Surg Endosc
DOI 10.1007/s00464-015-4607-x

Face, content, construct, and concurrent validity of a novel robotic surgery patient-side simulator: the Xperience™ Team Trainer

Song Xu[1,2] · Manuela Perez[1,3] · Cyril Perrenot[1,3] · Nicolas Hubert[1,4] · Jacques Hubert[1,4]

Received: 31 July 2015 / Accepted: 29 September 2015

Abstract

Objectives To determine the face, content, construct, and concurrent validity of the Xperience™ Team Trainer (XTT) as an assessment tool of robotic surgical bed-assistance skills.

Methods Subjects were recruited during a robotic surgery curriculum. They were divided into three groups: the group RA with robotic bed-assistance experience, the group LS with laparoscopic surgical experience, and the control group without bed-assistance or laparoscopic experience. The subjects first performed two standard FLS exercises on a laparoscopic simulator for the assessment of basic laparoscopic skills. After that, they performed three virtual reality exercises on XTT, and then performed similar exercises on physical models on a da Vinci® box trainer.

Results Twenty-eight persons volunteered for and completed the tasks. Most expert subjects agreed on the realism of XTT and the three exercises, and also their interest for teamwork and bed-assistant training. The group RA and the group LS demonstrated a similar level of basic laparoscopic skills. Both groups performed better than the control group on the XTT exercises ($p < 0.05$). The performance superiority of the group RA over LS was observed but not statistically significant. Correlation of performance was determined between the tests on XTT and on da Vinci® box trainer.

Conclusions The introduction of XTT facilitates the training of bedside assistants and emphasizes the importance of teamwork, which may change the paradigm of robotic surgery training in the near future. As an assessment tool of bed-assistance skills, XTT proves face, content, and concurrent validity. However, these results should be qualified considering the potential limitations of this exploratory study with a relatively small sample size. The training modules remain to be developed, and more complex and discriminative exercises are expected. Other studies will be needed to further determine construct validity in the future.

Keywords Xperience Team Trainer · Validity · Bedside assistant · Robotic surgical training · Robotic simulator · Virtual reality simulation

Robot-assisted laparoscopic surgery has expanded rapidly in the last decade. Approximately 570,000 da Vinci® procedures were performed worldwide in 2014 [1]. The advantages of this technology have been widely documented, e.g., ergonomics, EndoWrist® instruments, high-resolution 3D vision, tremor filtration, and motion scaling [2]. The latest da Vinci® Xi system was launched in 2014. Competitive robotic systems are also under development around the world with the promise of lower costs or some improved techniques [3]. It seems promising that robots will play an increasingly important role in surgery.

Surgeons need adequate training to achieve proficiency in robotic technology. Standards should be established for

✉ Jacques Hubert
j.hubert@chu-nancy.fr

1 IADI/Inserm U947, Lorraine University, Nancy 54000, France

2 Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

3 Department of Emergency and General Surgery, CHU Nancy, Nancy 54000, France

4 Department of Urology, Brabois Hospital, University Hospital of Nancy, Allée du Morvan, 54511 Vandoeuvre-les-Nancy, France

Published online: 10 December 2015

the assessment of techniques and the credentialing of robotic surgeons. Currently, robotic surgical training programs exist in many formats in the world, and the process of standardization has begun [4]. With multiple advantages such as the relative cost-effectiveness, ease of setup, and real-time feedback of performance, the role of virtual reality simulators has been emphasized in robotic surgical training. Five virtual reality robotic simulators are currently available on the market, of which the Mimic dV-Trainer® (dVT) has benefited from a largest number of validation studies [5, 6].

In robotic procedures, the console surgeon is un-scrubbed without direct access to the patient. He could only interact with the group through audio communication, whereas in conventional surgery, more than 80 % of the information exchange is realized via visual contact. The success of robotic surgery thus relies on high-quality teamwork, wherein the bedside assistant plays an important role. He takes responsibility for the patient safety and the smooth proceeding of the operation. Moreover, it is believed that experienced bed assistants could help decrease the operative time, reduce blood loss, and may also reduce complications [7]. Currently, sufficient attention is not given to the education of bedside assistants, and a convenient and valid tool is needed for the training and assessment of robotic bed-assistance skills [8].

In 2014, the Mimic Technologies launched the Xperience® Team Trainer (XTT) to be used in tandem with dVT. This innovation facilitates the training of bed assistants and also emphasizes the importance of teamwork, which may change the paradigm of robotic surgery training in the near future. This study represents the first effort to determine the face, content, construct, and concurrent validity of XTT as an evaluation tool of robotic surgical bed-assistance skills. The value of XTT in teamwork training is also investigated.

Materials and methods

Equipment

The experiments were performed during the curriculum of *Diplôme Inter-Universitaire* (DIU, inter-university degree) in robotic surgery in Nancy, France. An XTT simulator connected with a dVT simulator, as well as a da Vinci® S robot, was used for the study.

The dVT-XTT platform simulated actual robotic surgery scenario: One user, in the role of console surgeon, sat on dVT and controlled robotic instruments. The other user, in the role of patient-side assistant, operated XTT through laparoscopic handles. The two roles shared the same operative view but were provided with 3D and 2D vision, respectively (Fig. 1).

