journal of surgical research xxx (2013) 1–9

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Development and validation of a composite scoring system for robot-assisted surgical training—the Robotic Skills Assessment Score

Ashirwad J. Chowriappa, PhD,[a,b] Yi Shi, MS,[b] Syed Johar Raza, MD,[a] Kamran Ahmed, PhD,[c] Andrew Stegemann, BS,[a] Gregory Wilding, PhD,[b] Jihad Kaouk, MD,[d] James O. Peabody, MD,[e] Mani Menon, MD,[e] James M. Hassett, MD,[b] Thenkurussi Kesavadas, PhD,[b] and Khurshid A. Guru, MD[a,*]

[a] Roswell Park Cancer Institute, Buffalo, New York
[b] University at Buffalo, Buffalo, New York
[c] MRC Centre for Transplantation, King's College London, Guy's Hospital, London, United Kingdom
[d] Cleveland Clinic Foundation, Cleveland, Ohio
[e] Henry Ford Health System, Detroit, Michigan

ARTICLE INFO

Article history:
Received 1 May 2013
Received in revised form
10 June 2013
Accepted 24 June 2013
Available online xxx

Keywords:
Scoring system
Surgical training
Robotic skills assessment
Score
Robotic training

ABSTRACT

Background: A standardized scoring system does not exist in virtual reality-based assessment metrics to describe safe and crucial surgical skills in robot-assisted surgery. This study aims to develop an assessment score along with its construct validation.

Materials and methods: All subjects performed key tasks on previously validated Fundamental Skills of Robotic Surgery curriculum, which were recorded, and metrics were stored. After an expert consensus for the purpose of content validation (Delphi), critical safety determining procedural steps were identified from the Fundamental Skills of Robotic Surgery curriculum and a hierarchical task decomposition of multiple parameters using a variety of metrics was used to develop Robotic Skills Assessment Score (RSA-Score). Robotic Skills Assessment mainly focuses on safety in operative field, critical error, economy, bimanual dexterity, and time. Following, the RSA-Score was further evaluated for construct validation and feasibility. Spearman correlation tests performed between tasks using the RSA-Scores indicate no cross correlation. Wilcoxon rank sum tests were performed between the two groups.

Results: The proposed RSA-Score was evaluated on non-robotic surgeons ($n = 15$) and on expert-robotic surgeons ($n = 12$). The expert group demonstrated significantly better performance on all four tasks in comparison to the novice group. Validation of the RSA-Score in this study was carried out on the Robotic Surgical Simulator.

Conclusion: The RSA-Score is a valid scoring system that could be incorporated in any virtual reality-based surgical simulator to achieve standardized assessment of fundamental surgical tents during robot-assisted surgery.

* *Corresponding author.* Applied Technology Laboratory for Advanced Surgery (A.T.L.A.S.) Project, Department of Urology, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263. Tel.: +1 716 845 7612; fax: +1 716 845 3300.
E-mail address: khurshid.guru@roswellpark.org (K.A. Guru).

0022-4804/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2013.06.054

1. Introduction

Over the recent years, several robotic surgical simulation platforms have emerged as a useful adjunct to training in and outside the operating room [1]. Studies have shown that simulation-based training (SBT) can increase the safety profile for beginners by bridging the gap between safe acquisition of surgical skills and effective operative performance [2,3]. Moreover, the need to minimize surgical errors and the ethics of training on patients makes SBT platforms timely and critical [4,5]. Based on a recent survey of 297 urologic surgeons, 78% felt it was required or beneficial to have robot-assisted surgical training (RAST). Eighty-three percent would consider a robot-assisted fellowship, indicating a significant shift in favor of RAST [6].

Commercially available simulators such as the Robotic Surgical Simulator (RoSS) [7], da Vinci Skills Simulator, and MIMIC da Vinci Trainer have previously been validated for realism (face validity) and content and construct validity [8]. These simulators provide a safe training environment for basic robotic skills while providing objective feedback on the trainee's performance. However, the built-in metrics for evaluation of technical skills within these simulators are not consistent.

In SBT, the ability to assess trainee performance accurately is critical and requires metrics (score) to benchmark training outcomes. Such metrics used by the first generation of simulators [9] were based on criterions such as time to completion or number of errors committed. Mackay *et al.* [10] found that the process of assessing technical abilities is more robust if candidates are tested on multiple parameters using a variety of measures. van Dongen *et al.* [11] concluded that the implementation of a standardized scoring system could enable the assessment metrics to assess further aspects of performance.

To our knowledge, no standardized scoring system exists to evaluate, validate, and integrate virtual reality-based metrics during RAST. Such a standardized evaluation tool for technical skill assessment in SBT is likely to involve determining critical procedural steps, following an expert consensus (Delphi), and establishing performance scores and baseline assessment tools targeted to certify competency in the desired skills [12,13].

