown reasonably good validity (Young & Stanton, 2004). These techniques tend to be minimally intrusive to the primary task. However, implementation can require expensive equipment and past research has demonstrated inconsistency regarding their sensitivity (O’Donnell & Eggemeier, 1986; Wierwille & Connor, 1983; Young & Stanton, 2004). Requiring users to be connected to or physically restricted by the recording devices can also affect user acceptance and may make the use of such a technique difficult in real world scenarios. Many of these measures can also vary in their reliability, demonstrating low sensitivity to changes in task demands. Due to these reliability issues, they are often used as complementary measures with other more reliable measures of mental workload.

The last category includes performance-based measures. Performance measures are classified into two subcategories. The first are primary task measures which record the performance of an operator on the task of interest. Primary task performance measures are gathered in most workload studies and are critical for measuring operator workload (Wierwille & Eggemeier, 1993). Two fundamental metrics of primary task performance are often speed and accuracy. Performance is expected to decline as resource demands exceed the availability of cognitive resources necessary for unimpaired performance (Wierwille & Eggemeier, 1993). However, this is not always the case. If the primary task is too easy, the operator may be underloaded. In this case, the operator will have sufficient resources available to complete the primary task and the measure will be insensitive to the workload change (O'Donnell & Eggemeier, 1986; Wierwille & Eggemeier, 1993). Conversely, an operator may be overloaded with task demands, leading to very poor performance (O'Donnell & Eggemeier, 1986). Primary measures of task performance are valuable in that they are a direct index of performance on the task of interest and can be crosschecked with other workload measurements to assess validity (Lysaght et al., 1989; O'Donnell & Eggemeier, 1986).

Secondary task performance measures are often paired with those gathered from primary task performance. Secondary tasks reflect the operator's ability to perform an additional task in tandem with the primary task. Secondary tasks can be categorized in one of two ways. The loading task method requires the operator to maintain secondary task performance regardless of its effect on the primary task. On the other hand, subsidiary tasks require the operator to maintain primary task performance regardless of impact on the secondary task (O'Donnell & Eggemeier, 1986). Additionally, secondary tasks can be independent of the primary task or embedded into it (Wierwille & Eggemeier, 1993). An independent secondary task is characterized as not being part of the normal procedure associated with the primary task. For example, performing a mental arithmetic task while in a driving simulator would be considered independent from the typical operations associated with driving. Embedded secondary tasks differ in that the secondary task is actually part of the normal operations or procedure used in for the primary task. The embedded secondary task can be advantageous in that it minimizes primary task intrusions. Further, the secondary task should already occupy an accepted role within the system being assessed. However, embedded tasks do not always satisfy sensitivity and diagnosticity requirements necessary for workload measures. In this case, an independent secondary task may be a better option as they can be designed to be sufficiently sensitive and diagnostic of mental workload changes in the task of interest.

The ability to successfully perform the secondary task is thought to reflect excess resource capacity after the necessary resources demanded by the primary task are allocated (Ogden, Levine, & Eisner, 1979). Secondary tasks are usually more sensitive to changes in available resources and can be highly diagnostic (O’Donnell & Eggemeier, 1986). However, selection of an appropriate secondary task is paramount in retaining a high degree of diagnosticity. An effective secondary task should be sensitive to changes in primary task demand by competing for the same mental resources as the primary task (Carswell et al., 2005; O’Donnell & Eggemeier, 1986; Wickens, 1984, 2008; Wierwille & Eggemeier, 1993). If the secondary task fails to do so, it may not be sensitive enough to detect changes in mental workload. For example, Young and Stanton (2004) sought to assess workload changes in simulated driving tasks. Given that the driving task demanded visual-spatial processing resources, a secondary target identification task was developed in which the participant had to determine if a rotated figure was identical to the example provided. Both the primary and secondary tasks required visual processing of information and spatial coding for determining distance or rotation. As such, the target identification task should compete for the same attentional mental resources as the driving task and be sufficiently diagnostic. Results indicated that secondary task performance was poorer in the more demanding manual driving conditions compared to less difficult automated driving conditions, indicating that the secondary task was sensitive to changes primary task workload and selectively diagnostic of visual-spatial resource demands.

Although resource overlap is necessary for an effective secondary task, it can increase the possibility that the secondary task will intrude upon the performance of the primary task. If the secondary task demands overload the operator or performance of the secondary task requires the primary task to be discontinued, the amount of available resources for the primary task are reduced and may artificially decrease performance (Carswell et al., 2005; O'Donnell & Eggemeier, 1986; Wierwille & Eggemeier, 1993). Although this balance between sufficient resource overlap and reducing intrusiveness can be troublesome, it is important to recognize that primary and secondary tasks that draw upon the same resource pools will create a degree of intrusion in any case. Multiple resource theory would predict that if two tasks are performed together and utilize the same processing resources, they will interfere with the processing of each other to some degree (Wickens, 2002; 2008). As such, an optimal secondary task may not avoid intrusion all together but should minimize intrusion while maximizing resource overlap. Selection of an appropriate secondary task class can help ensure a balance between primary task intrusion and resource overlap.

As an example, consider choice reaction time, a classification task in which the participant is presented with two or more stimuli and required to respond differently to each. The method of presentation can be visual (e.g., target identification) or auditory (e.g., tonal discrimination) and response modalities can be verbal or motor in nature (O’Donnell & Eggemeier, 1986). When used to assess residual workload, secondary tasks of this nature are typically designed to demand the same resources as the primary task. An example of a choice reaction time task is the previously mentioned rotating figure task developed by Young and Stanton (2004).

Another class is interval production, which requires participants to produce a response whenever they believe that a set interval of time has transpired. This method is based on the attentional-gate control model of prospective duration judgment (Thomas & Weaver, 1975; Zakay, 1989; Zakay & Block, 1997; Zakay & Shub, 1998). According to this model, non-calculated temporal judgments are made via a mental accumulator that gathers pulses. These pulses are emitted at a constant rate and represent conceptual units of time. The contents of the accumulator are compared to a reference in memory containing a representation of past accumulated pulses. A cognitive mechanism then compares the present accumulated pulses with the reference memory to determine the amount of time that has passed. The pulse accumulator is operated by a switch that is related to mental workload capacity and is thus proposed to be a sufficient secondary task since the time estimation task and the primary task should be competing for resources (Brown, 1997). When primary task demand is high, the accumulator should take longer to store a sufficient number of pulses and the result is an overestimation of the interval duration (Zakay & Shub, 1998). Interval duration has been used in a number of workload studies and seems to be effective as a secondary task in many cases. Baldauf, Burgard, and Wittmann (2009) found that the length of intervals produced did increase as simulated driving tasks became more complex with no significant effects on primary task performance. Zakay and Shub (1998) found a similar effect in card sorting tasks, Stroop tests, and flight simulation as did Liu and Wickens (1994) using a customer assignment task. However, Zakay and Shub (1998) caution that time estimation secondary tasks, while apparently non-invasive and sensitive to changes in mental workload, are not appropriate for all primary tasks. For example, Wierwille, Rahimi, & Casali (1985) found that the interval production secondary task intruded on one of the primary tasks during flight navigation task. Pilots in this study demonstrated a much higher error rate in relation to answering navigation-based questions when the primary task was paired with the time estimation secondary task. Of 16 measures of workload assessed in their study, interval production was the only workload measure that demonstrated this effect. The inexplicable invasiveness of the task in this study seems to suggest that, despite previous successes, interval production tasks may not be ideal in all domains.

CHAPTER III

WORKLOAD IN LAPAROSCOPY

The workload experienced during laparoscopy has received some attention in the past (Cuschieri, 1995; Tendick & Cavusoglu, 1997; van Det et al., 2009). However, many of the methods used to assess laparoscopic workload have not been designed to the standards recommended by past literature (O’Donnell & Eggemeier, 1986; Wierwille & Eggemeier, 1993). Much of the research lacks the inclusion of a secondary task utilizing the same resources as the laparoscopic primary task, relying instead on physiological (Berguer, Smith, & Chung, 2001) or subjective measures (Klein et al., 2004; Klein et al., 2008). Of the research studies that employ a secondary task, some are difficult to define and may not be the ideal measure of laparoscopic workload. For example, Zheng, Cassera, Martinec, Spaun, and Swanstrom (2010) used a visual secondary task to measure residual workload during a laparoscopic simulation task. However, participants were required to respond verbally to the task which may not be similar enough to a motor response.

The past success of interval production as a secondary task suggests that it possesses sensitivity and some researchers have begun to recommend using the method for measuring workload experienced by surgeons (Grant, 2010; Grant, Carswell, Lio, & Seales, 2013; Lio et al., 2006; Lio et al., 2007). Carswell et al. (2005) have indicated that time estimation may already be embedded into laparoscopy due to the need to monitor the passage of time during surgery, providing the measure greater operator acceptance compared to other methods, as well as easy implementation through a variety of response modalities. Indeed, research using interval production has found the task to be sensitive to changes in workload when paired with simulated laparoscopic tasks (Grant, 2010).

However, the degree to which interval estimation shares the same resources as laparoscopy is unknown and, due to the lack of knowledge on how the accumulator mechanism functions in attentional-gate control, multiple resource theory makes no predictions regarding the resource demand. As a result, levels of diagnosticity or invasiveness of interval estimation are difficult to identify. Thus, it is possible that time estimation may simply be a global measure of mental workload and may not actually require the same resources as laparoscopy.

