Evaluating Mental Workload of Two-Dimensional and Three-Dimensional Visualization for Anatomical Structure Localization

Jung-Leng Foo, PhD,[1] Marisol Martinez-Escobar, MS,[1] Bethany Juhnke, BS,[1] Keely Cassidy, BA,[2] Kenneth Hisley, PhD,[3] Thom Lobe, MD,[4] and Eliot Winer, PhD[1]

Abstract

Visualization of medical data in three-dimensional (3D) or two-dimensional (2D) views is a complex area of research. In many fields 3D views are used to understand the shape of an object, and 2D views are used to understand spatial relationships. It is unclear how 2D/3D views play a role in the medical field. Using 3D views can potentially decrease the learning curve experienced with traditional 2D views by providing a whole representation of the patient's anatomy. However, there are challenges with 3D views compared with 2D. This current study expands on a previous study to evaluate the mental workload associated with both 2D and 3D views. Twenty-five first-year medical students were asked to localize three anatomical structures—gallbladder, celiac trunk, and superior mesenteric artery—in either 2D or 3D environments. Accuracy and time were taken as the objective measures for mental workload. The NASA Task Load Index (NASA-TLX) was used as a subjective measure for mental workload. Results showed that participants viewing in 3D had higher localization accuracy and a lower subjective measure of mental workload, specifically, the mental demand component of the NASA-TLX. Results from this study may prove useful for designing curricula in anatomy education and improving training procedures for surgeons.

Background

MEDICAL IMAGING RESEARCH has evolved rapidly in the last decade, because of the improvements in computer graphics rendering technology. As these technologies become easily available to the medical community, medical image data visualization has begun to shift from two-dimensional (2D) images to three-dimensional (3D) volume for higher-fidelity representations. Commercial radiological software packages such as Vitrea,[1] Amira,[2] and OsiriX[3] have been largely successful, providing the medical community with a vast array of tools to examine, analyze, and interact with 3D representations of patient medical image data obtained from computed tomography and magnetic resonance imaging scans.

It has been suggested that 3D visualization is beneficial to surgical planning and diagnosis as it facilitates understanding the shape of structures.[4] 3D visualization is accomplished by generating volume representations using the existing 2D medical image data of the patient. A preliminary study[5] was performed with a group of seven participants identifying anatomical structures with 2D or 3D visualization software. The study suggested that visualizing anatomical features in 3D had value over the traditional images viewed in 2D visualization software. The group was composed of two surgeons who had 9–10 years of experience and five residents who had 1–3 years of experience. Although the findings suggested the potential of 3D visualization tools, they also hinted at the challenges of moving from a 2D visualization environment into a 3D environment.

Extending from the preliminary results, the study presented in this article evaluated the mental workload of the participants while localizing anatomical structures in both 2D and 3D visualization environments. The participant sample size was expanded, and all participants were of the same level in experience and knowledge of anatomy. The study included four steps in the procedure: pre-test survey, software training, localization of structures, and post-test survey. The results discuss the participant's accuracy when localizing anatomy

[1]Virtual Reality Applications Center, Iowa State University, Ames, Iowa.
[2]Department of Anatomy, Des Moines University, Des Moines, Iowa.
[3]Department of Basic Sciences, Touro University, Valejo, California.
[4]Department of Pediatric Surgery, Blank Children's Hospital, Des Moines, Iowa.

using 2D and 3D medical visualization software, the amount of time required to localize the anatomy, and the mental workload required of the participant to complete the task. The findings from the study are concluded with recommendations for future work.

Mental workload

Understanding the mental capabilities of individuals during performance of tasks has become a growing interest in clinical practice research. Various tests and questionnaires have been developed to quantify mental work as it relates to human performance,[6–9] and increase in mental workload leads to an increase in errors as an individual completes a task.[6,7,10,11] Such errors can have a significant impact on the medical field by directly impacting patient safety, recovery time, and overall success of a procedure.[9,11]

The definition of mental workload assumes that humans have a finite mental capacity. When resources are allocated to a primary task this depletes the cognitive resources available to successfully monitor and complete secondary tasks.[7,9] In some cases primary and secondary tasks are both critical to successful clinical practice. Primary tasks for laparoscopic surgeons include manipulation of surgery instruments, whereas secondary tasks include monitoring and surgery planning, both critical to a successful surgery. Although the medical research community is citing the importance of attentional resources research, currently there is a lack of research to fully understand how attentional resources affects performance.[9,10]