To standardize assessment of technical skills in SBT platforms, we developed a Robotic Skills Assessment Score (RSA-Score) that can be incorporated into any virtual or augmented reality-based surgical simulator. This study aims at development and validation of this novel skill assessment scoring system.

2. Materials and methods

This observational comparative study was carried out at Roswell Park Cancer Institute, Buffalo, New York, after institutional review board approval. The RSA-Score was developed using a validated Fundamental Skills of Robotic Surgery (FSRS) curriculum [14]. This section has been explained under the following subsection headings: (1) identification of components, (2) development of RSA-Score, (3) validation of the RSA-Score, and (4) statistical analysis (Fig. 1).

2.1. *Identification of technical skills and critical components from FSRS curriculum*

A consensus was achieved (Delphi methodology) using an expert robotic surgery panel (surgeon experience of over 1000 procedures with >50% console time). Components that are critical to safety of technical skills were identified from the FSRS curriculum [14].

The FSRS curriculum is a sequential, modular criterion-based structured curriculum for acquiring basic skills of robot-assisted surgery [14]. This curriculum was incorporated into a previously validated virtual reality simulator, RoSS [15]. The curriculum contains four modules (Orientation, Motor Skills, Basic Surgical Skills, and Intermediate Surgical Skills) with a total of 16 tasks. Metrics related to performance on the RoSS were measured, recorded, and archived by the software as seen with high fidelity virtual reality simulators. Based on these measurements, the RSA-Score methodology used in this manuscript was applied to obtain a final score based on the weightage system (i.e., safety in operative field, economy, and bimanual dexterity).

2.2. *Development of the RSA-Score*

2.2.1. *Hierarchical metric decomposition*

Critical safety determining procedural steps were identified and a hierarchical task decomposition of multiple parameters using a variety of metrics was used to develop the RSA multimetric scoring system (Table 1). Following an expert consensus (Delphi), the expert robotic surgery panel was responsible for defining, weighting, integration, and configuration of the metrics into the hierarchical scoring system (Tables 1 and 2). Ten relevant metrics used on the FSRS curriculum were used to assess technical skills. These metrics were then integrated based on their definition into the following essential parameters: safety in operative field, economy of motion, bimanual dexterity, and critical error (Table 1).

2.2.2. *Safety in operative field*

Awareness of operative environment in which tasks are performed. Left Tool Out Of View, Right Tool Out Of View, Tool–Tool Collision, and Tissue Damage are the metrics used.

2.2.3. *Economy of motion*

Indicates effectiveness and the level of skill in which the task is performed. Distance By Camera, Number Of Errors, and Clutch Usage are the metrics used.

2.2.4. *Bimanual dexterity*

Indicates how an expert uses both hands in a complimentary way to provide best exposure in performing the task. Distance By Left Hand and Distance By Right Hand are the metrics used.

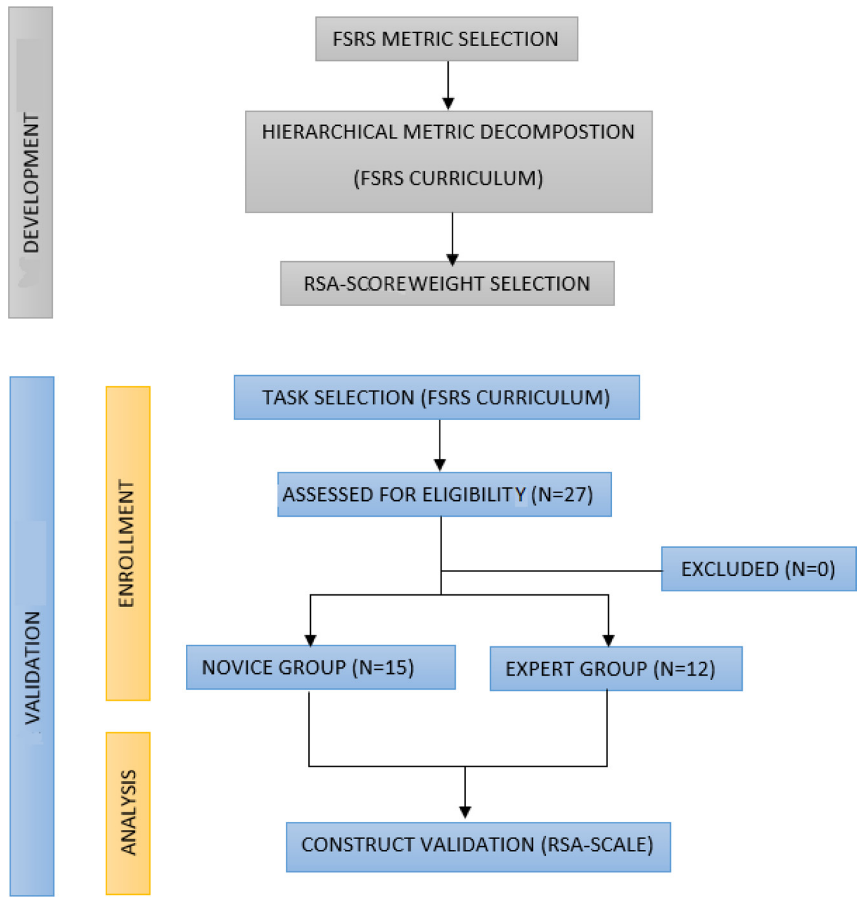


Fig. 1 – Development and validation of the RSA-Score.

