Surg Endosc (2003) 17: 803–810
DOI: 10.1007/s00464-002-9151-9

SURGICAL ENDOSCOPY
and Other Interventional Techniques

Construct validity

Experts and novices performing on the Xitact LS500 laparoscopy simulator

M. Schijven, J. Jakimowicz

Department of Surgery, Catharina Hospital Eindhoven, Michelangelolaan 2, PO Box 1350, 5602 ZA Eindhoven, The Netherlands

Received: 22 July 2002/Accepted: 12 September 2002/Online publication: 17 February 2003

Abstract
Background: This study was undertaken to establish construct validity of the Xitact LS500, a virtual reality laparoscopic cholecystectomy simulator. The primary research statement is: "The clip-and-cut task on the Xitact LS500 virtual reality laparoscopic cholecystectomy simulator mimics the surgical procedure of the clipping and cutting of the cystic duct and artery during the laparoscopic cholecystectomy adequately."
Methods: According to the level of experience of the surgeon, an "expert group opinion" was formed resulting from 37 surgeons having performed over 100 laparoscopic cholecystectomies, and a "novice group opinion" was formed resulting of 37 surgeons having no experience at all with performing laparoscopic cholecystectomy. Participants received an instructed hands-on "tour" on the Xitact simulator and performed three formal simulation runs.
Results: The "novice group" is younger and more surgeons are female. Performance scores in the "expert" group are significantly higher on the second (p value 0.011) and third (p value 0.005) run, compared to the novices' scores. Experts are significantly faster on completion of all three runs. There is an increase in score over runs in both groups, which is statistically significant in the "expert group." Less than one-third of surgeons in either group are able to correctly predict their performance score as generated by Xitact. Both "experts" and "novices" feel it is useful to train with Xitact LS500 in the surgical curriculum.
Conclusions: Three hypotheses, formulated to operationalize the primary research statement, could be answered affirmatively. Although further validation studies are needed, the Xitact LS500 simulator seems to be able to discriminate between expert and novice surgeons in this research setting, and thus the construct for this setting is considered to be valid.

Correspondence to: M. Schijven

Key words: Surgical skills — Surgical simulation — Virtual reality — Laparoscopy — Assessment — Construct validity — Validation study

Surgical skills, and in particular, complex psychomotor skills as needed in laparoscopic surgery, are in part innate, and can in part be learned from extensive, repetitive practice [25]. Although many skills and traits are needed to be a competent surgeon, the element of technical competence is eminent. However, the teaching and testing of technical skill is known to be of the least systematic or standardized components in the classic surgical curriculum [21]. Shortened working weeks, tight surgical curricula, and, more recently, medicolegal considerations stress the acquisition of surgical skill. This dilemma, combined with the fast evolving capacities and implementation of the computer in medicine, makes virtual reality in surgical training and education a phenomenon gaining rapid interest of the surgical community. The greatest power of virtual reality medical simulation is the opportunity to try and fail without consequence for the patient [23]. Furthermore, it may satisfy the need for accurate and objective assessment of technical process and skill acquisition, as for the moment, there is in fact none [19]. The nature of laparoscopic surgery makes it likely to benefit from developments in virtual reality [3]. Previous studies have shown potential and interest for virtual reality simulators in the field of laparoscopic surgery in terms of tutoring, training, and assessment of skill and performance [2, 4, 6, 7, 10, 11, 13, 15, 26, 28, 30].

A *valid* virtual reality simulator provides an environment that closely approximates the characteristics of the environment in which the task eventually will be performed [20]. It must be able to mimic visual–spatial and real-time characteristics of the procedure and to provide realistic haptic feedback. Besides, it must be able to evaluate the performance of the procedure under study objectively. Only then will the nonstressful environment of the virtual training setting enhance both level of skill and level of confidence of the trainee and lead him or her gently into a more advanced state of performance. The Xitact LS500, being one of the latest virtual reality training simulators combining realistic anatomical scenery with haptic feedback, may be an effective tool for teaching and testing laparoscopic surgical skills.

