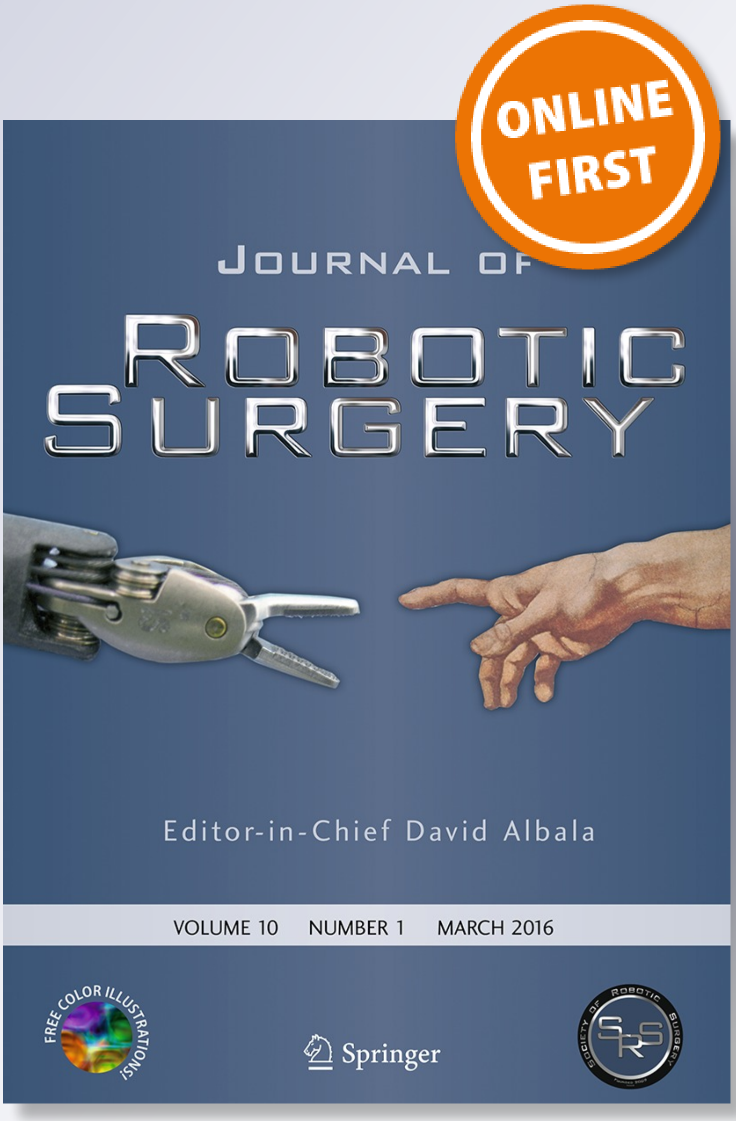
Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS)

Renata Sánchez, Omaira Rodríguez, José Rosciano, Liumariel Vegas, Verónica Bond, Aram Rojas & Alexis Sanchez-Ismayel

Journal of Robotic Surgery

ISSN 1863-2483

J Robotic Surg
DOI 10.1007/s11701-016-0572-1



J Robotic Surg
DOI 10.1007/s11701-016-0572-1

ORIGINAL ARTICLE

Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS)

Renata Sánchez[1,2] · **Omaira Rodríguez**[1,2] · **José Rosciano**[1] · **Liumariel Vegas**[1] · **Verónica Bond**[1] · **Aram Rojas**[1] · **Alexis Sanchez-Ismayel**[1,2]

Received: 22 January 2016 / Accepted: 6 March 2016

Abstract The objective of this study is to determine the ability of the GEARS scale (Global Evaluative Assessment of Robotic Skills) to differentiate individuals with different levels of experience in robotic surgery, as a fundamental validation. This is a cross-sectional study that included three groups of individuals with different levels of experience in robotic surgery (expert, intermediate, novice) their performance were assessed by GEARS applied by two reviewers. The difference between groups was determined by Mann–Whitney test and the consistency between the reviewers was studied by Kendall W coefficient. The agreement between the reviewers of the scale GEARS was 0.96. The score was 29.8 ± 0.4 to experts, 24 ± 2.8 to intermediates and 16 ± 3 to novices, with a statistically significant difference between all of them ($p < 0.05$). All parameters from the scale allow discriminating between different levels of experience, with exception of the depth perception item. We conclude that the scale GEARS was able to differentiate between individuals with different levels of experience in robotic surgery and, therefore, is a validated and useful tool to evaluate surgeons in training.