Mental workload can be measured with primary tasks, secondary tasks, physiological measures, and subjective measures. Primary and secondary task measurements are dependent on the specific application, whereas physiological and subjective measurements are more generic across different domains. Primary and secondary tasks measure the performance of the task under the assumption that if the primary task exceeds the attentional resources, the secondary task will show performance degradation.[6,7] Physiological measures take body measurements such as heart rate and eye movements as indices of mental workload. Subjective measures are self-reported measurements that can be sensitive to different dimensions of mental workload.[6,7]

NASA-Task Load Index

A leading workload assessment is the NASA-Task Load Index (NASA-TLX). The NASA-TLX breaks mental workload into six independent components:

1. *Mental demand.* How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
2. *Physical demand.* How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
3. *Temporal demand.* How much time pressure did you feel due to the rate or pace at which the tasks or task element occurred? Was the pace slow and leisurely or rapid and frantic?

4. *Frustration level.* How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?
5. *Effort.* How hard did you have to work (mentally and physically) to accomplish your level of performance?
6. *Performance.* How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Each component can be rated on a scale (0–20 points), and the final mental workload score indicates how much attentional resources an individual has focused on a task, to ensure performance and limited errors.[7–9] This workload scale was developed to study stress endured by flight crews and the related implications on performance during a flight.[8,9] Throughout the years, individuals have adapted the original NASA-TLX for aviation to other domains.[8] As technology becomes more prevalent within clinical practice, understanding how technology affects performance in the operating room is vital to the well-being of all patients.[7,9]

Overall, studies have found that NASA-TLX is the most comprehensive evaluation of subjective mental workload for clinical practice.[9] This type of evaluation appears to be the most easy to use and nonintrusive and has been validated in several user-accepted texts.[7]

Mental workload in medicine

NASA-TLX has been applied to clinical practice, to improve safety and minimize errors.[6,9,12] Multiple studies have been conducted in clinical practice to decrease the mental workload on doctors and surgeons to reduce errors and increase performance in their practice. Byrne et al.[6] analyzed the mental workload of anesthetists, where they used a wireless vibrotactile device to analyze the anesthetist's reaction time to the vibration and NASA-TLX to determine the workload during the task. They found that increased mental workload is likely to be a common problem in clinical practice. One participant, in particular, stopped responding to the vibration during a patient's unexpected cardiac arrest, indicating that during this time frame, the anesthetist was at the maximum mental capacity, increasing the chances for errors in the participants.[6]

Other studies looked at first-year students as they learned clinical techniques[11] analyzed the mental workload of students learning how to suture with varying levels of instruction and feedback. The study found that students with no formal feedback of results or performance of task had much higher NASA-TLX mental workload scores than those students who had feedback about their results and performance.[11]

Carswell et al.[7] studied laparoscopic mental workload including mental workload effects on errors. The study found that during the evaluation of the addition of technology into clinical practice, mental workload should be evaluated to ensure the primary task at hand could be completed. This study also stressed the concern to avoid complete reduction of workload through technical advancements to avoid the "potential for underload," which can cause a psychological stress that appears in the task as boredom or fatigue.[7,8]

Students' interactions with the endoscopic/laparoscopic and robot-assisted surgical techniques were also studied.[12]

Stress and workload were measured with the Dundee Stress State Questionnaire and the Multiple Resources Questionnaire, respectively. The results concluded that both techniques had equally high levels of stress and workload. Although the robotic system appeared to have the advantage over the traditional approach, the decrease in mental workload from having the robot assist in the procedure was offset by the increase in mental workload from maneuvering the robot. Another study[9] looked at workloads of students performing tasks with 2D and 3D views. The study found that less mental workload appeared in students who were performing tasks with a 3D view compared with a 2D view of the same task. However, this study was not able to differentiate between the most significant types of workload on the students.

The stress and pressure of time can also affect the performance during surgery. A study compared temporal demand in students estimating intervals of time during laparoscopic skills training.[10] Students were asked (1) to estimate the time to complete a task as the end of each trial or (2) to indicate when every 31 seconds had elapsed during the training. The study found an increase in temporal demand for the students in the second group (indicate when 31 seconds had elapsed) compared with students from the first group (estimating time to complete trial).