In addition, the presumption that time estimation, as it is described in prior research (Block & Zakay, 1997; Zakay & Block, 2004; Zakay & Shub, 1998), is already embedded into surgical simulation is debatable. The metacognitive study on which this argument is founded at no point suggests that surgeons do monitor time via this mechanism (Dominguez, 2001). Rather, Dominguez (2001) conducted a field study in which twenty surgeons were asked to watch a video of a laparoscopic cholecystectomy featuring significant amounts of blood and bile distorting the visual field and obscuring an exposed artery. At seven different points through the procedure, they were asked if they would convert to open surgery or continue laparoscopy and measures of comfort were taken. At no point was a cognitive estimation of time cited as a reason for conversion. Thus, the validity of time estimation as a sufficient secondary task in laparoscopic workload research may be questionable.

Since laparoscopy is highly demanding of visual and spatial resources (Berguer, Smith, & Chung, 2001; Cuschieri, 1995; Klein et al., 2004), a potentially superior secondary task may be one specifically designed to draw from visual and spatial resource pools. To address this issue, Stefanidis, Scerbo, Korndorffer, & Scott (2007) developed a visual-spatial secondary task that required participants to monitor a series of squares that would appear randomly on either side of the display. Participants were asked to press a foot pedal whenever three squares appeared in succession on the right side of the screen. In addition, participants were instructed to give priority to the primary suturing task and attend to the secondary task on a separate monitor whenever they were able. The use of a visual-spatial secondary task such as the squares task not only allowed for the demands on attentional resources to overlap but also provided an additional metric of performance to complement the more traditional measures of time and errors. Results indicated that all participants were able to perform the secondary task perfectly by itself, but significantly worse when paired with the laparoscopic suturing task. These findings indicate that the visual-spatial secondary task appears to be sufficiently sensitive and diagnostic, satisfying two of the requirements of a satisfactory workload measure. However, it was found that large numbers of participants did not or were unable to attend to secondary task in the dual task condition at the start of the session or just after beginning. One problem with using a second monitor for the squares task is that it forced participants to redirect their gaze from the primary task display to attend to the secondary task. Given that participants were instructed to prioritize the primary task, they may have been hesitant to divert attention away from the suturing display to monitor the squares.

Kennedy (2010) used a visual-spatial task that differed from the Stefanidis et al. (2007) sequential squares task in a few important ways. First, the task was no longer a sequential identification task with two-dimensional squares but rather a set of four multicolored balls presented within a simulated tunnel conveying depth. The balls in their standard orientation were located at the 3, 6, 9, and 12 o'clock positions in the tunnel. Depending on condition, one of the four balls could either rotate clockwise or counterclockwise on the same plane or move closer or further down the tunnel. Participants were asked to indicate a change in position by pressing one foot pedal or that the orientation of the balls had remained the same by pressing a different pedal allowing for measures of both target and non target identification. By doing so, it was now possible to identify cases in which the participant was not attending to the task. Second, the visual-spatial secondary task was presented on the same screen as the primary task. By integrating the secondary task onto the same display as the primary task, participants would not need to divert their eyes away from the primary task as was the case in the Stefanidis et al. study. Additionally, both tasks would now demand the same visual resources (i.e., focal), further enhancing the diagnostic quality of the secondary task. Results suggested that secondary task performance declined in dual-task conditions as compared to completion of the ball-and-tunnel task alone, showing that the secondary task was sensitive to changes in mental resources in relation to the laparoscopic simulation task.

More recent research using the ball-and-tunnel task has examined its effectiveness in detecting workload differences between laparoscopic tasks of varying difficulty levels and individual ability. Prytz et al. (2012) asked a group of novices to perform three different simulated laparoscopic primary tasks: peg transfer, circle cutting, and suturing. The ball-and-tunnel task was first completed by itself and then together with each of the laparoscopic tasks. The results indicated that secondary task performance declined when paired with any primary task. However, it was also found that secondary task performance fell to lower levels when performed concurrently with the most demanding laparoscopic procedures, indicating that the ball-and-tunnel task was sensitive enough to detect changes in workload as the primary task increased in difficulty. In a related study, Scerbo et al. (2013) tested the ability of the ball-and-tunnel task to discriminate between participants with varying levels of surgical expertise. To do so, a peg transfer primary task was paired with the ball-and-tunnel task and participants were asked to first perform the ball-and-tunnel task followed by both tasks together. Participants were classified as novices, intermediates, or experts. Results revealed that ball-and-tunnel performance again declined when performed with the primary task. More importantly, it was found that novices were significantly poorer at the ball-and-tunnel task than were the intermediates and experts, demonstrating the technique did possess some ability to differentiate between levels of expertise.

CHAPTER IV

PRESENT STUDY

The results obtained using the ball-and-tunnel task seem to indicate that it may be more valid than interval production for assessing mental workload in laparoscopy as the ball-and-tunnel task share common resource demands with laparoscopic procedures. However, additional research comparing the two methods must be undertaken to determine if one is superior to the other. The present study was designed to determine which of these secondary tasks is better for assessing of mental workload in relation to laparoscopic surgical simulation. To do so, a 2x3 mixed-model design was employed. The between-groups factor was the use of either a time estimation task or the ball-and-tunnel task as secondary measure of performance. The within-groups factor was the level of workload. Split-plot analyses of variance were used to test differences in each measure of workload.

To create varying levels of workload, the camera angle was manipulated. Research indicates that as the camera angle moves away from a frontal view, workload experienced by the operator becomes much higher (Conrad et al., 2006; Hanna, Shimi, & Cuschieri, 1998; Klein et al., 2004, 2008). By changing the angle from a normal frontal view of 0° to a 90° view during dual task conditions, the demand on visual-spatial resources should increase and result in higher levels of workload experienced. Additionally, a single task pretest was included to provide a measure of secondary task performance in the absence of primary task demands.

Hypotheses

For a secondary task measure to be sensitive and diagnostic, it must require the same resources as the primary task. If laparoscopy is visual and spatial in nature, then a secondary task that is visual and spatial in design should require the same resources and provide a better indication of residual resource availability than a secondary task that does not. As such, the first hypothesis was that ball-and-tunnel task performance will decline more than the time estimation task in the 90° condition. The second hypothesis was that the interval estimates will not significantly differ between the 0° and 90° camera angle conditions. Because time estimation does not appear to share an identifiable common resource with the highly visual-spatial demands of laparoscopy, the increase in mental workload caused by an increase in camera angle should have a negligible to non-existent effect on such a measure. The third and final hypothesis was that subjective reports of workload as measured by the NASA-TLX will reflect the manipulations on camera angle and demonstrate increases in perceived workload.

CHAPTER V

METHOD

Participants

A power analysis was conducted in G*Power to determine the appropriate number of participants necessary to achieve the same effect size as had been found in previous studies (Kennedy, 2011; Scerbo et al., 2013; partial $\eta^2 = .347$). The results suggested a sample size of 24. Sample recruitment exceeded 24 to ensure the detection of more moderate effects. A total of 37 Old Dominion University undergraduate students participated in the study to fulfill a course requirement or to receive extra credit. All participants were at least 18 years of age, with a mean age of 24. Twenty-two participants were female (62.86%) and thirteen were male (37.14%). All participants had normal or corrected to normal vision. Twenty-one participants (60%) reported playing video games with a reported average of 2.17 hours of gameplay per week.

Material and Equipment

A laparoscopic trainer box was used in all conditions. The trainer was constructed of a plastic box with a drawer approximately 42 cm x 36 cm x 25 cm. The trainer box was used to obscure the participant’s direct vision of the primary tasks. A Mircosoft LifeCam VX-5000 USB video camera was adhered to the inside of the box at fixed locations (0° and 90° respectively) and used to project images from inside the box to an Alienware OPTX AW2210 monitor placed on top of the box. Dell desktops were used to run the ball-and-tunnel task. Separate Toshiba and Alienware laptops were used to run the interval production task.

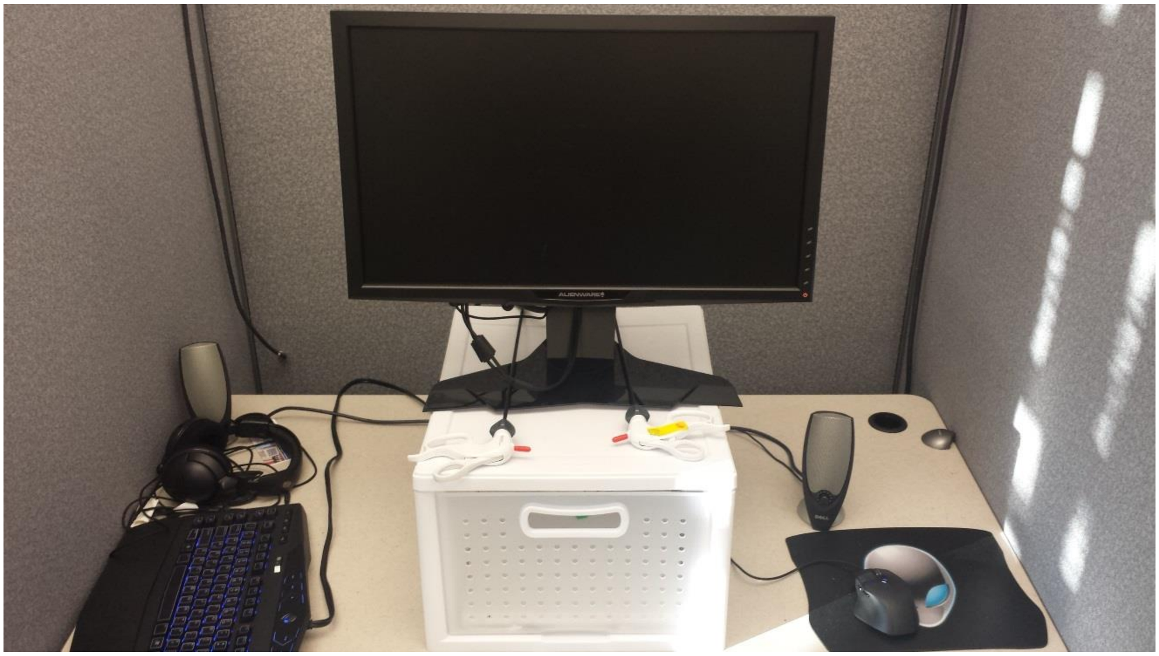

Figure 1. Laparoscopic box trainer.