Keywords Robotic surgery · Training · GEARS · Evaluation

Introduction

Incorporation of new technology to modern surgery has brought the need for changes in the teaching of current surgery practice; hence, training to safe environments is necessary to guarantee our patients quality care, with high success rates and low morbidity. The introduction of laparoscopic surgery provides great advantages over open surgery [1–3], however, is also truth that surgeons face difficulties: loss of depth perception, reduced range of motion of the instruments, loss of tactility and fulcrum effect [4, 5].

Among the latest developments in the field of minimally invasive surgery is incorporation of robotic technology; the Da Vinci System® (Intuitive Surgical Inc.) is the only system available to perform robot-assisted laparoscopic surgery, which has an optical binocular that provides optimum tridimensional vision and instruments specially designed with great maneuverability with seven degrees of freedom which in turn overcomes the difficulties of the laparoscopic approach [6, 7].

The teaching of robotic surgery should start with a theoretical introduction related to basic principles and study of the functioning and components of the system, then move to the implementation of practices aimed mainly the domain of the instruments from the console [8]. The practices should be distributed (with chronological pattern training), structured (defining the objectives and planning protocols in order to achieve specific goals) [9] and deliberate (related to experts' training in a given area) [9, 10].

With the purpose of achieving structured and deliberate practices, incorporation of objective methods of assessment is needed to define the degree of skill to be achieved before moving to the next level of instruction. Among the

---

✉ Omaira Rodríguez
orodriguez@unic.com.ve

[1] Robotic Surgery Program. University Hospital of Caracas, Medicine Faculty, Central University of Venezuela, Caracas, Venezuela

[2] Robotic and Minimally Invasive Surgery (UNIC), Caracas, Venezuela

evaluation methods, the time in which a task is performed is one of the most used parameters, nevertheless, multiple studies show that the realization of certain task in less time is not an expression of suitable dominion of the technique because it could be done with numerous and imprecise movements [11]. The incorporation of rating scales that include elements directly related to the mastery of technique is needed.

Vassiliou et al. in 2005 proposed the introduction of a global scale for performance evaluation in laparoscopic surgery (GOALS: global assessment of laparoscopic skills) [12], which has been used in multiple studies demonstrating its reliability, consistency and validity, being able to differentiate between individuals with different level of experience in minimally invasive surgery [13]. Recently, Goh et al. from the Urology Department of the *Baylor College of Medicine*, have proposed a modification of the original GOALS scale, adapting to it, elements of robotic surgery, including the parameters "robotic control" (related to the domain of the instruments and camera from the console) and "force control" (because the lack of haptic may lead to rupture of sutures or tissues and it is a factor directly related to morbidity). This scale, created under modifications described, has been named GEARS (global evaluative assessment of robotic skills), which it evaluate six parameters for a minimum score of 6 and a maximum of 30 (Fig. 1) [14].

Incorporation of a scale as an evaluation or certification method requires its previous assessment. One of the most relevant evaluation elements is the ability of the evaluation method to differentiate between subjects with different skill levels, in other words, to distinguish those individuals skilled and novice, what it is called "construct validity", and this is the kind of validation which will be determined in the present investigation. This is one of the most valuable features of evaluation method, because if is not possible to detect difference between novices and experts, the improvements made by the novices throughout the training would be undetectable and could not set goals regarding the desired performance [15].

Methods

This is a cross-sectional study, where 15 individuals were taken and divided into three groups based on the level of experience in robotic surgery: expert, intermediate and novice. In the experts group were included individuals with superior experience to 15 cases of robotic surgery, the intermediate group was formed by surgeons who have received formal training in the system console Da Vinci®, but have not been involved in in vivo surgeries, and finally, the novices group included individuals without experience on robotic surgery. Similar samples have been widely used in validation studies described in the literature and have shown be enough for appropriate models validation, simulators and scales [16, 17].