One of the most important research lines in the development of the Xitact is the attention to the psychometric properties of the instrument, that is, to its validity and reliability. The concept of validity addresses the question, "Do we measure what we intend to measure?" [16]. Basically, valid innovations are innovations with low nonrandom (systematic) errors. Validity can be categorized into different types: face, content, construct, criterion-related, and concurrent validity. The type of validity referred to must relate to the purpose of the concept of interest [29]. The concept of construct validity is often regarded as the central theme in validation studies [5]. Ultimately, a new virtual reality simulator should pass multiple aspects of validity to become a reliable skills trainer and predictor of performance. Previous studies with the Xitact LS500 laparoscopy simulator successfully addressed the issues of face and content validity of the laparoscopic cholecystectomy scene.

The face validity study addressed the question, "To what extent does the Xitact LS500 simulate what it is supposed to simulate, e.g., the procedure of the laparoscopic cholecystectomy?" [24]. Although not a formal validity concept, it refers to the subjective opinion about a test, e.g., about its appropriateness for its intended use of purpose within the target population. Face validity must be considered extremely important for a test's practical utility and success of implementation [14]. The content validation of the clipping-and-cutting task of the Xitact LS500 addressed the question, "Does the simulation measure all relevant dimensions of the construct under study?" Content validation was obtained by thorough search of the literature and by interviewing expert laparoscopic surgeons.

The next step in the validation of the Xitact LS500 laparoscopic simulator is the establishment of the construct validity of the system. Construct validity refers to the concept that a novelty actually mimics what it intends to mimic, by direct or indirect objective standards. It is satisfied when test performance is logical and consistent with parameters of interest [14]. Fundamentally, it is concerned with explaining individual differences in scores among subjects by relating the various outcomes with anticipated ones. A valid system should be able to differentiate between different levels of skill. One way to establish content validity for the Xitact LS500 is comparing scores within and between experts and novices in laparoscopic cholecystectomy.

The current study focuses on the concept of construct validity of the Xitact LS500 laparoscopic cholecystectomy simulator. Translated to our setting, the primary research statement is: "The clip-and-cut task on the Xitact LS500 virtual reality laparoscopic cholecystectomy simulator mimics the surgical procedure of the clipping and cutting of the cystic duct and artery during the laparoscopic cholecystectomy adequately."

Fig. 1. Xitact LS500 Simulator.

Three hypotheses were formulated to explore the primary research statement.

First hypothesis: Performance scores derived from the Xitact cholecystectomy clip-and-cut task by experts in clinical laparoscopic cholecystectomy are significantly higher than the performance score derived by novices in clinical laparoscopic cholecystectomy.

Second hypothesis: Performance scores derived from the Xitact cholecystectomy clip-and-cut task are related to the clinical laparoscopic cholecystectomy experience of the participant.

Third hypothesis: Performance scores of experts and novices in clinical laparoscopic cholecystectomy improve over runs of the Xitact cholecystectomy clip-and-cut task.

Materials and methods

Subjects

Fifty-seven attendants of the 8th World Congress of Endoscopic Surgery, held 13–16 March 2002 in New York, participated. Also, 34 surgical residents attending the Basic Surgical Skills Courses in the Academic Hospital of Leiden or the Catharina Hospital in Eindhoven, The Netherlands, held May 2002, participated. All participants were given an instructed one-on-one "tour" to familiarize them with Xitact LS500, its features, and the laparoscopic cholecystectomy simulation. The "tour" also featured a voice-over instruction video of the task under study, e.g., the clipping and cutting of the cystic artery and cystic duct. Trained instructors gave a detailed explanation of possible errors