Primary Task

The primary laparoscopic task was the peg transfer task from the Fundamentals of Laparoscopic Surgery (FLS) training and assessment module. The task was performed with two Johnson & Johnson Ethicon dissector/graspers, a pegboard, and a set of six rubber ring objects. The board was placed on a Velco® strip in the center of the laparoscopic box trainer. The peg transfer task required the participant to lift each rubber ring object with their non-dominant hand, transfer the object in mid-air to the dominant hand, and then place the rubber ring on a peg on the opposite side of the board. No importance was placed on the color of the object, the order of movements, or where the peg was placed. Once all six objects had been transferred, the process was reversed by lifting the rubber ring objects with the dominant hand, transferring each one to the non-dominant hand, and placing them on the original side of the board. The timing of the exercise began when the first rubber ring object was grasped and ended when the last object was placed. The transfer had to take place in mid-air. If a rubber ring object was dropped out of view of the camera, it was left alone and not counted in the total number of movements. Every instance of a rubber ring object being dropped was recorded.

Secondary Task

The ball-and-tunnel task presented an image of four spheres in a representation of a 3D tunnel displayed at 50% transparency. Depth perspective was conveyed in the tunnel using small dots that decrease in size and relative distance toward the center of the image. Images were presented every 2 to 4 s with a mean of 3 s. In the standard configuration, balls were located at the 12, 3, 6, and 9 o'clock positions. Participants were asked to attend to successive images to determine if any of the balls have "moved" from a standard configuration. A change from the standard configuration consisted of one ball appearing to move either closer or farther in the tunnel. Depth changes were represented by a change in the ball's diameter and shift in location. The diameter of the balls in the standard position was 26 mm. If a ball moved closer, the diameter increased to 53 mm and shifted 53 mm from the center. If a ball moved further, the diameter decreased to 11mm and shifted 11mm from the center. Only one of the four balls changed position at any given time, while the other three remained in their standard positions. The average visual angle of the ball-and-tunnel display was 37.25°. Performance in the ball-and-tunnel task was assessed by response time (RT), the proportion of correct responses, and the proportion of false alarms. All measures were recorded by the ball-and-tunnel software.

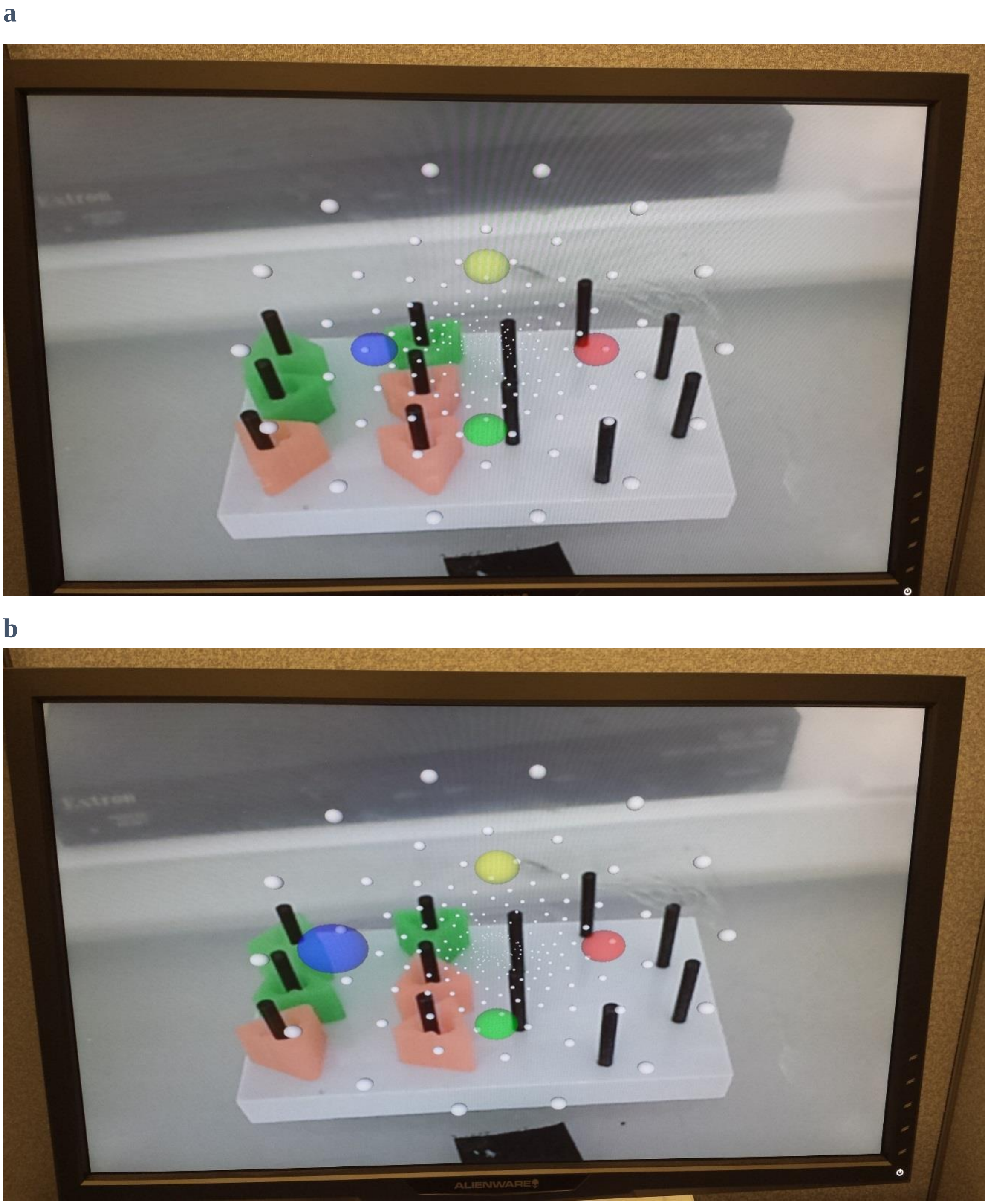

Figure 2. Two images from the ball-and-tunnel task. The top image (a) is the standard configuration to which each test image is compared. The bottom image (b) is an example of a change from the standard orientation, with the leftmost ball changing in depth.

The time estimation task required the participant to indicate every time they believed a 21 s interval had elapsed. The 21 s interval was selected based upon suggestions offered by Grant et al (2013). Some research has suggested that intervals as brief as 3 s may provide adequate sensitivity (Grant, 2010). However, a longer interval was adopted to minimize the possibility that participants would maintain intervals by using a simple counting strategy. The time estimation task began at the beginning of the primary task. Participants pushed a foot pedal to begin the trial. Time estimation performance was assessed using percent absolute error (PAE) and coefficient of variation (COV). PAE was calculated as M_D/TI, where TI was the target interval of 21 s and M_D was the mean absolute difference of the actual produced intervals and the TI. COV was calculated as SD_{PI}/M_{PI}, where M_{PI} was the mean of produced intervals and SD_{PI} was the standard deviation of produced intervals. Intervals were recorded by computer software.

Subjective Measures

The NASA-TLX (Hart & Staveland, 1988) required participants to report their perceived levels of mental demand, physical demand, temporal demand, performance, effort, and frustration on an interval scale from 1 to 20. The NASA-TLX has been validated in past studies (Hart & Staveland, 1988; Rubio, Díaz, Martín, & Puente, 2004) and has been indicated to possess greater sensitivity to changes in workload compared to other methods, such as the SWAT (Nygren, 1991; Luximon & Goonetilleke, 2001). The TLX has also been demonstrated to possess a higher degree of concurrent validity (Rubio et al., 2004) and more consistent estimates of workload (Reid & Nygren, 1988) than its counterparts, as well as greater resolution within its scales (Hill et al., 1992). Additionally, the TLX technique has been shown to have lower between-rater variability than other methods (Hart & Staveland, 1988). The version of the NASA-TLX used in the present research was adapted from the original version (see Appendix C).

Procedure

Participants first read and signed a consent form (see Appendix A) and completed a demographic survey (see Appendix B). Following this, they were randomly assigned to either the interval estimation task or ball-and-tunnel task conditions and completed a baseline measure of secondary task performance. This served as a single-task pretest measure that was compared to dual-task measures gathered in the later trials. Participants completed the NASA-TLX (see Appendix C) after performing the baseline task.

Participants were given 10 minutes to practice the peg transfer task during which they were permitted to ask questions, receive feedback, and familiarize themselves with the procedure. Once completed, no further feedback on performance was be provided.

Once participants were familiar with the both the primary and secondary tasks, they performed the experimental trials. The peg transfer task was completed in both the 0° and 90° angle conditions. The order of conditions was counterbalanced. In each trial, participants were instructed to perform both the peg transfer and secondary task to the best of their abilities, giving greater priority to the peg transfer task. Participants performed each set of tasks for a total of 300 s. Each trial was followed by the completion of the NASA-TLX measure. After completing all trials, participants were debriefed and thanked for their time. Each session took ~50 minutes. No sessions exceeded one hour.

