



Validation of a virtual reality laparoscopic appendicectomy simulator: a novel process using cognitive task analysis

Sandeep Krishan Nayar^{1,2} · Liam Musto¹ · Roland Fernandes¹ · Rasiyah Bharathan³

Received: 14 October 2018 / Accepted: 12 November 2018 / Published online: 19 November 2018
© Royal Academy of Medicine in Ireland 2018

Abstract

Background Virtual reality (VR) simulation is a vital component of surgical training with demonstrated improvements in surgical quality and clinical outcome.

Aims To validate the LAP Mentor (Simbionix™) laparoscopic appendicectomy (LA) VR simulator with inclusion of a novel tool, Cognitive Task Analysis (CTA).

Methods Thirty-two novices and nine experienced surgeons performed two simulated LAs. An expert-consensus questionnaire guided face validity assessment. Content validity was assessed using CTA-derived questions encompassing eight operative steps and four decision points. Construct validity was evaluated using dexterity metrics, masked assessment of surgical quality using the OSATS global rating scale, and mental workload from two validated tools: the NASA-TLX and SMEQ. Ten novices performed eight further LAs for learning curve assessment.

Results Face validity was demonstrated across all domains. Considering content validity, the essential technical and non-technical steps were evident. The experienced group performed the procedure quicker (median time 361 vs. 538 s, $P = 0.0039$) with fewer total movements (426 vs. 641, $P < 0.0001$) and shorter idle time (131 vs. 199 s, $P = 0.0006$). This correlated with higher OSATS scores (median 33.5 vs. 22.2, $P < 0.0001$) and lower mental demand (NASA-TLX: 9.0 vs. 13.75, $P = 0.012$; SMEQ: 60 vs. 80, $P = 0.0025$), indicating construct validity. Learning curve data showed statistically significant improvements after the 7th session for procedure time, total movements and idle time, which correlated with reduction in mental demand.

Conclusions The LAP Mentor demonstrates face, content and construct validity for LA; thus, it can be used as an effective tool in surgical training. Task repetition leads to achievement of expert benchmarks.

Keywords Cognitive task analysis · Laparoscopic appendicectomy · Learning curve assessment · Virtual reality simulation

Introduction

Appendicectomy is the most commonly performed procedure in general surgery, with over 34,600 cases performed in the UK in 2006 to 2007, and over 250,000 performed each year in the USA [1, 2]. Laparoscopic appendicectomy offers a number of advantages over the open procedure such as shorter hospital

stays, reduced need for post-operative analgesia, earlier return to work and lower rate of wound infection [3–5]. In addition, the laparoscopic approach has the benefit of better visualisation of other intra-abdominal organs, should the appendix be macroscopically normal. This is a common and acute entity [6]. The procedure is commonly performed by trainees progressing through their learning curve with appropriate supervision. The learning curve for laparoscopic appendicectomies has been found to be approximately 30 cases before stability [7]. It is also widely recognised that errors are more common during the learning curve of a procedure [8]. Simulation-based training can yield a “pre-trained novice”, who will be in a better position to learn and perform in the operating room [9]. It is therefore imperative that there is a structured approach to training to ensure competence at the earliest opportunity.

Virtual reality (VR) simulation-based training is a concept adopted from the training of airline pilots. VR simulation offers

✉ Sandeep Krishan Nayar
sandeep.nayar@doctors.org.uk

¹ Department of General Surgery, Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK

² Maidstone Hospital, Hermitage Lane, Maidstone ME16 9QQ, UK

³ Department of Gynaecological Oncology, Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK

a controlled environment to practice essential surgical skills and rehearse whole procedures [10]. In addition, immediate detailed feedback can be given, as well as data collected to establish common errors and help improve the quality of training [10]. It is important that validation studies are carried out for such simulation models in order to establish whether the VR simulator is functioning as intended [11, 12]. Validation entails establishment of face, content and construct validity. Face validity relates to how realistic the simulator is, content validity is an evaluation of the rigour with which the simulator replicates the task and construct validity is whether the simulator is able to distinguish between surgeons of varying experience [12].

During surgery, cognition comprises 75% of the task and motor skills account for the remaining 25% [13]. Therefore, in ensuring content validity, simulation should demonstrate non-technical aspects as well. Non-technical skills, such as intra-operative decision-making, are complex and challenging to teach. This is largely due to experts becoming automated when performing surgery [14]. One way to address this challenge is to use Cognitive Task Analysis (CTA), which is a method to analyse, understand and map complex tasks that involve a number of higher functions such as decision-making, problem solving and judgement [15]. This method has been demonstrated to improve learning and retention of both technical and non-technical skills in procedures such as central venous catheterisation and flexor tendon repair [16–18]. A study by Smink et al. (2012) used CTA as a method to create a teaching aid for laparoscopic appendicectomies, forming a CTA script detailing both the operative steps as well as the decision-making points of the procedure [14]. To our knowledge, a CTA-derived script has not been used in evaluating the content validity of a training or an assessing device.

