

Surgical suturing training with virtual reality simulation versus dry lab practice: an evaluation of performance improvement, content, and face validity

Michael J. Amirian · Samuel M. Lindner ·
Edouard J. Trabulsi · Costas D. Lallas

Received: 4 February 2014 / Accepted: 16 June 2014 / Published online: 19 July 2014
© Springer-Verlag London 2014

Abstract The purpose of this study is to evaluate the effectiveness of virtual reality (VR) simulation versus dry lab suturing practice at improving suturing performance in robotic surgery. Nineteen novice participants with no prior robotic suturing experience were randomized to two groups, VR simulation and dry lab, which consisted of inanimate training on a da Vinci Si surgical system. Each group underwent baseline suturing evaluation, then trained on the Symbionix™ Suturing Module (SSM) or undertook suturing practice using the da Vinci Surgical System in a dry lab. Final suturing performance was evaluated using the objective suture scoring method. Participants in the VR simulation group were surveyed to assess the face and content validity of the SSM. Both groups experienced significant improvement after training (VR simulation group $p = 0.0078$; dry lab group $p = 0.0039$). There was no significant difference in improvement between the two groups after undergoing training with either SSM or in the dry lab. Improvements in composite timing scores were 123 and 172 in the VR simulation and dry lab test groups, respectively ($p = 0.36$). Face validation varied with respect to the category assessed, but participants confirmed

content validity of the SSM in all categories. In this sample of novice operators, there was no significant advantage in training with VR simulation using the SSM over dry lab training in improving suturing performance. Users of the SSM found it useful and relevant as a training tool for improving suturing performance.

Keywords Surgical training · Simulation · Virtual reality · Validation

Introduction

Studies in the literature comparing outcomes between experienced and less-experienced robotic surgeons have demonstrated worse outcomes in robotic-assisted radical prostatectomy (RARP) with more frequent surgical complications and higher rates of positive surgical margins in the surgeons with relatively less experience [1, 2]. Longer operative times for surgeons with less experience increase surgical costs by reducing operating room efficiency and reducing surgical volume [3]. Because of the learning curve associated with gaining proficiency in robotic surgery, emphasis has been placed on the benefit of virtual reality (VR) simulation in advancing robotic surgical skills and reducing the risk to patients by providing an effective and efficient alternative to the intraoperative learning experience. The Fundamentals of Laparoscopic Surgery (FLS) curriculum is endorsed by the American College of Surgeons and is widely implemented in laparoscopic surgery to certify surgeon competency in performance metrics required for laparoscopic surgery [4]. Recent studies investigating VR simulation have confirmed face and content validity of VR simulation for performing tasks used for FLS certification [5]. Research to investigate the

M. J. Amirian · S. M. Lindner · E. J. Trabulsi · C. D. Lallas
Department of Urology, Thomas Jefferson University Hospital,
1015 Walnut Street, Suite 1100, Philadelphia, PA 19107, USA
e-mail: mjamirian@gmail.com

S. M. Lindner
e-mail: Samuel.Lindner@jefferson.edu

E. J. Trabulsi
e-mail: edouard.trabulsi@jefferson.edu

E. J. Trabulsi · C. D. Lallas (✉)
Kimmel Cancer Center, Thomas Jefferson University Hospital,
1015 Walnut Street, Suite 1100, Philadelphia, USA 19107
e-mail: costas.lallas@jefferson.edu

feasibility and effectiveness of a similar simulation-based curriculum for robotic surgery indicates such programs may be on the horizon for robotic surgeons [6].

The da Vinci Skills Simulator (dVSS) is a VR simulation *backpack* produced by Intuitive Surgical (Sunnyvale, CA, USA) that attaches to the back of the da Vinci Si surgeon console. In a prior publication, our group demonstrated the face, content, and construct validity of the dVSS [7]. In a press release in late 2012, the Symbionix™ Suturing Module (SSM) was announced for release [8]. This module adds suturing and knot-tying simulation to the dVSS simulation package.

Other studies have demonstrated that, in the absence of dedicated suturing simulation, VR simulation training to increase robotic surgery proficiency has not translated to increases in real-world suturing performance [9]. The aim of this study is to investigate the effectiveness of targeted suturing training using the SSM compared to suturing practice in a dry lab using the da Vinci Surgical System in improving suturing performance, as measured by a validated objective suture evaluation method described by Derossis et al. [10]. Following final performance evaluation, the face validity and content validity of the SSM were evaluated by surveying simulation users to assess the realism of the interface and the usefulness of the simulator as a training tool.