CHAPTER VI

RESULTS

Thirty-seven undergraduate students took part in the study. Two participants were incapable of completing all three conditions. Their data were excluded from the analysis, leaving a total of thirty-five participants (18 participants in the ball-and-tunnel condition and 17 participants in the interval production condition). Additionally, three participants in the interval production condition were found to have data deviating significantly from the mean (3 standard deviations or greater). All of the data from these participants was removed and replaced with data obtained from three new participants. To assess secondary task performance, a repeated-measure analysis of variance (ANOVA) was performed for each secondary task. The dependent measures for the ball-and-tunnel task performance were proportion of correct responses (hits), proportion of incorrect responses (false alarms), and response time. The dependent measures for the interval production task were comprised of percent absolute error (PAE) and coefficient of variation (COV).

Post hoc analyses were used to analyze significant results. Simple main effects were complimented with pairwise comparisons of the mean differences and analyzed with Bonferroni-corrected degrees of freedom. Statistical significance was assessed at the .05 level unless otherwise noted.

Primary Task Performance Results

An ANOVA was performed to assess primary task performance between the levels of dual task conditions and secondary task type. The results from the analysis can be seen in Table 1.

Table 1
Results of the Analysis of Variance for Primary Task Performance

	SS	df	MS	*F*	*p*	partial η^2
Moves						
Workload	516.446	1	516.446	42.554	.000*	.563
Workload x Task	25.703	1	25.703	2.118	.155	.060
Error	400.497	33	12.136			
Drops						
Workload	7.695	1	7.695	2.787	.104	.078
Workload x Task	.838	1	.838	.304	.585	.009
Error	91.105	33	91.105			

Note. *$p < .001$

Peg Transfer Moves. On average, each participant moved 7.80 rubber rings per trial. The analysis revealed a main effect for number of moves between workload levels, $F(1, 33) = 42.55$, $p < .001$, partial $\eta^2 = .563$. Post hoc comparisons indicated that significantly more rings were moved in the 0° visual condition than in the 90° visual condition (see Table 2).

Peg Transfer Drops. The average number of drops per trial for each participant was found to be 1.91. No significant difference was found for drops between workload levels (see Table 1).

Table 2
Means, Standard Errors, and Confidence Intervals for Primary Task Performance Measures by Task Type

			Mean	Std. Error	95% Confidence Interval	
					Lower	Upper
Ball-and-Tunnel						
	Moves					
		0°	9.444	1.055	7.299	11.590
		90°	5.222	.923	3.344	7.100
		Average	7.333	.803	5.700	8.967
	Drops					
		0°	1.944	.573	.778	3.111
		90°	1.500	.354	.780	2.220
		Average	1.722	.388	.934	2.511

Table 2 Continued

		Mean	Std. Error	95% Confidence Interval	
				Lower	Upper
Interval Production					
	Moves				
	0°	11.588	1.085	9.381	13.796
	90°	4.941	.950	3.009	6.873
	Average	8.265	.826	6.584	9.946
	Drops				
	0°	2.529	.590	1.329	3.729
	90°	1.647	.364	.906	2.388
	Average	2.088	.399	1.277	2.900

Ball-and-Tunnel Secondary Task Results

Proportion of Correct Detections. The mean proportion of correct detections per trial for all conditions was .60. A repeated measures ANOVA indicated the proportion of correct detections was significantly lower in more demanding conditions than in less demanding conditions, $F(2, 32) = 63.38$, $p < .001$, partial $\eta^2 = .798$. See Table 3 for all ball-and-tunnel ANOVA results.

Table 3

Results of the Analysis of Variance for Ball-and-Tunnel Proportion of Correct Detections

	SS	df	MS	*F*	*p*	partial η^2
P(Correct Detections)	2.635	2	1.137	63.378	.000*	.798
Error(Correct Detections)	.665	32	.021			

Note. **p* < .001

Pairwise comparisons were performed to indicate which workload levels differed. The results indicated that the proportion of correct detections was significantly higher during the pretest than in either the 0° or 90° angle visual conditions. The proportion of correct detections was also found to be significantly higher in the 0° visual condition than in the 90° condition (see Table 4).

Proportion of False Alarms. The mean proportion of false alarms per trial for all conditions was .19. No significant differences were found among workload levels (see Table 4 for descriptive statistics; see Table 5 for ANOVA results).

Response Time. The mean response time per trial for all conditions was 0.92. The repeated measures ANOVA revealed a significant difference between workload conditions, $F(2, 32) = 24.11$, $p < .001$, partial $\eta^2 = .601$. See Table 6 for results of ANOVA.

Table 4

Means, Standard Errors, and Confidence Intervals for Ball-and-Tunnel Dependent Measures

	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
P(Correct Detections)				
Pretest	.902	.028	.843	.961
0°	.534	.038	.453	.615
90°	.356	.063	.222	.490
P(False Alarm)				
Pretest	.223	.047	.123	.324
0°	.173	.042	.084	.263
90°	.177	.050	.070	.284
Response Time				
Pretest	.728	.035	.654	.801
0°	.980	.043	.889	1.070
90°	1.063	.066	.924	1.203

Table 5

Results of the Analysis of Variance for Ball-and-Tunnel Proportion of False Alarms

	SS	df	MS	*F*	*p*	partial η^2
P(False Alarm)	.260	2	.013	1.460	.247	.084
Error(False Alarm)	.289	32	.009			

Table 6

Results of the Analysis of Variance for Ball-and-Tunnel Response Time

	SS	df	MS	*F*	*p*	partial η^2
Response Time	1.039	2	.520	24.114	.000*	.601
Error	.689	32	.022			

Note. *$p < .001$

Pairwise comparisons revealed a significantly faster response times in the pretest condition than in either the 0° or 90° visual angle conditions (see Table 4). No significant difference was found between the 0° and 90° conditions.

Interval Production Secondary Task Results

Percent Absolute Error. The mean PAE per trial across all interval estimation conditions was .46. No significant difference was found for PAE between workload levels. See Table 7 for all interval production ANOVA results.

Table 7
Results of the Analysis of Variance for Interval Production Percent Absolute Error

	SS	df	MS	*F*	*p*	partial η^2
PAE	.324	1.478**	.219	1.251	.294	.073
Error(PAE)	4.139	23.650**	.175			

** Greenhouse-Geisser Corrected

Coefficient of Variation. The mean COV per trial for all workload conditions was 0.230. Results indicated that a significant effect for COV across workload levels, $F(2, 32) = 8.54$, $p = .001$., partial $\eta^2 = .348$. See Table 9 for ANOVA results.

Follow up comparisons indicated significantly higher COV in the pretest condition than in the 0° and 90° dual task conditions. No difference observed between the 0° and 90° degree conditions (see Table 8).

Table 8
Means, Standard Errors, and Confidence Intervals for Interval Production Dependent Measures

		Mean	Std. Error	95% Confidence Interval	
				Lower	Upper
PAE					
	Pretest	.385	.098	.177	.593
	0°	.431	.085	.250	.611
	90°	.572	.109	.342	.803
COV					
	Pretest	.157	.022	.111	.203
	0°	.246	.021	.202	.291
	90°	.285	.027	.229	.342

Table 9
Results of the Analysis of Variance for Interval Production Coefficient of Variation

	SS	df	MS	*F*	*p*	partial η^2
COV	.147	2	.073	8.535	.001*	.348
Error	134.489	32	4.203			

Note. *p < .001

NASA-TLX Score Results

An ANOVA was conducted to assess subjective workload scores under different workload levels and secondary task types. The mean total workload score per trial for all conditions was 70.54. The analysis revealed that total TLX scores were significantly different between workload levels, $F(2, 66) = 90.72$, $p < .001$, partial $\eta^2 = .733$. See Table 10 for ANOVA results.

Follow up tests showed significantly higher total scores in both the 0°and 90° visual conditions comparable to the pretest as well as a significantly higher total scores for the 90° compared to the 0°condition (see Table 11).

Table 10
Results of the Analysis of Variance for NASA-TLX Subscale and Total Scores

	SS	df	MS	*F*	*P*	partial η^2
Total						
Workload	33838.579	2	16919.289	90.720	.000**	.733
Workload x Task	1684.217	2	842.108	4.515	.015*	.120
Error	12309.002	66	186.500			
Mental						
Workload	1333.132	1.478***	902.018	70.518	.000**	.681
Workload x Task	70.465	1.478***	47.678	3.727	.043*	.101

Table 10 Continued

	SS	df	MS	*F*	*P*	partial η^2
Error	623.858	48.772***	12.791			
Physical						
Workload	2381.313	1.677***	1420.094	106.862	.000**	.764
Workload x Task	41.427	1.677***	24.705	1.859	.171	.053
Error	735.373	55.337***	13.289			
Temporal						
Workload	239.483	2	119.742	9.285	.000**	.220
Workload x Task	222.264	2	111.132	8.618	.000**	.207
Error	851.126	66	12.896			
Performance						
Workload	456.582	2	228.291	13.795	.000**	.295
Workload x Task	14.067	2	7.034	.425	.656	.013
Error	1092.237	66	16.549			
Effort						
Workload	1198.170	2	599.085	43.462	.000**	.568
Workload x Task	33.522	2	16.761	1.216	.303	.036
Error	909.754	66	13.784			
Frustration						

Table 10 Continued

	SS	df	MS	*F*	*P*	partial η^2
Workload	1044.177	2	522.089	38.900	.000**	.541
Workload x Task	105.663	2	52.831	3.936	.024*	.107
Error	885.804	66	13.421			

Note. * p <.05, **p < .001; *** Greenhouse-Geisser Corrected

NASA-TLX Subscale Results. Each TLX subscale was found to be significant at the .01 level or lower (see Table 10 for details). Follow up comparisons indicated that mental demand, physical demand, and frustration scores were all found to be significantly different between conditions, with scores significantly lower in the pretest condition than in either the 0° or 90° visual conditions and with lower scores in the 0° condition compared to the 90° condition. Both temporal demand and performance score were found to be significantly higher in the 90° condition than in either the pretest or 0° conditions but no difference was found between the pretest and 0° condition scores. Effort scores were significantly lower in the pretest than in either the 0° or 90° conditions but scores did not differ between the two dual task conditions (See Table 11 for all means and descriptives).