2.2.5. *Critical error*

Errors specific to performance of the task, for example, in the task Ball Placement, if the ball drops then it would constitute an error.

2.2.6. *RSA-Score methodology*

The RSA scoring schema was established using the FSRS metrics. Not all metrics contribute equally, it is important to note that some metrics although meaningful need not affect the quality of the task performed, for example. It is less desirable to attain quicker task completion time, but what is more critical to performing it is safety. In the proposed RSA-Score, weighting is appropriately set such that critical composite metrics that enhance technical skills and patient safety are given more weighting (Table 2). We obtain baseline performance scores for each of the FSRS tasks on the RoSS, established by expert robotic surgeons. Expert scores on each of the FSRS metrics ($i = 10$) are then normalized $|S_{iE}|$ and rescaled. According to the new RSA-Score, the maximum and minimum scores of each metric ranged from 1 to 5 (5, best and 1, worst), and its weight ω_i (contribution) determined by an expert consensus (Delphi) are given in Table 2. Metrics such as time and error (where higher scores express lower performance) were normalized and set to $1 - |S_{iE}|$. The mean metric value $\overline{S}_{iE}$ for the experts was used to establish baseline performance scores. Upper and lower limits that indicate expert performance were set as $\overline{S}_{iE} \pm 1SD$. A final composite score $\widehat{S}$ (overall score) was obtained as the weighted sum of individual composite scores.

$$\widehat{S} = \sum_i \omega_i \times S_{iE}$$

2.3. *Validation of the RSA-Score*

The construct validation of the RSA-Score was performed as a multi-institutional study with participants enrolled from the following institutions: Roswell Park Cancer Institute (Buffalo, NY), Cleveland Clinic Foundation (Cleveland, OH), and Henry Ford Health System (Detroit, MI) in the United States. The participants were divided into an Expert Group (EG) and Novice Group (NG), depending on their level of robotic surgical experience. EG included participants with >150 procedures of cumulative robotic console experience. This group participated in the validation study as the expert arm and did not participate in the development of the score. The NG participants had no prior robotics or simulator experience. All participants completed a basic demographic questionnaire before participating as EG or NG. Participants were then given a short didactic session to introduce them to the features of the RoSS. This was followed by a 15-min period of acclimation on the simulator console during which the participants could

Table 1 – FSRS metrics, description of metrics, and hierarchical metric integration.

S/no	FSRS metrics	Description
1	Task time	Time in seconds to complete the task
2	Distance By Left Hand	The total distance in millimeters moved by the left instrument
3	Distance By Right Hand	The total distance in millimeters moved by the right instrument
4	Left Tool Out Of View	The number of instances in which the left instrument was out of camera view
5	Right Tool Out Of View	The number of instances in which the right instrument was out of camera view
6	Tool–Tool Collision	Number of collisions caused by the instrument-stem, wrist and jaws colliding with each other
7	Tissue Damage	Damage caused to the surrounding tissue by collisions with instruments
8	Distance By Camera	The total distance in millimeters moved by the virtual camera
9	Number Of Critical Errors	Number of task-specific errors
10	Clutch Usage	Number of times clutch is used

FSRS metrics	Composite metrics	Skills	
Time to complete task	TT	Decision making	
Left Tool Out Of View	SOF		
Right Tool Out Of View	SOF	Situation awareness	Cognitive ability
Tool–Tool Collision	SOF		
Tissue Damage	SOF		
Distance By Camera	EC		
Clutch Usage	EC		
Distance By Left Hand	BD	Surgical skills	Technical assessment
Distance By Right Hand	SOF		
Task-Specific Error	ER		

BD = bimanual deterity; EC = Economy; ER = critical errors; SOF = safety in operative field; TT = task time.

ask questions. The participants then completed the selected tasks, namely the Fourth Arm Manipulation, Coordinated Tool Control, Ball Placement, and Needle Handling and Exchange tasks. Validation of RSA-Score in this study was carried out on the RoSS.

2.4. *Statistical analysis*

Measured outcome variables were summarized overall and by relevant demographic and baseline variables. Descriptive statistics such as frequencies and relative frequencies were computed for all categorical variables. Numeric variables were summarized using simple descriptive statistics such as the mean, standard deviation, and range. The Wilcoxon rank sum tests were used to statistically compare outcomes across independent groups (Expert *versus* Novice). Spearman correlation tests were used to statistically assess the dependence between tasks. All tests were two sided and tested at a 0.05 nominal significance level.