Thirteen team exercises were available on the platform for the training of three bed-assistance skills. Based on expert opinions, four exercises were chosen for the experiments: "Team Pick & Place 2" (PP2) for the instruction and warm-up; "Team Match Board 1" (MB1), "Team Ring Walk Retraction" (RWR), and "Team Clipping Ring Walk 1" (CRW) for the test. Each test exercise teaches one of the three bed-assistance skills: MB1 for transfers and handoffs, RWR for retraction, and CRW for clip application.

In PP2, the operator or the assistant picked up one jack and handed off it to the other, the latter then dropped the jack into its matching colored container. The exercise was complete when all the jacks were successfully placed in the containers. In MB1, the assistant picked up each time one (character) object and transferred it to the operator. The latter then placed it in its matching location on the board. In RWR, the operator grasped the ring and moved it along the vessel-like structure. Two wall loops covered by tissue flaps would be encountered. The assistant needed to retract the flaps up and away to reveal the loops. In CRW, the operator also moved the ring along the vessel-like structure. Four blue sections would be sequentially highlighted on the structure after the passing of the ring. The assistant placed clips over these blue sections and attached them to the vessel structure (Fig. 2A). Physical models were made to reproduce these four virtual reality exercises for the use on da Vinci® S box trainer (Fig. 2B).

The dVT-XTT platform automatically recorded the performance of assistants using a built-in scoring system. Multiple metrics were provided and have been previously described, such as the task completion time, the instrument motion, and several error metrics [9]. Two new metrics were available for CRW: The "clipping distance" refers to the total distance of applied clips to their corresponding marked application regions; the "incorrect clips" indicates the number of incorrectly applied clips. An overall score was generated by synthesizing the results of all metrics.

Subjects

The DIU trainees, robotic surgeons, residents, and medical students without previous XTT experience were recruited for the study. Informed consents were obtained from all participants. Each subject was assigned a unique identity number under which all his/her data would be collected.

Experimental procedures

Demographic data collection

Each subject fulfilled a demographic questionnaire. Questions about the numbers of robotic (as bed assistant or console surgeon) and laparoscopic cases were involved.

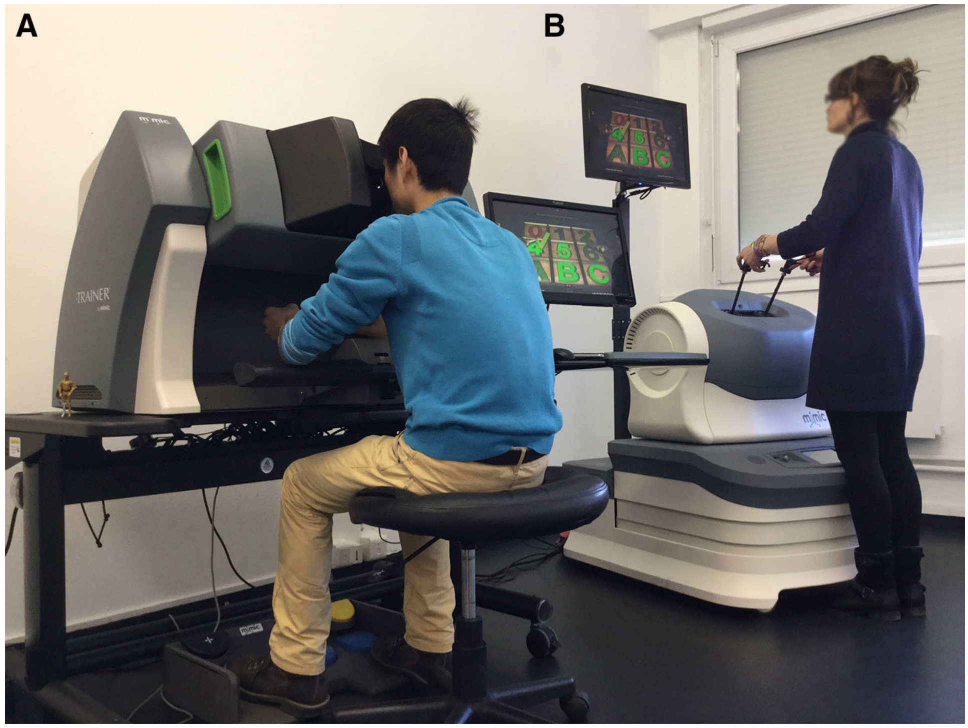

Fig. 1 dV-Trainer (**A**) and the Xperience Team Trainer (**B**)

