Optimal Timing of Assessment Tasks Depending on Experience Level of Surgical Trainees

Cecilie Våpenstad (MSc)[a,b], Erlend Fagertun Hofstad (MSc)[b], Tor Eivind Bernstein[d] (MD), Petter Aadahl (PhD)[a], Gjermund Johnsen (MD)[c,d], Ronald Mårvik (PhD)[a, c, d]

[a]Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
[b]Department of Health Research, SINTEF Digital AS, Trondheim, Norway,
[c]National Advisory Unit for Advanced Laparoscopic Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[d]Department of Gastrointestinal Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway

Corresponding author:
Cecilie Våpenstad
SINTEF Digital AS
Department of Health Research
PB 4760 Torgarden,
7465 Trondheim, Norway
Phone: +47 93 22 74 31
e-mail: cecilie.vapenstad@sintef.no

Abstract

Introduction: Box trainers with motion analysis are important add-ons to surgical training and skills assessment outside the operating room, given that they exhibit construct validity.

Material and Methods: Four different tasks were tested for construct validity on a new laparoscopic box trainer with integrated motion analysis. Tracking data from the simulator were analyzed for eighteen parameters per task using an in-house software comparing participants with three different experience levels.

Results: In total, 10 novices, 22 intermediates and 16 experts enrolled. No or limited significant difference were found for the *peg picker* and *rope race*. For the *precision cutting* task 12 parameters showed significant difference between novices and intermediates, 14 between novices and experts and 1 between intermediates and experts. For the *suture* task the corresponding results were 1, 15 and 6.

Conclusions: The *precision cutting* and *suture* task both showed construct validity for many of the parameters. While the *precision cutting* task distinguished best between novices and the other two groups, the *suture* task distinguished best between experts and the other two groups. These results show the importance of the timing of an assessment task, and that an assessment task might have limited value if experience levels are not considered.

KEY WORDS: Surgical Training, Laparoscopy, Box Trainer, Motion Tracking, Construct Validity

Introduction

Traditional attendance based surgical training has been criticised for no longer ensuring sufficient exposure to, nor training on, the different aspects of technical skills that make a competent surgeon [1-3]. Thus, training outside of the operating room (OR), using risk-free environments like simulators has been introduced [1, 4-8]. An important advantage of simulators is their assessment tools, an essential part of proficiency-based or competency-based education [2, 9]. Skills assessment tools can be used to evaluate the effect of the training and the competences of the surgical trainees, under the condition that they show construct validity [10]. A simulator that shows construct validity is a tool that can be used to measure surgical skills, and give both formative and summative feedback. Formative feedback, by evaluating if the trainee progresses and acquires surgical skills, using the simulator to check that the trainer improves his score on the simulator. Summative feedback, in the sense of a quality assurance to verify that the trainer attains proficiency-based skills using the simulator as an objective assessment tool. By giving formative and summative feedback to the trainee or the trainer on reference expert levels and what skills to improve, training can be optimized and quality assured [2, 11].

Compared to video box trainers, a key advantage of virtual reality (VR) simulators has been their assessment tools [12]. However, in recent years some box trainers have also been equipped with assessment tools, and at the same time their costs have been kept low compared to VR simulators [13]. The Simball® box (Gothenburg, Sweden) is a box trainer with an integrated assessment tool that tracks instruments movements using the Simball joystick [14, 15].

Prior to introducing a competency-based training program at three hospitals in Norway, we performed a construct validity test on the Simball® box from Surgical Science (Gothenburg, Sweden). To our knowledge, one previous study has investigated construct validity for a similar set-up on the suture task [15]. This is the first study that examines construct validity for four of the tasks on the Simball® box. Construct validity was tested by comparing time and motion metrics for surgeons and medical students with three different levels of experience in laparoscopy [10]. The four tasks *peg picker*, *rope race*, *precision cutting* and *suture* require multiple laparoscopic skills acquired throughout different stages of surgical education. By comparing time and motion metrics for three different levels of experience we could investigate whether certain tasks were more appropriate at certain levels of experience than others.

Materials and Methods

Participants

All aspects of the study were approved by the Norwegian Data Protection Agency, and all participants gave written informed consent to participate. Participants with different levels of experience in laparoscopy, from three hospitals in Norway were recruited.

The Video Box Trainer

The simulator used in this study was the Simball® Box from Surgical Science AB (Gothenburg, Sweden) (Figure 1) with the Simball® joystick (Figure 2). Regular 5 mm laparoscopic instruments were inserted into the joystick. The joystick consisted of a ball, with a laser marked dot pattern code that was unique for each angular position of the instrument, and an instrument holder with a linear potentiometer that measured linear motion of the instrument (Figure 2) [15]. The simulator had an integrated web-camera and could, as an option, attach a laparoscopic camera, as we did in our set-up. The system consisted of the Simball® Training Box software (SW) version 2.3.2.1 running on a laptop with Windows 8, the laparoscopic camera connected to Exera II CV-180 (Olympus GmbH, Hamburg, Germany) (Figure 1), and the following instruments from Olympus GmbH (Hamburg, Germany): graspers (dissect), scissors (Metzenbaum, Ergo handle, 19 mm jaws) and needle holders (30 cm). The participants operated the scissors in their dominant hand during the precision cutting task and used two needle holders during the suture task.