Numerous studies have assessed validity of surgical simulator models across multiple surgical specialties. However, in previous surgical simulator validation studies, the content validity has been assessed by questionnaires derived from expert consensus rather than a guided task analysis derivative such as one resulting from CTA [19–27].

The primary objective of this study is to validate the Symbionix™ laparoscopic appendicectomy simulator. This includes face validity, novel content validation with technical and non-technical elements derived from a CTA script, and construct validation. Indeed, the study will define construct validity in terms three categories of metrics; these are surgical quality, mental load metrics and simulator dexterity metrics. The secondary objective is to define the learning curve in terms of simulator dexterity metrics and mental load metrics.

Materials and methods

Participants and setting

Thirty-two novice participants (performed < 5 operative laparoscopic appendicectomies) were enrolled to perform two simulated laparoscopic appendicectomies. In addition, nine experienced surgeons (performed > 50 operative laparoscopic appendicectomies) participated in simulator validation. A written informed consent was obtained prior to participation.

The participants completed a baseline questionnaire to establish demographic information, surgical experience and other factors that may influence performance at laparoscopic surgery (Table 1).

Table 1 Demographic information

Demographic information	Novice (<i>n</i> = 32)	Experienced (<i>n</i> = 9)
Mean age (years) ± SD	26 ± 3.4	39 ± 6
Gender, <i>n</i> (%)		
Male	62.5%	89%
Female	37.5%	11%
Hand dominance, <i>n</i> (%)		
Right-handed	87.5%	89%
Left-handed	12.5%	11%
Do you play video games on a regular basis? <i>n</i> (%)		
Yes	34.4%	22%
No	65.6%	78%
If yes, how many hours per week? <i>n</i> (mean hours)	3	1.5
Do you play a musical instrument? <i>n</i> (%)		
Yes	50%	22%
No	50%	78%
Previous experience of laparoscopic surgery, <i>n</i> (%)		
Yes	34.4%	100%
No	65.6%	0%
Median number of laparoscopic appendicectomies observed	2	N/A
Median number of laparoscopic appendicectomies assisted	1	N/A
Median number of laparoscopic appendicectomies performed	0	80
Previous experience of a laparoscopy virtual reality simulator? <i>n</i> (%)		
Yes	34.4%	89%
No	65.6%	11%

An expert mentor demonstrated the steps of the procedure on the VR simulator. Each participant subsequently performed the procedure twice. The first attempt was to permit familiarisation with the task and simulator. The second session served as the assessment session. After the task, the participants completed feedback and assessment tools, namely a face validity questionnaire, a CTA-derived questionnaire completed by the experienced group only for ascertaining content validity and workload assessment questionnaires.

Virtual reality simulator

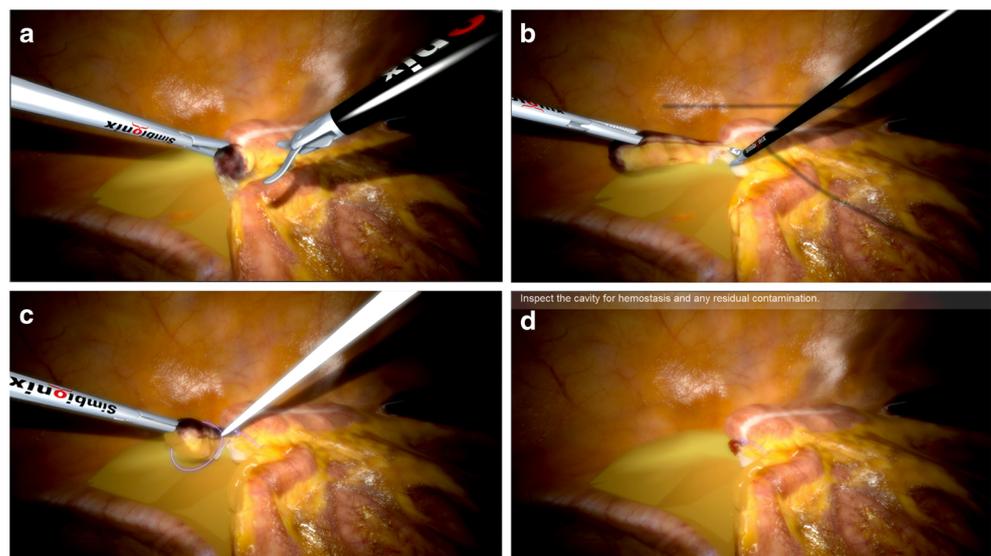
The LAP Mentor (Symbionix Corporation, Cleveland, OH, USA) is a VR laparoscopic simulator with a library of laparoscopic basic tasks, procedural tasks and whole procedures. In this study, participants performed a laparoscopic appendectomy on a gangrenous appendix that had not ruptured (Fig. 1a–d). The on-screen procedure starts following insertion of the laparoscopic ports and identification of the appendix and ends after endobag removal. A menu screen offers a selection of laparoscopic instruments that participants can select for the task. During the demonstration, participants were instructed on using an atraumatic grasper for manipulating the appendix, curved cold scissors for adhesiolysis and a bipolar diathermy device for resection of the mesoappendix, endoloops to ligate the appendix and an endobag for specimen retrieval. The simulator can generate a range of dexterity and safety metrics as performance outcome measures.