Table 11

Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale and Total Scores by Task Type

			Mean	Std. Error	95% Confidence Interval	
					Lower	Upper
Ball-and-Tunnel						
	Total					
		Pretest	54.167	4.997	44.000	64.333
		0°	74.389	5.162	63.887	84.891
		90°	88.500	4.192	79.971	97.029
	Mental					
		Pretest	9.333	1.350	6.587	12.079
		0°	13.167	1.118	10.892	15.441
		90°	16.278	.834	14.580	17.975
	Physical					
		Pretest	4.333	.767	2.772	5.895
		0°	12.667	1.267	10.090	15.244
		90°	13.667	1.203	11.219	16.114
	Temporal					
		Pretest	13.111	1.163	10.746	15.476
		0°	12.500	1.184	10.092	14.908
		90°	13.611	.978	11.621	15.601
	Performance					
		Pretest	8.556	1.090	6.338	10.774
		0°	10.056	1.089	7.841	12.270
		90°	13.500	1.081	11.302	15.698
	Effort					
		Pretest	9.611	1.249	7.070	12.152
		0°	14.444	.998	12.414	16.474
		90°	16.889	1.025	14.804	18.973

Table 11 Continued

		Mean	Std. Error	95% Confidence Interval	
				Lower	Upper
	Frustration				
	Pretest	9.222	1.232	6.716	11.728
	0°	11.556	1.128	9.260	13.851
	90°	14.556	1.192	12.130	16.981
Interval Production					
	Total				
	Pretest	39.294	5.142	28.833	49.755
	0°	75.176	5.311	64.370	85.983
	90°	91.706	4.314	82.929	100.483
	Mental				
	Pretest	7.412	1.389	4.586	10.237
	0°	14.882	1.150	12.542	17.223
	90°	17.647	.859	15.900	19.394
	Physical				
	Pretest	2.882	.790	1.276	4.489
	0°	12.765	1.303	10.113	15.416
	90°	15.294	1.238	12.776	17.812
	Temporal				
	Pretest	7.235	1.196	4.801	9.669
	0°	12.706	1.218	10.228	15.184
	90°	14.000	1.006	11.952	16.048
	Performance				
	Pretest	9.294	1.122	7.012	11.576
	0°	9.059	1.120	6.780	11.338
	90°	13.765	1.112	11.503	16.027
	Effort				

Table 11 Continued

	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
Pretest	7.471	1.285	4.856	10.085
0°	15.059	1.027	12.970	17.148
90°	15.882	1.054	13.737	18.027
Frustration				
Pretest	5.000	1.267	2.421	7.579
0°	10.706	1.161	8.344	13.068
90°	15.118	1.227	12.622	17.613

CHAPTER VII

DISCUSSION

The purpose of the current study was to compare two secondary tasks used to measure laparoscopic workload and determine which demonstrated greater sensitivity to changes in mental workload. More specifically, a visual-spatial and an interval production secondary task were used to measure spare attentional capacity during a laparoscopic peg transfer task. Each participant performed one of the secondary tasks alone and in dual task conditions at both 0° and 90° visual angles. The different visual angles were implemented to manipulate workload levels, with the 90° visual condition intended to create higher workload than the 0° visual condition.

Primary Task

Results indicated that participants moved significantly fewer pegs in the 90° visual condition than in the 0° condition, supporting the assumption that workload was higher in the 90° condition. Additionally, there were slightly fewer drops in the 90° condition than in the 0° condition, but this difference was not significant. However, the lower number of drops in the higher workload condition was likely due to participants making fewer moves altogether which subsequently limited the number of opportunities for drops.

Secondary Task

A good secondary task measure of residual workload should address the same resource demands as the primary task to be both sensitive to changes in workload and be diagnostic of the resources allocated during periods of higher workload (O'Donnell & Eggemeier, 1986; Carswell, Clarke, & Seales, 2005). The first goal of the study was to examine the sensitivity of a visual-spatial and an interval production secondary task to determine changes in workload. Specifically, it was predicted that secondary task performance would be significantly worse in the 90° condition for the ball-and-tunnel task as compared to the interval production task, because the ball-and-tunnel task requires resources that are more similar to those needed for the laparoscopic task than the interval production task. According to Multiple Resource Theory (MRT, Wickens, 1980, 1984, 2002, 2008), two tasks that demand the same mental resources will interfere with one another. As it has been suggested that laparoscopy is primarily visual and spatial in nature (Cuschieri, 1995, 2006; Tendick, Bhoyrul, & Way, 1997; Way et al., 2003), a secondary task that demands similar resources should demonstrate a greater level of diagnosticity over a different secondary task drawing from a different pool of resources. The results supported this hypothesis. Regarding the ball-and-tunnel task, participants were significantly less accurate and slower to respond in the 90° dual task condition than in either the single task pretest or the 0° dual task conditions, reflecting sensitivity to workload changes. The effect sizes observed for both the proportion of hits and response times were large with a partial η^2 of .80 and .60, respectively.

A significant result for the COV indicated that interval productions were also sensitive to changes in workload, with the consistency of productions declining in the dual-task conditions compared to the single-task pretest. However, no significant difference was observed between the 90° and 0° dual-task conditions, demonstrating an inability of time estimation to detect a change in resource demands between the visual angles. Additionally, the effect size observed for COV was moderate at .35 and much smaller than that of the ball-and-tunnel measures. Since the magnitude of difference between ball-and-tunnel measures was larger than those of the interval production task, the hypothesis that the ball-and-tunnel task would be more sensitive in high workload situations was supported. Ball-and-tunnel performance did become less accurate than interval production performance in the 90° visual condition as compared to the single task pretest.

The second goal of the study was to determine if the interval production method was capable of distinguishing lower and higher levels of mental workload. Specifically, it was hypothesized that interval estimate measures would not differ significantly between the 0° and 90° visual conditions. Since the interval production task does not share common resources with laparoscopy, it was believed that a temporally-based secondary task would not be sensitive to small changes in mental workload. This hypothesis was also supported. Neither PAE nor COV were able to detect differences between the 0° and 90° visual conditions, indicating a lack of discriminability between less extreme shifts in mental workload. However, COV was able to discriminate between the single-task pretest and the dual-task conditions, suggesting some sensitivity to workload changes.

It is important to remember that a critical difference between the ball-and-tunnel and interval production tasks was observed. In the ball-and-tunnel condition, each successive increase in task demand resulted in a significant difference. For example, the proportion of hits was significantly higher and mean response time was significantly lower in the single task pretest than in the 0° visual dual task condition. Similarly, the proportion of hits was higher and mean response time were lower in the 0° dual task condition than in the 90° dual task condition. Such was not the case for the interval production task measured by COV. MRT (Wickens, 1980, 1984, 2002, 2008) would predict a secondary task measure with high sensitivity and diagnosticity would be able to differentiate between all changes in workload due to the resource demands common to both the primary and secondary tasks. When workload increases and demands for a specific set of resources grows, sensitive and diagnostic measures should reflect the change regardless of the magnitude of change (O'Donnell & Eggemeier, 1986). Interval production did not demonstrate these characteristics. More specifically, the interval production task seemed to be sensitive only to extreme changes in workload, such as those between the pretest and dual-task conditions, and not to more moderate changes, such as that between the 0° and the 90° conditions. On the other hand, the ball-and-tunnel task demonstrated sensitivity and was able to differentiate between workload levels in all conditions. The increased sensitivity of the ball-and-tunnel task to changes in workload over the interval production task may be due a higher degree of diagnosticity. Unlike the interval production task, the ball-and-tunnel task was able to detect the difference between visual angles during visual-spatial tasks. Thus, it is reasonable to conclude that the ball-and-tunnel task specifically addressed differences in visual and spatial resource demands.

On the contrary, interval production likely possesses a lesser degree of diagnosticity to visual-spatial resource demands. Still, since the interval production task demonstrated the ability to differentiate between more exaggerated shifts in task demands, it is clearly sensitive to changes in workload but on a global level. Possessing sensitivity to workload changes gives interval production some value as a workload measure. However, it is important to note that the resource demands of laparoscopy are largely visual and spatial (Cuschieri, 1995, 2006; Tendick, Bhoyrul, & Way, 1997; Way et al., 2003). As such, a secondary task that measures resource pools shared by the primary task should produce a keener estimate of residual resource availability. Indeed, such an effect was observed between ball-and-tunnel and interval estimation performance. Ball-and-tunnel performance measures resulted in a much larger magnitude of change between workload levels than did interval production.

The difference between the two workload measures is important. As Carswell, Clarke, and Seales (2005) indicate, a relatively small change in workload may have dire consequences regarding successful performance on surgical tasks. Any method of measuring mental workload in the domain of laparoscopy would need to possess the ability to capture minute shifts resource demand to be considered an effective choice. As such, the ball-and-tunnel task is likely a more appropriate secondary task for mental workload assessment in the domain of laparoscopy.