The participants tested four tasks: *peg picker,* a modified *rope race, precision cutting and suture* (Figure 1 and 3). The suture task was one stitch with one double, one single and one double knot. The participants performed each task twice, i.e. a trial run to get acquainted with the task, and thereafter the assessment run. Motion data was calculated based on the assessment run.

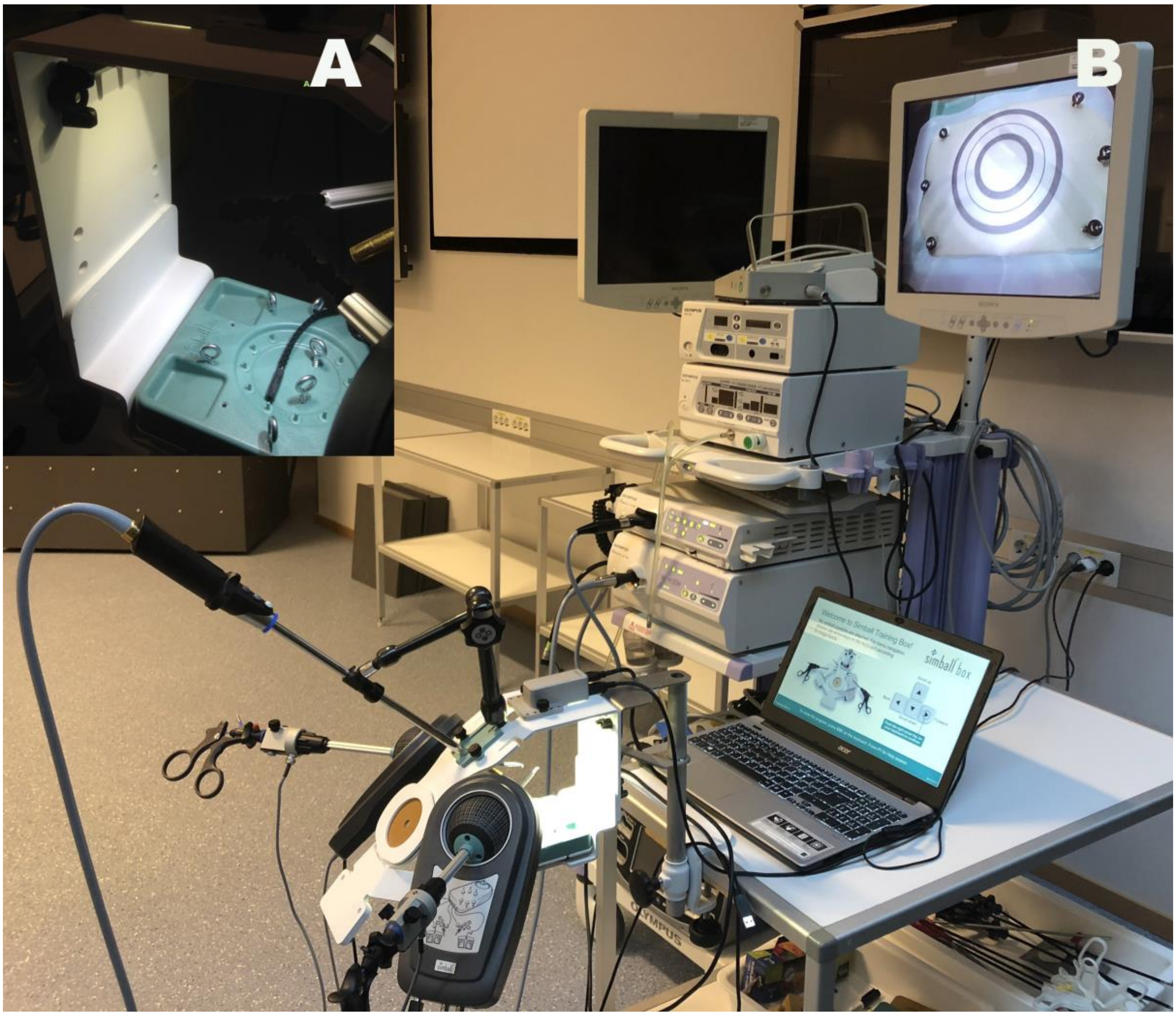

Figure 1. (A) The surgical space with the modified rope race. (B) The Simball® simulator with two cameras: 1) the webcam connected to the laptop showing the simulator software and 2) the laparoscope projecting on a screen above the simulator.

