REVIEW

Validation and implementation of surgical simulators: a critical review of present, past, and future

B. M. A. Schout · A. J. M. Hendrikx ·
F. Scheele · B. L. H. Bemelmans · A. J. J. A. Scherpbier

Abstract

Background In the past 20 years the surgical simulator market has seen substantial growth. Simulators are useful for teaching surgical skills effectively and with minimal harm and discomfort to patients. Before a simulator can be integrated into an educational program, it is recommended that its validity be determined. This study aims to provide a critical review of the literature and the main experiences and efforts relating to the validation of simulators during the last two decades.

Methods Subjective and objective validity studies between 1980 and 2008 were identified by searches in Pubmed, Cochrane, and Web of Science.

Results Although several papers have described definitions of various subjective types of validity, the literature does not offer any general guidelines concerning methods, settings, and data interpretation. Objective validation studies on endourological simulators were mainly characterized by a large variety of methods and parameters used to assess validity and in the definition and identification of expert and novice levels of performance.

Conclusion Validity research is hampered by a paucity of widely accepted definitions and measurement methods of validity. It would be helpful to those considering the use of simulators in training programs if there were consensus on guidelines for validating surgical simulators and the development of training programs. Before undertaking a study to validate a simulator, researchers would be well advised to conduct a training needs analysis (TNA) to evaluate the existing need for training and to determine program requirements in a training program design (TPD), methods that are also used by designers of military simulation programs. Development and validation of training models should be based on a multidisciplinary approach involving specialists (teachers), residents (learners), educationalists (teaching the teachers), and industrial designers (providers of teaching facilities). In addition to technical skills, attention should be paid to contextual, interpersonal, and task-related factors.

B. M. A. Schout · B. L. H. Bemelmans
Department of Urology, VU University Medical Centre
Amsterdam, Amsterdam, The Netherlands

B. M. A. Schout (✉) · A. J. M. Hendrikx
Department of Urology, Catharina Hospital Eindhoven,
P.O. Box 1350, Eindhoven, ZA 5602, The Netherlands
e-mail: barbara.schout@cze.nl

B. M. A. Schout · F. Scheele
Institute for Education and Training, VU University Medical
Centre Amsterdam, Amsterdam, The Netherlands

A. J. M. Hendrikx · A. J. J. A. Scherpbier
Institute for Medical Education, Faculty of Health, Medicine,
and Life Sciences, Maastricht University, Maastricht,
The Netherlands

F. Scheele
Department of Gynaecology, Sint Lucas Andreas Hospital,
Amsterdam, The Netherlands

Abbreviations

AR	Augmented reality
Lap	Laparoscopic
Chol	Cholecystectomy
Perc	Percutaneous procedure(s)
TMS	Training media specification
TNA	Training needs analysis
TPD	Training program design

UCS	Urethrocystoscopy
URS	Ureterorenoscopy
VR	Virtual reality
OSATS	Objective structured assessment of technical skills

Validity Terms

| Subjective validity | Novices' (referents) and/or experts' opinions, e.g., face content expert and referent validity |
| Objective validity | Prospective experimental studies e.g. construct discriminative concurrent criterion and predictive validity |

Validation of surgical simulators in the last two decades

While simulation and simulators have a long history in training programs in various domains, such as the military and aviation, their appearance on the scene of surgical training is more recent [1]. Simulators offer various important advantages over both didactic teaching and learning by performing procedures in patients. They have been shown to prevent harm and discomfort to patients and shorten learning curves, the latter implying that they also offer cost benefits [2–11]. They are tailored to individual learners, enabling them to progress at their own rate [6]. Additionally, learning on simulators in a skillslab environment allows learners to make mistakes. This is important considering that learning from one's errors is a key component of skills development [4, 8, 11]. Apart from their worth as training instruments, simulators can also be valuable for formative and summative assessment [3, 6, 12] because they enable standardized training and repeated practice of procedures under standardized conditions [13].

These potential benefits are widely recognized and there is considerable interest in the implementation of simulators in training programs. It is also generally accepted, however, that simulators need to be validated before they can be effectively integrated into educational programs [5, 6, 14, 15]. Validation studies address different kinds of validity, such as "face," "content," "expert," "referent," "discriminative," "construct," "concurrent," "criterion," and/or "predictive" validity. There is no uniformity in how these types of validity are defined in different papers [15–18]. Additionally, a literature search failed to identify any description of guidelines on how to define and measure different types of validity. Nevertheless, most papers report positive results in respect of all kinds of validity of various simulators. However, what do these results actually reflect?