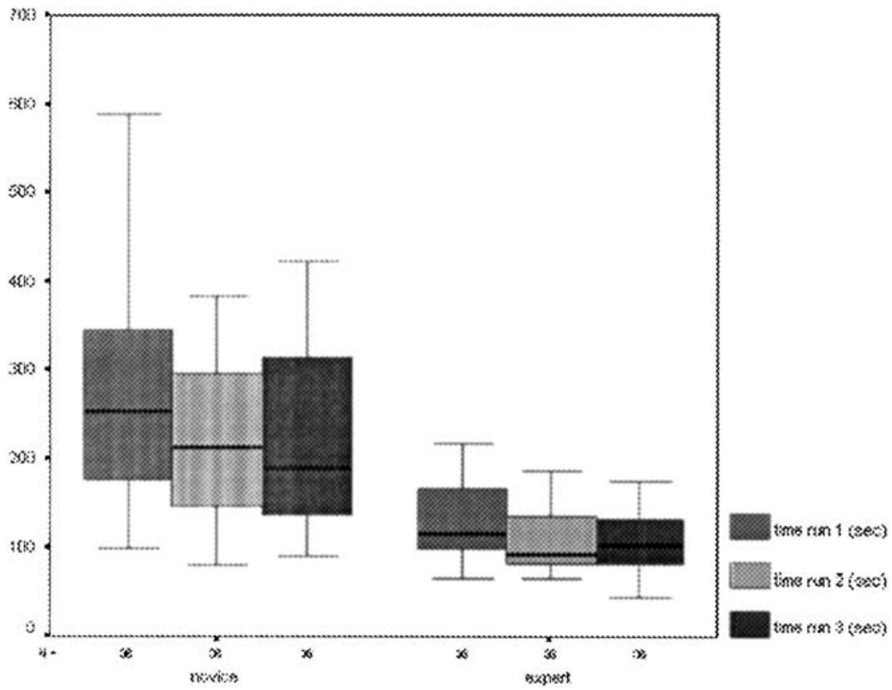

Fig. 2. Time for completion of runs.

in the procedure, as well as hands-on instruction in the task under study. Three runs were performed by each participant, as earlier studies with laparoscopic surgical virtual reality trainers showed an early familiarization with the scenery within three runs [2]. Afterward, participants were asked to complete a questionnaire on virtual skills training and testing. For the purpose of this study, only surgeons having performed over 100 clinical laparoscopic cholecystectomies were selected for the "expert" group ($n = 37$), and only the residents/interns with no experience in laparoscopic cholecystectomy were selected for the "novice" group ($n = 37$).

Apparatus

The Xitact LS500 Laparoscopy Simulator (Fig. 1) is a modular virtual reality training program developed for training and education of a variety of laparoscopic skills. It is a hybrid simulator, combining a physical object (the OpTable, or "virtual abdomen") with a computer program providing the visual image and haptic feedback.

The program featured and under study is the clipping-and-cutting task of the laparoscopic cholecystectomy simulation. The Xitact LS500 is developed and registered by Xitact SA, Morges, Switzerland.

Composition of sum score

A sum score was computed to estimate the "end result" of participants' performance. Items of interest were time needed for completion of run; number of clips used and placed on cystic artery or duct; number of clips lost; partial closure and overlap for artery and duct; clipping on clips; clipping multiple or in-adverse structures; correct placement of first medial, second medial, and lateral clips on artery and duct; correct location of cut for cystic artery and duct; cutting on clips; and finally, the cutting of multiple or in-adverse structures. The sum score was computed by subtracting weighted and expert-rated error scores for each possible outcome of the procedure, of a maximum score of 100%. Most severely weighted error—resulting in a subtraction of 100%—was clipping or cutting of the common bile duct. Clipping or cutting of the hepatic artery resulted in a subtraction of 80%. The scale was checked for internal consistency using Cronbach's alpha, resulting in an overall score of 0.47 with no negative correlations for the 15 items.

Data were analyzed using the Statistical Package for the Social Sciences, version 9.0.

Questionnaire

A questionnaire consisting of a two-page survey was given to each participant. Next to the participant's demographics and surgical laparoscopic experience, questions concerning the global realism, visual scene, haptic feedback, movement, and surgical techniques of the clip-cut task on the simulator were presented. A 5-point ordinal Likert answering scale was constructed. Value 1 was assigned to "not realistic/good/useful," and value 5 to "very realistic/good/useful." The usefulness of the simulator in terms of training/teaching capacities and error reduction was questioned. Also, participants were asked to score themselves for performance on Xitact.

