TRAINING ANATOMY RECOGNITION THROUGH REPETITIVE VIEWING OF LAPAROSCOPIC SURGERY VIDEO CLIPS

Stephanie Guerlain, K. Brook Green, Maranda LaFollette, Todd Mersch, Brian Mitchell, G. Reed Poole, Reid Adams, J. Forrest Calland, Viktor Bovbjerg, and Edward Chekan
University of Virginia, Charlottesville, VA USA
www.sys.virginia.edu/hci

This study evaluated the use of a video-based training environment, to support training the perceptual “rules” necessary for situation assessment during laparoscopic surgery. The training system shows multiple examples of procedural steps using edited laparoscopic surgery videos to enforce absorption, expose the learner to varied anatomy, and stress crucial maneuvers. A between-subjects experiment with 30 medical students showed a ten percent increase in cognitive skills as measured by pretest/posttest sets of test questions as compared to a control group who had access to the same videos but in an unstructured format for the same amount of time (pc.05). This difference was attributed primarily to improvements in perceptual judgments, as opposed to improvements in procedural, strategic, declarative or counterexample knowledge. This study shows that video repetition is a potential means for training perceptual rule-based skills in an effective and efficient manner.

INTRODUCTION

For this study, we were interested in a safe and efficient means to train novice surgeons on anatomy recognition during minimally invasive (laparoscopic) surgery. Currently, novice surgeons learn these skills primarily by participating first as an observer, but then as an increasingly prominent participant in actual surgeries, a time-honored tradition, but a risky and not necessarily most pedagogically effective method.

We chose laparoscopic cholecystectomy (LC) as our test case, as it is one of the most common surgeries currently performed in the United States (Ress, Sarr, Nagorney, Farnell, Donohue, and McIlrath, 1993) and because LC is often used as the “training case” for laparoscopy due to its high frequency and perceived low risk. However, there is strong evidence of a learning curve for performing this surgery (Morgenstern, McGrath, Carroll, Paz-Partlow, and Berci, 1995; Moore and Bennett, 1995; Voitk, Tsao, and Ignatius, 2001).

Technical training systems are available for training the physical dexterity skills for manipulating instruments (e.g., Derossis, Fried, Abrahamowicz, Sigman, Barkun and Meakins, 1998; Melvin, Johnson, and Ellison, 1996). Such systems work by having students control the instruments to perform computer-based tasks such as moving objects and turning and twisting the laparoscopic instruments appropriately. Sufficiently realistic simulation models, however, have yet to emerge for training cognitive-perceptual skills such as anatomy pattern recognition, understanding and recognizing appropriate lines of dissection, and determining the safest and most efficient means for progressing with a dissection.

Kellman and Kaiser (1994), Walker, Fisk, Phipps and Kirlik (1994), Walker and Fisk (1995) and Bailey (1994) have all studied methods for training novices on task-relevant perceptual patterns (in the domains of pilot navigation and instrument reading, recognition of offensive and defensive plays in football, and learning the hand movements involved in American Sign Language, respectively). These studies show that appropriately structured perceptual learning modules can condense perceptual learning processes that occur with extended experience. Perceptual learning modules (PLMs) isolate $r^{-1}$
condense relevant perceptual experiences through repetitive exposure to variations in stimuli. For example, the Walker and Fisk (1995) study used a technique of showing novice subjects patterns of movement in football plays using video footage of actual plays, recorded from the quarterback’s perspective to train players to recognize the “correct” wide receiver to whom a quarterback should pass the ball.

These studies and ours are based on the theories of Gibson (1969), that perceptual learning takes place by recognizing relevant cues in a pattern through repeated exposure to exemplars of a pattern. This hopefully leads people to a more expert, “recognition-primed” (Klein, 1989) decision making approach, where relevant patterns are detected as opportunities for action. We hypothesized that using this part-task training technique would be a beneficial and efficient means to train novice surgeons on relevant perceptual-based information extraction, where attention can be focused on patterns of psychologically or task-relevant information, rather than just on those features that are visually most salient.

METHOD

Performance was measured using a pre-test/post-test methodology, using two sets of question banks, designed to be equivalent in difficulty and subject matter. The order of the test banks was randomized for each subject. Comparisons were made on improvement from the pre-test to the post-test for the two groups. The pretest and posttest questions were varied in terms of the types of knowledge tested. Questions were classified a priori as requiring one of more of the following kinds of knowledge: perceptual (e.g., relevant anatomy and lines of dissection), declarative (facts), strategic (various approaches and when to apply them), procedural (the order in which events should occur), or counterexample (how not to do something). It was hypothesized that the Treatment Group would significantly improve on perceptual-based questions and procedural questions, but not on declarative, strategic or counterexample questions, whereas the Control Group would not significantly improve on any of these categories. These hypotheses are based on the fact that repetitively showing examples of each procedure step would enforce both procedural aspects as well as perceptual aspects. Declarative knowledge, strategic knowledge and counterexamples were not expected to improve, as none of our videos showed poor examples of technique, nor was there any audio or graphical overlays that would highlight declarative or strategic knowledge. In addition to this more general, theoretical classification of questions, we explicitly noted those questions that we hypothesized would be aided by the Treatment version of the system. The questions and our grouping of these questions are displayed in Table 1 in the Results section.