All individuals were explained with a demonstration, the task which they had to do. The evaluated exercise consisted in the realization of continuous suturing, with an initial stitch and knot, followed by three passes over the incision before make the last knot. The model used was a rubber sheet EVA (ethylene vinyl acetate) placed in tubular form on which an incision has been made previously. This kind of inanimate simulators have been used previously by us in the evaluation of learning curves in robotic surgery and correspond with task number nine of the training program proposed by Dulan et al., who have shown that it is an exercise with a high capacity to differentiate between different levels of skill in using the robotic system [18].

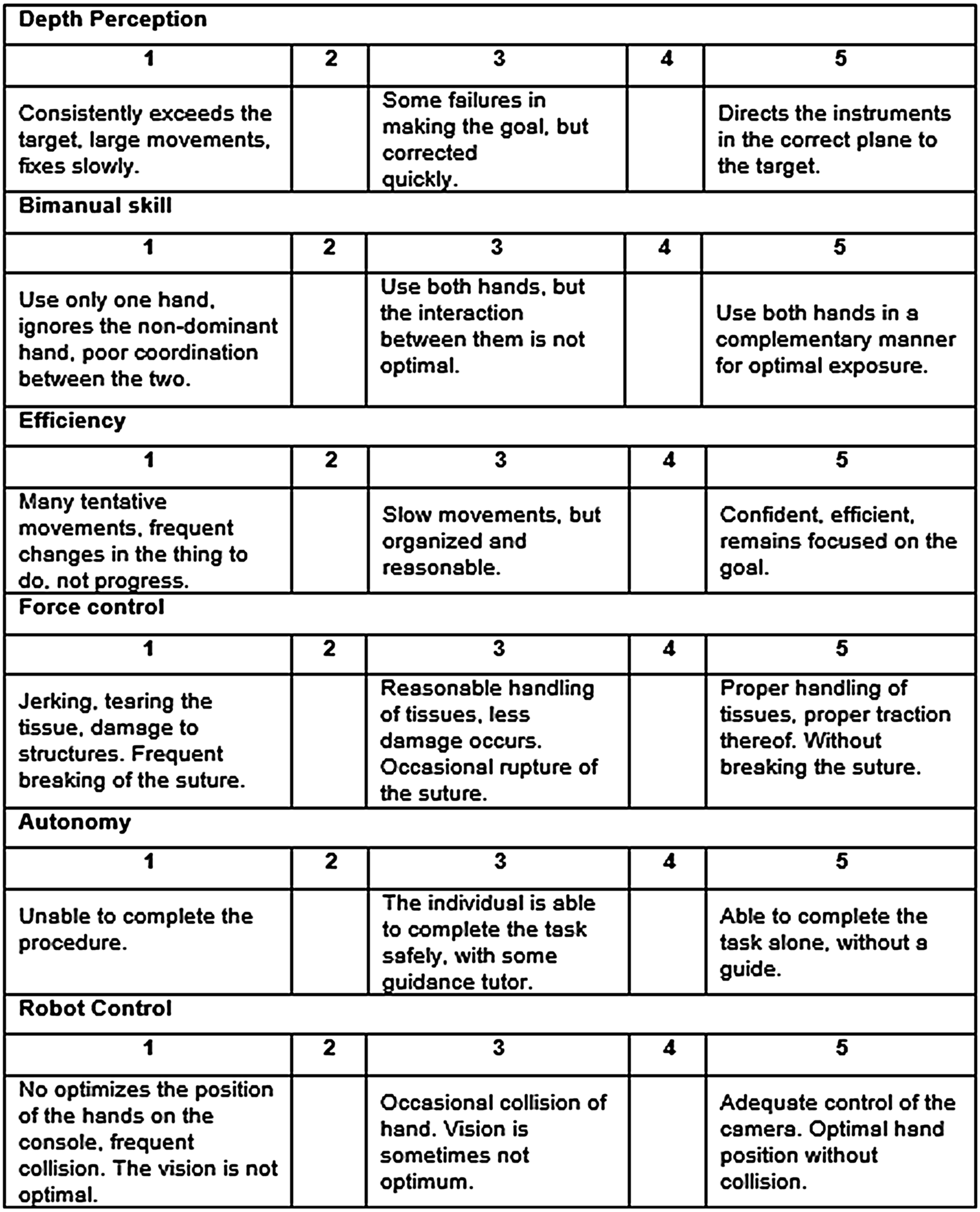
The tasks were performed in the operating room of robotic surgery at the Hospital Universitario de Caracas. For novices, they received a short introductory course in relation to console controls Da Vinci System® S^{HD}. The exercises were digitally recorded for later rating by two assessors, who were not aware of the level of experience of the individual that was being evaluated. Data were stored in tables designed for this purpose in Excel (Microsoft Office®) and then analyzed statistically. The purpose of using two evaluators was to determine the interobserver variability, the correlation between these was assessed using the coefficient W of Kendall.