Evaluating mental workload in clinical practice can decrease errors and increase performance during medical procedures. Many sources have found mental workload testing to assist in the development of clinical procedures and techniques. Understanding how technology can help or hinder performance in the operating room can have major implications on the overall well-being of patients.

Subjects and Methods

The purpose of this study aimed to answer the following research question: "What is the impact on human factors such as the mental workload of 3D visualization over 2D visualization in localization of anatomical features in medical students?"

Participants

In total, 25 participants, all first-year medical students with the same level of experience and knowledge, participated in the study. All participants had prior experience observing demonstrations of 2D and 3D representations during anatomy classes. Three of the 25 participants had prior experience interacting with 2D medical visualization software, whereas the remaining 22 of the 25 participants had no prior experience interacting with 2D medical visualization software. None of the participants had prior experience interacting with 3D medical visualization software.

Procedure

The participants were randomly divided into two groups and asked to identify three anatomical structures: a control group that only used the 2D representations and a second group that only used the 3D representations. The OsiriX visualization software was used for both groups to ensure consistency in testing conditions between the two participant groups. The user study procedure consisted of four parts:

1. *Pre-test survey.* The pre-test survey dealt with questions about experience with medical (both 2D and 3D) visualization tools and computer gaming experience. These questions were used to examine if correlations existed between previous experience and the study findings.
2. *Software training.* Participants were presented a quick introduction of the tools and features of the respective software, and they had up to 5 minutes to interact with the software. For participants using the 2D environment (Fig. 1), they were introduced to features such as windowing of tissue types and three orthogonal views of the data. Windowing allows the participants to interactively adjust the types of tissue shown based on the tissue density by holding down the left mouse button and moving the mouse up and down. Participants were also able to view the image slices along all three axes (axial, coronal, sagittal) simply by moving the crosshair
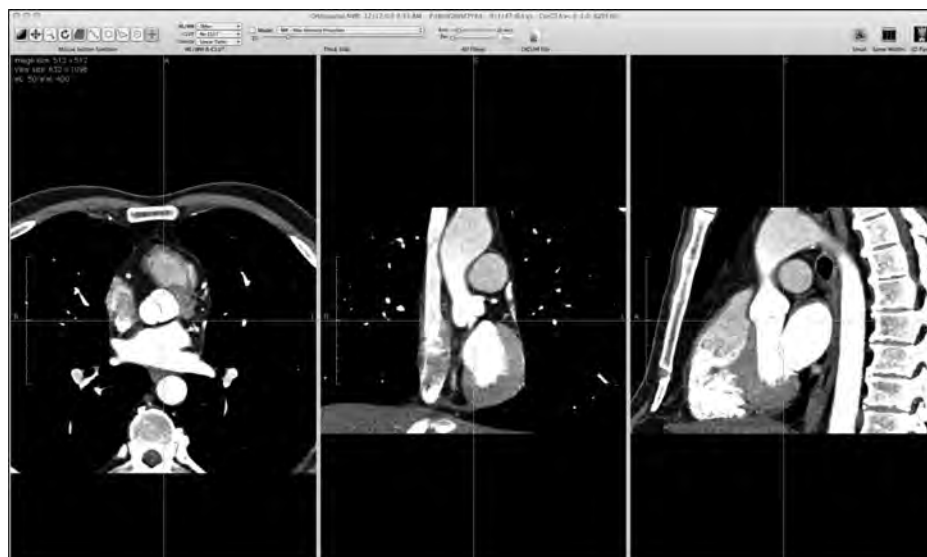

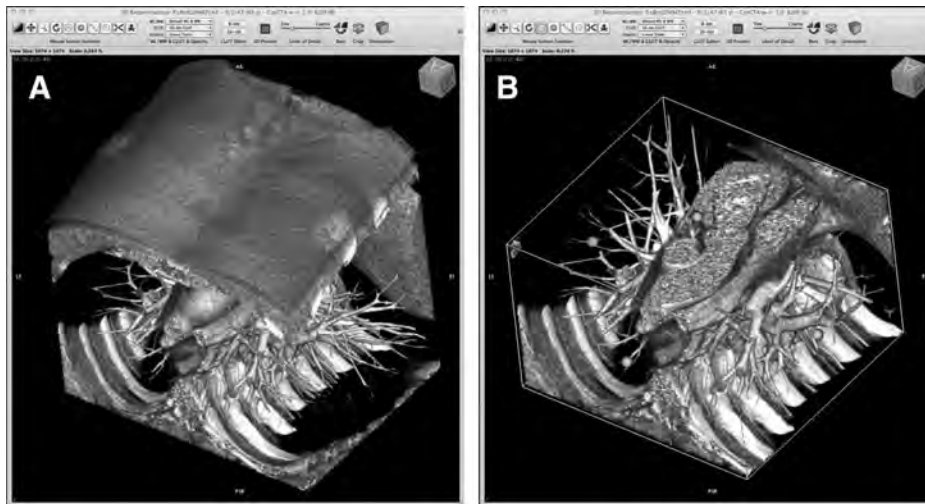**FIG. 1.** Screen capture of a typical two-dimensional environment.