Table 2 – RSA-Score, weighted composite metric schema.

Composite metrics*	Score range	Mean score	Weight
TT	1–5	3.5	0.1 (10%)
SOF	1–5	3.5	0.35 (35%)
EC	1–5	3.5	0.1 (10%)
BD	1–5	3.5	0.1 (10%)
ER	1–5	3.5	0.35 (35%)
Overall score (S)*	1–5	3.5	∑1 (100%)
Score ranges	Retake (<2.0), poor (2.0–2.5), average (2.5–3.0), good (3.0–3.5), expert (3.5–4.0), superior (>4.0)		

BD = bimanual dexterity; EC = Economy; ER = critical errors; SOF = safety in operative field; TT = task time.
* Overall score (S) = (0.1 × TT) + (0.35 × SOF) + (0.1 × EC) + (0.1 × BD) + (0.35 × ER).

3. Results

3.1. *Demographics*

A cohort of 27 surgeons participated in this prospective study (Table 3). Participants were classified as novice (0 robot-assisted cases) and experts (>150 robot-assisted surgical procedures). Novice training levels included fellows, medical students, and residents. The NG had no previous robotic console or simulator experience. Seven had <5 y of post-graduate training (PGY) and eight with >5 y of PGY. One novice was left-hand dominant, whereas the remaining 14 had right-hand dominance. The EG ($n = 12$) consisted of surgeons with >150 prior robotic cases experience. All the participants in the EG were male and right-handed dominant. Five of the experts had >10 y of experience with RAS. The remaining seven experts had <10 y prior experience on RAS.

3.2. *Relevance of selected tasks*

Subsets of tasks from the FSRS curriculum were selected to develop the RSA-Score. Four tasks were used to assess technical skills performance using the RSA-Score on the

Table 3 – Demographics of expert (n = 12) and novice (n = 15) groups.

Participant characteristics	Options	Novice ($n = 15$)	Expert ($n = 12$)
A. Demographics			
Age (y)	<40	15	7
	>40		5
Gender	Female	4	
	Male	11	12
Dominant hand	Left	1	
	Right	14	12
B. Education			
Level of training	Faculty/practicing surgeon		12
	Fellow/medical student/resident	15	
PGY level	<5	7	
	>5	8	
	N/A		12
Duration of practice (y)	<10		7
	>10		5
	N/A	15	
C. da Vinci robotic experience			
Prior robotic console experience	No	15	
	Yes		12

N/A = not applicable.

curriculum: Ball Placement, Coordinated Tool Control, Fourth Arm Manipulation, and Needle Handling and Exchange (Fig. 2). Relevance of the selected tasks was demonstrated through task independence. Spearman correlation tests were performed between tasks using the RSA-Score; tests indicated no cross-correlation (Table 4).

Following this expert consensus (Delphi), the above-mentioned critical procedural steps were identified and hierarchical task integration of multiple parameters using a variety of metrics was used to develop the composite metric scoring system. The RSA-Score focuses on time, safety in operative field, economy, bimanual dexterity, and critical error (Table 1).

3.3. *Construct validation of the RSA-Score*

Participant performance was assessed based on the RSA-Scores attained for each of the four tasks (fourth arm control, coordinated tool control, ball placement, and needle handling and exchange). Construct validity was analyzed in terms of task performance using the RSA-Score (shown in Table 5) as follows.

Table 4 – Spearman correlation coefficients between task cross-correlation.

Task	FAC	CTC	BP	NHE
FAC	1.0	0.0069	−0.2587	0.3076
		0.9828	0.4168	0.3306
CTC	0.0069	1.0	0.0349	0.5314
	0.9828		0.9141	0.0754
BP	−0.2587	0.0349	1.0	−0.2307
	0.4168	0.9141		0.4705
NHE	0.3076	0.5314	−0.2307	1.0
	0.3306	0.0754	0.4705	

BP = Ball Placement; CTC = Coordinated Tool Control; FAC = Fourth Arm Control; NHE = Needle Handling and Exchange.
Shows all four FSRS tasks equally relevant.

3.4. *Fourth arm control task*

EG completed the task quicker (3.8 *versus* 3.3; $P = 0.026$) while demonstrating better bimanual dexterity in instrument utilization (3.6 *versus* 2.4; $P = 0.002$). In context of safety in operative field, EG performance was significantly better than NG (3.5 *versus* 3.1; $P = 0.043$). In terms of economy of motion (3.7 *versus* 2.9; $P = 0.032$), experts used the camera and clutch more while performing the task. The experts were more precise (minimum error) ($P = 0.051$). The overall task performance of the EG was reflected in the overall score and showed significantly better RSA-Score ($P = 0.02$) in comparison to NG.