The difference in scores obtained by individuals from each group (expert, intermediate and novice) was determined using the nonparametric Mann–Whitney method with a significance level of 5 %. The overall score and score of each item in particular were studied, in order to determine which parameters are truly differentiators and the internal consistency of the test, that is to say, the similarity between the overall result and the result of each variable of the scale.

Results

A total of 15 assessments corresponding to five individuals in each group (expert, intermediate and novice) were performed by each of the evaluators. The distribution by sex and age, as well as time spent and score obtained by each one of the groups are shown in Table 1. Obtained interobserver agreement was high ($r = 0.96$).

The time used in performing the task was 311 ± 58 s for the experts, 578 ± 198″ for the intermediates and 873 ± 282″ for the novice. The difference was statistically significant between experts and intermediates ($p = 0.02$)



Depth Perception

1	2	3	4	5
Consistently exceeds the target, large movements, fixes slowly.		Some failures in making the goal, but corrected quickly.		Directs the instruments in the correct plane to the target.

Bimanual skill

1	2	3	4	5
Use only one hand, ignores the non-dominant hand, poor coordination between the two.		Use both hands, but the interaction between them is not optimal.		Use both hands in a complementary manner for optimal exposure.

Efficiency

1	2	3	4	5
Many tentative movements, frequent changes in the thing to do, not progress.		Slow movements, but organized and reasonable.		Confident, efficient, remains focused on the goal.

Force control

1	2	3	4	5
Jerking, tearing the tissue, damage to structures. Frequent breaking of the suture.		Reasonable handling of tissues, less damage occurs. Occasional rupture of the suture.		Proper handling of tissues, proper traction thereof. Without breaking the suture.

Autonomy

1	2	3	4	5
Unable to complete the procedure.		The individual is able to complete the task safely, with some guidance tutor.		Able to complete the task alone, without a guide.

Robot Control

1	2	3	4	5
No optimizes the position of the hands on the console, frequent collision. The vision is not optimal.		Occasional collision of hand. Vision is sometimes not optimum.		Adequate control of the camera. Optimal hand position without collision.

Fig. 1 GEARS scale (Global Evaluative Assessment of Robotic Skills)

however, it was not between intermediates and novices ($p = 0.09$) (Fig. 2).

The average score (GEARS) obtained for each group was 29.8 ± 0.4; 24 ± 2.8 and 16 ± 3; for experts, intermediate and novice, respectively. The difference between each evaluated group was statistically significant, being the experts' performance superior compared to the intermediates ($p = 0.008$) and these, in turn, higher than novices ($p = 0.016$), with a significance level of 5 % (Fig. 3).

The individual analysis of the parameters of the scale showed that the item “depth perception” is not a differentiator between individuals for any of the three levels of experience, because they all had the highest rating (5

Table 1 Sample characteristics and evaluation according to time spent on the task and scale GEARS

	n	Sex (F:M)	Ages (years)	Time (seconds)	GEARS* score	
					Evaluator 1	Evaluator 2
Experts	5	1:4	33.6 ± 4	311 ± 58	29.8 ± 0.4	29.8 ± 0.4
Intermediates	5	2:3	33.6 ± 6	578 ± 198	24 ± 2.8	23.6 ± 3.3
Novices	5	4:1	25.8 ± 1	873 ± 282	16 ± 3	15.8 ± 2.9