Subjects

Forty University of Virginia medical students with normal or corrected to normal vision enrolled in the experiment. After the first 10 trials, a software recording error was discovered. Thus, the data from these subjects was discarded. The data reported here include the final 30 subjects (14 Control, 16 Experimental).

Procedure

Subjects filled out a demographic form then watched and listened to a 15-minute introductory video which briefly demonstrated one example of each pertinent segment of the surgery (10 steps in all). Subjects then completed one of two 20-question tests (randomly selected at run time). The answer and time to answer were automatically recorded.

Following the pretest, the Experimental Group watched, in chronological order, three examples each of the 10 major surgical steps performed laparoscopically (see Figure 1 for an example frame from one of these videos). The Control Group had the same amount of time (27 minutes) to watch, unguided, three entire case videos (the same ones used to generate the experimental video segments). The control subjects retained the ability to switch among the three videos at their discretion, and also had t
ability to fast-forward or rewind within a particular video. This set-up almost exactly replicates the current available technology; students can make the effort to find and retrieve entire case studies documented on videotape and use their own time to navigate through them in an attempt to learn.

Figure 1. Laparoscopic surgery video

After the training section, subjects completed the 2nd test bank (the one not completed as the pre-test) as a posttest and the software once again recorded both their answers and answer times.

RESULTS

Subjects ranged in age from 22 to 26, with a mean age of 23.6. Thirty percent of the subjects were female. The Treatment Group had 50% of the female participants and 55% of the male participants. Sixty-three percent of the participants claimed no experience with LC (56% of the Treatment Group and 71% of the Control Group); the rest had observed at least one LC. Three participants (1 Control, 2 Treatment) had operated the LC camera, and one had acted as a junior surgeon in a case (the same Control Group participant who had operated a camera).

The difference in improvement within subjects based on test bank order was analyzed using a two-sample t-test assuming unequal variances. This resulted in a two-tailed p-value of .53 for the Control Group and .47 for the Treatment Group implying that the test banks were of equal difficulty value.

Scores were calculated as the number of questions the user answered correctly, out of twenty (unanswered questions within the time allowed are considered incorrect). There is no significant difference between the scores of the two groups on the pretest or the posttest. Comparing improvement from pretest to posttest across groups showed that the Control Group showed no significant improvement (11.43 vs. 11.93 points), while the Treatment Group increased their score by more than fourteen percent (10.88 vs. 12.94 points). Using a two-tailed t-test, this difference in scores from pretest to posttest is significant for the Treatment Group (p < 0.05), but not the Control Group (p = 0.27). Control subjects decreased their times by an average of 39.07 seconds, while experimental subjects decreased by 45.94 (a six percent greater reduction). Both within-subjects and between-subjects analyses show that these differences in times are not significant.

The data were also analyzed using difference in proportions tests by summing the number of correct questions for the Treatment vs. Control Group for pretest vs. posttest. Across all questions, similar to our previous result, the Treatment Group had significant improvement in scores (p < 0.008) whereas the Control Group had no significant improvement. Analyzing that subset of questions predicted a priori to be aided by our system, the Treatment Group again improved significantly (p = 0.03), whereas the Control Group did not. Our hypothesis that perceptual-based questions are aided by this type of training was confirmed (p = .024 for the Treatment Group, not significant for Control Group). However, our hypothesis that procedural information would be aided by this type of training was not met, although one could consider a trend in that direction (p = 0.095 for the Treatment Group, p = .826 for the Control Group). As hypothesized, neither the Treatment Group nor the Control Group was aided on Strategic, Declarative, or Counterexample question groups.
CONCLUSIONS

Part-task training using videos to selectively train practitioners on critical cues to be recognized has a theoretical basis in judgment and decision making literature (Bisantz, Kirlik, Gay, Phipps, Walker and Fisk, 2000; Klein, 1989), and has been demonstrated to be an effective technique for training football quarterbacks to recognize appropriate passing plays (Walker, Fisk, Phipps and Kirlik, 1994) and to train pilots to recognize critical visual cues during visual landings (Kellman and Kaiser, 1989). Our study has shown that this technique applies to surgery, and has demonstrated a potentially safer and more efficient method for training medical students and residents on at least some of the key perceptual and cognitive skills prior to beginning “training” for laparoscopic surgery on live patients.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0092985. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