FIG. 2. Screen capture of a typical three-dimensional environment: **(A)** full view of the three-dimensional volume representation and **(B)** cropped three-dimensional volume.

in any of the three orthogonal views or by scrolling on the mouse wheel. Participants using the 3D environment (Fig. 2) would see a 3D volume representation of the medical image data. Participants would interact with the 3D representation such as moving, rotating, and zooming, by holding down the left mouse button. Icons on the toolbar will activate additional tools such as windowing and cropping. Similar to the 2D environment, windowing when activated can be performed by holding down the left mouse button and moving the mouse up and down. For cropping, participants can click on the green spheres (Fig. 2) to move and crop the 3D representation.

3. *Localization of structures.* Participants were asked to localize three anatomical structures of varying complexities in 2D or 3D: the gallbladder, celiac trunk, and the

superior mesenteric artery. Participants were encouraged to think out loud as they performed the tasks, and an observer recorded the participant's comments. The primary task measurement of mental workload was the accuracy of localizing the structures. Participants would mark where they thought the structure is located, as shown in Figure 3. Another primary task measurement was the time taken for the tasks to be completed. Participants had a maximum of 20 minutes to complete these tasks and would be given a 2-minute warning if needed.

4. *Post-test survey.* This section of the study insisted of two parts. The first part was to ask participants to complete a NASA-TLX evaluation, used as a subjective measure of mental workload. The second part was a brief questionnaire about software usability, to lead to future
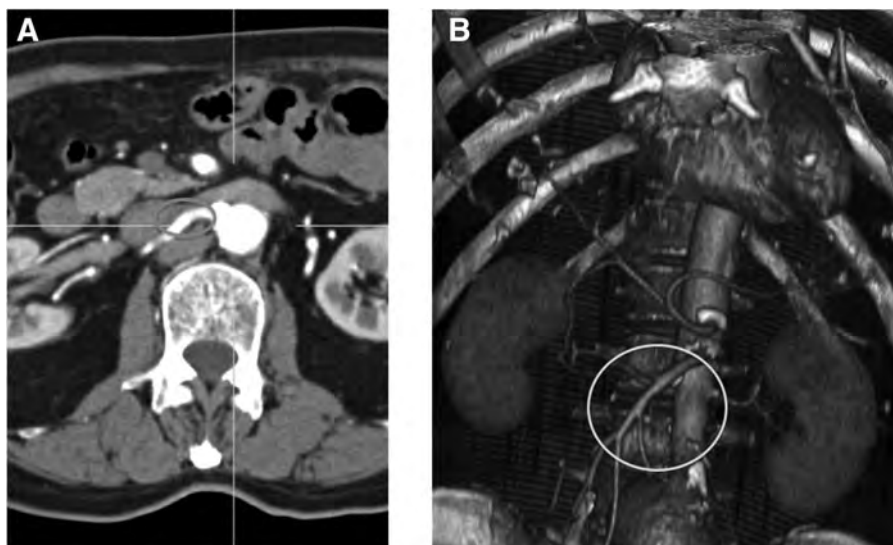