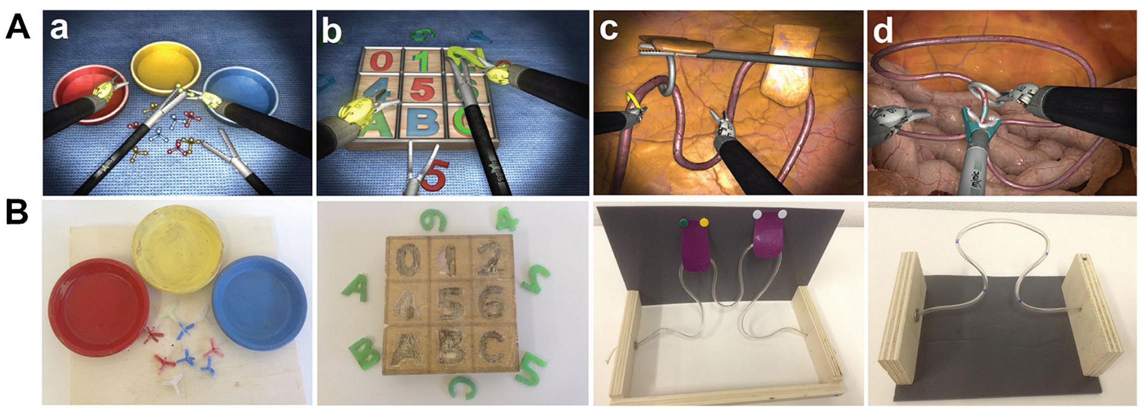

Fig. 2 Four virtual reality exercises available on Xperience Team Trainer (**A**) and the corresponding physical models (**B**): Team Pick & Place (*a*), Team Match Board 1 (*b*), Team Ring Walk Retraction (*c*), and Team Clipping Ring Walk (*d*)

Basic laparoscopic skills assessment

Each subject performed sequentially two standard FLS exercises–"Peg Transfer" and "Precision Cutting"–on a laparoscopic box trainer. The performances in both exercises were filmed and scored using standard FLS metrics [10]. The two scores were then averaged as the basic laparoscopic skills of the subjects.

XTT session

Subjects performed the exercises on XTT with a same expert console operator. After three attempts of PP2 within a XTT-instruction period, the subjects performed sequentially the test exercises of MB1, RWR, and CRW. Each exercise was repeated thrice, with the first attempt for familiarization and the last two attempts averaged as the performance. Immediately after the third attempt of an exercise, the subjects evaluated its realism and interest for training (of the technique that it aims at) on a five-point Likert scale. When all the exercises were finished, they evaluated the realism and interest of the global system, and fulfilled a questionnaire of NASA Task Load Index (NASA-TLX).

The NASA-TLX is a multi-dimensional rating procedure that provides an overall workload score based on a

Table 1 Modified GOALS

	1	2	3	4	5
Depth perception	Constantly overshooting target, hits backstop, wide swings, slow to correct		Some overshooting or missing plane but corrects quickly		Accurately directs instruments in correct plane to target
Bimanual dexterity (in MB1)	Always trying to use one hand, ignoring the other hand		Use of both hands but giving priority to one hand		Expertly utilizes both hands
Efficiency	Much wasted effort, many tentative motions, persisting at a task without progress		Slow, but able to complete the tasks with reasonable efforts		Confident, efficient and safe conduct of the tasks
Tissue handling	Rough, tears tissue by excessive traction, injures adjacent structures, grasper frequently slip off		Handles tissues reasonably well, with some minor trauma to adjacent tissues, occasional slipping of grasper		Handles tissues very well with appropriate traction on tissues and negligible injury of adjacent structures.
Assistance effects	Unable to assist effectively, frequently makes troubles. The operator has to make additional efforts to cooperate with him/her		Able to provide effective assistance, does not make troubles to the console operator		Intelligently and appropriately provides assistance, facilitating the manipulation of the console operator

weighted average of ratings on six subscales: mental demands, physical demands, temporal demands, own performance, effort, and frustration.

da Vinci session

The subjects performed the four exercises in the same order with an expert da Vinci console operator. Each exercise was repeated twice due to the limited access to the robot. The first attempt allowed for familiarization, and the second attempt was filmed for the subsequent performance evaluation. After all the exercises, subjects described the advantages and drawbacks of XTT compared to the da Vinci box trainer. They fulfilled also a NASA-TLX questionnaire.

Validation strategies

Face and content validity

Face validity refers to the realism of the simulator. Content validity examines the appropriateness of the system as a training tool. Only the opinions (collected in the XTT session) of the subjects with robotic bed-assistance or laparoscopic experience were used to determine face validity. Only the opinions of the subjects with robotic surgical or bed-assistance experience were used to determine content validity.

Construct validity

Construct validity assesses the simulator's ability to differentiate subjects possessing different levels of skills. The subjects were divided into three groups: (a) group RA—subjects with robotic bed-assistance experience; (b) group LS—subjects with laparoscopic surgical experience; (c) control group—subjects without robotic or laparoscopic experience. Performances on the XTT exercises were compared between the three groups.

Concurrent validity

Concurrent validity indicates the extent to which performance on a new test corresponds to performance on an established test of the same construct. Correlation of performances on the XTT and the da Vinci exercises was tested. The de-identified videos of da Vinci exercises were transferred to two evaluators. They assessed the subjects' performances independently using a modified version of the "Global Operative Assessment of Laparoscopic Skills" (GOALS) (Table 1).

Test–retest reliability

Test–retest reliability examines the correlation of performances on the exercises that are performed more than once. Correlation of the performances in the second and third attempts of each XTT exercise was tested.