Assessment tools

Face validity

A face validity assessment tool was based on local expert consensus. Participants rated each statement on a five-point Likert scale; results are summarised in Table 2.

Fig. 1 a–d Laparoscopic appendectomy images from LAP Mentor. **a** (top left) Prior to starting adhesiolysis. **b** (top right) Division of the mesoappendix. **c** (bottom left) Endoloop application. **d** (bottom right) Following completion of appendectomy



Content validity

In the surgical education literature, content validity of simulators is usually analysed by a series of features as extolled by experts [12, 26, 27]. In an attempt to improve on this process, we adopted a novel approach. We examined a previously published CTA script for laparoscopic appendectomy [14]. Specific steps were taken to avoid potential assessment bias associated with selecting the CTA-derived questions. This was achieved by adhering to the consensus demonstrated within the published CTA script. By definition, we excluded preparation, peritoneal access and closure because the simulator was not designed to exhibit those features. Following the second repetition of laparoscopic appendectomy, the experienced group completed our CTA-derived questionnaire. This encompassed eight essential operative steps and four decision points of a laparoscopic appendectomy, allowing for assessment of content validation (Table 3).

Construct validity

The construct validity was assessed using three categories of outcome measures, namely quality of surgery, mental workload and dexterity metrics. The outcome metrics are detailed below.

Surgical quality The surgical quality was defined by the global rating scale component of Objective Structured Assessment of Technical Skills (OSATS) [28]. Two masked experts assessed videos of laparoscopic appendectomy from the second session of all participants. The mean value of the two assessors was used for analysis.

Mental workload assessment Mental workload was assessed using two validated tools: the National Aeronautics and Space

Table 2 Face validity

Simulator features	Median score		P value
	Novice (n = 32)	Experienced (n = 9)	
Appearance of instruments	5	4	0.46
Manoeuvring of the instruments	4	4	0.35
Function of the instruments	4	4	0.93
Appearance of the tissue	4	3	0.011*
Response to manipulation	4	3	0.040*
Depth perception training	4	4	0.52
Hand-eye co-ordination training	4	4	0.41
Two-handed co-ordination training	4	4	0.99
Overall utility as a training tool	5	4	0.88
Overall ergonomics of the simulator	4	4	0.44

Administration (NASA) Task Load index [Appendix 1] and the Subjective Mental Effort Questionnaire (SMEQ) [Appendix 2]. The NASA Task Load Index encompasses the participant's subjective experience by assessing six dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration, giving a score between 0 and 20 for each domain [29, 30]. The SMEQ focuses on the subjective mental effort of executing the task, giving a score between 0 and 100, and has been previously used in validation studies [31].

Dexterity metrics A range of dexterity metric data was generated from the laparoscopic simulator. This included economy of motion, use of electro-surgery, idle time, injuries to the appendix, length of appendix stump, total number of movements, total path length and total time taken to carry out the procedure.

Learning curve assessment

Ten of the novice participants performed eight further simulated laparoscopic appendicectomies for learning curve

assessment. The NASA and SMEQ workload assessments were completed after each attempt.

Statistical analysis

Power calculations from previous studies on laparoscopic virtual reality simulation have shown that a minimum of eight participants is required [32]. Data were analysed with GraphPad® Prism 7 (GraphPad Software, California, USA). Statistical analysis of the non-parametric data was performed using a Mann-Whitney *U* test for validity data, and a Kruskal-Wallis test for learning curve data. A Spearman's rank correlation coefficient analysis was performed to assess the degree of congruence between dexterity metrics, workload dynamics and surgical quality. A *P* value of <0.05 was considered statistically significant.

Ethical approval

Participation was voluntary, and the participants provided informed written consent. All simulator data were kept

Table 3 Content validity from CTA-derived questions

Content validity statements	Experienced only (n = 9)	
	Yes	No
There was adequate view of the appendix	100%	0%
<i>The junction of the appendix and caecum was adequately visualised</i>	89.9%	11.1%
<i>You were able to mobilise the appendix to locate appendiceal base</i>	100%	0%
<i>You were able to determine the anatomy of the mesoappendix and base of appendix</i>	89.9%	11.1%
<i>You were able to divide the mesoappendix</i>	100%	0%
<i>You were able to divide the appendiceal base</i>	100%	0%
<i>You were able to remove the appendix</i>	100%	0%
You were able to inspect the operative bed	100%	0%
<i>You were able to identify the location of the appendix</i>	100%	0%
You were able to inspect for other pathology	66.7%	33.3%
<i>You were able to choose an appropriate instrument</i>	100%	0%
You were able to determine for contamination	89.9%	11.1%

confidential. All local regulations from the Maidstone and Tunbridge Wells NHS Trust Education Centre were met.

Disclosures

All authors have no conflicts of interest or financial ties to disclose. This research was carried out independently of the Symbionix Corporation.