Figure 2. The Simball® joystick, courtesy: Surgical Science AB

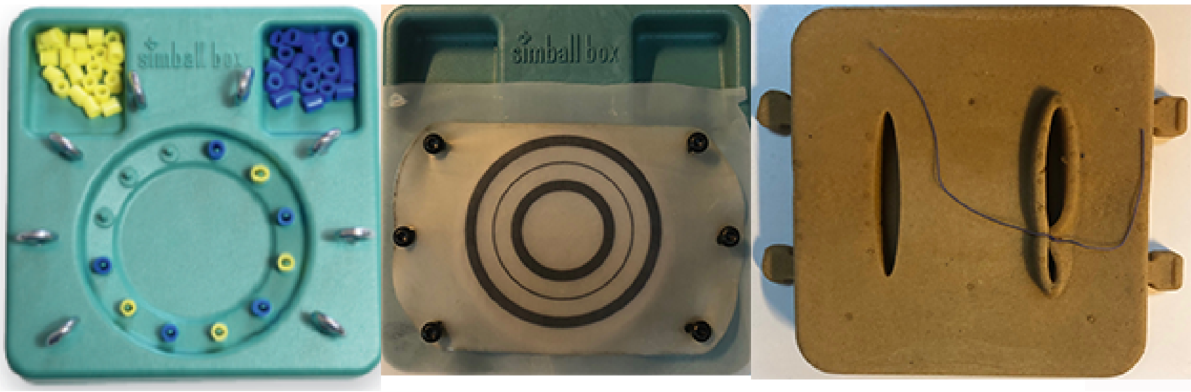

Figure 3. The tasks: Peg picker, precision cutting, suture. The modified rope race is found in figure 1A, courtesy: Surgical Science AB (the peg picker)

Motion Analysis

The acquired motion data was analysed using a motion analysis software, presented in detail in a previous paper [17]. The software was developed in MATLAB 7 (The MathWorks Inc., MA, USA).

All motion-related metrics were derived from the position and the orientation of the instruments as tracked by the Simball® joystick, where the position was defined by: $\mathrm{r}\,(\mathrm{t}) = [\mathrm{x}(\mathrm{t}), \mathrm{y}(\mathrm{t}), \mathrm{z}(\mathrm{t})]_{\mathrm{t}=0}^{\mathrm{T}}$, and the orientation was defined by the three angles $[\alpha(\mathrm{t}), \beta(\mathrm{t}), \gamma(\mathrm{t})]_{\mathrm{t}=0}^{\mathrm{T}}$, where α and β measured the orientation in two planes perpendicular to the instrument's axis and γ measured the rotation around the instrument's axis. The following metrics were used in the evaluation of the performance of the participants:

1. *Time* (T): Time was measured from start to completion of the task.

2. *Bimanual dexterity* (BD) was the participant's ability to control and manipulate two instruments at the same time. BD was found by calculating the correlation between the velocity of the tip of the instruments controlled by the left and the right hand:

$$BD = \frac{\int_0^T (v_{left}(t) - \bar{v}_{left})(v_{right}(t) - \bar{v}_{right})dt}{\sqrt{\int_0^T (v_{left}(t) - \bar{v}_{left})^2\, dt \cdot \int_0^T (v_{right}(t) - \bar{v}_{right})^2\, dt}}$$

   where v is the velocity of the instruments and $\bar{v}$ denotes the average velocity over the duration of the task.

3. *Path length* (PL) was the movement of the tip of the instrument, integrated over the duration of the task, measured in meters:

$$PL = \int_0^T \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\ dt$$

4. *Angular length* (AL) was the change in angle of the tip of the instrument in the plane perpendicular to the instrument's axis, integrated over the duration of the task measured in degrees:

$$AL = \int_0^T \sqrt{\left(\frac{d\alpha}{dt}\right)^2 + \left(\frac{d\beta}{dt}\right)^2}\ dt$$

5. *Depth perception* (DP) was calculated by the total distance traveled by the tip of the instrument in the instrument's axis direction, measured in meters:

$$DP = \int_0^T \left|\frac{dz}{dt}\right|\ dt$$

6. *Response orientation* (RO) was the total amount of instrument rotation around its axis, measured in degrees:

$$RO = \int_0^T \left|\frac{d\gamma}{dt}\right|\ dt$$

7. *Motion smoothness* (MS) was the change in acceleration of the tip of the instrument integrated over the duration of the task and normalized by the duration of the task. The motion smoothness was measured in m/s^3

$$MS = \sqrt{\frac{1}{2T}\int_0^T \left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 + \left(\frac{d^3z}{dt^3}\right)^2\ dt}$$

8. *Number of submovements* (NoS) was the number of times a movement of the tip of the instrument contained a velocity peak of at least 10 mm/s.

9. *Average velocity* (AV) was the average velocity of the tip of the instrument, measured in mm/s.

10. *Idle percentage* (IDLE) was the percentage of time that the instrument was moved at a speed below 2 mm/s.

Metric number 3–10 were measured for both hands separately. In total 18 metrics were calculated.

Statistical Analysis

The data were analysed using SPSS 24 (IBM Corporation, Armonk, USA). The Mann–Whitney U test was used to test for pairwise statistical differences in the distribution's central tendencies ($p < 0.05$) between different levels of laparoscopic experience. The related samples Spearman's rank order correlation was computed to investigate correlation between time and the motion metrics.

