BJUI
BJU International

Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool

Patrick Ramos, Jeremy Montez, Adrian Tripp, Casey K. Ng, Inderbir S. Gill and Andrew J. Hung

USC Institute of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Objectives

To evaluate robotic dry laboratory (dry lab) exercises in terms of their face, content, construct and concurrent validities.

To evaluate the applicability of the Global Evaluative Assessment of Robotic Skills (GEARS) tool to assess dry lab performance.

Materials and Methods

Participants were prospectively categorized into two groups: robotic novice (no cases as primary surgeon) and robotic expert (≥30 cases).

Participants completed three virtual reality (VR) exercises using the da Vinci Skills Simulator (Intuitive Surgical, Sunnyvale, CA, USA), as well as corresponding dry lab versions of each exercise (Mimic Technologies, Seattle, WA, USA) on the da Vinci Surgical System.

Simulator performance was assessed by metrics measured on the simulator. Dry lab performance was blindly video-evaluated by expert review using the six-metric GEARS tool.

Participants completed a post-study questionnaire (to evaluate face and content validity).

A Wilcoxon non-parametric test was used to compare performance between groups (construct validity) and Spearman's correlation coefficient was used to assess simulation to dry lab performance (concurrent validity).

Results

The mean number of robotic cases experienced for novices was 0 and for experts the mean (range) was 200 (30–2000) cases.

Expert surgeons found the dry lab exercises both 'realistic' (median [range] score 8 [4–10] out of 10) and 'very useful' for training of residents (median [range] score 9 [5–10] out of 10).

Overall, expert surgeons completed all dry lab tasks more efficiently ($P < 0.001$) and effectively (GEARS total score $P < 0.001$) than novices. In addition, experts outperformed novices in each individual GEARS metric ($P < 0.001$).

Finally, in comparing dry lab with simulator performance, there was a moderate correlation overall ($r = 0.54$, $P < 0.001$). Most simulator metrics correlated moderately to strongly with corresponding GEARS metrics ($r = 0.54$, $P < 0.001$).

Conclusions

The robotic dry lab exercises in the present study have face, content, construct and concurrent validity with the corresponding VR tasks.

Until now, the assessment of dry lab exercises has been limited to basic metrics (i.e. time to completion and error avoidance). For the first time, we have shown it is feasibile to apply a global assessment tool (GEARS) to dry lab training.

Keywords

educational models, robotics training, validation studies, dry lab models

Introduction

In the age of minimally invasive surgery, developing adequate surgical ability not only involves time in the operating room, but also training in the laboratory to fine tune technical skills. With recent changes to resident work hours, surgical residents are now exposed to fewer cases than before and, with an increased emphasis on patient safety, the traditional apprenticeship training model is no longer sufficient for surgical training [1,2]. Surgical training now requires trainees to obtain basic robotic and laparoscopic skill sets before entering the operating room, including handling suture, knot-tying, cutting, etc. Surgical programmes have addressed this with training laboratories using specialized equipment, dedicated practice time and adoption of specialized education curricula for a minimally invasive technique.

 | doi:10.1111/bju.12559
Published by John Wiley & Sons Ltd. www.bjui.org

In addition to conventional laparoscopic surgery, proficiency in minimally invasive surgery now requires experience using the da Vinci Surgical System (Intuitive Surgical, Sunnyvale, CA, USA). As a result of its widespread popularity and incorporation into surgical practice, contemporary surgeons must incorporate robotic training to remain competitive. Various training and assessment tools have been developed to aid surgical education in this regard. In general, they consist of a series of exercises that test specific minimally invasive skills that are clinically relevant, and have a scoring system to assess trainee proficiency as well as to track improvement of skills over time. A few validated protocols for training in robot-assisted laparoscopic surgery exist [3]; however, these are still in their infancy.

While there is extensive literature regarding validated virtual simulation training tools for robotic surgery [4,5], to our knowledge, only a few studies have examined the utility and validity of dry laboratory (dry lab) exercises for robotic surgical training [6]. Additionally, while the Global Evaluative Assessment of Robotic skills (GEARS) tool has been used to assess *in vivo* performance [6,7], it has never been applied to dry lab exercises.