Results

Demographic information

Information regarding demographics of the participants are summarised in Table 1. The experienced group consisted of eight senior residents (specialist registrars) and one consultant in surgery. The novice group consisted of 25 interns (foundation year doctors) and 7 first year residents (core surgical trainees) at Maidstone and Tunbridge Wells Hospitals NHS Trust. There were no significant differences between novices and experienced surgeons with regard to gender or handedness.

Face validity

Face validity scores of the laparoscopic simulator are summarised in Table 2. For all but two features (appearance of the tissue and response to manipulation), there were no significant differences between the median scores of the two groups.

Content validity

Content validity from CTA-derived questions encompassing the eight essential operative steps (technical skills in italics) and four decision points (non-technical skills) of a laparoscopic appendicectomy is summarised in Table 3. Content validity was evaluated by the experienced surgeons only.

Construct validity

Surgical quality

Two masked experts objectively scored participants’ surgical performance using video playback. The novice group scored significantly lower than the experienced surgeons (median 22.2 vs. 33.5, $P < 0.0001$) (Fig. 2).

Workload assessment

The mental workload results between the two groups from the SMEQ and NASA Task Load Index are summarised in Table 4. There were statistically significant differences

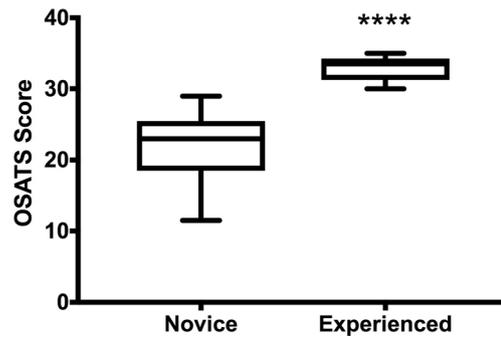


Fig. 2 Masked expert rating of surgical quality

between the two groups in mental demand scale of NASA-TLX ($P = 0.012$) and in the SMEQ scale ($P = 0.0025$), with the novice surgeons experiencing higher mental loading.

Dexterity metrics

The VR laparoscopic simulator generated a range of dexterity metrics, which are summarised in Table 5. The simulator was able to differentiate between the experienced surgeons and novices in a number of parameters, thus conferring construct validity. The experienced surgeons completed the procedure quicker (median time 361 vs. 538 s, $P = 0.0039$) (Fig. 3), and with fewer total numbers of movements in each hand (right hand: 242 vs. 380 movements, $P < 0.0001$; left hand: 184 vs. 261 movements, $P < 0.0001$) (Fig. 4). Furthermore, the median number of injuries to the appendix was zero in the experienced group compared to one in the novice group ($P = 0.0022$) and they completed the procedure with shorter idle time (median time 130.86 vs. 198.95 s, $P = 0.0006$). There were no statistically significant differences between the two groups in economy of motion or electro-cautery use.

Learning curve data

Assessment of the learning curve data revealed statistically significant improvement compared with the initial attempt by the 7th session for total time taken ($P = 0.0038$) (Fig. 5a), and idle

Table 4 Workload assessment outcomes

Domain	Median score		P value
	Novice (n = 32)	Experienced (n = 9)	
NASA-TLX			
Mental demand	13.75	9.0	0.012*
Physical demand	10	6.0	0.27
Temporal demand	8.5	7.0	0.63
Performance	6.75	6.0	0.82
Effort	13.0	9.0	0.17
Frustration	7.5	6.0	0.63
SMEQ	80	60	0.0025**

Table 5 Metric data outcomes

Metric data	Median score		P value
	Novice (n = 32)	Experienced (n = 9)	
Economy of motion (left) (average of path length/path time)	2.62	2.94	0.13
Economy of motion (right) (average of path length/path time)	3.05	3.22	0.77
Electro-surgery or energy activation time (total time in seconds)	15.65	12.14	0.95
Electro-surgery or energy usage while not in contact with tissue	0.32	0.79	0.54
Idle time (s)	198.95	130.86	0.0006***
Injury to the appendix	1	0	0.0022**
Number of movements of left instrument	261	184	< 0.0001****
Number of movements of right instrument	380	242	< 0.0001****
Total path length of left instrument (cm)	450.55	360.86	0.031*
Total path length of right instrument (cm)	708.94	481.20	0.0001****
Total procedure time (s)	537.68	361.11	0.0039**

time ($P = 0.0048$) (Fig. 5b). Regarding total number of movements, statistically significant improvement was seen at the 6th session for the left hand ($P = 0.016$) (Fig. 5c) and at the 7th session for right hand ($P = 0.0041$) (Fig. 5d). Furthermore, benchmarks set by the experienced surgeons were met at the 6th session for total time, idle time and total number of movements.

Mental workload, as identified by the SMEQ and NASA-TLX mental demand scores, showed statistically significant reduction at the 7th session with associated plateau ($P = 0.0017$) (Fig. 6a, b). Spearman’s rank correlation coefficient between total time taken and mental demand as defined by the SMEQ score and NASA-TLX was 0.95 ($P = 0.0001$) and 0.90 ($P = 0.0009$), respectively, indicating a strong correlation between time taken to perform the task and mental demand. Similar correlation was seen with total number of movements (SMEQ: $r = 0.95$, $P = 0.0001$; NASA-TLX: $r = 0.90$, $P = 0.0009$) and idle time (SMEQ: $r = 0.94$, $P = 0.0002$; NASA-TLX: $r = 0.89$, $P = 0.0012$).