3.5. *Coordinated tool control task*

Instrument safety in operative field was significantly better in the EG in terms of RSA-Score (3.6 *versus* 2.5; $P < 0.001$), the EG moved the instruments out of view less often and had fewer instrument–instrument collisions. The experts experienced significantly less tissue damage. Within this task, EG and NG exhibited similar results with respect to economy of motion. RSA-Scores for bimanual dexterity and errors were significantly higher ($P < 0.001$). The final score for the task showed a significantly greater RSA overall score in the EG.

3.6. *Ball placement task*

Experts demonstrated superior economy of motion using the clutch and the camera more frequently (3.5 *versus* 2.1;

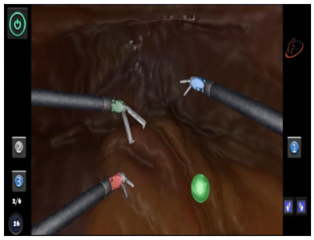
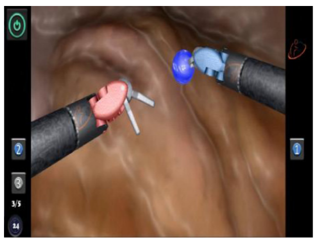
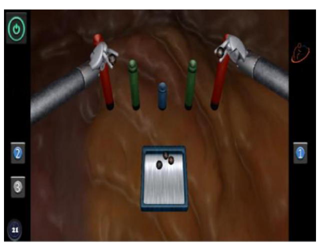
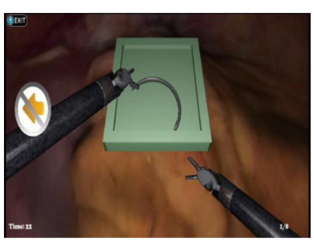
Task 1: Fourth arm control 	**Skills tested:** The DVSS™ provides the convenience of using a third tool for retraction purposes, which can be used as a substitute for a table side surgical assistant. The fourth tool is activated using the pedals on the console and it is essential that the surgeon be able to use the pedals to comfortably alternate between the arms as the situation demands. **Task description: The trainee is** presented with a virtual pelvis containing a red, blue or green sphere. The trainee is required to navigate the instrument accurately and grasp the spheres. **Evaluation metrics:** The time taken to complete the task, number of times tool-tool collision occurs, and the accuracy in performing the task are the main metrics that evaluate the competence of the trainee to use the arm in a coordinated fashion. **Task specific error:** Accuracy in locating and grasping the ball.
Task 2: Coordinated tool control 	**Skills tested:** As a prerequisite to working efficiently on the robotic console, the surgeon is expected to maneuver arms to a comfortable position with the use of the clutch. This task helps expand the workspace of master console and is key to comfortably performing surgical procedures. **Task description:** The trainee is presented with a virtual pelvis containing red and blue spheres. He/she is required to navigate the camera moving either instrument until the red or blue sphere is centered and zoomed in close. The needle driver is navigated accurately and positioned within the red sphere with its jaws open. The instrument is pinched until the sphere completely disappears. **Evaluation metrics:** The time taken to complete the task, number of times tool-tool collision occurs, and the accuracy in performing the task are the main metrics that evaluate the competence of the trainee to use the arm in a coordinated fashion. **Task specific error:** Accuracy in locating and grasping the ball.
Task 3: Ball placement 	**Skills tested:** Fine motor skills with precision control is required to be an effective robotic surgeon. This task is designed to challenge the motor skills of the trainee, allowing them to be honed to perfection over time and will be able to translate them into performing safe error free surgery. **Task description:** The trainee is presented with five balls and five columns. He/she is required to pick up the balls from the tray using your instruments and place them on top of the columns. The task is complete when there is a ball placed on top of every column. **Evaluation metrics:** The time taken to complete the task, number of times tool-tool collision occurs, the number of times the camera, and clutch pedal are pressed, are the main metrics that evaluate the competence of the trainee for this task. **Task specific error:** Accuracy in grasping and placing the ball on the column.
Task 4: Needle handling and exchange 	**Skills tested:** Suture and needle handling is a key part of performing any form of surgery. Incorrect handling can lead to lost sutures, inability to accurately anastomose, tearing of tissue and breaking of needle. To avoid this, these tasks try to reinforce certain safe needle handling practices. **Description: The trainee is** presented with a virtual needle and a tray. The trainee is required to use his/her instruments to pick up the needle transfer it to the other instrument and place it into the tray. **Evaluation:** The time taken to complete the task, accuracy in grasping, damage to needle, number of times tool-tool collision occurs, and the number of times the camera, and clutch pedal are pressed, are the main metrics that evaluate the competence of the trainee for this task. **Task specific error:** Accuracy in grasping the needle and placing the needle into the tray.

Fig. 2 – FSRS task description.

$P < 0.001$). In regard to safety in operative field, significance was observed in the EG ($P < 0.01$). No significance was noted in task time. Precision in completing the task (in terms of errors) by the EG was significantly higher than the NG ($P < 0.007$). The overall RSA-Scores of the experts were significantly higher than the novices ($P < 0.001$).