Recently, Mimic Technologies (Seattle, WA, USA) has created prototype dry lab modules derived from their virtual reality (VR) exercises. Before integrating these modules into minimally invasive surgical curricula, the exercises require independent validation [8]. Validation includes assessment for the overall feel/controls/interface realism of the training tool to expert users (face validity), utility as a training tool as viewed by experts (content validity), ability to differentiate levels of surgical experience (construct validity), and degree to which dry lab exercises correlate with previously validated virtual exercises (concurrent validity).

In the present study, we evaluate dry lab modules derived from previously validated Mimic VR exercises for their face, content, construct and concurrent validity. We also evaluate the applicability of the GEARS tool to assess dry lab performance.

Materials and Methods

Study Design

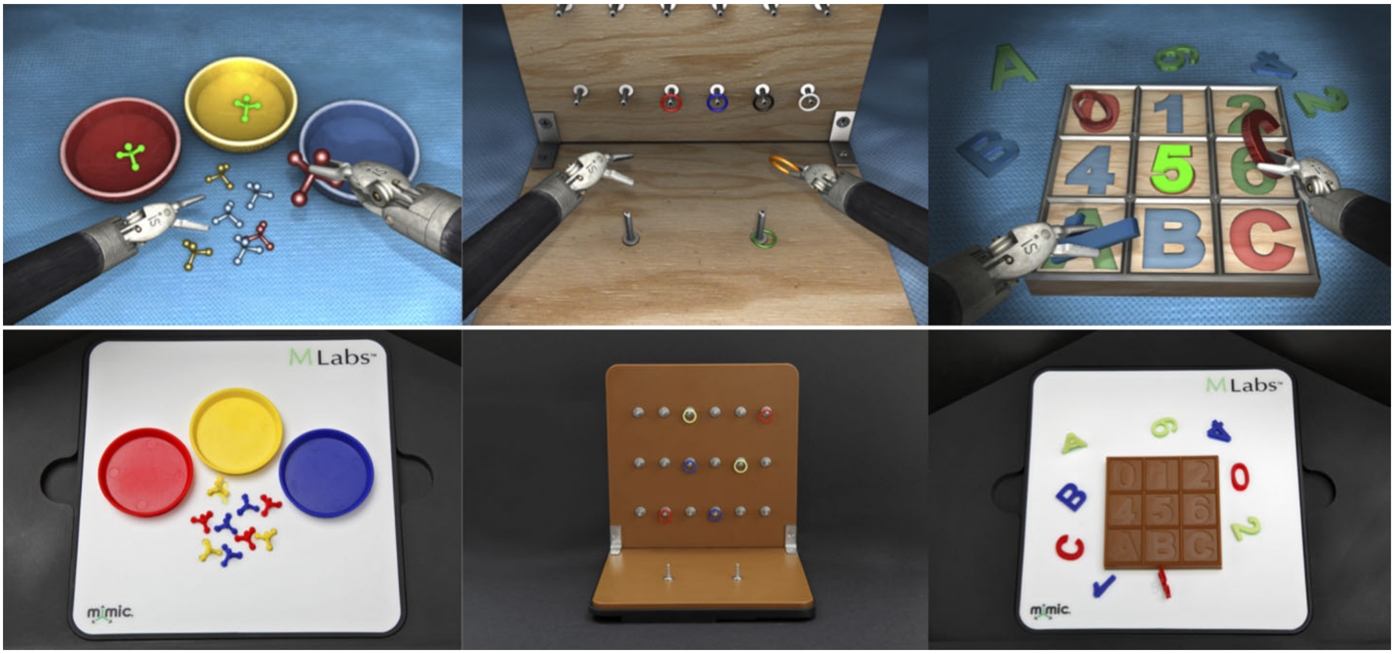
The dry lab model exercises used for the present study were prototypes developed by Mimic Technologies, each of which were reversed-engineered from previously validated VR exercises. Three dry lab modules were developed: Pick-and-Place, Peg Board and Match Board (Fig. 1). The da Vinci Skills Simulator was used for the corresponding VR exercises.