Discussion

In the current clinical climate, there are multiple challenges to delivering optimal surgical training. These include limitations of work-time regulations, non-surgical management, imaging

techniques and the medico-legal environment to mention a few. It is therefore vital that trainees are able to gain sufficient surgical experience through enhanced training [33]. Furthermore, the laparoscopic approach to appendicectomy has become increasingly common in recent years due to advantages including reduced length of hospital stay, faster time to discharge and reduction in post-operative morbidity [34]. As a consequence, simulation has become an effective modality in developing surgical skills. Simulation-based training provides a proficiency-based stepwise approach to skill learning in a safe and conducive environment. Furthermore, immediate detailed feedback can be given, as well as data collected to establish common errors and help improve the quality of training [10]. Simulation enhances patient safety by shifting the learning curve from the operating room to the skills laboratory, thus improving surgical performance in the operating room [35]. Therefore, it is imperative to ensure these training systems are validated to ensure they are appropriate and effective training tools.

The objectives of this study, validation and learning curve assessment, have been achieved. The authors have demonstrated face, content and construct validity of the Symbionix™ virtual reality laparoscopic appendicectomy simulator. The validation and learning curve assessment results signify that it can be used effectively as a surgical training tool. This complements a recent study published by Sirimanna et al. (2017), which evaluated the same device [36].

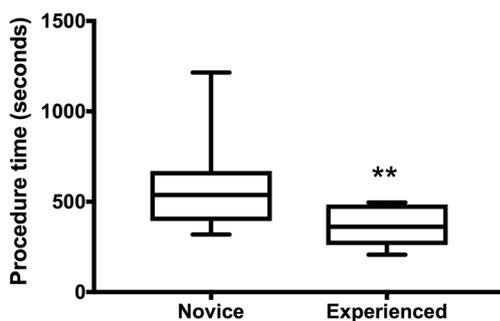


Fig. 3 Total procedure time

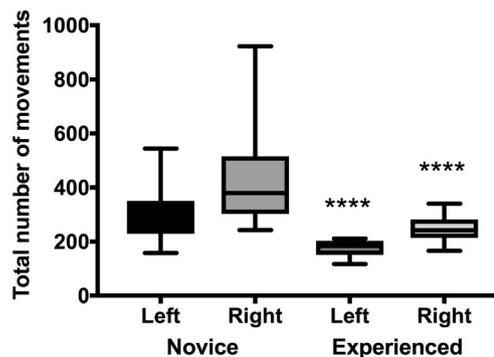
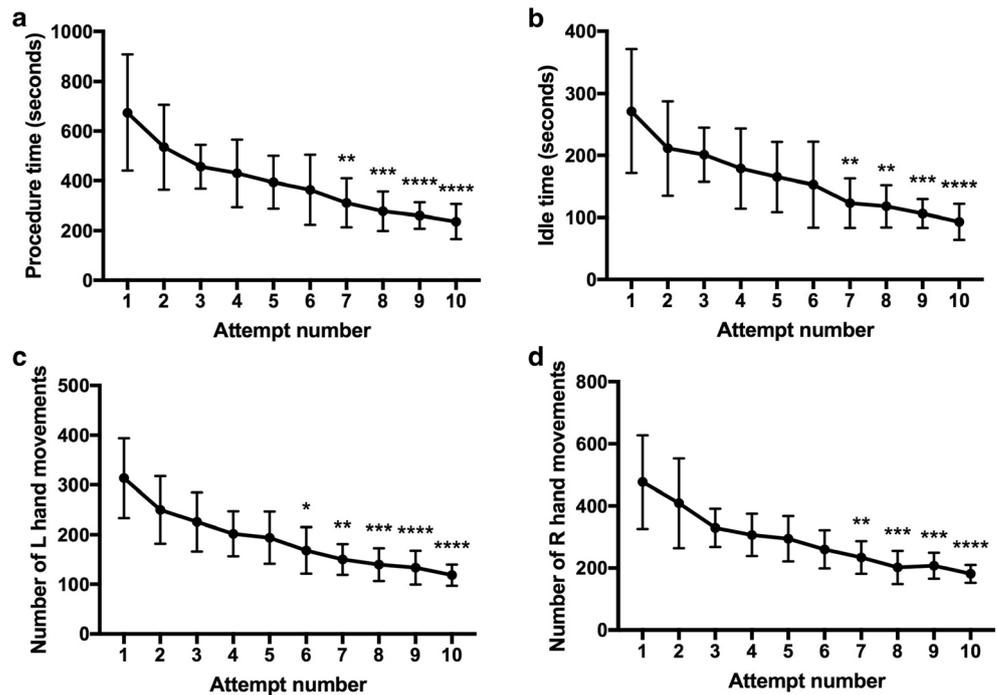


Fig. 4 Total number of movements in each hand

Fig. 5 a–d Learning curves for dexterity metrics. **a** (top left) Total procedure time. **b** (top right) Idle time. **c** (bottom left) Number of left hand movements. **d** (bottom right) Number of right hand movements



In the Sirimanna et al. (2017) study, face validity was based on a post-study questionnaire assessing domains such as whether the simulator was visually realistic, had accurate anatomy, authentic instruments and camera angles as well as tissue handling and behaviour. Similar results were found in our study, with globally high rating was given across the face validity criteria, including overall utility as a training tool (median score 5 in the novice group and 4 in the experienced group) and ergonomics of the simulator (median score 4 in both groups).

