Measuring intra-operative decision-making during laparoscopic cholecystectomy: validity evidence for a novel interactive Web-based assessment tool

Amin Madani$^{1,2}$ · Yusuke Watanabe$^{3}$ · Elif Bilgic$^{2}$ · Philip H. Pucher$^{4}$ · Melina C. Vassiliou$^{1,2}$ · Rajesh Aggarwal$^{1,2,5}$ · Gerald M. Fried$^{1,2,5}$ · Elliot J. Mitmaker$^{1}$ · Liane S. Feldman$^{1,2}$

Received: 22 December 2015 / Accepted: 5 July 2016 © Springer Science+Business Media New York 2016

Abstract

Background Errors in judgment during laparoscopic cholecystectomy can lead to bile duct injuries and other complications. Despite correlations between outcomes, expertise and advanced cognitive skills, current methods to evaluate these skills remain subjective, rater- and situation-dependent and non-systematic. The purpose of this study was to develop objective metrics using a Web-based platform and to obtain validity evidence for their assessment of decision-making during laparoscopic cholecystectomy.

Methods An interactive online learning platform was developed (www.thinklikeasurgeon.com). Trainees and surgeons from six institutions completed a 12-item assessment, developed based on a cognitive task analysis. Five items required subjects to draw their answer on the surgical field, and accuracy scores were calculated based on an algorithm derived from experts’ responses ("visual concordance test", VCT). Test–retest reliability, internal consistency, and correlation with self-reported experience, Global Operative Assessment of Laparoscopic Skills (GOALS) score and Objective Performance Rating Scale (OPRS) score were calculated. Questionnaires were administered to evaluate the platform’s usability, feasibility and educational value.

Results Thirty-nine subjects (17 surgeons, 22 trainees) participated. There was high test–retest reliability (intra-class correlation coefficient $= 0.95; n = 10$) and internal consistency (Cronbach’s $\alpha = 0.87$). The assessment...
demonstrated significant differences between novices, intermediates and experts in total score (p < 0.01) and VCT score (p < 0.01). There was high correlation between total case number and total score (ρ = 0.83, p < 0.01) and between total case number and VCT (ρ = 0.82, p < 0.01), and moderate to high correlations between total score and GOALS (ρ = 0.66, p = 0.05), VCT and GOALS (ρ = 0.83, p < 0.01), total score and OPRS (ρ = 0.67, p = 0.04), and VCT and OPRS (ρ = 0.78, p = 0.01). Most subjects agreed or strongly agreed that the platform and assessment was easy to use [n = 29 (78 %)], facilitates learning intra-operative decision-making [n = 28 (81 %)], and should be integrated into surgical training [n = 28 (76 %)].

**Conclusion** This study provides preliminary validity evidence for a novel interactive platform to objectively assess decision-making during laparoscopic cholecystectomy.

**Keywords** Visual concordance test · Assessment · Competency · Education · Performance · Cholecystectomy

### Materials and methods

This was a multi-institutional prospective study. The first phase included the development of an e-learning Web-accessible platform and novel metrics to assess intra-operative decision-making during laparoscopic cholecystectomy, and the second phase attempted to obtain validity evidence for them. The study protocol was approved by the institutional review board and conforms to the Canadian Tri-Council Policy Statement of Ethical Conduct.

#### Development of platform and assessment tool

A password-protected e-learning platform (http://www.thinklikeasurgeon.com) was developed to provide users with remote access to curricular content in an immersive and interactive environment. Design features include the use of multimedia with video modules, assessments, immediate feedback on performance and ongoing and repeated access for spaced education.