This paper is based on a review of the literature and the main experiences and efforts relating to the validation of simulators during the last two decades. Based on these,

suggestions are made for future research into the use of simulators in surgical skills training.

Terminology of validation

What exactly is validation and what types of validity can be distinguished? There is general agreement in the literature that a distinction can be made between subjective and objective approaches to validation [15–18]. Subjective approaches examine novices' (referents') and/or experts' opinions, while objective approaches are used in prospective experimental studies. Face, content, expert, and referent validity concern subjective approaches of validity. These types of validity studies generally require experts (usually specialists) and novices (usually residents or students) to perform a procedure on a simulator, after which both groups are asked to complete a questionnaire about their experience with the simulator. Objective approaches concern construct, discriminative, concurrent, criterion, and predictive validity, and these studies generally involve experiments to ascertain whether a simulator can discriminate between different levels of expertise or to evaluate the effects of simulator training (transfer) by measuring real-time performance, for example, on a patient, cadaver or a substitute real-time model.

Subjective approaches to validity (expert and novice views)

A literature search for guidelines on face and content validity yielded several definitions of validity [15–18] but no guidelines on how it should be established. As illustrated in Table 1, studies on face and content validity have used rather arbitrary cutoff points to determine the appropriateness and value of simulators [16, 19–24]. The variety in scales and interpretations in the literature suggests a lack of consensus regarding criteria for validity.

It is not only important to decide how validity is to be determined; it is also important to decide who is best suited to undertake this task. The literature offers no detailed answers in this regard. It may be advisable to entrust this task to focus groups of specialists who are experts in the procedure in question and in judging simulators. Perhaps judges should also be required to possess good background knowledge on simulators and simulator development. Preferred settings of validation studies need to be considered as well. So far, most tests of face and content validity of surgical simulators have been conducted at conferences (Table 1), where participants are easily distracted by other people and events. Selection bias may also be inherent in this setting, because those who do not believe in simulator training are unlikely to volunteer to practice on a simulator, let alone participate in a validation study.

Table 1 Methods used to quantify and interpret face and content validity

Type of questionnaire	Def. by no. of procedures expert (E)/novice (N)	Research setting	Interpretation	Cutoff point
Yes/no questions on realism and appropriateness [20]	E: >30/year (specialists) N: <30/year (specialists or industry employees)	Conference/course	% of yes/no answers	No actual cutoff point
Four-point Likert scale on realism and usefulness [21]	E: >1,000 Experienced: 200–1,000 Intermediate experienced: <200 N: no endoscopy experience (interns, residents or specialists)	Conference/course + in a hospital during daily work time	Realism: 1 = very unrealistic 4 = very realistic Usefulness: not mentioned	2.95 = "good" 2.57 = "doubtful"
Five-point ordinal scale on realism and usefulness [22]	E: >100 N: ≤100 (surgeons and residents)	Conference	1 = not realistic/good/useful 5 = very realistic/good/useful	>3.5 "favorable", 4.0 "quite well"
Five-point Likert scale on first impression and training capacities [24]	E: ≥100 N: <100 (surgeons and surgical trainees)	In a hospital during daily work time	1 = very bad/useless 5 = excellent/very useful	Mean score 4.0 "good", mean score 3.3 "relatively low"
Five-point modified Likert scale on acceptability of the simulator [23]	E: Board certified urologists	Conference	0.0 = totally unacceptable, 1.0 = moderately acceptable, 2.0 = slightly unacceptable, 3.0 = slightly acceptable, 4.0 = moderately acceptable, 5.0 = totally acceptable	3.0 = slightly acceptable
Seven-point Likert scale on realism and usefulness [16]	E: >50 N: ≤50 (gynecological surgeons)	Conference/course	Realism: 1 = absolutely not realistic 2 = not realistic 3 = somewhat not realistic 4 = undecided 5 = somewhat realistic 6 = realistic 7 = absolutely realistic Usefulness: 1 = strongly disagree 7 = strongly agree	Realism: 5 = somewhat realistic Usefulness: 6 = useful
Ten-point scale on realism, effectiveness and applicability [19]	E: >50 N: ≤50 (surgeons and surgical residents)	Conference/course + in a hospital during daily work time	10 = very positive	≥8 "positive"

See file "BarbaraSchout validation critical review_submission_Table 1"

Objective approaches (experimental studies)

Experimental studies on the simulator

Several studies have examined the construct (discriminative) validity of simulators for endourological procedures [25, 26]. Although the concept of construct validity is somewhat clearer than that of subjective studies of validity, there was substantial variation in methods, data analysis, participants, and outcome parameters. Between 1980 and 2008 several studies examined construct validity in relation to endourological simulators [25]. Table 2 presents the methods used in these studies, in which medical students and residents were the novices, and specialists fulfilled the role of experts, unless mentioned otherwise. Time taken to complete a procedure was a parameter used in all the studies. Time is considered a parameter of importance, but it is not necessarily indicative of achievement of the desired outcome [27]. An exclusive focus on decreasing performance time may eventually result in decreased quality of outcome, suggesting that, besides time, other parameters should be taken into account in measuring validity.