* Interobserver agreement (coefficient of Kendall W) = 0.96

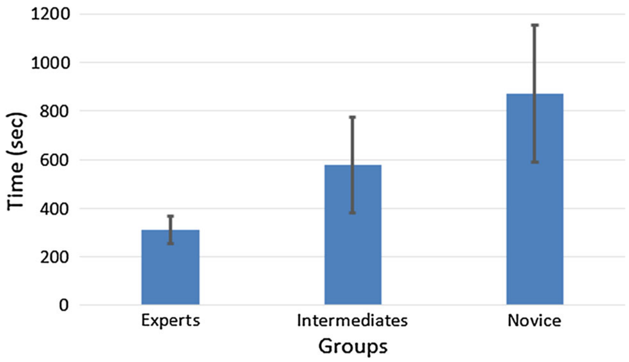


Fig. 2 Time taken to perform the task for each of the evaluated groups

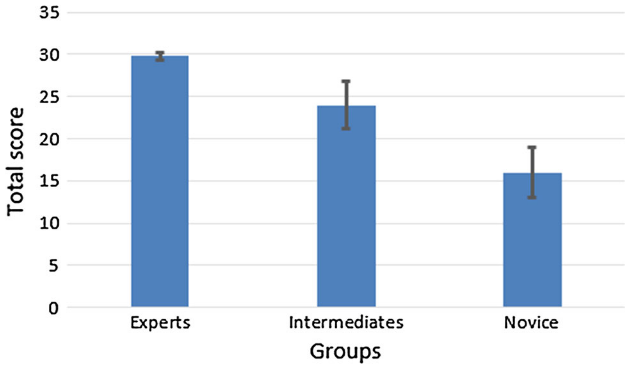


Fig. 3 Performance evaluation by the GEARS scale

points). The assessment of "autonomy" did not allow to differentiate the performance between experts and intermediates ($p = 1.00$), while all the other parameters of the scale proved to be variables capable of differentiating between individuals with different levels of experience.

Discussion

The incorporation of minimally invasive robotic surgery technology has been a breakthrough in allowing to overcome technical limitations, which has led to regarding it as an excellent choice for many processes, but in turn requires an appropriate training in surgical equipment as a critical factor to ensure effective and low mortality procedures [8, 19]. Learning of minimally invasive surgery should be scheduled and conducted in labs designed for that purpose, using a simulation [9].

Supervised practices in real environments are no longer considered as the first choice because time consumption, increased costs and ethical implications are obvious. Surgical trainees should acquire skills in labs designed for that purpose, to enable teachers to focus on key points without compromising patient comfort. Multiple studies show that the skills learned outside the operating room are actually transferred to surgeries in vivo, allowing to bring forward learning curve [20]. The practice led to the formation of highly trained specialists who should be distributed and structured, with special emphasis on the realization of deliberate practice in which supervised sessions where specific needs of each individual required to achieve the established goals are identified, winning expert training. Deliberate practice involves not only the acquisition of a skill but also improving it [9, 10].

The assessment objective parameters allows immediate feedback, which is considered an important element in the development of skills, and finally its application results in a greater motivation for surgeons in training, as well as to determine the progress in the acquisition of skills as successive practices are carried out, allowing to set a target level of skills.

The expert group scored higher than the intermediate and novice, also observed minimal variability between individuals of the group (29.8 ± 0.4). This data homogeneity is characteristic of the trained groups, and clearly differs from the groups with less experience. The analysis of the recorded data checks that score obtained by GEARS scale is able to distinguish different levels of experience in robotic surgery, which makes it a useful tool for evaluation and feedback during the execution of the training program, while that allows to monitor the progress in acquiring skills and determine goals to be achieved for each level of training.

The results indicate that the scale has a high reliability, with an interobserver concordance of $r = 0.96$. It is generally accepted that a higher value $r = 0.8$ catalogs the method as an objective evaluator not dependent tool [15].

The internal reliability with most of the parameters that constitute it was excellent, however, the variable "depth perception" proved not to be an element capable of differentiating between individuals with different level of expertise, which is explained by the fact that the optimum tridimensional vision provided by the Da Vinci System® enables even untrained surgeons have an excellent score on this item. Likewise, in the case of "automatic" parameter, i.e., the ability to complete the task alone showed no difference between the intermediate and expert levels, probably due to the versatility of the robotic system allows moderately trained individuals to complete the task with minimal instruction.

As described, it seems reasonable to omit the "depth perception" scale parameter, making it a scale of five items with a maximum score of 25 points, which could be called "modified GEARS". The evaluation of this change in the scale proposed will be part of future protocols of this line of research in the program of robotic surgery at the Hospital Universitario de Caracas.