Subjective Workload

The third hypothesis was that NASA-TLX scores would reflect the workload differences between the single task pretest, 0° visual dual task condition, and the 90° visual dual task condition, corroborating that workload differed among conditions. This hypothesis was supported. The total workload scores were found to be sensitive to changes in workload in all conditions and possessed a large partial η^2 of .73. Each subscale was also sensitive to the workload conditions. Differences were found between each workload level for the mental demand, physical demand, and frustration subscales. Both the temporal demand and performance subscales were found to have significantly higher scores in the 90° condition than in the 0° condition, but no such difference was observed between the 0° condition and the pretest. The effort subscale score was found to be significantly lower in the pretest than in the 0° condition, but no difference was found between the 0° and 90° conditions.

The lack of differences for temporal demand and performance scores between the pretest and the 0° conditions suggest that progressing from the single to dual task situation did not make participants feel more rushed or less able to perform the task. It is also possible that the lack of differences between the pretest and 0° dual task condition may be have been due to experimental design. All participants completed the pretest first, after which they moved on to the 0° or 90° conditions. The relative ease of the 0° degree condition after experiencing the demands of the 90° condition may have biased participant responses to the temporal demand and performance scales. Additionally, the time allotted for the pretest was 90 s compared to 300 s for the dual task conditions. The greater amount of time given for the dual task conditions may have also affected the way participants responded to these subscales. On the other hand, the lack of a significant difference for effort between the 0° and 90° conditions indicates that participants did not feel they needed to expend any additional effort to complete the tasks. These results are somewhat unexpected. Since participants had indicated that mental and physical demands were greater as they progressed from the single task pretest to the dual task conditions and from the 0° visual to the 90° visual, it would stand to reason that the effort experienced would increase as well. However, increasing the demands of a task does not necessitate an increase in effort in the part of the participant. Indeed, the results obtained suggest that participants did not produce a greater amount of effort between the dual task conditions or, at least, believed the effort expended was equal. Overall, the differences found would seem to indicate that workload did, in fact, increase from single task to dual task and from 0° to 90° visual conditions.

Interestingly, an interaction effect was found for the temporal demand subscale score between workload level and task type. Follow up tests indicated that the effect was due to a large difference between temporal demand scores in the pretest, with the ball-and-tunnel task scoring significantly higher than the interval estimation task. Although this was not an anticipated result, one might reasonably assume that a time-based secondary task would be perceptually more temporally demanding than a visual-spatial task, yet this did not appear to be the case.

Theoretical Implications

The current experimental findings are consistent with the predictions made by Multiple Resource Theory (Wickens, 1980, 1984, 2002, 2008). The difference between workload conditions was more pronounced when measured with a visual-spatial secondary task than an interval estimation secondary task. The mean proportion of hits for the ball-and-tunnel tasks dropped from .90 in the single task pretest to .53 in the dual task 0° visual condition. Similarly, the mean proportion of hits dropped from .53 in the 0° condition to .36 in the 90° dual task condition. This result would seem to indicate resource demands between the two tasks were similar and, as more resources were allocated for the primary task, less were available for the ball-and-tunnel task, leading to poorer performance and demonstrating that a visual-spatial task is a sufficiently sensitive and diagnostic measure of residual workload in laparoscopy.

However, interval production measures were not expected to differ between conditions. Although interval production failed to detect more subtle changes in workload, the results indicated the interval production task possessed sufficient sensitivity to detect changes in workload between the lowest workload level and the higher two. This effect suggests that prospective interval production may be an acceptable secondary task to measure workload in some cases, specifically in domains where the resource demands are more globally distributed over multiple resource pools rather than isolated to specific resource types. However, as the task demonstrated less sensitivity than that of the ball-and-tunnel task, it does not appear to be a more superior measure for the domain of laparoscopic surgery.

Still, interval production demonstrated a level of sensitivity that was unexpected, implying that a temporally-based secondary task shares some degree of resource demand with a highly visual and spatial task such as laparoscopy. One possible explanation for this result may relate to Expanded Multiple Resource Theory (EMRT). Boles and his colleagues (Boles & Law, 1998; Boles, 2010) propose that mental resources are orthogonal resulting from orthogonal mental processes. For example, MRT would predict that a reading task and a target identification task would interfere with each other as visual resources must be expended to attend to both tasks. However, EMRT would predict the tasks would require separate resources, as the reading task is a visual-lexical process task and the target identification task is spatial-positional. EMRT has found some support for a visual-temporal process pool of resources, which may explain why some degree of sensitivity may exist between a visual-spatial primary task and an interval estimation secondary task. However, it is difficult to determine if this can explain the interval production results observed in the present study as additional research into

EMRT would be necessary to determine if this is the case.

Another possible explanation for the unexpected results observed may be in the design of the interval estimation task itself. For example, the ball-and-tunnel task is a forced-choice task based on signal detection wherein participants have a limited amount of time to make a decision on whether the standard configuration was present. If they do not respond, it is either a correct rejection or a miss. There is no ambiguity regarding the presentation of a target, only in the perception of the target. On the other hand, prospective interval production is not amenable to signal detection analysis since a person cannot perceive the passage of time without actively tracking it. Interval estimation requires constant awareness of time and maintaining this awareness is a demand of the task in its own right. Maintenance of this awareness may require the use of the central executive function (Baddeley & Hitch, 1974; Baddeley, 1996), which coordinates performance on multiple tasks and selection of attended stimuli. As participants are asked to attend to the absence of a stimuli instead of their presence, the central executive is more likely to be overtasked, leading to the inability of the participant to successfully estimate a more accurate interval. Indeed, the central executive function has been proposed to be similar to an individual's resource allocation policy (Wickens, 2008).

Limitations

One possible limitation in the present study may have been the time allotted for the pretest. The pretest measurement for each secondary task was limited to 90 s. Although this did not appear to impact the ball-and-tunnel measure, this may not have been an adequate amount of time on which to base the interval production baseline measure since participants were able to only produce 3 to 4 intervals in this time. Some recommendations have suggested that at least 5 to 8 productions are necessary to produce sufficient estimates of performance (Carswell et al., 2013). As a result, the baseline measures for the interval production tasks may have been too short to produce stable single task estimates of estimation performance. Additionally, past researchers have given participants 90 s practice trials through which to familiarize themselves with the task, followed by feedback and dual task practice (Carswell et al., 2013). Practice effects may explain differences in findings between the present experiment and past interval production studies. However, participants received no additional practice or instruction on the ball-and-tunnel task and were instructed on how to perform each task until they had indicated that they were comfortable enough to begin. Additionally, the COV is resistant to individual differences in interval variations so such an effect might be negligible if not irrelevant. Further research assessing the effects of practice on both tasks may be needed.

Future Work

A possible direction for future work might be to create a version of the time estimation task that is more similar to the ball-and-tunnel task so that the two might be more easily compared to each other. Although it is possible to compare the results of the two tasks statistically, they are qualitatively different. Creating a version of the time estimation task that can cue an individual to indicate if the target interval had passed may make comparing the two tasks possible. If the attentional-gate control model of duration judgment (Thomas & Weaver, 1975; Zakay, 1989; Zakay & Block, 1997; Zakay & Shub, 1998) is accurate, then it should be possible to ask participants to judge whether a cue occurs in concert with the target interval as an alternative measure of interval estimation.

The completion of such a task would still require the use of the mental accumulator and temporal reference memories but would allow the experimenter to use the same dependent measures for both the visual-spatial and interval production tasks, and directly compare hits and response times in both tasks.

CHAPTER VIII

CONCLUSION

In the present study, a laparoscopic peg transfer task was coupled with both a visual-spatial and temporal secondary task to compare sensitivity to changes in mental workload. Errors committed during laparoscopy can be largely attributed to visual and spatial distortions, indicating that the task is predominately visual and spatial in resource demand. Because laparoscopic surgery is visual and spatial in nature, assessing residual resource availability as a metric of mental workload requires a measure that is both diagnostic of the visual and spatial pools of resources and sensitive to the changes in resource availability within each. The ball-and-tunnel task was hypothesized to be an a superior measure of mental workload compared to interval production since it addresses the same resources as laparoscopy. The results supported this hypothesis, indicating that the ball-and-tunnel task was sensitive to both moderate and extreme changes in workload.

The temporally-based interval estimation task also demonstrated sensitivity but to a lesser extent. While the results did not indicate that interval estimation was as sensitive to changes in workload as the ball-and-tunnel task, time estimation may be sensitive to more extreme variations in workload distributed over a greater number of resource types. However, since laparoscopy is highly demanding on visual and spatial resource pools, interval estimation may be less suitable than the ball-and-tunnel task.

Subjective workload measures and number of pegs transfered in the primary task reflected changes in workload between conditions where predicted, lending support to the assumption that workload increased from single task to 0° dual task conditions and 0° to

90° dual task conditions. Since ball-and-tunnel task performance more closely matched the subjective reports, the hypothesis that visual-spatial secondary tasks are better measures of mental workload is supported further.

In general, the results indicate that mental workload in laparoscopy can be measured accurately through the use of a visual-spatial task. Current simulator-based methods of training laparoscopic skill have shown an ability to transfer to the actual surgical environment (Dawe et al., 2013). However, novices trained with these methods do not always perform as well in real surgery as those with more experience (Korndoffer et al., 2005). Other researchers have indicated that laparoscopic simulation is not as effective as traditional surgical training methods (Mansour et al., 2012). The disparity of results within the laparoscopic training literature seems to indicate that a deeper understanding of the demands placed on surgeons is necessary to assure that they are fully prepared for actual surgery. Measuring residual mental workload with a secondary task similar to the ball-and-tunnel task demonstrated here may ultimately prove to be a useful method for decreasing surgical error and improving patient outcomes for novice surgeons in shorter periods of time.