In general surgery there is a similar awareness of discrepancies in the usage and interpretation of construct validity and outcome parameters. Thijssen et al. conducted a systematic review of validation of virtual-reality (VR) laparoscopy metrics, searching two databases and including 40 publications out of 643 initial search results [28]. The data on construct validation were unequivocal for "time" in four simulators and for "score" in one simulator [28], but the results were contradictory for all the other VR metrics used. These findings led those authors to recommend that outcome parameters for measuring simulator validity should be reassessed and based on analysis of expert surgeons' motions, decisive actions during procedures, and situational adaptation.

Transfer of simulator-acquired skills to performance in patients

Only three studies have examined criterion validity of endourological simulators [25, 29–31]. Ogan et al. demonstrated that training on a VR ureterorenoscopy (URS) simulator improved performance on a male cadaver [31]. Knoll et al. trained five residents in the URS procedure on the URO Mentor and compared their performances on the simulator with performances in patients by five other residents by having unblinded supervisors rate the residents' performances [30]. Brehmer et al. compared experts' real-time performances with their performances on a simulator [29].

Transfer studies of laparoscopic and endoscopic simulators have shown very positive results regarding improvement of real-time performances [12, 29–37]. These results should be interpreted with caution, however, because of small sample sizes (frequently less than 30), lack of randomization, supervisors who were not blinded to type of training, groups with dissimilar backgrounds (e.g., surgical and nonsurgical residents), and/or studies limited to a comparison between experts' performances on a simulator and in the operating room but not between experts' and novices' performances. Also, some of these studies did not use real patients but human cadavers or animal models to measure real-time performance [31, 33].

Ethical and legal concerns may hamper transfer studies where the ideal study protocol would involve groups of trained and untrained participants performing the procedure of interest in a patient. However, even though today many residents learn procedures in patients without prior training on a simulator, this type of study is unlikely to gain the approval of Medical Review Ethics Committees, especially if a study tests the hypothesis that trained participants will outperform controls, implying that the patient is at risk when procedures are performed by controls.

Definition of novices and experts

An important issue in validity research is defining participants' levels of expertise. Generally, the term "novices" designates persons with no experience at all in performing the procedure under study, while the term "expert" refers to specialists with ample experience in performing the procedure in patients. However, some studies labeled participants with only some experience as "novices" while residents who had not yet completed the learning curve were considered "experts" (Tables 1 and 2). In the absence of clear standards for classifying experts and novices, researchers apparently use arbitrary cutoff points. With regard to urethrocystoscopy, for example, Gettman et al. classified those who had performed 100 procedures or more as experts [38], whereas Shah et al. required performance of > 1,000 procedures for qualification as an expert [39]. Apart from differences regarding the number of procedures used as the cutoff point between novice and expert, it is questionable whether it is at all defensible to use number of procedures performed as a measure of expertise. For one thing, self-estimated numbers are likely to be unreliable [40] and, furthermore, having performed more procedures does not automatically correlate with increased quality of performance. It might be better to focus on external assessment of expertise or a more objective standard to discriminate between experts and novices.