Results

Forty-eight participants were enrolled in the study. Ten had performed ten laparoscopic procedures or less (the novices), two of these participants were medical students. Twenty-two had performed between eleven and one hundred procedures (the intermediates), and sixteen had performed more than one hundred procedures (the experts). There were six left-handed and forty-two right-handed, evenly distributed between the three groups. Time and nine different motion parameters, of which eight were measured separately for each hand, were analyzed, i.e. a total of eighteen metrics. The eighteen metrics were compared pairwise between novices and intermediates, novices and experts, and intermediates and experts, i.e. a total of 54 tests using the Mann-Whitney U test. The results are presented in Figure 4-5. No statistical differences between groups were found for the *peg picker* task. Five out of fifty-four tests showed statistical significance for the *rope race*, twenty-seven out of fifty-four for the *precision cutting*, and twenty-two out of fifty-four for the *suture* task. For the *precision cutting* task most statistical significances were found between novices and intermediates and novices and experts, only one statistical significance was found between intermediates and experts (*idle percentage dominant hand*). For the *suture* task, most statistical significances were found between novices and experts and intermediates and experts, only one statistical significance was found between novices and intermediates (*time*) (Figure 4-5).

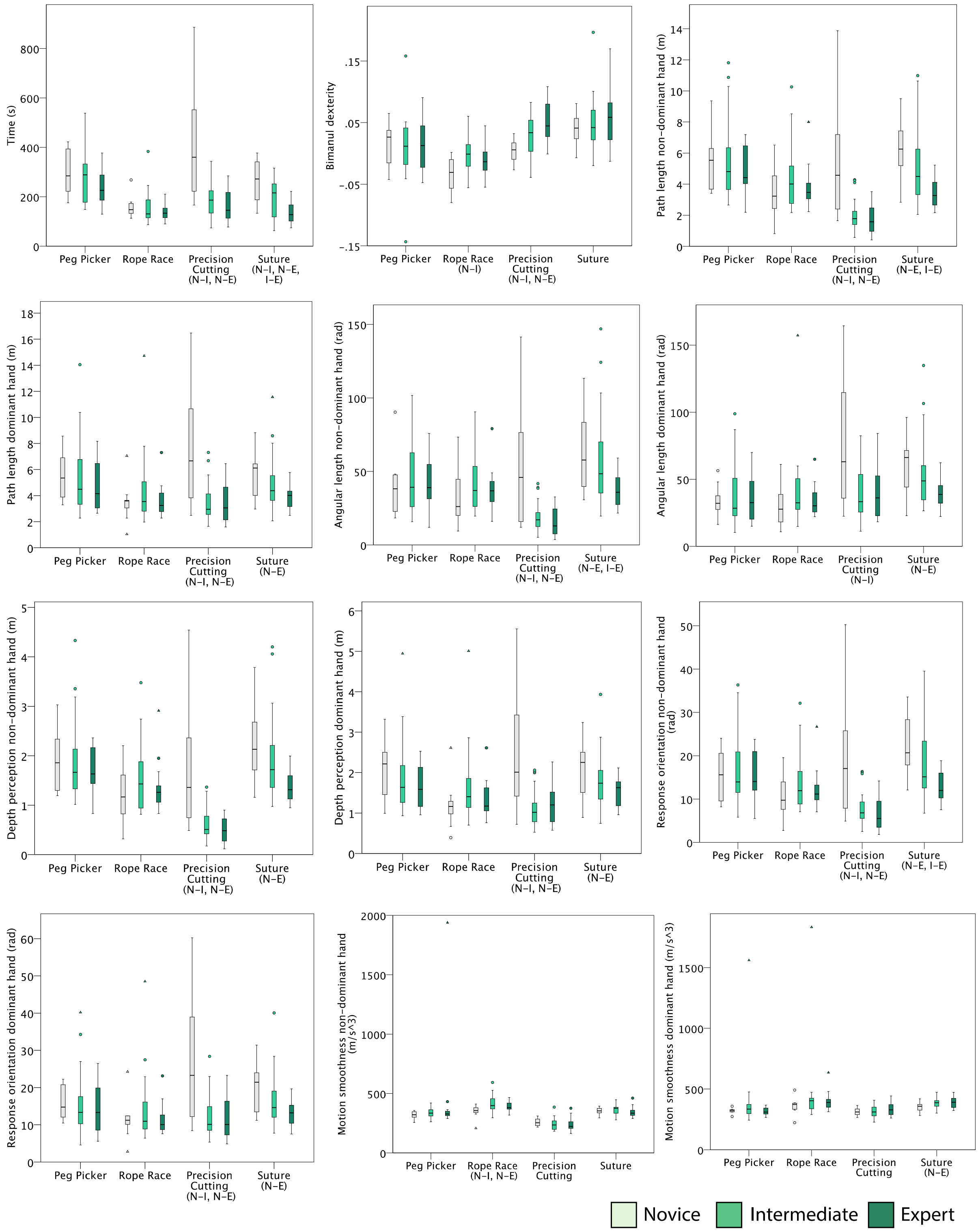

Figure 4. Boxplot presenting time, bimanual dexterity, path length, angular length, depth perception, response orientation and motion smoothness for dominant and non-dominant hand. Statistical significantly differences between novices and intermediates are marked with N-I, between novices and experts with N-E, and between intermediates and experts with I-E. The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as triangles.