REFERENCES

Alkhayal, A., Aldhukair, S., Alselaim, N., Aldekhayel, S., Alhabdan, S., Altaweel, W., Magzoub, M.E., & Zamakhshary, M. (2012). Toward an objective assessment of technical skills: a national survey of surgical program directors in Saudi Arabia. *Advances in Medical Education and Practice*, *3*, 97-105. doi: 10.2147/AMEP.S31720

Aziz, O., Constantinides, V., Tekkis, P. P., Athanasiou, T., Purkayastha, S., Paraskeva, P., Darzi, A.W., & Heriot, A. G. (2006). Laparoscopic versus open surgery for rectal cancer: a meta-analysis. *Annals of Surgical Oncology, 13*, 413-424. doi: 10.1245/ASO.2006.05.045

Baddley, A.D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47-89. doi: 10.1016/S0079-7421(08)60452-1

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49*, 5-28. doi:10.1080/027249896392784

Baldauf, D., Burgard, E., & Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied Ergonomics, 40,* 929-935. doi: 10.1016/j.apergo.2009.01.004

Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy, 15,* 1204-1207. doi: 10.1007/s004640080030

Bittner, J. G., Hathaway, C. A., & Brown, J. A. (2008). Three-dimensional visualisation and articulating instrumentation: Impact on simulated laparoscopic tasks. *Journal of Minimal Access Surgery, 4,* 31-38.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*, 184-197. doi: 10.3758/BF03209393

Boles, D. B. (2010). Multiple resources. *International Encyclopedia of Ergonomics and Human Factors, 1,* 445-450.

Boles, D. B., & Law, M. B. (1998). A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 204-215. doi: 10.1037/0096-1523.24.1.204

Braga, M., Vignali, A., Zuliani, W., Frasson, M., Di Serio, C., & Di Carlo, V. (2005). Laparoscopic versus open colorectal surgery. *Annals of Surgery, 242,* 890-896. doi: 10.1097/01.sla.0000189573.23744.59

Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY, US: Pergamon Press. doi: 10.1037/10037-000

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics, 59,* 1118-1140. doi: 10.3758/BF03205526

Carswell, C. M., Clarke, D., & Seales, W.B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical Innovation, 12,* 80-90. doi: 10.1177/155335060501200112

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*, 975-979. doi: 10.1121/1.1907229

Conrad, J., Shah, A., Divino, C., Schluender, S., Gurland, B., Shlasko, E., & Szold, A. (2006). The role of mental rotation and memory scanning on the performance of laparoscopic skills. *Surgical Endoscopy and Other Interventional Techniques, 20*, 504-510. doi: 10.1007/s00464-005-0363-7

Cuschieri, A. (1995). Visual displays and visual perception in minimal access surgery. *Surgical Innovation, 2*, 209-214. doi: 10.1177/155335069500200309

Cuschieri, A. (2006). Epistemology of visual imaging in endoscopic surgery. *Surgical Endoscopy and Other Interventional Techniques, 20*, 419-424. doi: 10.1007/s00464-006-0020-9

Dawe, S. R., Windsor, J. A., Broeders, J. A., Cregan, P. C., Hewett, P. J., & Maddern, G. J. (2014). A systematic review of surgical skills transfer after simulation-based training: Laparoscopic cholecystectomy and endoscopy. *Annals of Surgery, 259*, 236-248. doi: 10.1097/SLA.0000000000000245

Dominguez, C. O. (2001). Expertise and metacognition in laparoscopic surgery: A field study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*, 1298-1302. doi: 10.1177/154193120104501709

Gallagher, A., McClure, N., McGuigan, J., Ritchie, K., & Sheehy, N. (1998). An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy, 30*, 617-620. doi: 10.1055/s-2007-1001366

Gallagher, A. G., Al-Akash, M., Seymour, N. E., & Satava, R. M. (2009). An ergonomic analysis of the effects of camera rotation on laparoscopic performance. *Surgical Endoscopy, 23*, 2684-2691. doi: 10.1007/s00464-008-0261-x

Grant, R.C. (2010). *Time estimation errors as an index of task demand during laparoscopic skills training: Effects of target duration and attention allocation* (Unpublished master's thesis). University of Kentucky, Lexington, KY.

Grant, R. C., Carswell, C. M., Lio, C. H., & Seales, W. B. (2013). Measuring surgeons' mental workload with a time-based secondary task. *Ergonomics in Design: The Quarterly of Human Factors Applications, 21,* 7-11. doi: 10.1177/1064804612466068

Hanna, G. B., & Cuschieri, A. (2000). Influence of two-dimensional and three-dimensional imaging on endoscopic bowel suturing. *World Journal of Surgery, 24,* 444-448. doi: 10.1007/s002689910070

Hanna, G. B., Shimi, S. M., & Cuschieri, A. (1998). Task performance in endoscopic surgery is influenced by location of the image display. *Annals of Surgery, 227,* 481-484.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183. doi: 10.1016/S0166-4115(08)62386-9

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklade, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 34,* 429-439. doi: 10.1177/001872089203400405

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall.

Kennedy, R.A. (2011). *The effect of depth judgments on mental workload in laparoscopy measured by a visual-spatial secondary task* (Unpublished master's thesis). Old Dominion University, Norfolk, VA.

King, P.M., Blazeby, J.M., Ewings, P., Franks, P.J., Longman, R.J., Kendrick, A.H, Kipling, R.M., & Kennedy, R.H. (2006). Randomized clinical trial comparing laparoscopic and open surgery for colorectal cancer within an enhanced recovery programme. *British Journal of Surgery, 93,* 300-308. doi: 10.1002/bjs.5216

Klein, M. I., Warm, J. S., Riley, M. A., Matthews, G., Gaitonde, K., & Donovan, J. F. (2008). Perceptual distortions produce multidimensional stress profiles in novice users of an endoscopic surgery simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 291-300. doi: 10.1518/001872008x288312

Klein, M. I., Warm, J. S., Riley, M. A., Matthews, G., & Parsons, K. S. (2004). The multi-dimensional nature of stress in an endoscopic surgery simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48,* 1700-1703. doi: 10.1177/154193120404801515

Korndoffer Jr., J.R., Dunne, J.B., Sierra, R., Stefanidis, D., Touchard, C.L., & Scott, D.J. (2005). Simulator training for laparoscopic suturing using performance goals translates to the operating room. *Journal of the American College of Surgeons, 201,* 23-29. doi: 10.1016/j.jamcollsurg.2005.02.021

Lio, C. H., Bailey, K., Carswell, C. M., Seales, W. B., Clarke, D., & Payton, G. M. (2006). Time estimation as a measure of mental workload during the training of laparoscopic skills. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*, 1910-1913. doi: 10.1177/154193120605001741

Lio, C. H., Carswell, C. M., Seales, W. B., Clarke, D., Kurs, Y., & DeCuir, J. (2007). Using global implicit measurement strategies to assess situation awareness during the training of laparoscopic surgical skills. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51*, 1280-1282. doi: 10.1177/154193120705101849

Liu, Y., & Wickens, C. D. (1994). Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. *Ergonomics, 37*, 1843-1854. doi: 10.1080/00140139408964953

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44*, 229-243. doi: 10.1080/00140130010000901

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., & Wherry, R.J. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (DTIC Publication No. ADA212879). Retrieved from DTIC Online website: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA212879

Mansour, S., Din, N., Ratnasingham, K., Irukulla, S., Vasilikostas, G., Reddy, M., & Wan, A. (2012). Objective assessment of the core laparoscopic skills course. *Minimally Invasive Surgery, 2012*, 1-4. doi: 10.1155/2012/379625

Milsom, J. W., Böhm, B., Hammerhofer, K. A., Fazio, V., Steiger, E., & Elson, P. (1998). A prospective, randomized trial comparing laparoscopic versus conventional techniques in colorectal cancer surgery: a preliminary report. *Journal of the American College of Surgeons, 187*, 46-54. doi: 10.1016/S1072-7515(98)00132-X

Mohr, F. W., Falk, V., Diegeler, A., Walther, T., Gummert, J. F., Bucerius, J., Jacobs, S., & Autschbach, R. (2001). Computer-enhanced "robotic" cardiac surgery: experience in 148 patients. *Journal of Thoracic and Cardiovascular Surgery, 121*, 842-853. doi: 10.1067/mtc.2001.112625

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84-92. doi: 10.1016/0001-6918(67)90048-0

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*, 214-255. doi: 10.1037/0033-295X.86.3.214

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64. doi: 10.1016/0010-0285(75)90004-3

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 33*, 17-33. doi: 10.1177/001872089103300102

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance, vol. II: Cognitive processes and performance* (pp. 1-49). Oxford, England: John Wiley & Sons.

Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 21,* 529-548. doi: 10.1177/001872087902100502

Prytz, E, Montano, M., Kennedy, R., Scerbo, M., Britt, R., Davis, S., & Stefanidis, D. (2012). Using a spatial task to measure laparoscopic mental workload: Initial results. *Simulation in Healthcare, 7,* 511.