Materials and methods

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study.

Study design and subjects

A convenience cohort of 26 medical students responding to an institution-wide email were recruited to participate in this prospective, randomized, institutional review board-approved study to test the effectiveness of training with VR simulation versus dry lab suturing practice using the da Vinci Surgical System in improving suturing performance in two randomized groups of novice operators.

Participants were eligible for inclusion in the study as novice operators if they had no prior experience suturing with the SSM or the robotic surgical system. The cohort was stratified by medical school class and randomized to two groups: dry lab and VR simulation. Participant characteristics are included in Table 1. All participants were evaluated for baseline suturing performance using the

Table 1 Test group demographics

Medical school year	Dry lab	VR simulation
2016 (MS4)	2	3
2015 (MS3)	3	3
2014 (MS2)	4	4
Total	$n = 9$	$n = 10$

robotic surgical system; they utilized a 2-0 polyglactin suture trimmed to 13 cm and a prepared Penrose drain specimen, as described by a prior group in their evaluation method for intracorporeal knots [10]. After baseline testing, the two groups underwent training, as described below. At the conclusion of the training sessions, suturing performance was re-evaluated. Following post-training performance evaluation, the simulation group participated in a survey to rate the face and content validity of the SSM.

Dry lab group training

Following baseline testing, the dry lab group underwent a supervised practice session where subjects received a standardized explanation and demonstration of the suturing method using the robotic surgical system followed by 30 min of dry lab suturing practice using 3-0 polyglactin suture and a sponge practice pad. Following the practice session, final suturing performance was re-evaluated using the same method utilized in the baseline performance evaluation [10].

VR simulation group training

Following baseline testing, the simulation group underwent VR training on the dVSS using the SSM. Subjects viewed the SSM tutorial and participated in iterative training sessions on the vertical defect module until achieving competence. Following VR training and competence, final suturing performance was evaluated using the prior described method [10].

VR simulation competency criteria

As tested, the SSM monitors and records various performance parameters including wound entry/exit accuracy, needle entrance angle, and suture over-tensioning. However, the simulator lacks concrete pass/fail criteria for suturing skill competence. To overcome this limitation, minimum simulation performance requirements were adapted from the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) testing criteria currently used to evaluate suturing competency in laparoscopic surgery [11]. Based on this, the criteria for a

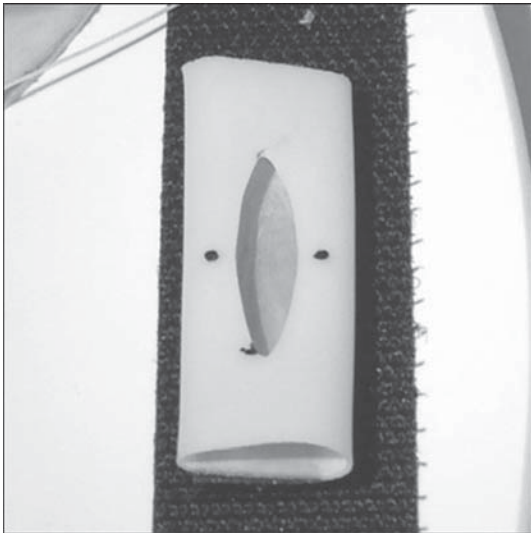


Fig. 1 Penrose drain suturing specimen

successful simulated knot were determined to include a maximum suture completion time of 112 s per suture, satisfactory wound entry/exit accuracy (as determined by the simulator), and zero over-tensioned knots. Test subjects were judged competent for re-evaluation after completing one simulation session meeting all of these criteria.

Suturing evaluation method

The suture evaluation method used in this study was adapted from the method for intracorporeal knot evaluation originally described by Derossis et al. [10] and later validated by Dauster et al. [12]. This method is an objective evaluation using a prepared section of Penrose drain that takes into account the amount of time required to tie a standard knot (a surgeon's knot followed by two additional single square throws), the accuracy for needle entrance and exit at pre-marked points on the Penrose drain edge, the measured gap in the approximated edges of the specimen's "wound" slit, and integrity of the knot (i.e. secure, loose, or coming apart). From this information, a composite timing score is calculated which gives points for speed and adds penalty time for inaccuracy, gap, and looseness. A higher timing score reflects higher performance (minimum score = 0, maximum score = 600). The maximum allowable time to complete the exercise was 10 min (600 s).