With approval from the institutional review board, participants were prospectively enrolled in this study at a single institution. The study used three available da Vinci robotic systems in the operating room as well as the da Vinci simulator.

Study Process

Subjects were categorized as novice (no cases as primary surgeon) or expert (≥30 cases as primary surgeon [Fig. 2]).

Fig. 1 Top row: virtual simulation exercises. Bottom row: corresponding dry lab exercises. From left to right: Pick and Place, Peg Board, Match Board (Credit: Mimic Technologies).



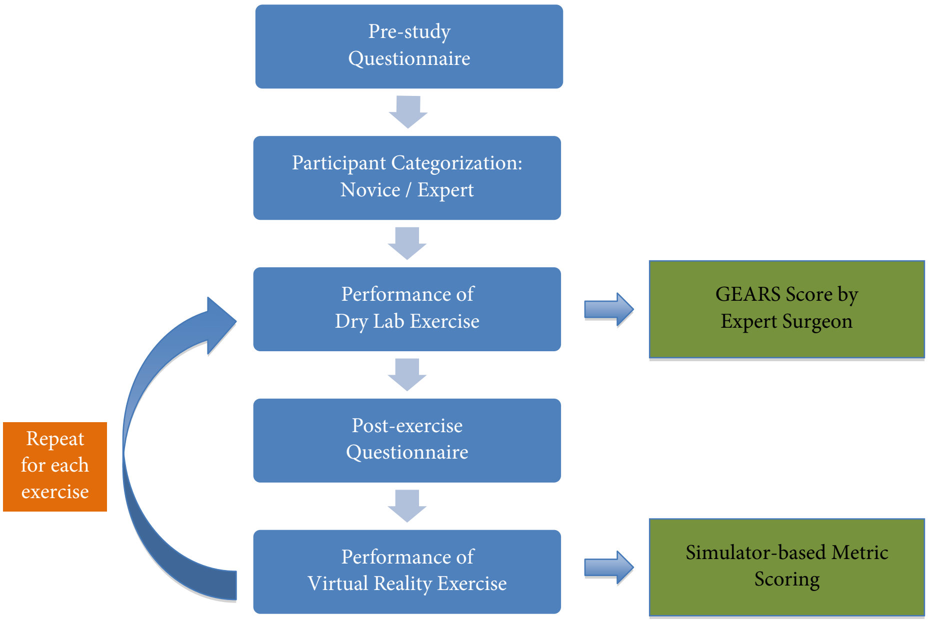


Fig. 2 Flow chart of testing process.

Before completing the modules, participants completed a pre-study questionnaire (on demographics and self-report of surgical experience). Study subjects were then given a standardized orientation of the da Vinci console controller functions to ensure basic understanding of console controller functions (i.e. camera control and clutch). Each subject subsequently completed three dry lab exercises in tandem with the corresponding VR exercises. Before each exercise, the proctor explained the exercise task.

Performance Assessment Tools

During each dry lab exercise, three objective metrics were obtained for each of the dry lab exercises: number of instrument collisions, number of drops and time to complete exercise. Additionally, an expert in robotic surgery (>200 cases as primary surgeon) blindly reviewed the video-taped performance of each dry lab exercise and evaluated performance using GEARS. After completion of the dry lab exercises, participants completed a post-study questionnaire. During performance of the corresponding simulator exercises, the simulator computer recorded participant performance based on specific metrics, described in Table 1. Finally, an overall score was also calculated at the completion of each exercise based on an algorithm accounting for all recorded simulator metrics.

Outcome Measures

For the dry lab exercises, post-study questionnaires were used to assess the realism of the simulator (face validity) and its utility as a training tool (content validity) based on a 1–10 visual analogue scale (1 being the worst and 10 being the best score). Construct validity was evaluated by comparing objective metrics and GEARS scores of novices and expert surgeons. Correlation of GEARS performance and simulator-based metrics was used to measure concurrent validity.