Reid, G.B. & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in Psychology, 52,* 185-218. doi: 10.1016/S0166-4115(08)62387-0

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology, 53,* 61-86. doi: 10.1111/j.1464-0597.2004.00161.x

Scerbo, M. W., Kennedy, R. A., Montano, M., Britt, R. C., Davis, S. S., & Stefanidis, D. (2013). A spatial secondary task for measuring laparoscopic mental workload: Differences in surgical experience. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*, 728-732. doi: 10.1177/1541931213571159

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

Stefanidis, D., Scerbo, M. W., Korndorffer, J. R., Jr., & Scott, D. J. (2007). Redefining simulator proficiency using automaticity theory. *American Journal of Surgery, 193*, 502-506. doi: 10.1016/j.amjsurg.2006.11.010

Tendick, F., Bhoyrul, S., & Way, L. W. (1997). Comparison of laparoscopic imaging systems and conditions using a knot-tying task. *Computer Aided Surgery, 2,* 24-33. doi:10.3109/10929089709149080

Tendick, F., & Cavusoglu, M. C. (1997). Human-machine interfaces for minimally invasive surgery. *Proceedings of the 19th Annual International Conference of the IEEE, 6,* 2771-2776. doi: 10.1109/IEMBS.1997.756904

Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M. C., Feygin, D., Wu, X., Eyal, R., Hegarty, M., & Way, L. W. (2000). A virtual environment testbed for training laparoscopic surgical skills. *Presence: Teleoperators and Virtual Environments, 9*, 236-255. doi:10.1162/105474600566772

Thomas, E. A., & Weaver, W. B. (1975). Cognitive processing and time perception. *Perception and Psychophysics, 17*, 363-367. doi: 10.3758/BF03199347

van Det, M. J., Meijerink, W. J., Hoff, C., Totte, E. R., & Pierie, J. P. (2009). Optimal ergonomics for laparoscopic surgery in minimally invasive surgery suites: a review and guidelines. *Surgical Endoscopy, 23*, 1279-1285. doi: 10.1007/s00464-008-0148-x

Way, L. W., Stewart, L., Gantert, W., Liu, K., Lee, C. M., Whang, K., & Hunter, J. G. (2003). Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. *Annals of Surgery, 237*, 460-469. doi: 10.1097/01.SLA.0000060680.92690.E9

Westebring-van der Putten, E. P., Goossens, R. H., Jakimowicz, J. J., & Dankelman, J. (2008). Haptics in minimally invasive surgery--a review. *Minimally Invasive Therapy and Allied Technologies, 17*, 3-16. doi: 10.1080/13645700701820242

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

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

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 449-455. doi: 10.1518/001872008x288394

Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 25*, 1-16. doi: 10.1177/001872088302500101

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*, 263-281. doi: 10.1177/001872089303500205

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 27*, 489-502. doi: 10.1177/001872088502700501

Young, M. S., & Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics, 47*, 1014-1035. doi: 10.1080/00140130410001686348

Zakay, D. (1989). Subjective time and attentional resource allocation: An integrated model of time estimation. In I. L. D. Zakay (Ed.), *Time and human cognition: A life-span perspective* (pp. 365-397). Oxford, England: North-Holland.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science, 6*, 12-16. doi: 10.1111/1467-8721.ep11512604

Zakay, D., & Block, R. A. (2004). Prospective and retrospective duration judgments: an executive-control perspective. *Acta Neurobiologiae Experimentalis, 64*, 319-328.

Zakay, D., & Shub, J. (1998). Concurrent duration production as a workload measure. *Ergonomics, 41*, 1115-1128. doi: 10.1080/001401398186423

Zheng, B., Cassera, M. A., Martinec, D. V., Spaun, G. O., & Swanstrom, L. L. (2010). Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy, 24*, 45-50. doi: 10.1007/s00464-009-0522-3

APPENDIX A

INFORMED CONSENT FORM

PROJECT TITLE: Assessment of Mental Workload during Laparoscopic Skill Acquisition on a Virtual Reality Simulator

RESEARCHERS:

Mark W. Scerbo, Ph.D., Responsible Project Investigator, Associate Professor, College of Sciences, Psychology Department

Co-investigators:

Rebecca A. Kennedy, Graduate Student, College of Sciences, Psychology Department
Erik G. Prytz, Graduate Student, College of Sciences, Psychology Department
Michael Montano, Graduate Student, College of Sciences, Psychology Department
Levi Warvel, Graduate Student, College of Sciences, Psychology Department

DESCRIPTION OF RESEARCH STUDY

Laparoscopic surgery is a type of surgery that is performed by inserting a small camera and surgical instruments through small incisions in the body. This technique is generally safer for the patient, but often more difficult for the surgeon to perform. Therefore, computer-based simulators are now being used to help surgeons acquire laparoscopic skills.

If you decide to participate, then you will be one of approximately 80 undergraduate students involved in a study designed to improve current methods for training future laparoscopic surgeons using a computer-based simulator. You will be instructed in how to perform several simulated surgical tasks on the computer using simulated surgical tools and a foot pedal and then given time to practice those tasks. In addition, you will be asked to perform another task that requires you to identify different targets in different areas of your display. Afterward, you will also be asked to complete two brief questionnaires that ask you to rate the ease or difficulty of the tasks. The total amount of time for participation is approximately one hour.

EXCLUSIONARY CRITERIA:

To participate in this study, you must be an undergraduate student at ODU. You must be 18 years of age or older. You also must have normal or corrected-to-normal vision. If you wear contacts or glasses, you must have these with you when you participate

In addition, in order to participate in this study you should not have any problems with your ability to physically use your right leg and right foot to press a foot pedal periodically. You should also not have any problem physically using both your right and left hands to interact with the simulated surgical instruments

RISKS:

If you decide to participate in this study, then you may face a risk of slight physical fatigue. Both your arms and hands may become tired from interacting with the simulator instrument device. The researchers have tried to reduce these risks by incorporating frequent breaks and resting periods. And, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS:
There are no direct benefits for participation. However, you will have the opportunity to learn how a surgical simulator is used for developing basic laparoscopic skills.

COSTS AND PAYMENTS:
If you decide to participate in the study, you will receive 1 Psychology department research credit, which may be applied to course requirements or extra credit in certain Psychology courses. Equivalent credits may be obtained in other ways, such as conducting library reports and online surveys. You do not have to participate in this study, or any Psychology Department study, in order to obtain this credit.

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

WITHDRAWAL PRIVILEGE:
It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study – at any time. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation. If at any point during the study you wish to stop, simply tell the researcher and you will not be penalized in any way. Any data that has already been collected will be destroyed and will not be included in the final analysis.

COMPENSATION FOR ILLNESS AND INJURY:
If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of injury, or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact the Faculty research advisor, and responsible principle investigator Dr. Mark W. Scerbo at 757-683-4217 or Dr. George Maihafer the current IRB chair at 757-683-4520 at Old Dominion University, who will be glad to review the matter with you.

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

Dr. Mark W. Scerbo, mscerbo@odu.edu, (757) 683-4217
Rebecca A. Kennedy, rkenn014@odu.edu

Erik G. Prytz, erik.prytz@gmail.com
Michael Montano, mmont033@odu.edu
Levi Warvel, lwarv001@odu.edu

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

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

-----------------------------------	----------------------------------	---------
Participant's Name	Participant's Signature	Date

INVESTIGATOR'S STATEMENT

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

-----------------------------------	----------------------------------	---------
Investigator's Name	Investigator's Signature	Date

APPENDIX B

PARTICIPANT BACKGROUN INFORMATION FORM

Participant #:_____ Group:_____ Date:_____ Time:_____

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

1. Age______
2. Gender______
    0 = Female
    1 = Male

3. Ethnicity______

    0 = Black, Non-Hispanic
    1 = Hispanic
    2 = Native American/Alaskan
    3 = Asian/Pacific Islander
    4 = Caucasian, Non-Hispanic
    5 = Other/Unknown

4. Do you have normal or corrected-to-normal vision?_____

    0 = Yes
    1 = No

5. What is your dominant hand?_____

    0 = Right
    1 = Left
    2 = Ambidextrous

6. Do you play video games?_____

    0 = Yes
    1 = No

    If yes: how many hours, on average, do you play each week?____

APPENDIX C

NASA-TLX WORKLOAD QUESTIONNAIRE

Participant #:______ Group:______

NASA-TASK LOAD INDEX (TLX) WORKLOAD QUESTIONNAIRE

(Hart & Staveland, 1988)

MENTAL DEMAND

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

PHYSICAL DEMAND

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

TEMPORAL DEMAND

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

PERFORMANCE

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

EFFORT

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

FRUSTRATION

Low High

|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

VITA

Levi P. Warvel
Department of Psychology
106 Mills Godwin Building
Old Dominion University
Norfolk, VA 23529-0267
E-mail: lwarv001@odu.edu

EDUCATION

PhD, Human Factors Psychology	Old Dominion University, VA May 2017 (expected)
MS, Experimental Psychology	Old Dominion University, VA May 2015
BA, Psychology (Cum Laude)	Gannon University, PA December 2008

WORK EXPERIENCE

Researcher, Old Dominion University	May 2015 – Present
Research Assistant, Old Dominion University	September 2012 – May 2015
Teaching Assistant, Old Dominion University	September 2012 – Present
Research Assistant, Gannon University	January 2006 – December 2008
Teaching Assistant, Gannon University	September 2007 – December 2008

RESEARCH EXPERIENCE

Warvel, L.P. (2014). Assessing the effectiveness of an interval estimation and a visual-spatial secondary task as measures of mental workload during laparoscopy. Unpublished master’s thesis. Old Dominion University, Norfolk, VA.

Leonard, R.C., Tingley, L.A., Townsend, B., & Warvel, L.P. (2011, January). *Maintenance of undergraduate attitudes toward older adults following an interview assignment.* Poster presented at the annual meeting of the National Institute on the Teaching of Psychology, St. Pete Beach, FL.

COMPUTER EXPERIENCE

Microsoft Windows
Microsoft Office
SONA
Adobe Connect
Statistical Package for the Statistical Sciences (SPSS)

HTML (entry-level)
CSS (entry-level)
PHP (entry-level)
MySQL (entry-level)
JavaScript (entry-level)