An objective method was utilized to judge knot integrity by re-measuring the gap in the Penrose specimen edge after the specimen has experienced a constant 10-s transverse tensile stress. A secure knot demonstrated no change in gap or ruptured the specimen without slipping the knot. A loose knot demonstrated increased post-stress gap, and a coming-

apart knot was incompletely tied or had gaps and spaces in the knot upon pre-stress visual inspection. We also inspected the specimens and recorded whether or not the knot was square. The prepared Penrose drain specimen is shown in Fig. 1. The timing score and knot integrity evaluation methodology are described in Fig. 2.

Face and content validity evaluation

As done in previous dVSS validation studies, simulation subjects were surveyed at the completion of the study using a visual analog scale (VAS) rating of 1–5 to describe similarity between the simulator and actual suturing (face validity) and usefulness of the simulation for improving actual suturing performance (content validity) [7]. VAS values were collected, ranging 1–5. For evaluation of face validity, a value of 1 indicated that the SSM was nothing like actual surgery, while a value of 5 indicated that using the SSM was exactly like actual surgery. For evaluation of content validity, a value of 1 indicated that the SSM was not relevant or useful for training, while a value of 5 indicated that using the SSM was very relevant and useful for training. A score of 3 indicates the midpoint of the VAS, and is neutral. Validity evaluation was considered confirmatory for a median rating >3. Validity was rejected for a median rating <3. A copy of the validity survey is included in the "Appendix".

Statistical analysis

All statistical analysis was performed by a consultant biostatistician from the Thomas Jefferson University Department of Pharmacology and Experimental Therapeutics, Division of Biostatistics. Summary statistics (frequencies and percents for categorical variables and means, standard deviations, medians, minimums, and maximums for continuous variables) were calculated for all available data. Accuracy, gap, integrity, and squareness summary statistics were calculated for subjects who had measurements at both initial and final measurement; eight in the dry lab group and nine in the simulator group, as one participant in each group was unable to complete the initial knot assessment. Where the scores were sufficiently continuous (time in seconds, time score, accuracy, and gap) differences were calculated between the initial and final scores and these difference scores were compared between the dry lab and simulator groups using Wilcoxon rank sum test. The difference between the groups with regard to final integrity scores was assessed. The integrity score was an ordinal variable, scored as 0, 1, and 2. The difference was taken between initial and final scores; this change score was tested between groups using Fisher's Exact test. Because all subjects in both groups had a value of "Not Square" for

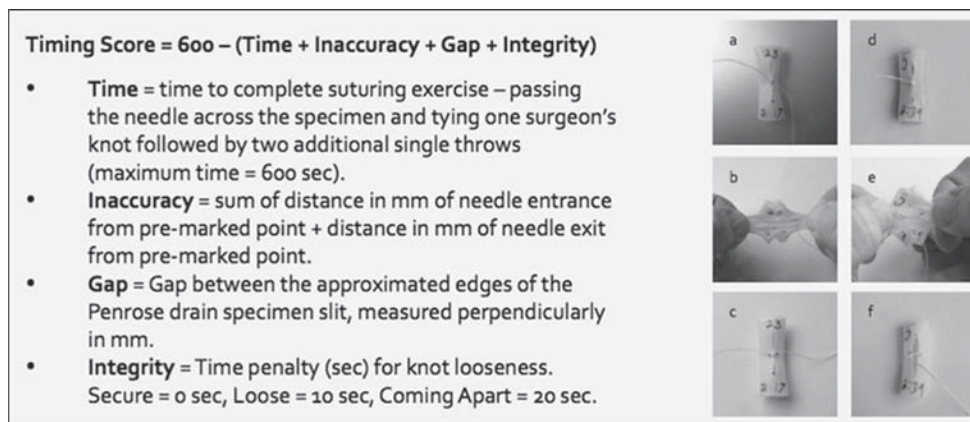


Fig. 2 Timing score and knot integrity evaluation

the square measure, only the final square measure was assessed using Fisher's Exact test. SAS software version 9.3 was utilized for the data analysis, and R version 3.0.1 was utilized for the construction of the validity assessment boxplots.