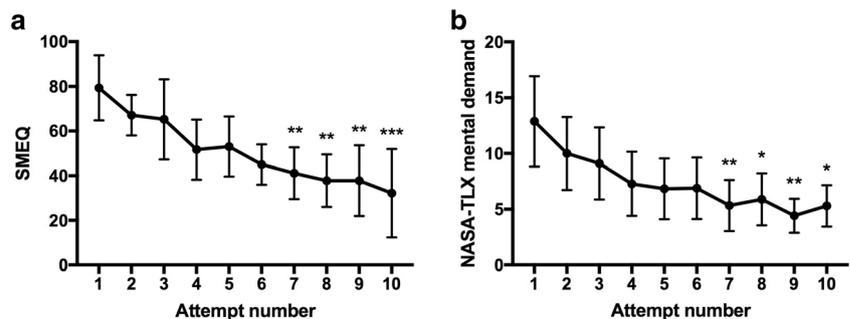
In addition to confirming the face demonstrated by the earlier study, we have introduced novel elements in terms of content validation and mental workload assessment. Our method of evaluating content validity is through the use of a CTA-derived questionnaire. To our knowledge no other simulator validation study has employed this approach. This provides a more objective and unbiased measure than traditional evaluation of content validity based on subjective assessment of the steps and skills involved [14]. Content validity from the CTA-derived questions was based on assessment from the experienced surgeons, who felt that the majority of agreed key steps of a

laparoscopic appendicectomy were covered during the VR simulation. Notably, three of the nine experienced surgeons felt the simulation did not allow for inspection for other pathology. The simulator was not designed to examine this aspect. Therefore, we recommend the manufacturer modify the task paradigm to include this feature in their next iteration.

Regarding construct validity, the study has demonstrated that the simulator can effectively discriminate novices from experienced surgeons. Dexterity metric data revealed that the median procedure time was significantly shorter for the experienced surgeons. Moreover, the experienced group did not commit errors leading to damage of the appendix and completed the tasks with fewer movements and shorter idle time. This demonstrates, as evidenced by OSATS score, that those with greater experience delivered a superior surgical performance in terms of quantity and quality. This complements and builds upon construct validity results from Sirimanna et al. based solely on dexterity metrics [36].

The experienced surgeons reported lower subjective mental demand and mental effort compared with the novices. This

Fig. 6 a, b Mental workload after repeated attempts. **a** (left) SMEQ. **b** (right) NASA-TLX mental demand



would infer that novices had to dedicate greater information processing capacity to execute the task, corresponding with the idea of rationing attention capacity as outlined by Gallagher et al. [9]. They state that for any individual, there is a finite amount of information capacity for a given task. Hence, when a novice is acquiring new skills such as those required for laparoscopic surgery, they must use this limited capacity to consciously monitor what their hands are doing in addition to the spatial judgments, whereas experts have developed automaticity in these areas and can focus on other areas such as non-technical skills. Simulation skills training, therefore, allows for the development of the “pre-trained novice”, who has been trained using simulation to the point where many of the psychomotor skills and spatial judgments become automated, and thus mental effort is reduced, freeing up processing capacity in theatre to focus on additional tasks such as decision-making and managing complications [9]. The benefits of training to automaticity have been shown with respect to skill transference [37].

This concept was further reinforced through our learning curve assessment, which demonstrated that task repetition at simulated laparoscopic appendectomy leads to reduction in mental load, correlating with improvement in dexterity metrics. Repeated workload assessment was novel in this regard and complements previous research demonstrating that increased workload is associated with inferior task performance and higher likelihood of errors [38]. Hence, simulation-based training can enhance performance and potentially reduce errors [39].

In terms of immediate implications for medical educators, the device can be used for training residents in the skill laboratory and potentially use this as a vehicle for warm-up immediately prior to surgery [40].

Strengths of our study include the reasonable numbers used to achieve adequate power for statistical analysis, multi-domain data (dexterity metrics, video assessment, mental load metrics) for construct validation and the use of two different tools to assess mental workload. In addition, the novel use of a CTA script for objective measurement of content validity. The main limitation to the study is the lack of an “intermediate” group, to evaluate the relevance of this task simulation for those trainees who are further on in their learning curve. Furthermore, the simulator does not cover the entire procedure and thus simulator competency alone is insufficient before independent practice. The next steps in advancing our findings would be through evaluation of automaticity rather than dexterity metrics, as well as to assess simulation-based training in skill transference to the operating room [37].