Results

Demographics

Seventy-four surgeons and surgical residents/interns were selected for in this study, originating from 17 different countries. Of the "expert" group, the majority (32%) were inhabitants of the United States; 10.8% were UK residents. Of the "novice" group, the majority (89.2%) was Dutch. The mean age of the participants in the "expert" group is 44 years, ranging from 28 to 61 years of age. The mean age of the participants in the "novice" group is 29 years, ranging from 23 to 58 years. Thus, there is a significant difference in age distribution among groups (p value Student's t-test: 0.000). There are 20 females and 54 males participating. In the "expert" group, the percentage of females is 10.8%; in the "novice" group, 43%. Therefore, sex is unequally represented among groups (p value Pearson's chi-square: 0.002). Dexterity is more or less equally distributed among groups, with 86% of right-handedness in the "expert" group and 73% of right-handedness in the "novice" group.

Of all experts, 97.2% work as general surgeons and 2.8% as pediatric surgeons. Of the novices, 17% are interns, 21.6% are in training to be general surgeons,

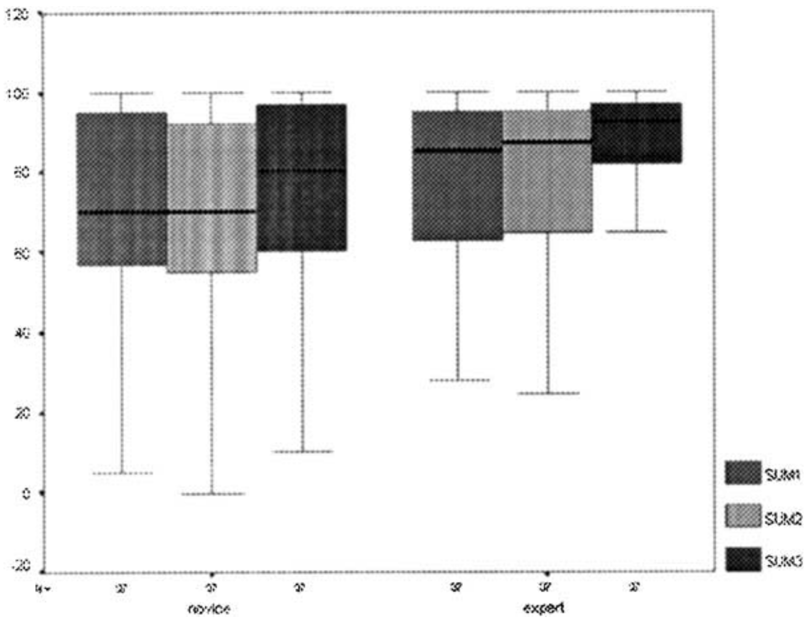

Fig. 3. Sum scores unclusted.

Table 1. Ratings (1, not realistic, to 5, very realistic)

	Total Mean	Novice group (0 lapchol) Mean	Novice group (0 lapchol) SD	Expert group (> 100 lapchol) Mean	Expert group (> 100 lapchol) SD	p-value[a] $n = 37/n = 37$
Global realism of the clipping-and-cutting task in Xitact	3.72	3.74	0.51	3.70	0.70	1.000
Visual realism of the clipping-and-cutting task in Xitact	3.62	3.55	0.73	3.68	0.75	1.000
Reality of haptic feedback (tactile sensation) of the clipping-and-cutting task in Xitact	3.42	3.67	0.76	3.17	1.08	0.124
Reality of movement of organs/structures in the clipping-and-cutting task in Xitact upon grasping or manipulating	3.63	3.65	0.88	3.62	0.79	0.988
Reality of exposure	3.92	3.93	0.61	3.92	0.77	0.947
Reality of clipping	3.92	3.91	0.67	3.92	0.72	1.000
Reality of cutting	3.87	3.82	0.72	3.92	0.86	0.997

[a] Kolmogorov–Smirnov test, two-tailed

10.8% are in training to be emergency medicine specialists, 8.1% are in training to be orthopedic surgeons, 2.7% are in training to be thoracic surgeons, and 2.7% are in training to become urologists. There are three qualified specialists in this group: a gynecologist, a general surgeon, and a thoracic surgeon.

Face validity

Table 1 refers to the similarity of the Xitact laparoscopic cholecystectomy environment to surgeon's actual experience with or perception toward clinical laparoscopic cholecystectomy. There is no significant difference in opinion regarding the perception of Xitact's reality.

Construct validity

Figure 2 refers to the time needed for completion of the clip-and-cut task. Comparing means using Student's t-testing, groups differ significantly in time needed for completion of each run (p value of 0.000, 0.000, and 0.000, respectively).