Results

Demographics

After stratification by medical school class and randomization, 26 medical student volunteers with no prior robotic suturing or SSM experience were divided among simulation and dry lab groups. Of the 26 original participants, seven were lost to follow-up for scheduling training sessions and final performance evaluation and were excluded from final analysis. The characteristics of the final 19 participants are described in Table 1.

Performance evaluation

There was no significant difference in baseline performance between the two groups, with baseline timing scores of 254 and 235 for simulator and dry lab groups, respectively ($p = 0.5928$). Following training, both groups demonstrated significant improvements in performance (VR simulation group $p = 0.0078$; dry lab group $p = 0.0039$). The difference between the simulator and dry lab groups in performance improvement favored the dry lab group, but this difference did not reach significance, with average improvements in the timing scores of 123 and 172 in the simulator and dry lab test groups, respectively ($p = 0.3602$). Timing scores improved after practice in 9/9 participants in the dry lab group and in 8/10 participants in the simulator group. Of the two students who did not improve with simulation training, one was a 3rd-year

student whose timing score was the same pre- and post-training (difference = 0), and the other was a 4th-year student whose timing score decreased by 34. Analysis of the factors influencing the timing score calculation reveals no significant difference in accuracy of entrance and exit points, edge gap, or knot integrity after training. There was an average improvement in accuracy of 0.63 and 1.33 mm in the dry lab and simulator groups, respectively ($p = 0.2726$). The amount of gap in approximated edges was slightly increased with post-training gap differences of +0.38 and +0.44 mm for the dry lab and simulation groups, respectively ($p = 0.1106$). Results are outlined in Table 2. Cliff's delta was used to calculate effect sizes for these variables.

Face and content validity

Based on the simulation group's post-training survey results, the participants rejected face validity (realism) of the simulator's tissue behavior (median 2.6) and confirmed face validity with regards to clutching (median 4.1), depth/spatial relationship (median 3.5), needle driving (median 3.5), and visual appearance (median 3.3). The participants confirmed content validity (usefulness as a training tool) in all categories: clutching (median 4.3), depth/spatial relationship (median 3.7), needle driving (median 4.2), tissue behavior (median 3.8), and visual appearance (median 3.8). Validity data are listed in Table 3.

Discussion

The performance evaluation indicates that both VR simulation and actual suturing practice with the robotic surgical system in a dry lab improve suturing performance in novice operators, as judged by increases in the objective suture evaluation timing scores.

Table 2 Performance data summary

	Dry lab (<i>n</i> = 9)		VR simulation (<i>n</i> = 10)		<i>p</i> value	Cliff's delta
	Mean (SD)	Median [Min, Max]	Mean (SD)	Median [Min, Max]		
Timing score						
Initial	235.22 (119.01)	265 [0, 353]	254.3 (144.75)	318.5 [0, 401]	0.3602	0.2444
Final	407.67 (74.22)	440 [263,476]	377.5 (51.52)	375 [302,453]		
Change	172.44 (114.38)	154 [18, 360]	123.20 (115.4)	98 [−34, 302]		
Time (s)						
Initial	355.67 (120.96)	313 [247,600]	337.7 (146.6)	270 [195, 600]	0.4023	0.2667
Final	185.22 (70.1)	159 [123, 327]	212.6 (53.13)	214.5 [143,291]		
Change	−170.44 (117.3)	−142 [−350, −15]	−125.10 (117)	−101.50 [−309, 27]		
Accuracy (mm)						
Initial	1.38 (1.3)	1 [0, 4]	2.11 (1.27)	2 [0, 4]	0.2726	0.3333
Final	0.75 (1.16)	0 [0, 3]	0.78 (1.09)	0 [0, 3]		
Change	−0.63 (1.06)	−0.50 [−2, 1]	−1.33 (1.32)	−1 [−3, 1]		
Gap (mm)						
Initial	0.13 (0.35)	0 [0, 1]	0.11 (0.33)	0 [0, 1]	0.1106	0.0556
Final	0.50 (0.76)	0 [0, 2]	0.56 (1.13)	0 [0, 3]		
Change	0.38 (0.92)	0 [−1, 2]	0.44 (1.24)	0 [−1, 3]		
	Dry lab (<i>n</i> = 9)		VR simulation (<i>n</i> = 10)		<i>p</i> value	Cliff's delta
	Initial	Final	Initial	Final		
Integrity						
Secure	3	5	3	3	0.0945 ^a	0.4583
Loose	3	3	6	6		
Coming apart	2	1	0	1		
Incomplete	1 ^a	0	1 ^a	0		
Total	9	9	10	10		
Square knots						
Square	0	3	0	2	0.6825 ^a	^b
Not square	8	6	9	8		
Incomplete	1 ^a	0	1 ^a	0		
Total	9	9	10	10		