For the laparoscopic cholecystectomy curriculum, a 12-item assessment tool was developed to provide formative feedback throughout the modules. The items were specifically designed to target cognitive processes that experts identified as essential for avoiding major bile duct injuries based on prior cognitive and hierarchical task analyses for achieving a critical view of safety [23]. The assessment consisted of a variety of question styles, including four multiple-choice items and three script concordance test (SCT) items. SCT is a method of assessment...
that applies script theory from cognitive psychology to evaluate the mental model of an individual by ranking their willingness to affirm or refute a hypothesis in relation to an ill-defined intra-operative scenario on an agreement scale [24–27]. Despite the evidence to support the psychometric properties of SCT as a method of evaluating intra-operative decision-making, there are some limitations. Items are restricted to written text, information from the operative field is already presented and interpreted for the learner, and the participants’ responses are scored according to a linear scale. We sought to develop a method that requires learners to synthesize visual data from the surgical field into an accurate understanding of the operative environment and to make decisions based on these interpretations. Therefore, the remaining five items required subjects to draw their answer on the surgical field and accuracy scores were calculated based on an algorithm derived from experts’ responses—the “visual concordance test” (VCT).

The VCT evaluates decision-making in relation to a graphical illustration and aims to assess cognitive skills, such as pattern recognition and situation awareness by providing learners the opportunity to describe their thought processes or how to carry out tasks in relation to the anatomy in a surgical field. Similar to the SCT, these are complex scenarios without a single correct solution, and responses are compared and scored against a distribution of expert responses. Users are initially presented with a surgical video and at defined time points, the video stops and the user is prompted to answer a specific question by drawing annotations on the still frame (single static) image (e.g. “where do you want to start your dissection?”). The image pixels from these annotations are subsequently mapped onto Cartesian coordinates for analysis (Fig. 1). Similar to SCT, in order to account for the heterogeneity of expert responses (as experts would seldom select the same set of pixels for each item), the score is calculated based on the distribution of expert annotations (Fig. 2). Pixels identified from experts’ annotations are grouped into weighted zones to create a topographical map based on the proportion of expert responses that highlighted each pixel. Pixels in the higher percentile groups are assigned a proportionally greater weight for score calculation.

Free-text feedback and five-point Likert scale questionnaires were also administered to evaluate and improve the platform’s usability, educational value and feasibility for adoption.

Participants

Participants from six institutions in Canada, USA, United Kingdom and Japan were enrolled in this study and completed the 12-item assessment without access to any curricular modules. Subjects included general surgery-trained residents, fellows and attending surgeons. There were no restrictions on either years in training, years in independent practice, total or annual case volume, or subspecialty training. Participants were categorized based on training level and prior experience (total laparoscopic cholecystectomies performed) into three groups: novice (<25 cases), intermediate (25–100 cases, or surgical residents or fellows with more than 100 cases) and expert (attending surgeons having performed more than 100 cases). All subjects received a unique username and password and completed the assessment remotely on a personal computer, tablet or mobile device. Scores were automatically calculated by the software and investigators were blinded.

Sample size and statistical analysis

A contemporary framework of validity was used to provide evidence for the assessment as a measure of intra-operative decision-making [28]. Inter-group score comparison between novices, intermediates and experts were made using analysis of variance. Internal consistency and Spearman’s rank correlation coefficient with self-reported experience (total case volume), Global Operative Assessment of Laparoscopic Skills (GOALS) score and Objective Performance Rating Scale (OPRS) score were calculated. GOALS and OPRS scores were previously obtained prior to this study and were only used if they were obtained within 6 months of the assessment. Test–retest reliability was also calculated based on a random computer-generated sample of 10 participants who repeated the assessment after 1 week. Power calculation was based on prior work with the GOALS and OPRS assessment tools for laparoscopic cholecystectomy [29–31]. Using an α of 0.05 and a power of 80 %, with 2-sided testing, more than 6–8 subjects were required in each group. A p value of <0.05 was considered statistically significant. All statistical analyses were performed using JMP 11 software (SAS Institute Inc., Cary, NC).