Figure 3 refers to the end-parameter ''score,'' which was calculated after completion of the clip-and-cut task. Mean scores of the novice group were 64.6 for run 1, 67.6 for run 2, and 71.8 for run 3. Mean scores of the expert group were 77.6 for run 1, 77.6 for run 2, and 86.4 for run 3. Scores were clustered in order to construct five categories of performance (Table 2). Over runs, scores between novices and experts seem to differ increasingly, and are statistically significant for the second (p value 0.011) and third (p value 0.005) run. Within groups, there is an increase in scores over the sequence of three runs (Table 3). In the expert group, because of smaller standard deviations, this increase is statistically significant (p value 0.041).

Questionnaire

Both novices and experts feel the clip-and-cut task of the Xitact laparoscopic cholecystectomy setting is very

Table 2. Comparine novices and experts over runs between groups

Sum score	Category	Label	Run 1 Novices	Run 1 Experts	Run 2 Novices	Run 2 Experts	Run 3 Novices	Run 3 Experts
0–20	I	Very inadequate performance	21.6	0.0	8.1	5.4	5.4	0.0
21–40	II	Inadequate performance	0.0	2.7	8.1	5.4	5.4	2.7
41–60	III	Questionable performance	10.8	21.6	16.2	5.4	21.6	2.7
61–80	IV	Adequate performance	24.3	21.6	32.4	16.2	16.2	18.9
81–100	V	Highly adequate performance	43.2	54.1	35.1	67.6	48.6	75.7
Kruskall–Wallis over categorized sum score Asymp. Sig. (2-tailed)				Sum1 **0.180**		Sum2 0.011*		Sum3 0.005*

Table 3. Comparing novices and experts over runs within groups

Run	Category	Novices, mean rank	Experts, mean rank
1	Sum 1 categorized	1.99	1.78
2	Sum 2 categorized	1.92	2.00
3	Sum 3 categorized	2.09	2.22
Friedman test over categorized sum score Asymp. Sig. (2-tailed)		**0.648**	**0.041***

useful for surgical residents' training and education. To a somewhat lesser extent, they feel the simulator is useful for medical students and surgical specialists. Xitact is considered to be a useful instrument for laparoscopic error reduction and skills enhancement (Table 4). Experts and novices are uniform in their opinion, as there seems to be no significant difference in responses. Asked for their own performance on the clip-and-cut task, the self-reported and categorized score in the novice group does not seem to differ significantly from the Xitact generated categorized score (Table 5). In the expert group, there is a difference, e.g., experts tend to underestimate their own performance. For both groups, less than one-third were able to predict their score accurately. Novice and expert groups do not differ in their predictive abilities for the Xitact score.

Discussion

Evaluating validity is an important and often neglected part of the development of innovations. It is very important to realize that we cannot study well what we cannot measure well. To date, limited data exist concerning the assessment of surgical simulators in the teaching environment. Before surgical simulators can be used to train and assess surgical skill formally, they must be clearly shown to be reliable and valid [19]. Fortunately, more and more studies are becoming available concerning the validation of new virtual reality and simulation system in teaching and training of surgical (laparoscopic) skills [11, 19].

Unique to the concept of Xitact's development is that the software is being build and refined not only by software engineers. Assumptions and boundaries of the simulation are determined by a board of experts within the laparoscopic surgical community. Validation of the simulation is being assessed per scene (inspection, clip-and-cut, dissection) so that important adjustments can be made before there is a fixed end product. Constructing a sum score for the purpose of this study and for evaluation purposes has been an arbitrary procedure in itself. The clip-and-cut exercise is a multidimensional procedure, as it demands multiple skills and insights of the participants. The internal consistency of the items making up the sum score, as measured by Cronbach's alpha, is therefore relatively low. Deletion of items out of the scale would have increased alpha, but items were considered to be essential for the scale and contributing to the end score.