^a Participants unable to complete a knot in the initial evaluation were excluded from calculation

^b Cliff's delta calculation not appropriate for completely dichotomous data. In the dry lab group, 33 % tied square knots, in the simulator group, 20 %

Table 3 Validity survey summary

	Face validity median [min, max]	Content validity median [min, max]
Clutching	4.09 [2.95, 4.63]	4.28 [2.98, 5.00]
Depth/spatial relationship	3.47 [2.61, 4.63]	3.73 [2.08, 5.00]
Needle driving	3.47 [2.08, 5.00]	4.21 [2.28, 5.00]
Tissue behavior	2.62 [1.00, 4.36]	3.79 [1.57, 4.36]
Visual appearance	3.32 [1.47, 4.36]	3.84 [1.61, 4.80]

While multiple past studies have demonstrated that VR training improves surgical performance, it has been shown that VR training programs that do not include dedicated suturing simulation have not produced reliable improvements in dry-lab robotic suturing performance [9]. This

study confirms that, indeed, simulated suturing translates to real-world improvements in suturing performance in novice operators.

Regarding validity assessment, the results of the validity survey's Likert ratings mirrored the written comments participants submitted with their surveys. The majority of participants felt that the interface did not realistically simulate suturing activities with regards to tissue behavior. However, most participants felt the simulator was a useful training tool and was instrumental in improving performance in the final evaluation. We had many philosophical debates regarding the veracity of having a student judge validation. Most validation studies for robotic surgery have experienced surgeons; they can comment on the tissue behavior of the simulator and compare it to real tissue

behavior, and their experience is not the same as someone new to the field. However, for the purposes of this study, this validation does not really matter, as they are not going to be training on a simulator, anyway. The perceived validity of the end user is the true marker, however, because they are the ones who are going to be actively learning from the simulator.

Based on the results on this study, training with the SSM offers no performance advantage over dry lab practice using the robotic surgical system. This was a surprising, but not unexpected result, and similar to what other groups have discovered [13]. Our two groups undertook disparate training regimens, one time-based (dry lab) and the other proficiency-based (VR simulation). One would expect that the proficiency-based model, with constant performance feedback, would be more beneficial than a training paradigm not supplying real time critique. Obviously, this was not the case, and the study instead demonstrated that a set dry-lab curriculum can be just as effective. Still, training with the SSM will save the expense of da Vinci consumable training instruments, suture, and practice pads, and it does not require the use of the da Vinci patient-side cart. However, it requires a da Vinci Si surgeon console, the dVSS *backpack*, and the acquisition cost of the SSM must be considered. And while this study confirms that the SSM is effective in improving performance in novice operators, it does not evaluate its effectiveness in improving surgical skills of operators with baseline proficiency using the da Vinci Surgical System. In fact, closer analysis of the individual participant performances in the simulation group indicates, indeed, student volunteers benefitted from simulation training, but average absolute benefit decreased as participants' year in medical school increased (improvement in timing scores for 2nd-year students = 224, 3rd-year students = 127, and 4th-year students = 45). Based on this trend, it is difficult to generalize the effectiveness of the SSM in increasing suturing performance in novice operators to residents, fellows, and attending surgeons with much greater relative experience.

Limitations

The largest drawback of this study was the small sample size. This resulted in a study that was underpowered to detect a difference in the performance improvement between the two study groups. We did try to correct for this by reporting an effect size, which provides an index of the strength of an effect without being influenced by sample size. Study participants were drawn from a convenience sample of students interested in participating in a surgical study. As such, there is likely some response bias, with this population of students possibly being more surgically

inclined than a truly random sample of medical student novice operators. The initially recruited sample size was small ($n = 26$), and the requirement for participants to follow up for initial evaluation, training sessions, and final evaluation resulted in a high rate of participants being lost to follow-up and possibly contributing to some sample bias.