FIG. 3. Sample results from **(A)** two-dimensional and **(B)** three-dimensional participants. Images were cropped and enlarged to show emphasis on the areas where participants indicated the structure was located.
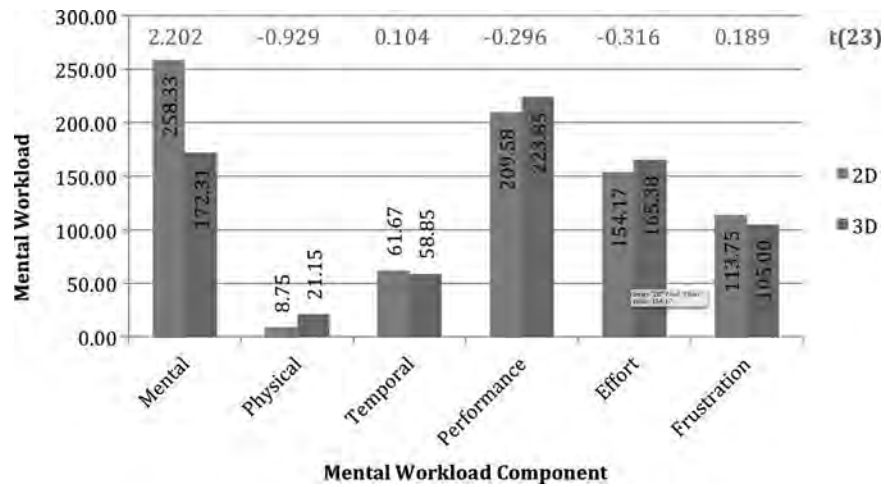
FIG. 4. Weighted mental workload component ratings by participant group and computed *t* test values above the respective components.

research questions, that included the following questions:

1. What is your experience with medical imaging software? How you used it personally? Seen it used? Where?
2. What was your favorite feature? What was easy to perform/complete?
3. What was your least favorite feature? What was difficult to perform/complete?
4. Can you give a brief summary of your thought process as you were completing the task?

Results And Discussion

Accuracy

Accuracy of the structures was the primary task measured on a 2-point scale: 0 for inaccurate, 1 if the participant correctly identified the structure. The participants were asked to locate three anatomical structures, for a total score ranging between 0 points (three wrong answers) and 3 points (three correct answers). The average accuracy for the 2D participant group was 2.08, and the 3D participant group scored higher at 2.54, which is not a statistically significant difference, testing at 95% of $P = .05$. Further studies need to be completed to determine statistical significance and determine if the sample size was too small or tasks were not complex enough. The improvement in 3D performance could be due to the additional information available to the participants, such as using spatial relationships between anatomical structures as landmarks during the localization process.

Time

Time to localize each structure, the second objective measure, was recorded for every participant. The average time for the 3D participant group was almost 11 minutes, and the average time for the 2D participant group was 6½ minutes. These results are statistically significant, testing at 95% of $P = .05$. Overall, the 3D participant group took 3.6 minutes longer to complete the task than the 2D participant group. An observation made during the user study procedure noticed that participants tend to explore the 3D volume representation more before confirming their decisions. The 3D volume representations allowed the participants to experience a complete representation of the human anatomy, which invoked more curiosity on the part of the participants, and they tended to search through the anatomy for anatomical structures more freely compared with the 2D participants. The 2D participants tend to use a "seek, find, confirm, move on" approach, where the 3D participants wanted to complete a more thorough search before confirming the structure and moving to the next anatomical structure.

Individual mental workload components

The NASA-TLX ratings were taken from every participant, and the average weighted ratings were obtained for both participant groups. An independent two-sample *t* test for each of the six mental workload components was also computed. Dependent on sample size 25, a degree of freedom of 23 was calculated, giving $t(23)$ for all statistical evaluations.

Of the six components, only mental demand showed a statistical significant difference (at 95% of $P = .05$) between the 2D participant group and the 3D participant group (Fig. 4). This indicates that the mental and perceptual activity (decision-making, memory, and searching) required for completing the task was more demanding for the 2D participant group.

Although there is no statistical difference for the physical demand rating between the participant groups, there is a difference based on the rating of almost three times higher for 3D than for 2D. This can be attributed to several causes that should be addressed in future studies. One of the reasons could be that the participants used the mouse more during the task to interact with the 3D representation, and participants also had to click on icons to switch between the various mouse functions such as basic interactions, windowing of tissue types, and cropping the 3D representation. Another explanation could be that the participants were less familiar with 3D and therefore interacted with 3D for longer periods of time, or that the participants were more engaged with the novelty of the 3D representations.

The other four components—temporal demand, performance, effort, and frustration—showed no statistical or practical difference because the weighted ratings for those components were very close for both participant groups.