Table 2 Methods and parameters used to assess construct validity

Type of procedure	Training model	Method: training program	Def. expert (E)/novice (N)	Parameters
UCS	Virtual reality	One task, ten repetitions, one occasion (unsupervised) [69]	N: no endoscopic experience (urology nurse practitioners)	Total time, no. of flags
		One task, ten repetitions, one occasion [39]	E: >1000 flexible cystoscopies; N: no endoscopic experience	Total time, no. of flags
		Three tasks, five times, one occasion [38]	E: ≥100 flexible and rigid cystoscopies; N: no cystoscopies performed (specialist assistents and extenders, nurses, technicians, office workers	Time
	Virtual reality	Two training sessions, two occasions [70]	E: residents; N: no endoscopic experience	Mucosa inspected
URS	Virtual reality	Individual 10-min practical mentoring [71]	E: medical students who had underwent training; N: no endoscopic experience	Time to bladder neck, time to ureteral orifice, time to cannulate orifice, time to calculus and total time, number of attempts at cannulation, number of times subject had to be reoriented, perforation rate, number of ureteral petechiae, OSATS
		5 h training in 2 weeks [72]	E: residents with varying degrees of endoscopic experience; N: no endoscopic experience	Total time, fluoroscopy time, instrument trauma, attempts at cannulation, OSATS
		Five 30-min training sessions over 2 weeks [73]	E: medical students who had underwent training (no exact no. mentioned); N: no endoscopic experience (no exact no. mentioned)	Total time, fragmentation time, trauma, perforation, insert guidewire, ability to perform task, overall performance, OSATS, self-evaluation
		No training [30]	E: >80 URS; N: <40 URS (urologists)	Mean time, time of progression to stone contact, X-ray exposure time, bleeding events, clearance of stone
		Ten 30-min training sessions in 2 weeks [31]	E: residents; N: no endoscopic experience	Overall time, fluoroscopy time, trauma, no. of URS attempts, OSATS
		No training [74]	E: senior residents (6.4 ± 2.3 URS performed); N: junior residents (1 ± 6 URS performed)	Time, scope trauma, instrument trauma, percent passing, guidewire insertion attempts, OSATS, checklist score
	Bench	1 h practice session [75]	E: senior residents; N: junior residents	Time, pass rating, checklist score, OSATS
	Bench	Two 2-day courses with 8 and 16 h practice, respectively [29]	N: 14 participants with no experience, 5 performed 6–10 URS, 7 performed 1-5 URS	Checklist, OSATS, total score
TURBT	Virtual reality	Two training sessions, two occasions [70]	E: residents; N: no endoscopic experience	Total time, tumors treated, blood loss

Table 2 continued

Type of procedure	Training model	Method: training program	Def. expert (E)/novice (N)	Parameters
TURP	Virtual reality	No training [23]	E: board-certified urologists N: with master level education or above	Orientation time, coagulation time, cutting time, grams resected, cuts at tissue, total fluid use, cut pedal presses, blood loss
		No training [76]	E: urologists N: master's degree educational level, Resident: in training	Orientation time, coagulation time, cutting time, blood loss, gm resected, tissue cuts
	Virtual reality	Six tasks, one occasion [77]	N: no endoscopic experience	Time during which there was high pressure resected volume, blood loss, distance the resectoscope tip was moved, amount of absorbed irrigation fluid
Perc	Virtual reality	Two 30-min training sessions separated by a minimum of 24 h [78]	31 students, 31 residents, 1 fellow	Total time, fluoroscopy time, attempted needle punctures, blood vessel injuries, collecting system perforation

See file "BarbaraSchout validation critical review_submission_Table 2"

Recommendations for validation and implementation of surgical training models

It is inadvisable to use training models before their validity as an educational tool has been proven by research [5, 6, 14, 15]. However, there is as yet no consensus on appropriate methods and parameters to be used in such studies. So far validity studies have mainly focused on technical skills. Although these skills are important they are not the only aspect of operating on patients. The problems concerning transfer studies and the diversity of study outcomes demonstrate that it may be better to design and evaluate a comprehensive training program instead of validating only one aspect or part of a procedure that can be performed on a simulator. This requires an understanding of educational theories and backgrounds and a multidisciplinary approach in which specialists, residents, educationalists, and industrial designers collaborate. In addition, we should learn from experiences in other domains, such as the military and aviation, where similar difficulties with regard to the use of simulators in training are encountered.

Integration of training needs analysis and training program design in developing training facilities

"For a long time, simulator procurement for military training purposes has been mainly a technology-pushed process driven by what is offered on the market. In short, the more sophisticated the simulator's capabilities, the more attractive it is to procure. Training programmes are later developed based upon the device procured, sometimes only for the training developers to conclude that the simulator "did not meet the requirements" or, even worse, that it was unusable because of a complete mismatch between the capabilities and limitations of the device on the one hand and the basic characteristics and needs of the trainees on the other" [41].

Nowadays, there is awareness of the mechanism described by Farmer et al. within surgical communities too, and there is also a growing realization of the need to reevaluate the methods and approaches used in developing surgical training programs. In military training in the 1990s there was a generally acknowledged need for an integrated framework as well as research and development of simulations based on the realization that the world was changing and conditions and constraints were evolving [41]. It was stated that "simulation by itself cannot teach" and this concept led to the Military Applications of Simulator and Training concepts based on Empirical Research (MASTER) project in 1994, in which 23 research and industrial organizations in five countries combined their knowledge to develop generic concepts and common guidelines for the

procurement, planning, and integration of simulators for use in training.