Inter-rater reliability

This modality determines the degree of agreement among raters. Correlation of the two set of GOALS scores of the two evaluators was tested.

Statistical analysis

Data were analyzed using the software R. Performances of the three groups on XTT exercises were compared using a

Table 2 Demographic data of the three groups

	RA	LS	Control
n	11	7	9
Age (years)	30.7 ± 5.2	41.9 ± 14.7	38.3 ± 10.2
Position (n)	Surgeon (10)	Surgeon (7)	Nurse (6)
	Nurse (1)		Student (3)
Median robotic bed-assistance cases (range), n	25 (3–100), $n = 11$	$n = 0$	$n = 0$
Median laparoscopic surgical cases (range), n	30 (5–300), $n = 10$	100 (6–1000), $n = 7$	$n = 0$

Kruskal–Wallis rank sum test followed by a post hoc Wilcoxon rank sum test. The p value was adjusted using Holm method. Correlation of performances was tested using the Spearman's correlation. Statistical significance was determined at $p < 0.05$.

Results

Demographics

Thirty persons volunteered for the experiments, whereas two persons stopped at the FLS tasks. The rest 28 subjects completed the XTT session, and 19 of them completed the da Vinci session. One subject performed the XTT exercises with a different console operator. His data were only used to examine face and content validity. Demographic data are presented in Table 2.

Face validity

Data were derived from 19 subjects. 68.4 % persons considered MB1 and CRW to be with a high or very high realism. The ratio was 89.5 % for RWR. More than 73.7 % subjects voted a high or very high realism for the global system as well as its hardware, visual display, virtual working space, and virtual instruments. In particular, the ratio was 94.7 % for the visual display.

Only eight (42.1 %) persons thought the interaction between instruments and objects were of high or very high realism. Nine others thought it to be moderately realistic (Table 3).

Content validity

Data were derived from 14 subjects. More than 78.6 % persons considered the three virtual reality exercises were of high or very high interest to teach the skills that they aimed at. Most subjects thought that XTT was interesting for bed assistant (85.7 %) and robotic surgical teamwork (100 %) training. Similarly, a large majority of subjects agreed that XTT should be implemented into the training programs of residents, operating room nurses, and novice robotic surgeons (Table 3).

Construct validity

The group RA and the group LS achieved similar scores on the FLS tasks (112.6 ± 55.2 vs. 100.9 ± 36.0), and both groups performed better than the control group (39.9 ± 31.0) ($p < 0.05$).

Both the groups RA and LS performed better than the control group in nearly all metrics in each of three XTT exercises, although statistical significance was not achieved in all comparisons (Table 4). The group RA performed better than the group LS in MB1 in all metrics but the instrument collisions, whereas statistical significance was not achieved (Table 4). Similar performance superiority of RA over LS was not clearly observed in RWR and CRW (Fig. 3).

Data were averaged cross the three XTT exercises. The combined data demonstrated a similar tendency: Both the groups RA and LS performed better than the control group; a slight superiority of performance of the group RA over LS was observed in nearly all metrics, but statistical significance was only achieved in the excessive instrument force ($p < 0.05$) (Table 4; Fig. 4).

The control group declared a heavier workload (with a weighted rating of 74.3 ± 13.3) versus the group RA (53.3 ± 22.5) and LS (57.1 ± 16.7), but statistical significance was not achieved (Fig. 5).

Concurrent validity

With the combined data cross exercises, each XTT measure demonstrated correlation with at least one GOALS item ($p < 0.05$) (Table 5). For the separate exercises, a strongest correlation of performances in the XTT and da Vinci tests was determined in MB1, followed by CRW. The

Table 3 Face validity (realism) and content validity (interest)

	Very high	High	Average	Low	Very low	≥High (%)
Face validity ($n = 19$)						
MB1	2	11	6	0	0	68.4
RWR	5	12	2	0	0	89.5
CRW	7	6	5	1	0	68.4
Global system	4	11	3	1	0	79.0
Hardware	8	6	4	1	0	73.7
Visual display	6	12	1	0	0	94.7
Virtual working space	5	11	2	1	0	84.2
Virtual instruments and objects	4	10	3	2	0	73.7
Interaction of instruments with objects	2	6	9	2	0	42.1
Content validity ($n = 14$)						
MB1	5	6	3	0	0	78.6
RWR	4	9	1	0	0	92.9
CRW	6	6	2	0	0	85.7
Interest of XTT for bedside assistant training	7	5	2	0	0	85.7
Interest of XTT for teamwork training	7	7	0	0	0	100.0

		Yes (%)	No	No opinion
XTT should be implemented into the training programs of:	Residents	13 (92.9)	1	0
	Operating room nurses	11 (78.6)	1	2
	Novice robotic surgeons	13 (92.9)	0	1