Statistical Analysis

We performed a statistical power estimation before beginning the study based on previous performance data in the literature regarding robotic simulator performance [3]; we expected experts to score on average 85%, while novices 60%. Considering a two-sided test at the significance level of 5% ($\alpha = 0.05$), a sample size of three participants was needed in each of the expert and novice groups to obtain statistical power of 80% ($\beta = 0.2$). In evaluation of construct validity, performance metrics of experts and novices were compared using the non-parametric Wilcoxon test for continuous variables and a chi-squared test for categorical variables. Spearman's analysis was used to determine the extent of correlation between dry lab and virtual simulation performance (concurrent validity).

Results

Thirty-six subjects participated in the present study, including 12 experts and 24 novices (Table 2). The novice cohort was composed primarily of medical students and surgical interns with no previous robotic experience. The expert cohort consisted primarily of urology attendings and fellows, who had performed a mean (range) of 200 (30–2015) robotics cases.

Table 1 Concurrent validity: extent metrics of dry lab performance correlated with corresponding metrics of virtual simulator performance.

Simulator metrics (computer-generated)	Dry lab performance metrics (GEARS; scored by expert judge)	Correlation coefficient r	P
Overall score (%)	**Total score**		
Weighted average: below individual metrics weighted 1.0, except for Master Controller Range weighted 0.2	Sum of all GEARS metrics	0.54	<0.001
Number of objects dropped	**Number of drops**		
Drops of grasped objects/needles		0.35	0.023
Economy of motion (cm)	**Efficiency* (1–5 points)**		
Distance travelled by robotic instruments	Confident, efficient and safe conduct, maintains focus on task	0.64	<0.001
	Depth perception* (1–5 points)		
	Accurately directs instrument in correct plane to target	0.75	<0.001
	Bimanual dexterity* (1–5 points)		
	Uses both hands in a complimentary manner to optimize exposure	0.62	<0.001
Excessive instrument force (s)	**Force sensitivity* (1–5 points)**		
Time that force over preset limit is applied	Applies appropriate tension, negligible injury to adjacent structures no suture breakage	NS	NS
Number of instrument collisions	**Force sensitivity* (1–5 points)**		
	Applies appropriate tension, negligible injury to adjacent structures no suture breakage	−0.48	0.002
Instruments out of view (cm)	**Robotic control* (1–5 points)**		
Distance travelled out of surgical view	Controls camera and hand position optimally and independently. Minimal collisions or obstruction of assistant	NS	NS
Master controller range (cm)	**Efficiency* (1–5 points)**		
Distance travelled by physical controllers		0.48	0.001
Time to completion (s)	**Time to completion (s)**	0.87	<0.001
	Efficiency* (1–5 points)		
	Confident, efficient and safe conduct, maintains focus on task	−0.79	<0.001
	Autonomy* (1–5 points)		
	Ability to complete task independently	NS	NS

**Validated GEARS. NS, nonsignificant relationship,* P *> 0.05.*

Table 2 Participant demographics.

	Expert	Novice	*P*
n	12	14	
Median (range), years	34 (30–46)	25 (22–37)	<0.001
Sex, %			
Male	92	67	
Female	8	33	
Mean (range) robotic experience			
Years	3.5 (0.5–8)	0 (0–3)	<0.001
Cases	200 (30–2015)	0 (0)	0.004

Table 3 Face and content validity.

	Median (range) score out of 10, from Expert group
Face validity (Realism)	8 (4–10)
Content validity (Utility)	
Resident	9 (5–10)
Fellow	7 (4–10)
Expert Surgeon	4 (1–8)

Face and Content Validity

Scores from the expert cohort showed that the overall dry lab and virtual simulation experience was 'very realistic' (median visual analogue scale score 8/10 [Table 3]). Moreover, expert robotic surgeons also rated the simulator as a 'very useful' training tool for residents (median score 9/10) and fellows (median score 7/10), although less so for experienced robotic surgeons (median score 4/10).