The MASTER project underlined the importance of three key phases of program development: training needs analysis (TNA), training program design (TPD), and training media (simulators, for example) specification (TMS) [41]. These phases have also been described in the medical education literature [2, 42]. TNA involves task analysis and listing the pitfalls of a procedure that need to be trained. When training needs and the place of a simulator in the curriculum are analyzed before a simulator is actually introduced, a major problem of validation studies can be avoided, namely the fact that some simulators train and measure different, not equally relevant, parameters [43]. TPD follows TNA, and is concerned with organizing the existing theoretical and practical knowledge about the use of simulators with a focus on outlining training program requirements. Following TPD, the TMS phase focuses on simulator requirements. Validation has its place in this phase. As Satava stated "Simulators are only of value within the context of a total educational curriculum" and "the technology must support the training goals" [44].

Figures 1 and 2 present a ten-step approach to developing surgical training programs. Figure 1 represents the preparation phase, consisting of training needs analyses. Figure 2 shows a recommended approach to evaluating and implementing surgical simulators in curricula. For every new training program it should be considered whether all the steps of the process are feasible and cost effective. New developments and improvement of education mostly require financial investments. However, in order to minimize costs it is important to consider the expected benefits
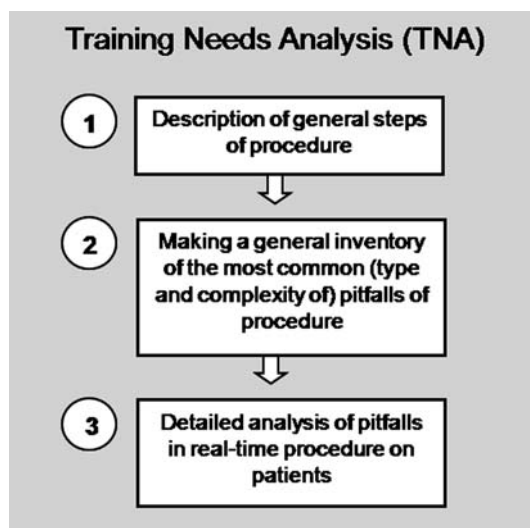
as well as possible drawbacks and the costs that go along with those.

Accreditation and certification are also very important aspects that need to be considered once the definitive training program has been designed. Because accreditation and certification follow program development, they are not included in Figs. 1 and 2.

Integration of nontechnical factors that influence practical skills performances

As early as 1978 Spencer et al. pointed out that a skillfully performed operation is 75% decision making and only 25% dexterity [45]. Nontechnical (human) factors strongly influence residents' and students' performances [3, 14, 46–57]. Moreover, research concerning safety in surgery has shown that adverse events are frequently preceded by individual errors, which are influenced by diverse (human) factors [9, 58].

Surgical training is still very much focused on technical skills, although a skillslab environment may be an ideal situation for integrating technical and nontechnical factors. There is still a gap between research into human factors and educational research [41]. Taking account of expertise on human factors early in the development of training programs and also in the specification of training media can make a considerable contribution to improved the validity and cost-effectiveness of training [41].

Effective surgical training depends on programs that are realistic, structured, and grounded in authentic clinical contexts that recreate key components of the clinical experience [8, 9, 14, 56, 59, 60]. Ringsted et al. showed that factors involved in the acquisition of technical skills can be divided into three main groups: task, person, and context [53]. The model of the acquisition of surgical practical skills shown in Fig. 3 is based on these groups. It illustrates the complexity of a learning process that is affected by various factors.

Collaboration of specialists, residents, educationalists, and industrial designers

Curriculum design is not a task that should be left to one individual. Preferably, it involves multidisciplinary consultations and research [41]. When specialists, residents, educationalists, and industrial designers collaborate and share their knowledge they will be able to make progress in developing and implementing simulator training in curricula [61].

Simulators can assist clinical teachers and relieve some of their burden. Not every specialist is a good teacher. Superior performance of procedures in patients does not automatically imply similar excellence in teaching others