Conclusion

The GEARS scale proved to be able to differentiate between individuals with different levels of experience in robotic surgery, validating as a useful tool in training and evaluation of the surgeon in training.

Compliance with ethical standards

Conflict of interest Omaira Rodriguez, Renata Sanchez, Jose Rosciano, Liumariel Vegas, Veronica Bond and Aram Rojas declare that they have no conflict of interest. Alexis Sanchez is proctor of robotic surgery.

References

1. Perissat J, Collet D, Belliart R, Desplantez J, Magne E (1992) Laparoscopic cholecystectomy. The state of the art. A report on 700 consecutives cases. World J Surg 16(6):1074–1082
2. Guller U, Hervey S, Purves H, Muhlbaier L, Peterson E, Eubank S et al (2004) Laparoscopic versus open appendectomy: outcomes comparison based on a large administrative database. Ann Surg 239(1):43–52
3. Staudacher C, Vignali A (2010) Laparoscopic surgery for rectal cancer: the state of the art. World J Gastrointest Surg 2(9):275–282
4. Fraser Klassen R, Feldman D, Ghitulescu D, Stanbridge D, Fried G (2003) Evaluating laparoscopic skills. Surg Endosc 17(6):964–967
5. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic skills training and assessment. Br J Surg 91(12):1549–1558
6. Corcione F, Esposito C, Cuccurullo D, Settembre A, Miranda N, Amato F et al (2005) Advantages and limits of robot-assisted laparoscopic surgery. Surg Endosc 19(1):117–119
7. Jayaraman S, Quan D, Al-Ghamdi I, El-Deen F, Schlachta C (2010) Does robotic assistance improve efficiency in performing complex minimally invasive surgical procedures? Surg Endosc 24(3):584–588
8. Herrom D, Marohn M, SAGES-MIRA Robotic surgery consensus group (2008) A consensus document on robotic surgery. Surg Endosc 22(2):313–325
9. Tsuda S, Scott D, Doyle J, Jones D (2009) Current problems in surgery: surgical skills training and simulation. Curr Probl Surg 46(4):271–370
10. Ericsson K, Krampe R, Tesch-Romer C (1993) The role of deliberated practice in the acquisition of expert performance. Psychol Rev 100(3):363–406
11. Mason J, Ansell J, Warren N, Torkington J (2013) Is motion analysis a valid tool for assessing laparoscopic skill? Surg Endosc 27(5):1468–1477
12. Vassiliou M, Feldman L, Andrew C, Bergman S, Leffondre K, Stanbridge D et al (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 190(1):107–113
13. Chang L, Hogle N, Moore B, Graham M, Sinanan M, Bailey R et al (2007) Reliable assessment of laparoscopic performance in the operating room using videotaped analysis. Surg Innov 14(2):122–126
14. Goh A, Goldfarb D, Sander J, Miles B, Dunkin B (2012) Global evaluative assessment of robotic skills: a clinical assessment tool to measure robotic surgical skills. J Urol 187:247–252
15. McDougall E (2007) Validation of surgical simulators. J Endourol 21(3):244–247
16. Van Empel P, Rijssen L, Commandeur J, Verdam M, Huirne J, Scheele F et al (2012) Validation of a new box trainer-related tracking device: the TrEndo. Surg Endosc 26(8):2346–2352
17. Sánchez A, Otaño N, Rodríguez O, Sánchez R, Benítez G, Schweitzer M (2012) Laparoscopic common bile duct exploration four-task training model: construct validity. JSLS 16(1):10–15
18. Dulan G, Rege R, Hogg D, Gilberg-Fisher K, Arain N, Tesfay S et al (2012) Proficiency-based training for robotic surgery: construct validity, workload and expert levels for nine inanimate exercises. Surg Endosc 26(6):1516–1521
19. Sachdeva A, Russell T (2007) Safe introduction of new procedures and emerging technologies in surgery: education, credentialing and privileging. Surg Clin N Am 87:853–866
20. Sturm L, Windsor J, Cosman P, Cregan P, Hewett P, Maddern G (2008) A systematic review of skills transfer after surgical simulation training. Ann Surg 248(2):166–179