Construct Validity

Dry lab assessment with GEARS found experts significantly outperforming novices in all but one of the individual metrics ($P < 0.001$; Table 4). For depth perception, experts scored a

Table 4 Construct validity.

	Expert	Novice	*P*
Construct validity			
Median (range) time, s	212 (164–286)	504 (309–885)	<0.001
Median (range) no. of collisions	0 (0)	0 (0–2)	0.33
Median (range) no. of drops	1.4 (0–6)	2.4 (0–7)	0.16
GEARS, median (range) score out of 5			
Depth perception	4.5 (4–5)	2.3 (1–5)	<0.001
Bimanual dexterity	4.2 (3–5)	2.6 (1–5)	<0.001
Efficiency	4.1 (3–5)	2.5 (1–5)	<0.001
Force sensitivity	4 (3–5)	2.7 (1–5)	<0.001
Autonomy	5 (5)	5 (5)	–
Robotic control	4.3 (3–5)	2.1 (1–4)	<0.001

median of 4.5/5, whereas novices scored 2.3/5, which meant that experts accurately directed their instruments in the correct plane to target, and novices overshot targets, had wider swings and were slower to correct. For bimanual dexterity, experts scored 4.2/5 and novices 2.6/5, which signified that experts used both hands in a complementary way and novices used one hand more, ignored the non-dominant hand and had poorer coordination. With respect to efficiency, experts scored 4.2/5 and novices 2.5/5, meaning experts maintained focus on task with fluid progression and novices had more uncertain movements, constantly changing focus and persisted without progression. For force sensitivity, experts scored 4/5 and novices 2.7/5, which suggested that experts applied appropriate tension with negligible injury to adjacent structures, whereas novices caused more trauma to adjacent structures. In terms of robotic control, experts scored 4.3/5 and novices 2.1/5, implying that experts controlled the camera and instrument positions more optimally and novices had less optimum views and instrument positions. There was no significant difference in points earned for autonomy between experts (5/5) and novices (5/5). In addition to better GEARS scores, experts completed the tasks in less than half the time of novices (212 vs 504 s, $P < 0.001$). There was no significant difference in the number of objects dropped or instrument collision between the two groups.

Concurrent Validity

The GEARS assessment metrics, along with the corresponding VR metrics, are described in Table 1. Simulator composite total score (sum of all metrics) moderately correlated with GEARS total score ($r = 0.54$, $P < 0.001$). Economy of motion, a simulator metric, moderately to highly correlated with GEARS scoring of efficiency, depth perception and bimanual dexterity ($r = 0.64, 0.75, 0.62$, respectively, $P < 0.001$). The number of instrument collisions moderately correlated ($r = -0.48$, $P = 0.002$) with GEARS force sensitivity. Master controller range, a simulator metric which measures distance travelled by physical controllers, correlated moderately ($r = 0.48$, $P = 0.001$) with GEARS efficiency. Lastly, VR time to completion strongly correlated with GEARS time to completion and efficiency ($r = 0.87, -0.79$, $P < 0.001$).