Fig. 1 The training needs analysis phase of training program development. See file "BarbaraSchout validation critical review_submission_Figure 1"
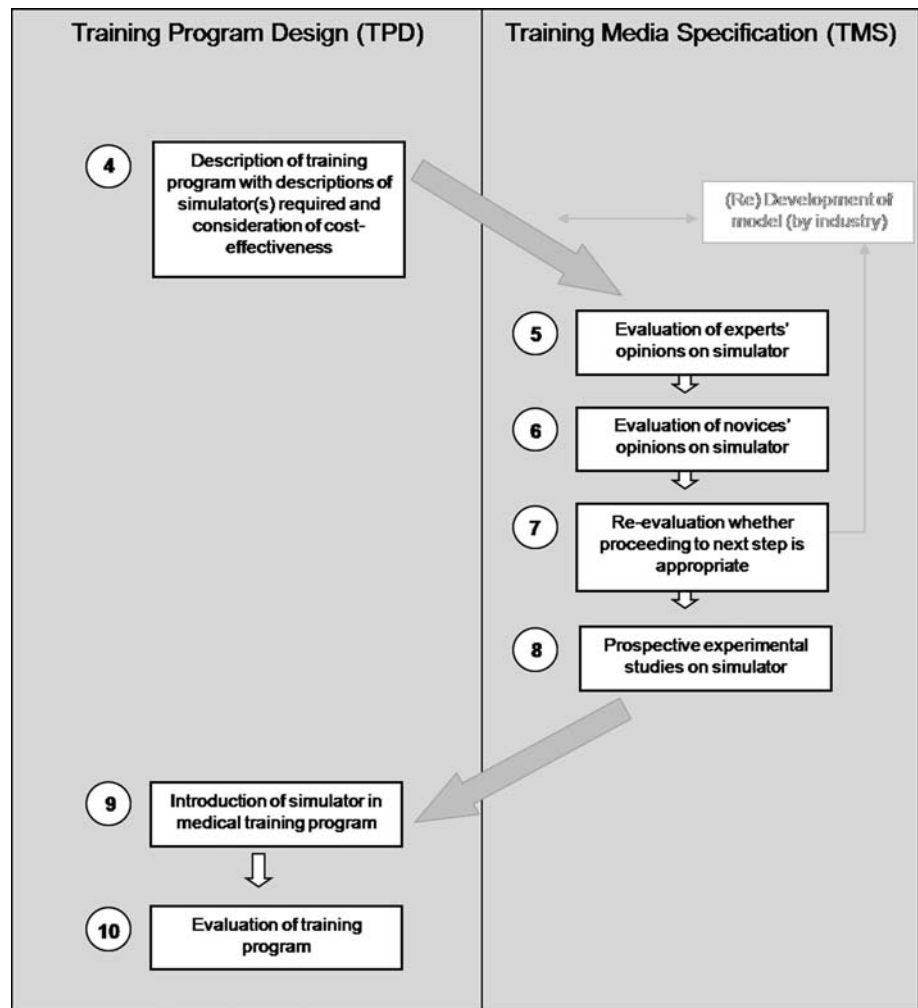
to do the same. Currently, training of medical skills during procedures on patients depends largely on the willingness of trainers to allow trainees to practice and improve their diagnostic and procedural skills. As a result, training is strongly teacher and patient centered [62]. A skillslab environment offers a much more learner-centered educational environment [8]. However, this can only be achieved if not only specialists (teachers), but also residents (learners), educationalists (teaching the teachers), and industrial designers (suppliers of teaching facilities) are allowed to contribute their expertise to developing the content of training programs.

Development and evaluation of assessment methods

Performance assessment tools are needed to evaluate and validate surgical simulators. Several methods that have been developed or are being developed involve the use of simulators not only to practice but also to assess skills. VR and augmented-reality (AR) simulators allow automatic gathering of objective data on performance [14, 17, 37].

However, the development of these metrics is itself an emerging field, and as we described earlier, there is no uniform approach to measuring performance with VR or AR simulators. Motion analysis, tracking how trainees move laparoscopic instruments, is a relatively new and important type of assessment [63]. Although this enables objective performance assessment, assessment methods based on data generated by VR/AR simulators and motion analysis offer limited possibilities because of their exclusive focus on technical skills and because many of these systems can only be used in training environments [63]. Another promising, upcoming factor in assessment is error analysis by means of video analysis [64–66].

Currently, the most commonly used and the only thoroughly validated method to assess technical as well as nontechnical skills is Objective Structured Assessment of Technical Skills (OSATS). OSATS can be used to assess performance on simulators as well as real-time performance in patients. Performance is usually scored by a supervisor on a five-point scale [67]. However, although OSATS has been thoroughly evaluated and validated, it has