Results

Thirty-nine subjects completed the assessment, including 19 residents, 17 attending surgeons and 3 fellows. Participants included 8 novices, 14 intermediates and 17 experts, and their characteristics are summarized in Table 1. All residents and fellows were general surgery-trained, with the majority (19 [86 %]) having performed <100 laparoscopic cholecystectomies. One chief resident and two fellows performed between 100 and 150 cases. Amongst experts, 13 (76 %) practise at an academic institution, 5 at a community hospital (29 %) and 2 (12 %) at a rural hospital. All but one attending surgeon work in public

\[ \text{Sample size and statistical analysis} \]

A contemporary framework of validity was used to provide evidence for the assessment as a measure of intra-operative decision-making [28]. Inter-group score comparison between novices, intermediates and experts were made using analysis of variance. Internal consistency and Spearman’s rank correlation coefficient with self-reported experience (total case volume), Global Operative Assessment of Laparoscopic Skills (GOALS) score and Objective Performance Rating Scale (OPRS) score were calculated. GOALS and OPRS scores were previously obtained prior to this study and were only used if they were obtained within 6 months of the assessment. Test–retest reliability was also calculated based on a random computer-generated sample of 10 participants who repeated the assessment after 1 week. Power calculation was based on prior work with the GOALS and OPRS assessment tools for laparoscopic cholecystectomy [29–31]. Using an α of 0.05 and a power of 80 %, with 2-sided testing, more than 6–8 subjects were required in each group. A p value of <0.05 was considered statistically significant. All statistical analyses were performed using JMP 11 software (SAS Institute Inc., Cary, NC).

\[ \text{Participants} \]

Participants from six institutions in Canada, USA, United Kingdom and Japan were enrolled in this study and completed the 12-item assessment without access to any curricular modules. Subjects included general surgery-trained residents, fellows and attending surgeons. There were no restrictions on either years in training, years in independent practice, total or annual case volume, or subspecialty training. Participants were categorized based on training level and prior experience (total laparoscopic cholecystectomies performed) into three groups: novice (<25 cases), intermediate (25–100 cases, or surgical residents or fellows with more than 100 cases) and expert (attending surgeons having performed more than 100 cases). All subjects received a unique username and password and completed the assessment remotely on a personal computer, tablet or mobile device. Scores were automatically calculated by the software and investigators were blinded.

\[ \text{Sample size and statistical analysis} \]

A contemporary framework of validity was used to provide evidence for the assessment as a measure of intra-operative decision-making [28]. Inter-group score comparison between novices, intermediates and experts were made using analysis of variance. Internal consistency and Spearman’s rank correlation coefficient with self-reported experience (total case volume), Global Operative Assessment of Laparoscopic Skills (GOALS) score and Objective Performance Rating Scale (OPRS) score were calculated. GOALS and OPRS scores were previously obtained prior to this study and were only used if they were obtained within 6 months of the assessment. Test–retest reliability was also calculated based on a random computer-generated sample of 10 participants who repeated the assessment after 1 week. Power calculation was based on prior work with the GOALS and OPRS assessment tools for laparoscopic cholecystectomy [29–31]. Using an α of 0.05 and a power of 80 %, with 2-sided testing, more than 6–8 subjects were required in each group. A p value of <0.05 was considered statistically significant. All statistical analyses were performed using JMP 11 software (SAS Institute Inc., Cary, NC).

\[ \text{Results} \]

Thirty-nine subjects completed the assessment, including 19 residents, 17 attending surgeons and 3 fellows. Participants included 8 novices, 14 intermediates and 17 experts, and their characteristics are summarized in Table 1. All residents and fellows were general surgery-trained, with the majority (19 [86 %]) having performed <100 laparoscopic cholecystectomies. One chief resident and two fellows performed between 100 and 150 cases. Amongst experts, 13 (76 %) practise at an academic institution, 5 at a community hospital (29 %) and 2 (12 %) at a rural hospital. All but one attending surgeon work in public

\[ \text{Participants} \]