Table 4 Construct validity

	Overall score	Task duration	Instrument motion	Instrument collisions	Excessive force	Out of view	Clipping distance	Incorrect clips
MB1								
RA versus LS	*N*[a]	*N*	*N*	*N*	*N*	*N*	–	–
LS versus control	*N*	0.010	*N*	*N*	*N*	*N*	–	–
RA versus control	0.004	0.0004	0.009	*N*	0.013	0.006	–	–
RWR								
RA versus LS	*N*	*N*	*N*	*N*	*N*	*N*	–	–
LS versus control	*N*	*N*	0.046	*N*	*N*	*N*	–	–
RA versus control	*N*	0.0069	0.029	0.022	0.036	*N*	–	–
CRW								
RA versus LS	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*N*
LS versus control	0.033	0.035	0.004	*N*	*N*	0.046	*N*	*N*
RA versus control	0.002	0.035	0.0007	0.025	*N*	0.002	*N*	0.009
Combined data								
RA versus LS	*N*	*N*	*N*	*N*	0.031	*N*	–	–
LS versus control	0.023	0.004	0.011	0.015	*N*	0.016	–	–
RA versus control	0.002	0.0004	0.002	0.015	0.013	0.0007	–	–

[a] Nonsignificant

correlation was poor in RWR. The weighted ratings of NASA-TLX in the two tests were significantly correlated ($r = 0.835$, $p = 5.83e{-}05$).

The advantages and drawbacks of XTT compared to the da Vinci box trainer described by the subjects are summarized in Table 6.

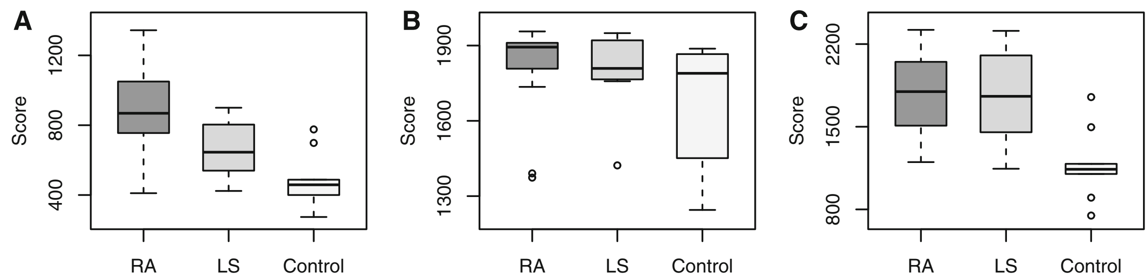

Fig. 3 Overall score of the three groups in the virtual reality exercises of MB1 (**A**), RWR (**B**), and CRW (**C**)

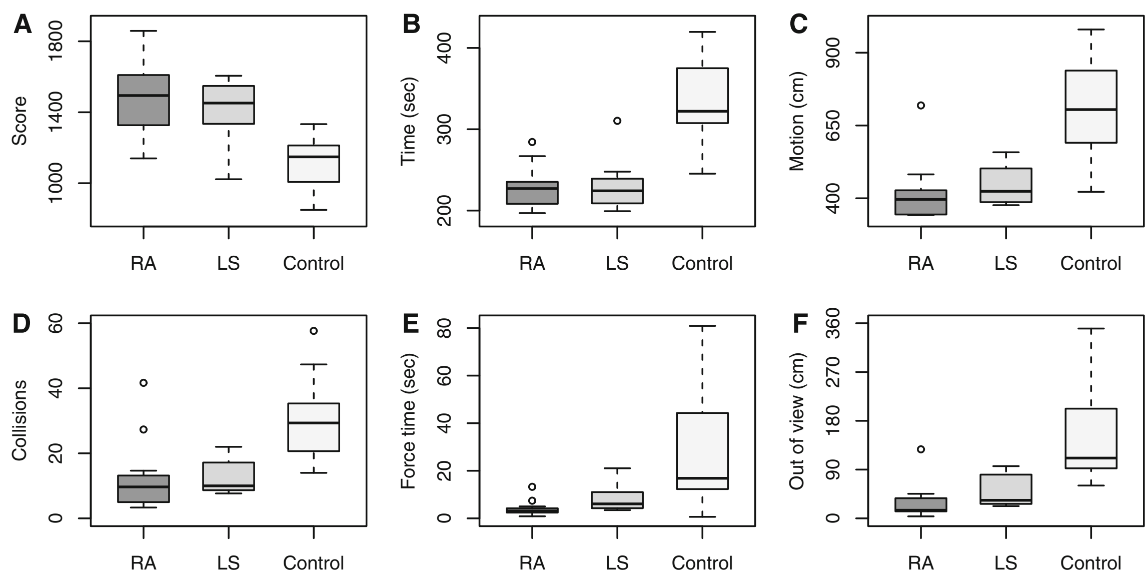

Fig. 4 Combined data cross the three virtual reality exercises: overall score (**A**), task completion time (**B**), instrument motion (**C**), instrument collisions (**D**), excessive instrument force (**E**), and instruments out of view (**F**)

Test–retest reliability

The correlation of performances in the second and third attempts was determined in all metrics of MB1, and in the metrics of task completion time and instrument collisions of all the three exercises ($p < 0.05$) (Table 7).

Inter-rater reliability

Correlation of the average GOALS scores evaluated by the two independent evaluators was determined ($r = 0.661$, $p = 7.249e{-}08$).

Discussion

This study represents the first effort to determine the face, content, construct, and concurrent validity of the new XTT simulator as an assessment tool of robotic surgical bed-assistance skills.