Fig. 3 Factors that influence performance of practical skills of trainees. See file "BarbaraSchout validation critical review_submission_figure 3"
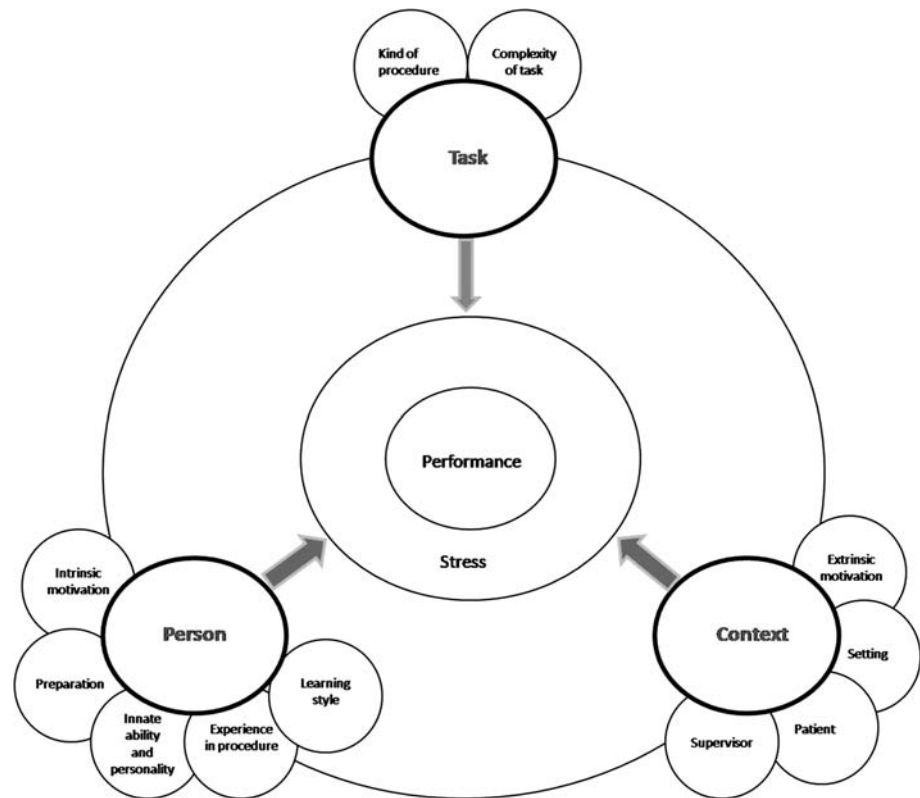
the disadvantage of being dependent on supervisors' subjective opinions. As Miller stated in 1990, "No single assessment method can provide all the data required for judgment of anything so complex as the delivery of professional services by a successful physician" [68]. It seems eminently desirable to further develop and thoroughly evaluate and validate these assessment methods, especially for assessment of real-time performance.

Conclusion

Studies examining the validity of surgical simulators are recommended for progress in the implementation of simulators in surgical education programs. The absence in the literature of general guidelines for interpreting the results of subjective validity studies points to a need to seek consensus, if possible, and perform research to identify appropriate methods for evaluating this type of validity and for interpreting results. A considerable number of studies have addressed objective construct (discriminative) validity of simulators. However, there is considerable variation in outcome parameters and it is questionable whether the measured parameters actually reflect those aspects that are most important for novices to learn on a simulator. Few objective studies have examined whether skills learned on

a simulator can be transferred successfully to patient care. This lack of studies is partly due to ethical and legal issues restricting these types of studies.

Validation and integration of surgical simulators in training programs may be more efficient if training needs analysis (TNA) is performed first and program requirements are set in a training program design (TPD) phase by a multidisciplinary team, consisting of specialists, residents, educationalists, and industrial designers. Furthermore, for successful transfer of skills from simulator to patient, it is important to consider and include the influence of contextual, (inter)personal, and task-related factors in training programs, rather than merely focusing on technical skills. Multiple validated assessment methods of practical performance are essential for evaluating training programs and individual performances. Current assessments methods are few, not yet thoroughly validated, and mostly focused on technical skills only. Educational and medical communities should join forces to promote further development and validation of the available assessment methods.