3.7. *Needle handling and exchange task*

The EG completed the task in shorter time ($P < 0.001$) and demonstrated better economy of motion ($P < 0.001$). The experts also performed the task with more precision (lesser error) ($P = 0.056$). Additionally, both safety and dexterity of the EG was significantly better. The overall task score showed significantly better performance in the expert group ($P < 0.001$).

3.8. *Overall scores*

The RSA-Scores were tabulated for each of the four tasks and mean values were established for comparison between the EG and NG. Overall RSA-Scores reported significantly better performance by the EG (Table 5). Box plots for the RSA overall scores between EG and NG for the four FSRS tasks clearly demonstrated that the experts outperformed the novices (Fig. 3).

4. Discussion

Recent demands for structured and effective training can be addressed by development of training curricula that implement standardized assessment tools for evaluation of surgical

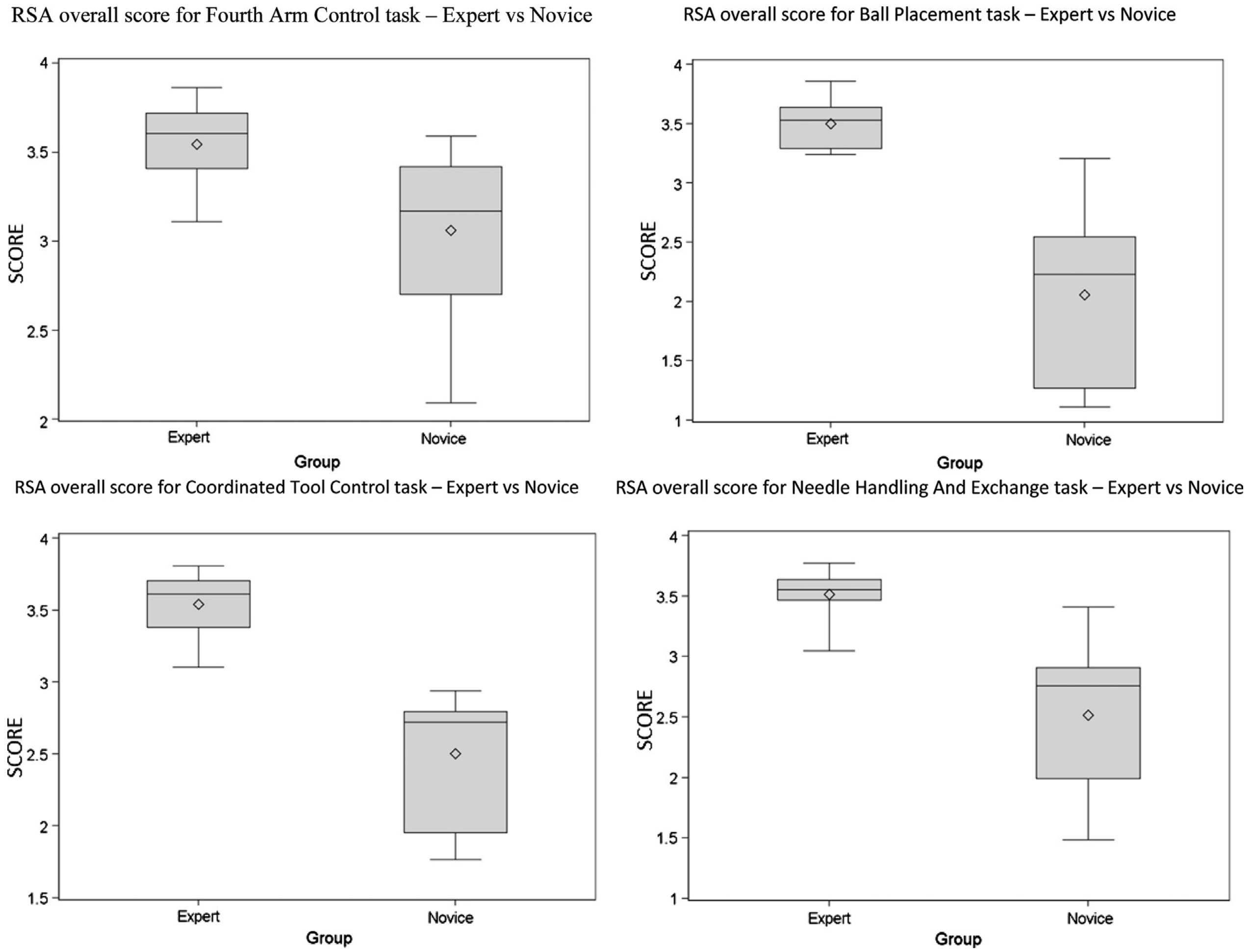


Fig. 3 – Box plots for RSA overall score for the four FSRS tasks. RSA overall score for Fourth Arm Control task—expert *versus* novice. RSA overall score for Coordinated Tool Control task—expert *versus* novice. RSA overall score for Ball Placement task—expert *versus* novice. RSA overall score for Needle Handling And Exchange task—expert *versus* novice.

skills. Our study aimed at developing and validating a standardized scoring system for RAST, using the metrics measured by a virtual reality simulator.