The XTT simulator demonstrates face validity, but the training modules remain to be improved. Except the interaction between instruments and objects, the realism of the global system and most of its functions was approved. Of the three exercises, RWR was thought to be the most realistic. In MB1, the motion of instruments was sensitive or somewhat amplified. The instrument tips also felt too soft. In CRW, bugs were occasionally encountered during the charge and the deployment of clips. That is why half of the subjects voted an average realism for the instrument–object interaction.

Most expert subjects agreed on the interest of XTT for bed-assistant training, and especially all the subjects appreciated its interest for teamwork training, and thus proved content validity. Unlike in open or laparoscopic surgery, robotic surgeons are un-scrubbed and physically separated to the patient. This setting has increased the importance and responsibility of bed assistants, and an effective teamwork between surgeons and assistants becomes crucial. Experienced robotic surgeons make full

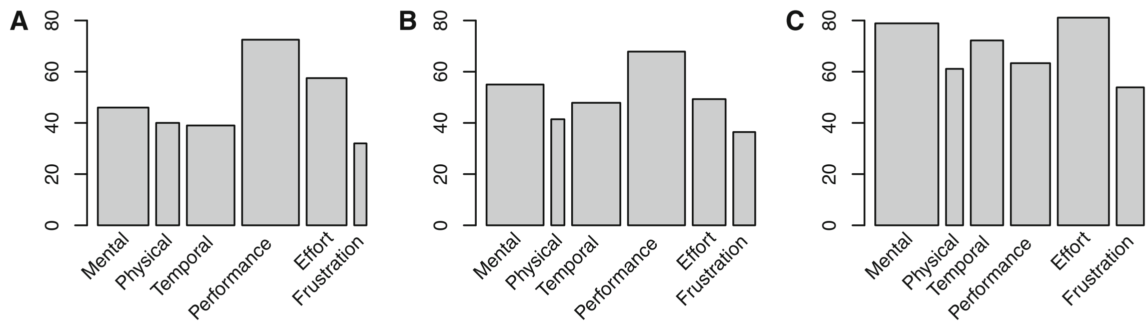

Fig. 5 Results of NASA-TLX in the XTT session of the group RA (**A**), LS (**B**), and control (**C**) The width of the subscale bars reflects the importance of each factor (weight), and the height represents the magnitude of each factor (rating)

Table 5 Concurrent validity (combined data cross exercises)

XTT measures	*r*	*p*	Items of GOALS
Overall score	0.476	0.054	Average performance
	0.496	0.043	Depth perception
	0.514	0.035	Tissue handling
Task completion time	−0.661	0.004	Average performance
	−0.733	0.001	Depth perception
	−0.586	0.013	Efficiency
	−0.691	0.002	Tissue handling
	−0.487	0.047	Effectiveness
Instrument motion	−0.700	0.002	Average performance
	−0.718	0.001	Depth perception
	−0.529	0.029	Efficiency
	−0.699	0.002	Tissue handling
	−0.604	0.010	Effectiveness
Instrument collisions	−0.577	0.015	Average performance
	−0.608	0.010	Depth perception
	−0.628	0.007	Tissue handling
	−0.492	0.045	Effectiveness
Excessive instrument force	−0.574	0.016	Average performance
	−0.631	0.007	Depth perception
	−0.566	0.018	Tissue handling
	−0.543	0.024	Effectiveness
Instruments out of view	−0.504	0.039	Tissue handling

use of their advantages of the 3D vision and the EndoWrist® instruments, and give appropriate commands to their assistants. Experienced bed assistants keep fine balance between over- and under-activity and provide suggestions at the right moment. The success of robotic procedures relies on the expertise of both surgeons and assistants, and this would be unrealizable without adequate training. Current robotic curricula usually provide independent training for surgeons and bed assistants. The wet-laboratory (animal) procedures permit cooperation of the two roles, but the access is limited with considerable costs. The connection of XTT with dVT provides a valuable platform for the robotic surgeons and their assistants to train together. This is an innovation that indicates a new paradigm of robotic surgery training. In order to meet the needs of practical applications, the training modules of XTT need to be further developed. Some basic techniques should be included (e.g., introduction of instruments, suction and irrigation, needle control), and more complex exercises will be needed.

For construct validity, the XTT exercises succeeded to differentiate subjects with or without laparoscopic experience, but were not discriminative enough to highlight robotic assistants. The group RA and the group LS demonstrated a similar level of basic laparoscopic skills. The group RA then had additional robotic bed-assistance experience (with a median of 25 cases). There is no consensus about how many cases a bedside assistant would need to achieve proficiency, but a minimum of 10–20 cases was suggested [7, 11]. Both the groups RA and LS performed better than the control group on the XTT exercises, but significant difference was not determined between the former two groups themselves. Laparoscopic but not robotic bed-assistance experience seemed to be more important as a determinant of performance. In the respective exercises, differences between the groups RA and LS were actually observed in MB1. The statistical non-significance could be results of the small sample size. Similar differences were less clear in RWR and CRW. These two exercises demand little cooperation and may be too simple for surgeons with laparoscopic experience. Proficient basic laparoscopic skills do not guarantee a qualified bedside assistant. More complex and discriminative exercises should be developed for XTT and further validated in the future.