Conclusion

The LAP Mentor has demonstrated face, content and construct validity as a VR simulator for laparoscopic

appendectomies, indicating that it can be used as a safe and effective tool in surgical training. Furthermore, task repetition leads to achievement of expert benchmarks, associated with reduction in cognitive demand. Implications from these results include implementation of the simulation device in training residents in the skills laboratory as well as potential use of the device as a vehicle for warm-up immediately prior to surgery. Future steps will be to assess simulation-based training in skill transference to the operating room.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. HES Hospital episode statistics. Health and Social Care Information Centre. <http://www.hesonline.nhs.uk>. Accessed 19th July 2017
2. Owings MF, Kozak LJ (1998) Ambulatory and inpatient procedures in the United States, 1996. *Vital Health Stat* 13(139):1–119
3. Biondi A, Di Stefano C, Ferrara F, Bellia A, Vacante M, Piazza L (2016) Laparoscopic versus open appendectomy: a retrospective cohort study assessing outcomes and cost-effectiveness. *World J Emerg Surg* 11(1):44. <https://doi.org/10.1186/s13017-016-0102-5>
4. Yu MC, Feng YJ, Wang W, Fan W, Cheng HT, Xu J (2017) Is laparoscopic appendectomy feasible for complicated appendicitis? A systematic review and meta-analysis. *Int J Surg* 40:187–197. <https://doi.org/10.1016/j.ijso.2017.03.022>
5. Di Saverio S, Birindelli A, Kelly MD, Catena F, Weber DG, Sartelli M, Sugrue M, De Moya M, Gomes CA, Bhangu A, Agresta F, Moore EE, Soreide K, Griffiths E, De Castro S, Kashuk J, Kluger Y, Leppaniemi A, Ansaloni L, Andersson M, Coccolini F, Coimbra R, Gurusamy KS, Campanile FC, Biffl W, Chiara O, Moore F, Peitzman AB, Fraga GP, Costa D, Maier RV, Rizoli S, Balogh ZI, Bendinelli C, Cirocchi R, Tonini V, Piccinini A, Tugnoli G, Jovine E, Persiani R, Biondi A, Scalea T, Stahel P, Ivatury R, Velmahos G, Andersson R (2016) WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. *World J Emerg Surg* 11(34):34. <https://doi.org/10.1186/s13017-016-0090-5>
6. Agresta F, De Simone P, Leone L, Arezzo A, Biondi A, Bottero L, Catena F, Conzo G, Del Genio G, Fersini A, Guerrieri M, Illomei G, Tonelli P, Vitellaro M, Docimo G, Crucitti A (2004) Laparoscopic appendectomy in Italy: an appraisal of 26,863 cases. *J Laparoendosc Adv Surg Tech A* 14(1):1–8. <https://doi.org/10.1089/109264204322862270>
7. Jaffer U, Cameron AE (2008) Laparoscopic appendectomy: a junior trainee's learning curve. *JLS* 12(3):288–291
8. Thompson JE, Egger S, Bohm M, Haynes AM, Matthews J, Rasiah K, Stricker PD (2014) Superior quality of life and improved surgical margins are achievable with robotic radical prostatectomy after a long learning curve: a prospective single-surgeon study of 1552 consecutive cases. *Eur Urol* 65(3):521–531. <https://doi.org/10.1016/j.eururo.2013.10.030>
9. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 241(2):364–372