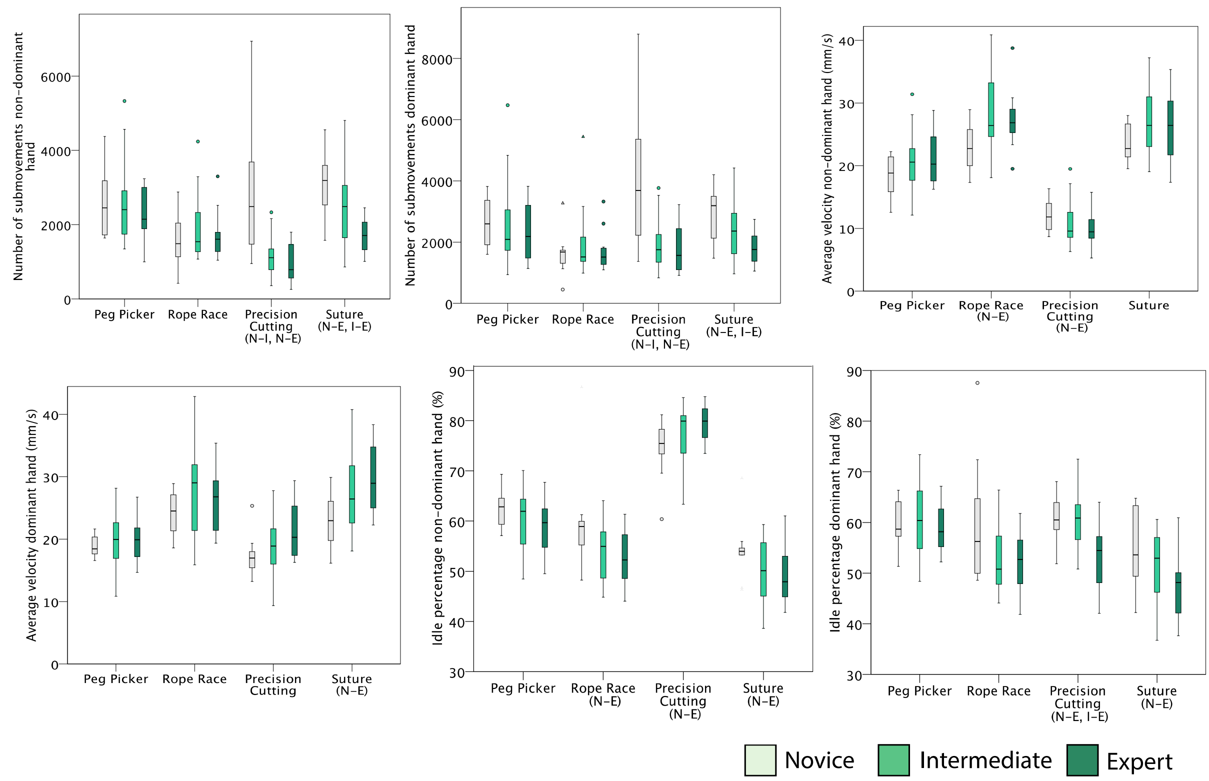

Figure 5. Boxplot presenting number of sub-movements, average velocity and idle percentage for dominant and non-dominant hand. Statistical significantly differences between novices and intermediates are marked with N-I, between novices and experts with N-E, and between intermediates and experts with I-E. The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as triangles.

The related samples Spearman's rank order correlation was computed to investigate correlation between time and the motion metrics for *precision cutting* and *suture*, and are listed in Table 1. A high correlation was found between time and the motion metrics *path length*, *angular length*, *depth perception*, *response orientation* and *number of sub-movements* with Spearman's ρ correlation coefficients above 0,79. *Bimanual dexterity*, *motion smoothness*, *average velocity* and *idle percentage* showed limited correlation with time.

Table 1. *Spearman's ρ correlation coefficients for the motion analysis metrics compared to Time. Statistically significant values marked with bold numbers. Metrics with a high correlation with Time are marked with italics.*

Task	Precision Cutting		Suturing	
Instrument	Grasper (Non-dominant hand)	Scissors (Dominant hand)	Needle holder (Non-dominant hand)	Needle holder (Dominant hand)
Bimanual dexterity	0.27		**0.47**	
Path Length	***0.93***	***0.88***	***0.91***	***0.90***
Angular length	***0.89***	***0.79***	***0.83***	***0.80***
Depth Perception	***0.91***	***0.86***	***0.85***	***0.86***

Response orientation	***0.93***	***0.88***	***0.88***	***0.91***
Motion smoothness	**0.30**	0.13	0.05	**0.51**
Number of sub-movements	***0.95***	***0.93***	**0.97**	***0.97***
Average velocity	**0.42**	0.22	**0.36**	**0.59**
Idle percentage	**0.37**	0.25	**0.47**	**0.64**