Overall mental workload

The overall mental workload was approximately 9.8% higher for the 2D participant group, at 53.75, compared with the 3D participant group, at 49.77, although overall there is no statistically significant difference, testing at 95% of $P=.05$. Thus, the NASA-TLX data indicate that localizing anatomical structures in 2D induces higher levels of mental workload than that in 3D. Correlating this difference in mental workload with the accuracy measurement for the primary task, we can also infer that lower mental workload levels will result in improved accuracy.

Other observations

Participants using 2D commented on how they would quickly scan through the different slices of images to find a familiar landmark (i.e., the liver for the gallbladder). Once this landmark was identified, they would move back and forth between the few slices to localize the structure in question. For participants using 3D, they commented on how the manipulation of the different tissue densities and the cropping tool were useful in removing non-relevant structures to get a better view of the target structure relative to other landmark structures.

Conclusions

Based on data collected for accuracy, task completion time, and NASA-TLX results, participants performed better when localizing anatomical structure and had lower cumulative mental workload when using the 3D environment compared with the 2D environment. The NASA-TLX test determined that the results of the mental demand component are statistically significant, where the other five components did not present a statistical significance between the 2D and 3D participant groups. It was surprising, however, that physical demand was rated three times higher for 3D participants than 2D participants.

This study showcased the benefits of 3D while finding anatomical structures. It provided better accuracy and lower mental workload. Future studies will address the following concerns: (1) higher number of localization tasks for accuracy and time information, (2) supplemental subjective measures such as the Multiple Resources Questionnaire,[13] and (3) less time allowed for each localization task.

By investigating the effects of mental workload while performing tasks that requires a high level of processing and precision, such as diagnosis or surgery planning, we can obtain beneficial information to improve the curriculum or training procedures for doctors and surgeons.

Acknowledgments

The authors would like to thank the students from the Department of Anatomy at Des Moines University for volunteering as participants for this study.

Disclosure Statement

No competing financial interests exist.

References

1. Vital Images. www.vitalimages.com/solutions/Vitrea_Enterprise_Suite.aspx (accessed March 4, 2011).
2. Amira. www.amira.com (accessed March 4, 2011).
3. OsiriX imaging software. www.osirix-viewer.com (accessed March 4, 2011).
4. St. John M, et al. The use of 2D and 3D displays for shape-understanding versus relative-position tasks. Hum Factors 2001;43:79–98.
5. Nekolny B, Holub J, Foo JL, Winer E. Evaluation of endoscopic surgical planning in an interactive 3D visualization environment. Presented at the International Pediatric Surgery Group 19th Annual Congress for Endosurgery in Children, Emerging Technologies Session, Waikoloa, HI, June 8–12, 2010.
6. Byrne AJ, Oliver M, Bodger O, Barnett WA, Williams D, Jones H, Murphy A. Novel method of measuring the mental workload of anaesthetists during clinical practice. Br J Anaesth 2010;105:767–771.
7. Carswell CM, Clarke D, Seales WB. Assessing mental workload during laparoscopic surgery. Surg Innov 2005;12:80–90.
8. Hart SG. Nasa-Task Load Index (Nasa-TLX): 20 years later. Proc Hum Factors Ergon Soc 2006;50:904–908.
9. Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. Proc Hum Factors Ergon Soc 2009;53:1186–1190.
10. Lio CH, Bailey K, Carswell CM, Seales WB, Clarke D, Payton GM. Time estimation as a measure of mental workload during the training of laparoscopic skills. Proc Hum Factors Ergon Soc 2006;50:1910–1913.
11. O'Connor A, Schwaitzberg SD, Cao CGL. How much feedback is necessary for learning to suture? Surg Endosc 2008;22:1614–1619.
12. Klein MI, Warm JS, Riley MA, Matthews G, Donovan JF, Doarn C. Performance, stress, workload, and coping profiles in 1st-year medical students' interaction with the endoscopic/laparoscopic and robot-assisted surgical techniques. Proc Hum Factors Ergon Soc 2008;52:885–889.
13. Boles DB, Bursk JH, Philips JB, Perdelwitz JR. Predicting dual-task performance with the Multiple Resource Questionnaire (MRQ). Hum Factors 2007;49:32–45.

Address correspondence to:
Jung-Leng Foo, PhD
Virtual Reality Applications Center
Iowa State University
1620 Howe Hall
Ames, IA 50011

E-mail: foo@iastate.edu