Discussion

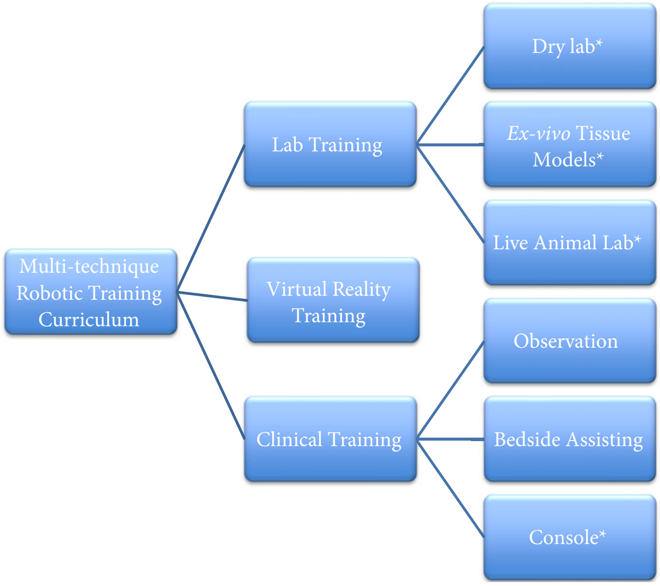
The present study meticulously documents the face, content, construct and concurrent validity of three dry lab exercises for robotic training, reverse-engineered from their original VR format. Expert surgeons found the featured dry lab exercises to be realistic (median score 8/10) and useful for training residents (median score 9/10). Furthermore, they were able to distinguish between expert and novice robotic surgeons using the GEARS metrics ($P < 0.001$). Finally, most GEARS metrics, reflecting dry lab performance, moderately to strongly correlated to corresponding simulator metrics on the original VR exercises (overall $r = 0.54$, $P < 0.001$). The latter two validation steps relied on analysis with the validated GEARS tool, suggesting that a global assessment tool such as GEARS can be used in the dry lab setting.

Few previous studies have validated dry lab exercises for robotic surgical skills training [6]. Even more lacking is an established and validated curriculum for robotic surgery training. Studies have shown that certification for laparoscopy is beneficial for ensuring basic surgical competence [9]; however, unlike basic laparoscopic surgery, there is no formal certification process for robotic surgery. To prepare surgeons, varying robotic training curricula have been developed using VR exercises that have been evaluated with respect to face, content, construct, concurrent and predictive validity [3–5,10]; however, unlike Fundamentals for Laparoscopic Surgery [9], no single, recognized robotics curriculum exists. Dry lab modules, as part of a multi-technique training curriculum, can add to the training armamentarium and allow the development of basic robotic skills without any risk to patients and at minimal additional cost, unlike VR simulations (cost: $1 450 for the featured dry lab exercises vs ~$100 000 for the dV-Trainer or Skill Simulator) Unlike VR simulators, the 'touch and feel' of the robot has 100% fidelity. And as shown for the first time in this study, a validated global assessment tool such as GEARS can be a complementary scoring partner to dry lab training. Even Fundamentals for Laparoscopic Surgery, a well-established dry lab platform, has limited and basic assessment metrics (i.e. time to completion, number of drops, etc.).

Expert participants in the present study reported excellent face and content validity for the dry lab modules. We requested that the expert subjects in the study judge the featured exercises against the overall 'feel' of the da Vinci interface and controls. Truly, we could not practically judge dry lab exercises for environmental realism against live clinical surgery. While experts scored the dry labs exercises as very useful for resident and fellow training, their ratings indicate that it probably has a limited role for experienced robotic surgeons. The present dry lab exercises test a very finite, basic skill set. Creation of other, more robust dry lab exercises, which could include suturing, knot-tying and procedure-specific simulation may prove even more useful in skills training for all subjects from novice to expert.

The ability of the dry lab exercises to distinguish between novice and expert, or construct validity, was also shown by the present study. With our study design, true robotics novices were compared with expert robotic surgeons. The definition of expert surgeon included all those with >30 robotic cases. We aimed to set the standard for minimum expertise as supported in the literature [11]. Presumably, if construct validity can be demonstrated at a low threshold for 'expert,' a more robust

Fig. 3 Multi-modality Robotic Training Curriculum. A comprehensive training curriculum ought to include various modalities, including lab training, virtual reality simulation, and clinic training. Ideally, trainee performance assessment across modalities is universal. For example, performance assessment in dry lab (present study), ex-vivo tissue models [12], live animal lab [6], and console training [7] have been validated utilizing the Global Evaluative Assessment of Robotic Surgery (GEARS) (marked with *).



requirement for expert would show an even greater difference between expert and novice performance. The dry lab results reaffirmed findings of VR validation results for the corresponding exercises [6], with expert surgeons significantly outperforming the novice group in almost all measured metrics.