Concerning the NASA-TLX results, the groups LS and especially RA tended to pursue better performance and suffered less stress. The control group had no much spare to care about the performance. They suffered significant mental pressure and made more efforts to complete the

Table 6 Advantages and drawbacks of XTT compared to the da Vinci® box trainer

Advantages	Drawbacks
Easily assessable, without requirement of robot; easier and less time to implement; needs less space	Lack of force feedback; difficult to discover instrument collisions
Real-time and graphic feedback of performance with multiple measures; comprehends better the temporal and mental demands	Less realistic:
	does not reproduce easily the real condition; difference of manipulation with actual experience;
Reproducibility of the exercises	amplified instrument motion in MB1;
More intuitive with visual effects	the instrument tips feel too soft in MB1;
Comfort of installation and manipulation; less tiring	some anomalies in the behaviors of instruments or objects in CRW;
High-resolution and beautiful images	Some errors in visual representation in CRW
Less physical constraints (due to robotic arms)	Some resistance during the movement of grips
Easier to communicate with the console operator	Instruments already in place, without the step of introduction
More exercises and the capacity of further development	

Table 7 Test–retest reliability between the second and third attempts

	Overall score	Task duration	Instrument motion	Instrument collisions	Excessive force	Out of view	Clipping distance	Incorrect clips
MB1	2.25e−07*	0.0003*	3.83e−06*	1.29e−08*	0.010*	4.16e−08*	–	–
	(0.864)	(0.713)	(0.910)	(0.899)	(0.538)	(0.886)		
RWR	0.189	0.004*	0.080	0.042*	0.238	0.155	–	–
	(0.284)	(0.586)	(0.374)	(0.427)	(0.256)	(0.307)		
CRW	0.073	0.002*	0.184	0.005*	0.066	0.157	0.080	0.497
	(0.400)	(0.645)	(0.301)	(0.590)	(0.409)	(0.320)	(0.392)	(0.157)

Results presented in forms of p value (r)

Asterisk indicate the correlation is statistically significant ($p < 0.05$)

tasks (Fig. 5). Similarly, significant difference was not determined between the groups RA and LS.

In terms of concurrent validity, correlation of performance was determined between the tests on XTT and da Vinci box trainer. It is the task completion time and the instrument motion but not the overall score that proved a better correlation with the GOALS items. This is not surprising since, according to our experience with dVT, the former two measures are more sensitive and precise in the assessment of performance. For the respective exercises, the correlation was strongest in MB1 and poor in RWR. Still, RWR is too simple to correctly reveal skills or tell differences.

Strictly speaking, we may not define this correlation as "concurrent validity" because the test on da Vinci® box trainer using GOALS is not actually the "gold standard" to assess robotic bed-assistance skills. Both XTT and da Vinci box trainer potentially could become the standard tool, but XTT would be a preferable choice. As described by the subjects, the XTT platform presents some specific advantages: better access, real-time feedback of performance, reproducibility of exercises, etc. These advantages have been widely documented in previous studies concerning virtual reality simulators [12, 13]. The most mentioned disadvantage of XTT was the absence of force feedback, which is currently under development.

The exercise MB1 demonstrates good test–retest reliability. Performance was less stable in RWR and CRW—the two relatively simple exercises in which occasional errors may cause significant differences. Of course, the performance did not reach plateau within three attempts. Results could be better after the achievement of proficiency. The task completion time and the instrument collisions proved to be the most reliable metrics in these exercises. The overall score may not be quite stable since it is calculated in two different ways depending on the existence or not of critical errors.

During robotic surgery, the bedside assistant is directly responsible for the security of patients, especially in case of accidents. Moreover, an experienced assistant could help to optimize workflow, reduce operative time, and improve surgical outcomes [14]. Currently, few institutes possess specialized or permanent bedside assistants. In many cases, it is the residents or fellows who fill the vacancy. The time as bed assistant is believed to be an essential period during the training of robotic surgeons [15, 16]. The training of bed assistants is often informal and inadequate, and the standards of credentialing are absent. Some robotic surgery curricula also take responsibility to train assistants, from docking to practice in animal procedures. However, as mentioned above, the access to wet-laboratory training is limited with considerable costs. The introduction of XTT provides a relatively compact and accessible platform that would facilitate the training and assessment of robotic surgical bed-assistance skills.

Van der Vleuten has defined the utility of an assessment method as a multiplicative function of its reliability, validity, educational impact, cost-effectiveness, and acceptability [17]. This concept emphasizes that each element of the utility should be taken into account when designing an assessment, although the trade-offs could be made depending on the context and purpose of the assessment. We have focused on the validity of XTT in this study, and the reliability and acceptability were also assessed. The results seem positive, while additional attention should be paid to explore the educational impact and cost-effectiveness in the future.

We did not investigate the opinions of console surgeons in this study since a same operator should be on dVT to ensure the comparability. This may be a limitation of the study. Only three XTT exercises were tested due to the workload. Other studies covering more exercises are expected. The relatively small sample size with limited statistical power is another shortcoming. In addition, predictive validity should be determined to prove the translation of virtual reality experience into clinical benefits.