The literature shows a wide range of scoring systems that express common clinical skills such as economy of movements [16] or precision [11]. Although it is useful to know whether the different metrics show construct validity, the final goal is to judge the performance and assess the development of critical surgical tenets in SBT. However, to our knowledge, no scoring system exists to evaluate, validate, and integrate SBT metrics into a scoring system to assess comprehensive surgical performance at robot-assisted surgery. This article serves as a guide to standardize incorporation of skill components into any virtual or augmented reality-based simulator. However, in our study, we used RoSS to validate the methodology of the RSA-Score.

Optimal use of the camera and robotic instruments plays a critical role in assuming control of the robotic surgical console. Metrics that are able to identify skills such as Safety and Economy also impact human–machine interface learning and finally help in acquisition of safe surgical skills on the da Vinci Surgical System. The RSA-Score is not only able to demonstrate construct validation in the overall weighted score of the novices and experts but also identify skills that need improvement. This can be seen in lack of proficiency of the novices in skills such as Economy of Motion (demonstrating significantly greater camera and clutch usage) in the Fourth Arm Control task and Safety in Operative Field (fewer instrument–instrument collisions and loss of instruments) in the Ball Placement task.

Our composite metric methodology not only assesses technical skills but also evaluates cognitive assessment (Table 1). Because the fundamental technical skill tasks remain the same, RSA-Score could possibly be used across a range of simulators for RAS. Furthermore, the RSA-Score may improve trainee results by identifying areas in need of improvement (such as Dexterity, Economy, and Safety) and tailoring the curriculum design to targeted individual skills. Existing surgical simulators currently provide the trainee with assessment scales created by the industry, while the RSA-Score addresses the assessment based on data from expert and novice surgeons.

Table 5 – Comparison of RSA-Scores; expert (n = 12) and novice (n = 15) groups performance metrics on four FSRS tasks.

Composite metrics	Novice (mean/SD)	Expert (mean/SD)	P value
RSA-Scores for Fourth Arm Control*expert versus novice*			
Safety in operative field	3.1/0.1	3.5/0.1	0.043
Critical errors	2.9/0.2	3.5/0.1	0.051
Bimanual dexterity	2.4/0.3	3.6/0.1	0.002
Task time	3.3/0.1	3.8/0.1	0.026
Economy	2.1/0.2	3.7/0.1	0.032
Overall score	3.1/0.1	3.5/0.1	0.002
RSA-Scores for Coordinated Tool Control—expert *versus* novice			
Safety in operative field	2.5/0.1	3.6/0.1	<0.001
Critical errors	2.1/0.2	3.5/0.1	<0.001
Bimanual dexterity	1.4/0.3	3.5/0.1	<0.001
Task time	3.8/0.1	3.5/0.1	0.018
Economy	3.5/0.1	3.5/0.1	0.751
Overall score	2.5/0.1	3.5/0.1	<0.001
RSA-Scores for Ball Placement—expert *versus* novice			
Safety in operative field	1.4/0.2	3.5/0.1	<0.001
Critical errors	1.8/0.4	3.5/0.1	0.007
Bimanual dexterity	3.3/0.1	3.5/0.1	0.164
Task time	3.5/0.1	3.6/0.1	0.788
Economy	2.4/0.1	3.5/0.1	<0.001
Overall score	2.1/0.2	3.5/0.1	<0.001
RSA-Scores for Needle Handling and Exchange—expert *versus novice*			
Safety in operative field	2.8/0.2	3.4/0.1	0.007
Critical errors	2.6/0.3	3.5/0.1	0.056
Bimanual dexterity	1.4/0.3	3.7/0.1	<0.001
Task time	2.0/0.2	3.7/0.1	<0.001
Economy	3.0/0.1	3.4/0.1	0.001
Overall score	2.5/0.1	3.5/0.1	<0.001
Tasks	**Novice (mean/SD)**	**Expert (mean/SD)**	**P value**
Overall RSA-Score for the FSRS tasks—expert *versus* novice			
Fourth Arm Control	3.1/0.1	3.5/0.1	0.002
Coordinated Tool Control	2.5/0.1	3.5/0.1	<0.001
Ball Placement	2.1/0.2	3.5/0.1	<0.001
Needle Handling and Exchange	2.5/0.1	3.5/0.1	<0.001

SD = standard deviation.