10. Yiannakopoulou E, Nikiteas N, Perrea D, Tsigris C (2015) Virtual reality simulators and training in laparoscopic surgery. *Int J Surg* 13: 60–64. <https://doi.org/10.1016/j.ijssu.2014.11.014>
11. Aucaer JA, Groch NR, Troxel SA, Eubanks SW (2005) A review of surgical simulation with attention to validation methodology. *Surg Laparosc Endosc Percutan Tech* 15(2):82–89
12. McDougall EM (2007) Validation of surgical simulators. *J Endourol* 21(3):244–247. <https://doi.org/10.1089/end.2007.9985>
13. Spencer FC (1978) Teaching and measuring surgical techniques: the technical evaluation of competence. *Bull Am Coll Surg* 64:9–12
14. Smink DS, Peyre SE, Soybel DI, Tavakkolizadeh A, Vernon AH, Anastakis DJ (2012) Utilization of a cognitive task analysis for laparoscopic appendectomy to identify differentiated intraoperative teaching objectives. *Am J Surg* 203(4):540–545. <https://doi.org/10.1016/j.amjsurg.2011.11.002>
15. Pugh CM, DaRosa DA, Santacaterina S, Clark RE (2011) Faculty evaluation of simulation-based modules for assessment of intraoperative decision making. *Surgery* 149(4):534–542. <https://doi.org/10.1016/j.surg.2010.10.010>
16. Crandall B, Klein GA, Hoffman RR (2006) Working minds: a practitioner's guide to cognitive task analysis. MIT Press, Cambridge
17. Velmahos GC, Toutouzas KG, Sillin LF, Chan L, Clark RE, Theodorou D, Maupin F (2004) Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *Am J Surg* 187(1):114–119
18. Luker KR, Sullivan ME, Peyre SE, Sherman R, Grunwald T (2008) The use of a cognitive task analysis-based multimedia program to teach surgical decision making in flexor tendon repair. *Am J Surg* 195(1):11–15. <https://doi.org/10.1016/j.amjsurg.2007.08.052>
19. Paro JAM, Luan A, Lee GK (2017) Validation of a cognitive task simulation and rehearsal tool for open carpal tunnel release. *Arch Plast Surg* 44(3):223–227. <https://doi.org/10.5999/aps.2017.44.3.223>
20. Xu S, Perez M, Perrenot C, Hubert N, Hubert J (2016) Face, content, construct, and concurrent validity of a novel robotic surgery patient-side simulator: the Xperience Team Trainer. *Surg Endosc* 30(8):3334–3344. <https://doi.org/10.1007/s00464-015-4607-x>
21. Barsness KA, Rooney DM, Davis LM, O'Brien E (2015) Evaluation of three sources of validity evidence for a laparoscopic duodenal atresia repair simulator. *J Laparoendosc Adv Surg Tech A* 25(3):256–260. <https://doi.org/10.1089/lap.2014.0358>
22. Escamirosa FP, Flores RM, Garcia IO, Vidal CR, Martinez AM (2015) Face, content, and construct validity of the EndoViS training system for objective assessment of psychomotor skills of laparoscopic surgeons. *Surg Endosc* 29(11):3392–3403. <https://doi.org/10.1007/s00464-014-4032-6>
23. Kawaguchi K, Egi H, Hattori M, Sawada H, Suzuki T, Ohdan H (2014) Validation of a novel basic virtual reality simulator, the LAP-X, for training basic laparoscopic skills. *Minim Invasive Ther Allied Technol* 23(5):287–293. <https://doi.org/10.3109/13645706.2014.903853>
24. Xiao D, Jakimowicz JJ, Albayrak A, Buzink SN, Botden SM, Goossens RH (2014) Face, content, and construct validity of a novel portable ergonomic simulator for basic laparoscopic skills. *J Surg Educ* 71(1):65–72. <https://doi.org/10.1016/j.jsurg.2013.05.003>
25. Hennessey IA, Hewett P (2013) Construct, concurrent, and content validity of the eoSim laparoscopic simulator. *J Laparoendosc Adv Surg Tech A* 23(10):855–860. <https://doi.org/10.1089/lap.2013.0229>
26. Hung AJ, Zehnder P, Patil MB, Cai J, Ng CK, Aron M, Gill IS, Desai MM (2011) Face, content and construct validity of a novel robotic surgery simulator. *J Urol* 186(3):1019–1024. <https://doi.org/10.1016/j.juro.2011.04.064>
27. Gavazzi A, Bahsoun AN, Van Haute W, Ahmed K, Elhage O, Jaye P, Khan MS, Dasgupta P (2011) Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). *Ann R Coll Surg Engl* 93(2):152–156. <https://doi.org/10.1308/003588411x12851639108358>
28. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 84(2): 273–278
29. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) *Advances in psychology*, vol 52. North-Holland, pp 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
30. Hart SG (2006) NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (9):904–908. <https://doi.org/10.1177/154193120605000909>
31. Anvari M (2007) Remote telepresence surgery: the Canadian experience. *Surg Endosc* 21(4):537–541. <https://doi.org/10.1007/s00464-006-9040-8>
32. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, Darzi A (2006) An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg* 244(2):310–314. <https://doi.org/10.1097/01.sla.0000218094.92650.44>
33. Chikwe J, de Souza AC, Pepper JR (2004) No time to train the surgeons. *BMJ* 328(7437):418–419. <https://doi.org/10.1136/bmj.328.7437.418>
34. Guller U, Hervey S, Purves H, Muhlbaier LH, Peterson ED, Eubanks S, Pietrobon R (2004) Laparoscopic versus open appendectomy: outcomes comparison based on a large administrative database. *Ann Surg* 239(1):43–52. <https://doi.org/10.1097/01.sla.0000103071.35986.c1>
35. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 338:b1802. <https://doi.org/10.1136/bmj.b1802>
36. Sirimanna P, Gladman MA (2017) Development of a proficiency-based virtual reality simulation training curriculum for laparoscopic appendectomy. *ANZ J Surg* 87(10):760–766. <https://doi.org/10.1111/ans.14135>
37. Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD (2012) Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg* 255(1):30–37. <https://doi.org/10.1097/SLA.0b013e318220ef31>
38. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D (2010) Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc* 5(5): 267–271. <https://doi.org/10.1097/SIH.0b013e3181e3f329>
39. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 236(4):458–463; discussion 463–454. <https://doi.org/10.1097/01.sla.0000028969.51489.b4>
40. Calatayud D, Arora S, Aggarwal R, Kruglikova I, Schulze S, Funch-Jensen P, Grantcharov T (2010) Warm-up in a virtual reality environment improves performance in the operating room. *Ann Surg* 251(6):1181–1185. <https://doi.org/10.1097/SLA.0b013e3181deb630>