Dry lab exercise performance moderately to highly correlated with virtual simulation performance, establishing concurrent validity. Application of the GEARS assessment tool was essential in comparing dry lab and VR performances. We believed it was appropriate to determine concurrent validity of the dry lab exercises against performance with the original VR exercises, given that the original VR exercises have been extensively validated in multiple studies [3–5,10]; therefore, the VR exercises were the *de facto* 'gold standard' for the purposes of the present validation study.

Our expert judge had minimal difficulty evaluating dry lab performance using GEARS. The one exception was the inability to thoroughly assess the GEARS metric 'autonomy' based on blinded review of video clips because there was no audio recorded nor assessed. This is perhaps a limitation of the study design (blinded review to remove bias) as opposed to a limitation of the metric 'autonomy' as an assessment tool. Previously, it has been shown that GEARS scoring has excellent interobserver reliability amongst expert surgeons with at least 100 cases as primary surgeon [10].

Data from the present study can help provide the groundwork for creating a robust curriculum for training and credentialing surgeons for robotic surgery (Figure 3). While there is no true substitute for a formalized, structured training programme with live operative experience, dry lab exercises can serve as an effective and inexpensive component of a multiple-approach training programme that has been validated with performance data such as those reported in the present study. Global assessment ought to be incorporated into dry lab training, in addition to traditional dry lab metrics (i.e. time to completion and error avoidance).

In conclusion, the present study establishes the face, content, construct and concurrent validity of novel dry lab exercise modules. Additionally, we have shown the feasibility of using a global assessment tool to evaluate dry lab performance. Predictive validity studies, as well as continued development of dry lab exercises, which focus on advanced robotic skills, will further determine its utility in clinical practice.

Acknowledgements

Mimic Technologies provided the dry lab prototypes.

Conflict of Interest

None declared.

References

1 **Scalese RJ, Obeso VT, Issenberg SB.** Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med* 2008; 23 (Suppl. 1): 46–9

2 **Scott DJ, Dunnington GL.** The new ACS/APDS Skills Curriculum: moving the learning curve out of the operating room. *J Gastrointest Surg* 2008; 12: 213–21

3 **Hung AJ, Zehnder P, Patil MB et al.** Face, content and construct validity of a novel robotic surgery simulator. *J Urol* 2011; 186: 1019–24

4 **Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G, Chandrasekhar R, Wilding GE, Guru KA.** Face validation of a novel robotic surgical simulator. *Urology* 2010; 76: 357–60

5 **Seixas-Mikelus SA, Stegemann AP, Kesavadas T et al.** Content validation of a novel robotic surgical simulator. *BJU Int* 2011; 107: 1130–5

6 **Hung AJ, Jayaratna IS, Teruya K, Desai MM, Gill IS, Goh AC.** Comparative assessment of three standardized robotic surgery training methods. *BJU Int* 2013; 112: 864–71

7 **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: 247–52

8 **McDougall EM.** Validation of surgical simulators. *J Endourol* 2007; 21: 244–7

9 **Hafford ML, Van Sickle KR, Willis RE et al.** Ensuring competency: are fundamentals of laparoscopic surgery training and certification necessary for practicing surgeons and operating room personnel? *Surg Endosc* 2013; 27: 118–26. doi: 10.1007/s00464-012-2437-7. Epub 2012 Jul 7

10 **Hung AJ, Patil MB, Zehnder P et al.** Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. *J Urol* 2012; 187: 630–7

11 **Patel VR, Tully AS, Holmes R, Lindsay J.** Robotic radical prostatectomy in the community setting–the learning curve and beyond: initial 200 cases. *J Urol* 2005; 174: 269–72

12 **Hung AJ, Ng CK, Patil MB et al.** Validation of a novel robotic-assisted partial nephrectomy surgical training model. *BJU Int* 2012; 110: 870–4

Correspondence: Andrew J. Hung, USC Institute of Urology, 1441 Eastlake Ave, Suite 7416, Los Angeles, CA 90089, USA.

e-mail: ajhung@gmail.com

Abbreviations: dry lab, dry laboratory; GEARS, Global Evaluative Assessment of Robotic Skills; VR, virtual reality.