There are various standardized tools available for observation of practice in clinical and laboratory environments. The use of validated standardized tools has been widely used to assess performance on surgery. The Global Operative Assessment of Laparoscopic Skills [17] has been designed to assess technical skills in laparoscopic surgery, whereas the Global Assessment of Gastrointestinal Endoscopic Skills [18] assesses endoscopic skills. Global Operative Assessment of Laparoscopic Skills has shown construct validity to evaluate laparoscopic skills in multiple surgical procedures [17,19]. Global evaluative assessment of robotic skills is another standardized assessment tool for robotic surgical skills that has been validated in a clinical setting [20]. Despite the number of observational tools being available to assess surgical performance, there is still a scarcity of standardized tools available to measure the level of performance and assessment of technical skills on a simulator; particularly when studies have shown that SBT can improve technical and nontechnical skills in surgery [3]. This study could possibly be translated across to other non-robotic modalities. The basics of acquisition of technical skills remain the same. Consensus is needed to incorporate standardized curriculums into the virtual reality simulators.

Our study has a few limitations. First and most importantly, participants were only required to perform each of the four FSRS tasks once, regardless of their performance. Hence, scores used may not completely represent participant performance. Also, drawing conclusions based on a subset of FSRS tasks is not sufficient to characterize expert skills. Establishing true baseline performance on the RSA-Score requires each task of the FSRS curriculum to be performed several times by expert robotic surgeons. Nevertheless, the overall RSA-Scores provide some indication of participant familiarity with RAS skills.

More emphasis is needed to establish assessment methodologies such as the RSA-Score, to make learning in the simulated environment standardized, comparable, and effective. Moreover, to our knowledge, there is no governing body mandated credentialing guidelines for robot-assisted surgery. Certifying proficiency of robotic surgery skills remains in control of institutions and often includes industry-driven certification, a method that is neither standardized nor competency based. Our methodology could possibly help established a set of assessment guidelines in which robotic surgery credentialing is offered. Future efforts will aim to establish external reliability, efficacy, and validity of the RSA-Score by assessing it through wider multi-institutional studies. Studies are also underway to assess the concurrent and predictive validity using the RSA-Score. Because this tool measures technical skills in general and not the specifics of a particular task, further investigation can broaden the application of the RSA-Score for specific procedure-based training.

5. Conclusions

The RSA-Score is a valid scoring system that could be incorporated in any virtual reality-based surgical simulator to achieve standardized assessment of fundamental surgical skills required during robot-assisted surgery.

REFERENCES

[1] Abboudi H, Khan MS, Aboumarzouk O, et al. Current status of validation for robotic surgery simulators—a systematic review. BJU Int 2013;111:194.
[2] Lerner MA, Ayalew M, Peine WJ, Sundaram CP. Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? J Endourol 2010;24:467.
[3] Shamim Khan M, Ahmed K, Gavazzi A, et al. Development and implementation of centralized simulation training: evaluation of feasibility, acceptability and construct validity. BJU Int 2013;111:518
[4] McCulloch P, Mishra A, Handa A, Dale T, Hirst G, Catchpole K. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. Qual Saf Health Care 2009;18:109.
[5] Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001;322:517.
[6] Guru KA, Hussain A, Chandrasekhar R, et al. Current status of robot-assisted surgery in urology: a multi-national survey of 297 urologic surgeons. Can J Urol 2009;16:4736. discussion 4741.
[7] Kesavadas T, Stegemann A, Sathyaseelan G, et al. Validation of Robotic Surgery Simulator (RoSS). Stud Health Technol Inform 2011;163:274.
[8] Kelly DC, Margules AC, Kundavaram CR, et al. Face, content, and construct validation of the Da Vinci Skills Simulator. Urology 2012;79:1068.
[9] Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002;16:1746.
[10] Mackay S, Datta V, Chang A, Shah J, Kneebone R, Darzi A. Multiple Objective Measures of Skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. Ann Surg 2003;238:291.
[11] van Dongen KW, Tournoij E, van der Zee DC, Schijven MP, Broeders IA. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 2007;21:1413.
[12] Lee JY, Mucksavage P, Sundaram CP, McDougall EM. Best practices for robotic surgery training and credentialing. J Urol 2011;185:1191.
[13] Sweet RM, Hananel D, Lawrenz F. A unified approach to validation, reliability, and education study design for surgical technical skills training. Arch Surg 2010;145:197.
[14] Andrew P, Stegemann KA, Syed JR, et al. Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum. Urology 2013;81:767.
[15] Seixas-Mikelus SA, Stegemann AP, Kesavadas T, et al. Content validation of a novel robotic surgical simulator. BJU Int 2011;107:1130.
[16] Schijven M, Jakimowicz J. Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 2003;17:803.
[17] Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. J Am Coll Surg 2007;204:308.
[18] Vassiliou MC, Kaneva PA, Poulose BK, et al. Global Assessment of Gastrointestinal Endoscopic Skills (GAGES): a valid measurement tool for technical skills in flexible endoscopy. Surg Endosc 2010;24:1834.
[19] Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107.
[20] Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. J Urol 2012;187:247.