The current study compares the results of an expert group of laparoscopic surgeons with a novice group, performing the clip-and-cut task on the Xitact laparoscopic cholecystectomy simulator. Age and sex were unequally distributed among groups, but this is more or less inevitable as novices tend to be younger and nowadays, more females are in their surgical residences than ever before. Briefly, the face validity of the clip-and-cut task was assessed. A face validity study concerning the overall realism of the Xitact laparoscopic cholecystectomy scenery has been performed, but authors felt it would be appropriate for the subtask of the clip-and-cut to look into this concept even further [24]. Again, there is a favorable opinion on the realism of the scene, and ratings for the clip-and-cut task are even higher (3.72 for global realism of clip-cut versus 3.42 for global realism of the visual laparoscopic cholecystectomy scene). The parameter of haptic feedback is improved (3.42 for clip-cut in this study versus 3.07 in the previous study). Respondents of both groups value Xitact's use in the surgical resident training and education curriculum highly.

The main focus of this study was the concept of the content validity. For this, the hypotheses-of-research generated beforehand were as follows.

Table 4. Usefulness of Xitact's clip-and-cut scene for skills training/error reduction

Ratings (1, not useful, to 5, very useful)	Total mean	Novice group ≤ 100 Mean	SD	Expert group > 100 Mean	SD	p value $n = 37$ / $n = 37$
Medical student training/education	3.99	4.05	1.12	3.91	1.22	0.994
Surgical residents training/education	4.44	4.65	0.55	4.27	0.90	0.399
Surgical specialist training/continuing education	3.78	3.93	0.89	3.65	1.16	0.997
Overall laparoscopic error reduction	3.91	4.03	0.67	3.81	0.94	0.765
Overall laparoscopic skills enhancement	3.98	4.20	0.66	3.81	1.04	0.884

[a] Kolmogorov–Sminov test, two-tailed

Table 5. Comparing self-reported scores with categorized sum score

	Novices			Experts	
	$n = 35$	%		$n = 34$	%
Within groups					
Self-reported end score categorized < sum 3 categorized	17	48.6		21	61.8
Self-reported end score categorized = sum 3 categorized	8	22.9		10	29.4
Self-reported end score categorized > sum 3 categorized	10	28.6		3	8.8
Wilcoxon signed ranks test Asymp. Sig. (2-tailed)	0.214			0.001*	
Between groups					
Mann–Whitney *U*					
Asymp. Sig. (2-tailed)			0.170		

First hypothesis: Performance scores derived from the Xitact cholecystectomy clip-and-cut task by experts in clinical laparoscopic cholecystectomy are significantly higher than the performance score derived by novices in clinical laparoscopic cholecystectomy.

In other words: do experts score higher on the Xitact runs than novices do? There is indeed a difference in the summative score between experts and residents. Figure 3 and Table 2 provide evidence that on runs 2 and 3, groups differ significantly. It can be seen that the confidence intervals in both groups are wide, especially in the novice group. This phenomenon is to be expected, as in the novice group, surgeons are inexperienced with the laparoscopic environment. As procedural knowledge and instrumental familiarization was set at the same level for both groups by one-on-one training and a familiarization run, the difference in confidence intervals and mean score must be made up by a combination of abilities and experience.

Second hypothesis: Performance scores derived from Xitact cholecystectomy clip-and-cut task are related to the clinical laparoscopic cholecystectomy experience of the participant.

Unless younger or female surgical residents (more present in novice group) lack native abilities more than older and male surgeons (more present the experienced group), the actual experience with the clinical laparoscopic cholecystectomy must primarily be the explanation for the higher scores. The latter is supported by the significant difference in time needed to complete the runs, as is shown in Fig. 2.

Hypothesis three: Performance scores of experts and novices in clinical laparoscopic cholecystectomy improve over runs of the Xitact cholecystectomy clip-and-cut task.

Table 3 shows an increase in scores over the sequence of three runs, for both the expert and the novice groups. The expert group shows statistically significant improvement over runs. This means that, although there has been a familiarization protocol for both groups, the learning curve associated with the laparoscopic environment itself probably weighs heavily on the novice group.

The three hypotheses, generated for indirect operationalization of the primary research statement "The clip-and-cut task on the Xitact LS500 virtual reality laparoscopic cholecystectomy simulator mimics the surgical procedure of the clipping and cutting of the cystic duct and artery during the laparoscopic cholecystectomy adequately," can be answered affirmatively. Therefore, the assumption must be that the Xitact LS500 laparoscopic simulator is in fact mimicking the laparoscopic cholecystectomy.