Discussion

Simulation-based assessment is gaining in importance, as surgical education evolves into an era of competency-based training. There are many unanswered or partly answered questions related to simulation-based assessment and competency-based education: What is surgical expertise, how can it be measured, and what kind of surgical expertise can simulators measure [7, 17]. The Simball ® box is situated in a segment of low-cost simulators, yet with capabilities of motion tracking through the Simball® joystick [15]. Our research goal was to investigate whether the simulator could be used to measure laparoscopic surgical skills for assessment in competency-based training, i.e. investigating construct validity. A common way of doing so is by comparing the results from participants with different experience levels [18], the so-called known-groups technique [19]. We found that for the *peg picker* task the simulator was not capable of distinguishing between the different experience levels, and based on these results the *peg picker* task is not suited to assess surgical skills. The modified *rope race* showed significant differences for only five out of fifty-four tests, and we found that the task was not suited to assess surgical skills in our competency-based training set-up as it showed only very limited construct validity. Both the *precision cutting* task and the *suture* task showed significant differences for a large number of tests, thus showing construct validity. Both tasks were implemented as part of our competency-based training program. Hagelsteen et al. found similar results on the suture task when they compared results of participants throughout a course [15].

The *precision cutting* task was primarily capable of distinguishing between novices and the two other groups, but not between intermediates and experts (significant differences for 12 out of 18 metrics between novices and intermediates, 14 out of 18 between novices and experts and only 1 out of 18 between intermediates and experts). This indicated that the *precision cutting* task required skills mastered at the intermediate level, and plateaued there, with the result that the skills levels were comparable or not distinguishable after that level. Skills related to laparoscopic adhesiolysis, that are acquired in the clinic at the intermediate level of experience, might be one factor explaining the results. The *suture* task, on the other hand, was primarily capable of distinguishing between experts and the other two groups, but not between novices and intermediates (significant differences for 15 out of 18 metrics between novices and experts, 6 out of 18 between intermediates and only 1 out of 18 for novices and intermediates). This indicates that suture is a skill acquired at the late intermediate or early expert level. Volume clinical exposure to laparoscopic suturing probably correspond with late intermediate or early expert level. These results show that an assessment task might have limited value if applied on surgeons with experience levels outside of the task's proven validity range.

Time is a metric that is easy to measure without the need of much extra equipment, and as indicated in Table 1, it also correlates with five out of nine motion metrics. To measure motion metrics, more expensive equipment, in addition to SW to calculate them, is needed. If most metrics do correlate with time, then what is the added value of the Simball® joystick, the enabler of motion metrics measurements? First, although most of the parameters do correlate, the feedback that can be given to the trainer is far more detailed with motion metrics than without, e.g. helping the trainee identify areas for improvement. Secondly, there were four metrics that did not correlate well with time. These metrics seem to measure aspects that are different from the ability to perform a task fast, and have added value. What that added value is, should be further investigated [15, 20]. The motion smoothness and bimanual dexterity metrics had low correlation with time. We have seen in other studies that the ability to control the non-dominant hand and to coordinate the movements of both hands, measured by the bimanual dexterity metric, are skills acquired at a relatively high level of surgical experience [17, 21]. Both motion smoothness and bimanual dexterity seem to tell us something about the quality of the performance beyond efficiency [17, 21].

The Simball® Box is delivered with a motion analysis SW. We chose to use an in-house SW to analyse the motion data from the Simball® joystick, as our SW also calculated the bimanual dexterity, number of submovements and idle percentage metrics[17]. We verified that the values that we found were equivalent to the ones found by the Simball® Box SW. Our results are thus transferable for the metrics that our SW and the Simball® Box SW calculates.

The term “fidelity” has been much debated in the literature on technology enhanced learning [22], especially with regard to skills transfer. This debate has also value when discussing the use of simulators for skills assessment. Hamstra et al. argues that “*the field of simulation should shift emphasis away from structural properties of the simulator (i.e., physical resemblance) to functional properties of the entire simulation context that align with learning objectives (i.e., functional task alignment)*” [22]. To put it another way: regardless of physical resemblance, if the simulator is capable of measuring surgical skills, i.e. it has functional task alignment, it can be used in competency-based education. Nevertheless, to be able to measure surgical (psychomotor) skills the simulator must be able to recreate some essential aspects of a clinical setting. The Simball ® box comes with a web-cam, but can easily be connected to a laparoscope, as we did in our set-up. The laparoscope has a far better image quality than the web-cams that are usually used in box simulators, and resembles the video quality in a clinical setting. Another essential aspect of a clinical setting might be the use of physical instruments. Regular laparoscopic instruments can be used with the Simball ® box, giving it added physical resemblance compared to box simulators that uses specialized instruments or virtual reality simulators that are dependent on a virtual simulated haptic feedback. Virtual reality simulators either try to simulate haptic feedback, which they have struggled to succeed with [23, 24] or they do not, i.e. the instruments move with constant resistance independently of the existence of a virtual contact with a physical object [24, 25]. The instruments on the Simball® box were inserted in the Simball® joystick and attached in a fixed position resulting in the rotator being immobilized. This was a limitation on the physical resemblance that might have influenced the results. The instruments are firmly fixed in the joystick to enable motion tracking. More expert surgeons than novices complained about not being able to use the rotator. Our impression was, however that the candidates solved the limitation by turning their wrists and hands accordingly, and as the tests were relatively short, i.e. up to 30 minutes, including natural breaks in between the tasks, any potential ergonomic issues were negligible. The space in the simulator where the task was performed, i.e. the room where the instruments could move around, was relatively small compared to e.g. the surgical space during a cholecystectomy. We believe this was a limitation on the physical resemblance that restricted construct validity both for the peg picker task and the modified rope race. With the *peg picker* task, another important aspect was that the pegs, regular midi Hama beads, were hard and small, and therefore, unsuited for manipulation by the laparoscopic graspers. The result was that the *peg picker* task and the modified *rope race* did not, or only to a limited extent, require surgical skills that the experts or intermediates mastered, and where they could have exceled above the novices, resulting in no construct validity. In a previous study, we found that a similar task to the *rope race* exhibited construct validity. That simulator had a much larger space for instrument movement [17]. We therefore modified the original *rope race,* before the study, so that there would be more space for the instruments to move, but it might seem that it was not enough. The texture of the *precision cutting* tissue and the *suture* pad both got remarks from the surgeons regarding their natural touch and feel. These two examples of good physical resemblance might have influenced the functional task alignment. It can be difficult to point out exactly which aspects of the simulator and the tasks make them succeed or fail at construct validity tests. Physical resemblance is an indicator which can be leading or misleading, whereas functional task alignment is a measure of a specific capability, in this case the capability of distinguishing between levels of surgical experience or surgical skills.