Participants from six institutions in Canada, USA, United Kingdom and Japan were enrolled in this study and completed the 12-item assessment without access to any curricular modules. Subjects included general surgery-trained residents, fellows and attending surgeons. There were no restrictions on either years in training, years in independent practice, total or annual case volume, or subspecialty training. Participants were categorized based on training level and prior experience (total laparoscopic cholecystectomies performed) into three groups: novice (<25 cases), intermediate (25–100 cases, or surgical residents or fellows with more than 100 cases) and expert (attending surgeons having performed more than 100 cases). All subjects received a unique username and password and completed the assessment remotely on a personal computer, tablet or mobile device. Scores were automatically calculated by the software and investigators were blinded.

\[ \text{Sample size and statistical analysis} \]

A contemporary framework of validity was used to provide evidence for the assessment as a measure of intra-operative decision-making [28]. Inter-group score comparison between novices, intermediates and experts were made using analysis of variance. Internal consistency and Spearman’s rank correlation coefficient with self-reported experience (total case volume), Global Operative Assessment of Laparoscopic Skills (GOALS) score and Objective Performance Rating Scale (OPRS) score were calculated. GOALS and OPRS scores were previously obtained prior to this study and were only used if they were obtained within 6 months of the assessment. Test–retest reliability was also calculated based on a random computer-generated sample of 10 participants who repeated the assessment after 1 week. Power calculation was based on prior work with the GOALS and OPRS assessment tools for laparoscopic cholecystectomy [29–31]. Using an α of 0.05 and a power of 80 %, with 2-sided testing, more than 6–8 subjects were required in each group. A p value of <0.05 was considered statistically significant. All statistical analyses were performed using JMP 11 software (SAS Institute Inc., Cary, NC).

\[ \text{Results} \]

Thirty-nine subjects completed the assessment, including 19 residents, 17 attending surgeons and 3 fellows. Participants included 8 novices, 14 intermediates and 17 experts, and their characteristics are summarized in Table 1. All residents and fellows were general surgery-trained, with the majority (19 [86 %]) having performed <100 laparoscopic cholecystectomies. One chief resident and two fellows performed between 100 and 150 cases. Amongst experts, 13 (76 %) practise at an academic institution, 5 at a community hospital (29 %) and 2 (12 %) at a rural hospital. All but one attending surgeon work in public
practice. All attending surgeons are general surgeons who perform laparoscopic cholecystectomies when on call [17 (100 %)]. There were 6 (41 %) minimally invasive surgeons and 4 (27 %) hepatopancreatico-biliary surgeons. Median time in practice was 7 years (interquartile range: 3–10). All 17 surgeons performed greater than 100 total laparoscopic cholecystectomies [100–500 cases: 10 (59 %); 501–1000 cases: 3 (18 %); >1000 cases: 2 (12 %)]. Only two surgeons (12 %) perform more than 100 cases annually.

There was high test–retest reliability (intraclass correlation coefficient $= 0.95$, 95 % CI 0.80–0.99) and internal consistency for the assessment (Cronbach’s $= 0.87$). Total examination score (all 12 items) and VCT score (5 drawing items) were significantly different between novices, intermediates and experts (Fig. 3; $p < 0.01$), with significantly greater score variance amongst novices and intermediates (standard error: 0.79 and 0.55, respectively) compared to experts (standard error: 0.16). There was a high correlation between total case number and total score ($= 0.83$, $p < 0.01$) and between total case number and VCT score ($= 0.82$, $p < 0.01$; Fig. 4). Specifically, scores seemed to increase proportionally during the initial stage of the learning curve (0–250 total case volume), and eventually begin to plateau after 200–250 cases. Intra-operative assessments (GOALS and OPRS) were available for 9 residents, with moderate to high correlation between total score and GOALS score ($= 0.66$, $p = 0.05$), between VCT score and GOALS score ($= 0.83$, $p < 0.01$), between total score and OPRS score ($= 0.67$, $p = 0.04$) and between VCT score and OPRS score ($= 0.78$, $p = 0.01$; Fig. 5).

Thirty-seven participants completed the questionnaire (Table 2). Non-responders included one chief resident and one fellow, both in the intermediate group. Most subjects either agreed or strongly agreed that the assessment tool was easy to use [$n = 29$ (78 %)], facilitates development of intra-operative decision-making [$n = 28$ (81 %)], and should be integrated into surgical training [$n = 28$ (76 %)]. Average time to complete the assessment was 5–10 min.