The need for accurate and appropriate assessment for surgical trainees is well recognized [18]. Technical skill assessment in a residency program is mainly based upon subjective judgments [21], and is therefore in itself inconsistent. Summative assessments in terms of paper examinations tend to focus on theoretical knowledge. Self-reported progress in skill is extremely susceptible to bias. Our study shows that fewer than one-third of either population estimates performance correctly. Questionnaires, as a form of formative assessment, are more structured, but most of them do not seem to assess all areas in surgical performance, particular not in psychomotor skills. In general, questionnaires are seldom properly evaluated. The questionnaire that is tested for reliability and validity still has the subjective component of a surgeon that is ranking the score [22]. In general, because of the subjective character of a questionnaire, it is prone to many threats to validity. Therefore, although better than no means for structured evaluation at all, it is unjust to state firm conclusions regarding a surgical resident's performance based solely on a questionnaire. Xitact, being a computer, is able to assess a resident's performance objectively as it is relatively free of subjective assessment and nonrandom error.

Criterion-related validation studies address the presence of evidence of a relationship between scores on a research instrument, e.g., the Xitact LS500, and related concepts which can be considered the criterion, such as the MIST-VR. So far, the MIST-VR is the only laparoscopic virtual reality trainer that can act as a criterion, because it is the only surgical virtual reality system which has been reasonably validated. A study by Taffinder et al. established construct validity of the simulator by showing significant differences in score comparing experienced laparoscopic surgeons with trainee surgeons and nonsurgeons [27]. In other settings, similar conclusions were stated [2, 8].

Another study presenting a basic laparoscppic skills trainer the LapSim, has recently shown criterion validity in a small-scale study [12]. Although simulators are not fully alike (both MIST-VR and LapSim are simulators lacking haptic feedback and anatomical representation of a laparoscopic procedure) and probably more pure psychomotor testers than Xitact LS500, a study addressing criterion-related validity would be interesting.

Virtual reality simulation can become a promising and potent tool for surgical skills training and testing, but only when property validated [20]. The end goal of the validation process refers to the concept of instructional effectiveness, e.g., concurrent validity. Practice on a VR module must ultimately improve user's performance of the actual skill.

Recent studies with the MIST-VR both failed [19] and succeeded in showing concurrent validity of the simulator [9]. Concurrent validation requires correlation of performance in the Xitact LS500 with skill in vivo, that is, the performance of the laparoscopic cholecystectomy in the operating room. Validity itself is not an all-or-nothing statement, but merely reflects a gradual judgment, depending on the purpose of the measurement and the proper interpretation of the results. A common mistake is the assumption that validity is a characteristic of the system. The term "validity," in fact, refers to the proper interpretation and use of the measurement results of the system. A single instrument may be used for many different purposes, and resulting scores may be more valid for one purpose than for another [5]. For the Xitact LS500, this means that validity statements based on the evaluation of one task can, and probably will be, different on another task. Indeed, it is fundamental for creating a useful VR teaching environment to recognize that a surgical procedure has to be divided into series of steps that can be trained and measured separately [11]. It is unlikely, though, that within one systematically developed system such as the Xitact LS500, one application can be considered valid when another one is not.

It is important to remember that evidence derived from validation studies is never stable; it will in fact fluctuate between settings. In conclusion, repetitive sampling in various settings exploring the multiple dimensions of validity for the various scenes of the Xitact LS500 laparoscopic cholecystectomy simulator is needed for firmer positioning of this virtual reality simulator in the surgical curriculum.