Competency-based education is alluring by giving the impression that candidates can be educated to predefined and measurable levels of competency. In practice, the picture is much more complex [2, 3, 26]. This study addresses only a limited part of what would be expected in surgical competency-based education. One aspect that was not measured by our test was level of automaticity and thus attentional capacity, as understood by Fitts and Posner's three steps motor learning theory [27], a limitation which is common in psychomotor skills tests on simulators. We chose three different experience levels, with above 100 procedures as the expert level. It is reasonable to believe that surgical skills continue to increase also beyond 100 procedures. The distinction was well suited for the precision cutting task as skills levels seemed to plateau in the intermediate level. For the suture task, it could have been interesting to compare surgeons with higher experience levels to find out whether the suture task also show a similar plateau of skills.

In summary, we found that two out of four tasks, i.e. the *precision cutting* and the *suture* task, tested with the Simball box ® showed construct validity. These two tests can be implemented and used in a competency-based set-up to measure surgical skills. In addition, we found that the assessment tasks were more suitable at specific experience levels than others. It is therefore important to select the appropriate assessment task for a given participant's experience level.

Acknowledgements

We thank all those who participated in the study, in addition to Kirsten Rønning for help with performing the tests. The work was supported by a grant from the Central Norway Regional Health Authority, SINTEF, the Norwegian University of Science and Technology (NTNU), and the Norwegian National Advisory Unit for Advanced Laparoscopic Surgery at St. Olavs Hospital (all Trondheim, Norway).

Declaration of interest

Cecilie Våpenstad, Erlend Fagertun Hofstad, Tor Eivind Bernstein, Petter Aadahl, Gjermund Johnsen and Ronald Mårvik have no conflict of interest or financial ties to disclose.

References

1. Zendejas B, Brydges R, Hamstra SJ, Cook DA State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg* 2013; 257 4): 586-593. https://doi.org/10.1097/SLA.0b013e318288c40b
2. Sonnadara RR, Mui C, McQueen S, Mironova P, Nousiainen M, Safir O, et al. Reflections on competency-based education and training for surgical residents. *J Surg Educ* 2014; 71 1): 151-158. https://doi.org/10.1016/j.jsurg.2013.06.020
3. Fry H, Kneebone R, *Surgical Education: Theorising an emerging domain*. Vol. 2. 2011: Springer Science & Business Media. 268.
4. Scott DJ, Cendan JC, Pugh CM, Minter RM, Dunnington GL, Kozar RA The changing face of surgical education: Simulation as the new paradigm. *Journal of Surgical Research* 2008; 147 2): 189-193. https://doi.org/10.1016/J.Jss.2008.02.014
5. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR Virtual reality training for surgical trainees in laparoscopic surgery. *The Cochrane database of systematic reviews* 2009; 1): CD006575. https://doi.org/10.1002/14651858.CD006575.pub2
6. Gurusamy KS, Nagendran M, Toon CD, Davidson BR Laparoscopic surgical box model training for surgical trainees with limited prior laparoscopic experience. *The Cochrane database of systematic reviews* 2014; 3 CD010478. https://doi.org/10.1002/14651858.CD010478.pub2
7. Stefanidis D, Sevdalis N, Paige J, Zevin B, Aggarwal R, Grantcharov T, et al. Simulation in surgery: what's needed next? *Ann Surg* 2015; 261 5): 846-853. https://doi.org/10.1097/SLA.0000000000000826
8. Våpenstad C, Buzink SN Procedural virtual reality simulation in minimally invasive surgery. *Surg Endosc* 2013; 27 2): 364-377. https://doi.org/10.1007/s00464-012-2503-1
9. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005; 241 2): 364-372. https://doi.org/10.1097/01.sla.0000151982.85062.80
10. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 2005; 19 12): 1523-1532. https://doi.org/10.1007/s00464-005-0384-2
11. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, et al. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. *Ann Surg* 2011; 253 6): 1216-1222. https://doi.org/10.1097/SLA.0b013e3182197016

12. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A Laparoscopic virtual reality and box trainers: is one superior to the other? *Surg Endosc* 2004; 18 3): 485-494. https://doi.org/10.1007/s00464-003-9043-7
13. Li MM, George J A systematic review of low-cost laparoscopic simulators. *Surg Endosc* 2017; 31 1): 38-48. https://doi.org/10.1007/s00464-016-4953-3
14. Hagelsteen K, Langegard A, Lantz A, Ekelund M, Anderberg M, Bergenfelz A Faster acquisition of laparoscopic skills in virtual reality with haptic feedback and 3D vision. *Minim Invasive Ther Allied Technol* 2017; 1-8. https://doi.org/10.1080/13645706.2017.1305970
15. Hagelsteen K, Sevonius D, Bergenfelz A, Ekelund M Simball Box for Laparoscopic Training With Advanced 4D Motion Analysis of Skills. *Surg Innov* 2016; 23 3): 309-316. https://doi.org/10.1177/1553350616628678
16. Willaert W, Aggarwal R, Bicknell C, Hamady M, Darzi A, Vermassen F, et al. Patient-specific simulation in carotid artery stenting. *J Vasc Surg* 2010; 52 6): 1700-1705. https://doi.org/10.1016/j.jvs.2010.08.015
17. Hofstad EF, Vapenstad C, Chmarra MK, Lango T, Kuhry E, Marvik R A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. *Surg Endosc* 2013; 27 3): 854-863. https://doi.org/10.1007/s00464-012-2524-9
18. Fairhurst K, Strickland A, Maddern G The LapSim virtual reality simulator: promising but not yet proven. *Surg Endosc* 2011; 25 2): 343-355. https://doi.org/10.1007/s00464-010-1181-0
19. APA AN, *The Standards for Educational and Psychological Testing*. 2014, Joint Committee on Standards for Educational and Psychological Testing:
American Psychological Association, American Educational Research Association and National Council on Measurement in Education: Washington, USA.
20. Hiemstra E, Chmarra MK, Dankelman J, Jansen FW Intracorporeal suturing: economy of instrument movements using a box trainer model. *J Minim Invas Gyn* 2011; 18 4): 494-499. https://doi.org/10.1016/j.jmig.2011.04.003
21. Hofstad EF, Vapenstad C, Bo LE, Lango T, Kuhry E, Marvik R Psychomotor skills assessment by motion analysis in minimally invasive surgery on an animal organ. *Minim Invasive Ther Allied Technol* 2017; 26 4): 240-248. https://doi.org/10.1080/13645706.2017.1284131
22. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA Reconsidering fidelity in simulation-based training. *Acad Med* 2014; 89 3): 387-392. https://doi.org/10.1097/ACM.0000000000000130
23. Coles TR, Meglan D, John NW The Role of Haptics in Medical Training Simulators: A Survey of the State of the Art. *Ieee T Haptics* 2011; 4 1): 51-66. https://doi.org/10.1109/Toh.2010.19
24. Overtoom EM, Horeman T, Jansen FW, Dankelman J, Schreuder HWR Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview. *J Surg Educ* 2018; https://doi.org/10.1016/j.jsurg.2018.06.008
25. Våpenstad C, Hofstad EF, Lango T, Marvik R, Chmarra MK Perceiving haptic feedback in virtual reality simulators. *Surgical Endoscopy* 2013; 27 7): 2391-2397. https://doi.org/10.1007/s00464-012-2745-y
26. Prentice R, *Bodies in formation: An ethnography of anatomy and surgery education*. 2012: Duke University Press.
27. Sadideen H, Kneebone R Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? *Am J Surg* 2012; 204 3): 396-401. https://doi.org/10.1016/j.amjsurg.2011.12.020

Legends of figures and tables

Figure 1. (A) The surgical space with the modified rope race. (B) The Simball® simulator with two cameras: 1) the webcam connected to the laptop showing the simulator software and 2) the laparoscope projecting on a screen above the simulator.

Figure 2. The Simball® joystick, courtesy: Surgical Science AB

Figure 6. The tasks: Peg picker, precision cutting, suture. The modified rope race is found in figure 1A, courtesy: Surgical Science AB (the peg picker)

Figure 4. Boxplot presenting time, bimanual dexterity, path length, angular length, depth perception, response orientation and motion smoothness for dominant and non-dominant hand. Statistical significantly differences between novices and intermediates are marked with N-I, between novices and experts with N-E, and between intermediates and experts with I-E. The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as triangles.

Figure 5. Boxplot presenting number of sub-movements, average velocity and idle percentage for dominant and non-dominant hand. Statistical significantly differences between novices and intermediates are marked with N-I, between novices and experts with N-E, and between intermediates and experts with I-E. The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as triangles.

Table 1. Spearman's ρ correlation coefficients for the motion analysis metrics compared to Time. Statistically significant correlations marked with bold numbers.