**Discussion**

Bile duct injuries and other complications during laparoscopic cholecystectomy can be a significant source of morbidity, and avoiding such injuries relies heavily on complex mental processes that guide intra-operative

---

**Fig. 1** E-learning platform screenshot (http://www.thinklikeasurgeon.com)
decisions and behaviours. Yet, despite the paradigm shift in surgical education and emphasis on patient safety, most training programmes seldom reinforce these important skills in a systematic manner using objective and measurable methods. With the technological advances in computing, gaming and mobile devices of our generation, there is a plethora of opportunities to introduce innovative and cost-effective educational material into surgical training. 

Fig. 2 A Schematic example of score calculation for the visual concordance test. Pixels selected during annotation of the still frame image are captured and mapped on a Cartesian coordinate and compared to a distribution of expert responses. Pixels selected by experts are topographically grouped into zones and assigned a weight based on the proportion of experts who selected each pixel and the shape of the expert response distribution. 

B Screenshot of sample user response when asked to identify where they would start dissecting the hepatocystic triangle based on a video from a laparoscopic cholecystectomy. The user can subsequently receive feedback on their decision-making based on the distribution of expert responses for that same question.
training. In this study, an interactive multimedia e-learning platform (Think Like A Surgeon) was designed for experiential learning with remote access for ongoing learning, repeated performance assessment and immediate feedback. This study describes a novel metric for intra-operative decision-making using this platform and provides validity evidence to support its use as an assessment of advanced cognitive skills during laparoscopic cholecystectomy. Performance was better amongst higher levels of training, and there were strong correlations with self-reported experience and intra-operative performance.

The foundation of a competency-based training model is based on the achievement of measurable and observable competencies. This definition implicates the need for objective metrics that can evaluate such skills to determine if standards of competency are being met prior to...
practicing on patients. This creates significant challenges given the inherent difficulty in tapping into the minds of individuals to obtain an accurate depiction of their thought habits and cognitive processes, despite the strong reliance on these aptitudes for minimizing errors such as bile duct injuries. There is a need for better methods to appreciate what learners are thinking intraoperatively and to provide a forum to deliberately practise these skills—be it real or in a simulated environment. Most surgical training programmes today rely on a time-dependent approach, whereby intra-operative judgment and decision-making are acquired throughout training in a non-systematic, situation and instructor-dependent manner, despite the fact that most experts recognize their fundamental role in surgical expertise [32–34]. Assessment tools developed to date tend to rely on rating scales with generic items, such as “decision-making” and “situation awareness”, or are task-specific, which provide limited insight into the cognitive underpinnings of the procedure. Avoiding adverse events during laparoscopic cholecystectomy has become a primary concern, such as with the formation of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) Safety in Cholecystectomy Task Force [35]. While educational material becomes widely available and helps contribute to a culture of safety in the operating room, there currently exists no objective metric to evaluate these complex skills necessary for avoiding bile duct injuries.

The VCT is a new method of measuring performance on a user-friendly platform that is accessible remotely. This study demonstrates its psychometric properties and provides validity evidence with regard to content, internal structure and comparison to other variables [28] (such as years in training, total case volume and intra-operative GOALS and OPRS scores), ultimately supporting its integration into curricular modules for learning and assessment. The advantage of this methodology is that learners can compare their thoughts and cognitive behaviours to that of a panel of experts—a process that is normally restricted to watch a single expert operates in real life or watch a surgical video. These methods hardly engage the learner actively with opportunities for experiential learning (e.g. “practice by doing”), by deliberately training advanced cognitive skills or by providing feedback from multiple experts at once as opposed to a single expert in one sitting. While it may seem that plenty of opportunities exist to practise these skills throughout the course of a 5-year training programme [36], the practice is not done deliberately to target specific learning objectives in the cognitive domain, and immediate repetition and performance feedback are seldom available. The VCT score and Think Like A Surgeon platform attempt to overcome these challenges by embedding the assessment tool within the curricular modules to provide ample opportunities for learners to evaluate their performance of newly acquired skills on an entire library of case-specific examples. Specifically, it evaluates an individual’s ability to synthesize vast amounts of data from the surgical field into an understanding of the operative environment and compares the accuracy of their judgments and decisions to that of experts.