References

1. Ault G, et al. (2001) Exporting a technical skills evaluation technology to other sites. Am J Surg 182: 254–256
2. Chaudhry A, et al. (1999) Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human–computer interface. Ann R Coll Surg Engl 81: 281–286
3. Coleman J, Nduka CC, Darzi A (1994) Virtual reality and laparoscopic surgery. Br J Surg 81: 1709–1711
4. Derossis AM, Antoniuk M, Fried GM (1999) Evaluation of laparoscopic skills: a 2-year follow-up during residency training. Can J Surg 42: 293–296
5. Gaberson KB (1997) Measurement reliability and validity. AORN J 66: 1092–1094
6. Gallagher AG, et al. (1999) Virtual reality training in laparoscopic surgery: a preliminary assessment of Minimally Invasive Surgical Trainer Virtual Reality (MIST VR). Endoscopy 31: 310–313
7. Gallagher A, Crothers I, Satava R (2001) Comprehensive objective assessment of fundamental skills for laparoscopic surgery. In 19th International Congress of the European Association for Endoscopic Surgery, Maastricht, The Netherlands
8. Gallagher AG, et al. (2001) Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg DOI: 10.1007/s00268-001-0133-1
9. Gallagher AG, et al. (2002) From VR to OR: a prospective, randomized, double-blind controlled trial of virtual reality (VR) training for operating room (OR) performance of laparoscopic cholecystectomy. In 10th International Congress of the European Association for Endoscopic Surgery, Centro de Congresses, Lisboa, Portugal
10. Grantcharov TP, et al. (2001) Virtual reality computer simulation: an objective method for the evaluation of laparoscopic surgical skills. Surg Endosc 15: 242–244
11. Grantcharov TP (2002) Teaching and testing surgical skills on a VR laparoscopy simulator—learning curves and effect of previous operative experience on performance. In 10th International Congress of the European Association for Endoscopic Surgery, Centro de Congresses, Lisboa, Portugal
12. Hyltander A et al. (2002) The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc DOI: 10.1007/s00464-001-9184-5
13. Jordan J-A, et al. (2000) A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. Am J Surg 180: 208–211
14. Karras D (1997) Statistical methodology: II. Reliability and validity assessment in study design, part B. Acad Emerg Med 4: 144–147
15. McNatt SS, Smith SGT (2001) A computer-based laparoscopic skills assessment device differentiates experienced from novice surgeons. Surg Endosc 15: 1085–1089
16. Nelson AA (1980) Research design: measurement, reliability and validity. Am J Hosp Pharm 37: 851–857
17. Ota D, et al. (1995) Virtual reality in surgical education. Comput Biol Med 25: 127–137
18. Paisly AM, Baldwin P, Paterson-Brown S (2001) Feasibility, reliability and validity of new assessment form for use with basic surgical trainees. Am J Surg 182: 24–29
19. Paisly AM, Baldwin P, Paterson-Brown S (2001) Validity of surgical simulation for the assessment of operative skill. Br J Surg 88: 1525–1532
20. Prystowski JB, et al. (1999) A virtual reality module for intravenous catheter placement. Am J Surg 177: 171–175
21. Reznick RK (1993) Teaching and testing technical skills. Am J Surg 165: 358–361

22. Reznick R, et al. (1997) Testing technical skills via an innovative "Bench Examination." Am J Surg 173: 226–230
23. Satava RM (2001) Accomplishments and challenges of surgical simulation. Dawning of the next-generation surgical education. Surg Endosc 15: 232–241
24. Schijven MP, Jakimowicz J (2002) Face, expert, and referent validity of the Xitact LS500 Laparoscopy Simulator. Surg Endosc DOI: 10.1007/s00464-001-9229-9
25. Schijven MP, Jakimowicz J, Schot C (2002) The Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) objectifying subjective psychomotor test performance. Surg Endosc DOI: 10.1007/s00464-001-9146-y
26. Taffinder N, et al. (1998) Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomized controlled studies using the MIST VR laparoscopic simulator. In: Westwood JD, et al. (2000) (ed) Technology and Informatics 50: Proceedings of Medicine Meets Virtual Reality, 28–31 January 1998 IOS Press and Ohmsha Amsterdam, San Diego
27. Taffinder NJ, et al. (1998) An objective assessment of surgeons' psychomotor skills: validation of the MIST-VR laparoscopic simulator. Br J Surg 85 (Suppl 1): 75
28. Torkington J, et al. (2001) Skill transfer from virtual reality to a real laparoscopic task. Surg Endosc 15: 1076–1079
29. Whitney JD (1999) Notes on methodology. Validity and reliability: basic research concepts. J WOCN 26: 240–249
30. Wilson MS, et al. (1997) MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. Ann R Coll Surg Engl 79: 403–404