Lastly, our study established competency thresholds for the participants' simulator performance. While all the performance parameters considered for competence (suture tension and entrance/exit angle and accuracy) were tracked by the simulator, the maximum time allowed for exercise completion was drawn from studies that timed experienced surgeons tying knots using traditional laparoscopic equipment [11]. Timing for such an activity does not necessarily directly translate to experienced surgeons tying knots robotically. However, the purpose of the competence threshold was to give participants a concrete goal for performance and an incentive to improve suturing speed. Additionally, the structure of the training regimens of the two groups was intended to simulate the circumstances under which simulators and dry lab practice time is typically used, with simulators being used iteratively until performance increases to a "passing" level versus dry lab practice time usually being made available in divided blocks of time set aside for practice. It is possible that our cutoff for timing competence was not aggressive enough. However, more appropriate timing cutoffs for competency could be elucidated in future studies evaluating SSM construct validity and performance improvement in a group of experienced operators, including surgical residents, fellows, and attending surgeons.

Conclusions

In this sample of novice operators, there was no significant advantage to training with VR simulation using the SSM over dry lab practice in improving suturing performance. Furthermore, while users of the SSM did not find all aspects of the simulator realistic in their face validity assessment, they found it useful and relevant as a training tool for improving suturing performance.

Acknowledgments All statistical analysis performed by Jocelyn Andrei Sendeck, MSPH of Thomas Jefferson University Department of Pharmacology and Experimental Therapeutics, Division of Biostatistics.

Conflict of interest Costas D. Lallas declares that he has no conflict of interest. Michael J. Amirian declares that he has no conflict of interest. Samuel M. Lindner declares that he has no conflict of interest. Edouard J. Trabulsi declares that he has no conflict of interest.

Appendix—Validity Survey

Face Validity:

Draw a vertical line on the visual analog scale (VAS) for the following. (1 = totally unlike surgery. 5 = identical to actual surgery)

- Visual Appearance
1 |-----| 5
- Tissue Behavior
1 |-----| 5
- Clutching
1 |-----| 5
- Needle Driving
1 |-----| 5
- Depth/Spatial Relationships
1 |-----| 5

Commentary:

Content Validity:

Draw a vertical line on the VAS to indicate simulator usefulness in training for the following. (1 = not relevant or useful. 5 = very relevant and useful)

- Visual Appearance
1 |-----| 5
- Tissue Behavior
1 |-----| 5
- Clutching
1 |-----| 5
- Needle Driving
1 |-----| 5
- Depth/Spatial Relationships
1 |-----| 5

Commentary:

References

1. Raman JD, Dong S, Levinson A et al (2007) Robotic radical prostatectomy: operative technique, outcomes, and learning curve. *JLS* 11:1
2. Ou YC, Yang CR, Wang J et al (2011) The learning curve for reducing complications of robotic-assisted laparoscopic radical prostatectomy by a single surgeon. *BJU Int* 108(3):420–425
3. Steinberg PL, Merguerian PA, Bihrlle W III et al (2008) The cost of learning robotic-assisted prostatectomy. *Urology* 72:1068–1072
4. Society of American Gastrointestinal and Endoscopic Surgeons. FLS Program Description. <http://www.flsprogram.org/index/fls-program-description/>. Accessed 27 Sep 2013
5. Arikatla VS, Ahn W, Sankaranarayanan G, De S (2013) Towards virtual FLS: development of a peg transfer simulator. *Int J Med Robot.* doi:10.1002/rcs.1534
6. Stegemann AP, Ahmed K, Syed JR et al (2013) Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum. *Urology* 81(4):767–774
7. Kelly DC, Margules AC, Kundavaram CR et al (2012) Face, content, and construct validation of the da Vinci Skills Simulator. *Urology* 79(5):1068–1072
8. Symbionix™ Press Releases. <http://symbionix.com/2012/12/05/innovative-simulation-solutions-change-surgical-education-symbionix-launches-four-new-products-at-ac/>. Accessed 27 Sep 2013
9. Lerner MA, Ayalew M, Peine WJ, Sundaram CP (2010) Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? *J Endourol* 24(3):467–472
10. Derossis AM, Fried GM, Abrahamowicz M et al (1998) Development of a model for training and evaluation of laparoscopic skills. *Am J Surg* 175(6):482–487
11. Ritter EM, Scott DJ (2007) Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov* 14(2):107–112
12. Dauster B, Steinberg AP, Vassiliou MC et al (2005) Validity of the MISTELS simulator for laparoscopy training in urology. *J Endourol* 19(5):541–545
13. Tergas AI, Sheth SB, Green IC et al (2013) A pilot study of surgical training using a virtual robotic surgery simulator. *JLS* 17(2):219–226