Other innovative techniques that provide objective measures of situation awareness, pattern recognition and decision-making have also surfaced in the literature, including eye-tracking devices to map visual attention [37] and a variety of technologies for video-based performance analysis [38]. Schlachta et al. [39] have developed a similar metric specifically designed to evaluate visuospatial abilities by comparing the level of agreement between trainees’
perception of the dissection plane (drawn as a line with a stylus on a tablet computer) compared to the ideal plane. In spite of the heterogeneity and intricacies of mental processes, it is unlikely that a single methodology will provide a comprehensive assessment of advanced cognitive skills, and these innovations (including the VCT scoring system) are mostly complimentary.

An e-learning educational tool was specifically chosen for this assessment to provide a ubiquitous, cost-effective and secure platform that provides ample opportunities for simulation-based learning at the point of access and at the convenience of the learner when they are maximally engaged. Advantages of this pervasive technology are that it is widely accessible to a large audience in different geographic locations, is easily adapted into a training environment, avoids the need for experts, faculty, equipment and other resources, is not reliant on the availability of teaching moments and provides ongoing access for spaced education and longitudinal performance tracking. Most participants in this study supported the feasibility and educational value of *Think Like A Surgeon* and its assessment tool, as long as immediate feedback on performance was made available. While there is good evidence to suggest that technology-rich learning environments such as e-learning and gaming can be effective tools for surgical training [40], their value for learning remains highly contingent on adherence to a theory-driven instructional design and best practices in educational psychology [41].

Intra-operative judgment and decision-making are complex to understand, let alone measure, and there is considerable work still needed to achieve these objectives. Despite the validity evidence to support the interpretations of results from this assessment, there are also other complimentary methods that have similarly attempted to evaluate this construct. Other important thought habits may not be optimally expressed by drawing on a still frame of the surgical field—often they require to be articulated in order to be assessed. To address this limitation, the platform has since been updated with a new property that allows users to answer questions by typing free-text responses that are subsequently analysed for correctness against a repertoire of predetermined correct textual answers. Also, while the psychometric properties of this assessment tool seem positive, they remain restricted to a small sample size and larger studies evaluating a broad range of expertise are required.

### Conclusion

In summary, this study describes the development of novel metrics and a Web-based platform and provides preliminary validity evidence for their assessment of decision-making during laparoscopic cholecystectomy. Most participants perceive the platform as useful and feasible to use for learning and assessment. Given the consequences of intra-operative injuries, the implementation of this educational programme and assessment tool into competency-based curricula can provide objective and structured feedback and ultimately improve patient safety.

**Acknowledgments** We acknowledge the work of Robert Messina for developing *Think Like A Surgeon* and all the surgeons who participated in this study. Amin Madani is supported by a Quebec Health Science Research (FRQ-S) scholarship and the McGill Surgeon-
Scientist Program. The Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation is supported by an unrestricted educational grant from Medtronic. Rajesh Aggarwal is supported by the Chercheurs-Boursiers Cliniciens of the FRQ-S and the Montreal General Hospital Foundation.

Compliance with ethical standards

Disclosures  Amin Madani, Yusuke Watanabe, Elif Biligic, Philip H. Pucher, Melina C. Vassiliou, Rajesh Aggarwal, Gerald M. Fried, Elliot J. Mitmaker and Liane S. Feldman have no conflicts of interest or financial ties to disclose. Rajesh Aggarwal is a consultant for Applied Medical. Melina C. Vassiliou is a consultant for Coviden. The Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation is supported by an unrestricted educational grant from Medtronic.

References