Conclusions

The introduction of XTT facilitates the training of bed assistants and also emphasizes the importance of teamwork. This innovation may change the paradigm of robotic surgery training in the near future. As an assessment tool of bed-assistance skills, XTT proves face, content, and concurrent validity. However, these results should be qualified considering the potential limitations of this exploratory study with a relatively small sample size. The training modules of XTT remain to be developed, and more complex and discriminative exercises are expected. Other studies will be needed to further determine construct validity in the future.

Acknowledgments The authors would like to thank the Ecole de Chirurgie (surgery school) Nancy-Lorraine and the Stan Institute for their generous supports in this study. The first author is supported by the fellowship provided by the China Scholarship Council for the studies at Lorraine University.

Compliance with ethical standards

Disclosures Dr. Hubert J reports non-financial support from Mimic Technologies, during the conduct of the study: The XTT simulator was provided by Mimic Technologies for the education and the study during the DIU curriculum; Drs. Xu, Perez, Perrenot, and Hubert N have no conflicts of interest or financial ties to disclose.

References

1. Intuitive Surgical—Investors—Investor Relations Home. http://investor.intuitivesurgical.com/phoenix.zhtml?c=122359&p=irol-irhome. Accessed 13 May 2015
2. Autorino R, Zargar H, Kaouk JH (2014) Robotic-assisted laparoscopic surgery: recent advances in urology. Fertil Steril 102:939–949. doi:10.1016/j.fertnstert.2014.05.033
3. Diana M, Marescaux J (2015) Robotic surgery. Br J Surg 102:e15–e28. doi:10.1002/bjs.9711
4. Fisher RA, Dasgupta P, Mottrie A, Volpe A, Khan MS, Challacombe B, Ahmed K (2015) An over-view of robot assisted surgery curricula and the status of their validation. Int J Surg 13:115–123. doi:10.1016/j.ijsu.2014.11.033
5. Buchs NC, Pugin F, Volonté F, Morel P (2013) Learning tools and simulation in robotic surgery: state of the art. World J Surg 37:2812–2819. doi:10.1007/s00268-013-2065-y
6. Smith R, Truong M, Perez M (2015) Comparative analysis of the functionality of simulators of the da Vinci surgical robot. Surg Endosc 29:972–983. doi:10.1007/s00464-014-3748-7
7. Sheridan C, Bach C, Koupparis A (2014) How to train your surgeon! Experience of a patient side assistant. Arab J Urol 12:62–63. doi:10.1016/j.aju.2013.08.015
8. Sgarbura O, Vasilescu C (2010) The decisive role of the patient-side surgeon in robotic surgery. Surg Endosc 24:3149–3155. doi:10.1007/s00464-010-1108-9
9. Xu S, Perez M, Yang K, Perrenot C, Felblinger J, Hubert J (2014) Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer® simulator. Surg Endosc 28:2569–2576. doi:10.1007/s00464-014-3504-z
10. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM (2003) Evaluating laparoscopic skills. Surg Endosc Interv Tech 17:964–967. doi:10.1007/s00464-002-8828-4
11. Guzzo TJ, Gonzalgo ML (2009) Robotic surgical training of the urologic oncologist. Urol Oncol Semin Orig Investig 27:214–217. doi:10.1016/j.urolonc.2008.09.019
12. Nagendran M, Gurusamy KS, Aggarwal R, Loizidou M, Davidson BR (2013) Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 8:CD006575. doi:10.1002/14651858.CD006575.pub3
13. Abboudi H, Khan MS, Aboumarzouk O, Guru KA, Challacombe B, Dasgupta P, Ahmed K (2012) Current status of validation for robotic surgery simulators—a systematic review. BJU Int 111:194–205
14. Renaud M, Reibel N, Zarnegar R, Germain A, Quilliot D, Ayav A, Bresler L, Brunaud L (2013) Multifactorial analysis of the learning curve for totally robotic Roux-en-Y gastric bypass for morbid obesity. Obes Surg 23:1753–1760. doi:10.1007/s11695-013-1020-1
15. Lee JY, Mucksavage P, Sundaram CP, McDougall EM (2011) Best practices for robotic surgery training and credentialing. J Urol 185:1191–1197. doi:10.1016/j.juro.2010.11.067

16. Ahmed K, Khan R, Mottrie A, Lovegrove C, Abaza R, Ahlawat R, Ahlering T, Ahlgren G, Artibani W, Barret E, Cathelineau X, Challacombe B, Coloby P, Hubert J, Khan MS, Michel MS, Montorsi F, Murphy D, Palou J, Patel V, Piechaud P-T, Van Poppel H, Rischmann P, Sanchez-Salas R, Siemer S, Stoeckle M, Stolzenburg JU, Terrier JE, Thueroff J, Christophe CV, Van Der Poel HG, Van Cleynenbreugel B, Volpe A, Wagner C, Wiklund P, Wilson T, Wirth M, Witt J, Dasgupta P (2015) Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts. BJU Int 116:93–101. doi:10.1111/bju.12974
17. Van Der Vleuten CP (1996) The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1:41–67