References

1. Satava RM (2008) Historical review of surgical simulation–a personal perspective. World J Surg 32(2):141–148
2. Dent JA, Harden RM (2005) A practical guide for medical teachers, 2nd edn. Elsevier, Dundee
3. Aggarwal R, Undre S, Moorthy K, Vincent C, Darzi A (2004) The simulated operating theatre: comprehensive training for surgical teams. Qual Saf Health Care 13(Suppl 1):i27–i32
4. Blumenthal D (1994) Making medical errors into "medical treasures". JAMA 272(23):1867–1868
5. Bradley P, Bligh J (2005) Clinical skills centres: where are we going? Med Educ 39(7):649–650
6. Bradley P (2006) The history of simulation in medical education and possible future directions. Med Educ 40(3):254–262
7. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G et al (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241(2):364–372
8. Kneebone RL, Scott W, Darzi A, Horrocks M (2004) Simulation and clinical practice: strengthening the relationship. Med Educ 38(10):1095–1102
9. Kneebone RL, Nestel D, Vincent C, Darzi A (2007) Complexity, risk and simulation in learning procedural skills. Med Educ 41(8):808–814
10. Reznick RK, MacRae H (2006) Teaching surgical skills—changes in the wind. N Engl J Med 355(25):2664–2669
11. Ziv A, Wolpe PR, Small SD, Glick S (2003) Simulation-based medical education: an ethical imperative. Acad Med 78(8):783–788
12. Seymour NE (2008) VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg 32(2):182–188
13. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER et al (1999) Simulation technology for health care professional skills training and assessment. JAMA 282(9):861–866
14. Kneebone R (2003) Simulation in surgical training: educational issues and practical implications. Med Educ 37(3):267–277
15. McDougall EM (2007) Validation of surgical simulators. J Endourol 21(3):244–247
16. Bajka M, Tuchschmid S, Streich M, Fink D, Szekely G, Harders M (2008) Evaluation of a new virtual-reality training simulator for hysteroscopy. Surg Endosc. doi: 10.1007/s00464-008-9927-7, Apr 24, 2008
17. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB et al (2005) Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19(12):1523–1532
18. Gallagher AG, Ritter EM, Satava RM (2003) Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 17(10):1525–1529
19. Ayodeji ID, Schijven M, Jakimowicz J, Greve JW (2007) Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum. Surg Endosc 21(9):1641–1649
20. Hruby GW, Sprenkle PC, Abdelshehid C, Clayman RV, McDougall EM, Landman J (2008) The EZ trainer: validation of a portable and inexpensive simulator for training basic laparoscopic skills. J Urol 179(2):662–666
21. Koch AD, Buzink SN, Heemskerk J, Botden SM, Veenendaal R, Jakimowicz JJ et al (2008) Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 22(1):158–162
22. Schijven M, Jakimowicz J (2002) Face-, expert, and referent validity of the Xitact LS500 laparoscopy simulator. Surg Endosc 16(12):1764–1770
23. Sweet R, Kowalewski T, Oppenheimer P, Weghorst S, Satava R (2004) Face, content and construct validity of the University of Washington virtual reality transurethral prostate resection trainer. J Urol 172(5 Pt 1):1953–1957
24. Verdaasdonk EG, Stassen LP, Monteny LJ, Dankelman J (2006) Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. Surg Endosc 20(3):511–518
25. Schout BM, Hendrikx AJ, Scherpbier AJ, Bemelmans BL (2008) Update on training models in endourology: a qualitative systematic review of the literature between January 1980 and April 2008. Eur Urol 54(6):1247–1261
26. Wignall GR, Denstedt JD, Preminger GM, Cadeddu JA, Pearle MS, Sweet RM et al (2008) Surgical simulation: a urological perspective. J Urol 179(5):1690–1699
27. Schijven MP, Jakimowicz J (2004) The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc 18(1):121–127
28. Thijssen AS, Schijven MP (2008) Validation of performance metrics in virtual reality laparoscopy trainers: a systematic review of the literature. In 16th international conference of the EAES, Stockholm, Abstract no. Oo11
29. Brehmer M, Swartz R (2005) Training on bench models improves dexterity in ureteroscopy. Eur Urol 48(3):458–463
30. Knoll T, Trojan L, Haecker A, Alken P, Michel MS (2005) Validation of computer-based training in ureterorenoscopy. BJU Int 95(9):1276–1279
31. Ogan K, Jacomides L, Shulman MJ, Roehrborn CG, Cadeddu JA, Pearle MS (2004) Virtual ureteroscopy predicts ureteroscopic proficiency of medical students on a cadaver. J Urol 172(2):667–671
32. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DAIII et al (2007) Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 193(6):797–804
33. Ganai S, Donroe JA, St Louis MR, Lewis GM, Seymour NE (2007) Virtual-reality training improves angled telescope skills in novice laparoscopists. Am J Surg 193(2):260–265
34. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 91(2):146–150
35. Naik VN, Matsumoto ED, Houston PL, Hamstra SJ, Yeung RY, Mallon JS et al (2001) Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? Anesthesiology 95(2):343–348
36. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S et al (2007) Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg 194(2):205–211
37. Schijven MP, Jakimowicz JJ, Broeders IA, Tseng LN (2005) The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. Surg Endosc 19(9):1220–1226
38. Gettman MT, Le CQ, Rangel LJ, Slezak JM, Bergstralh EJ, Krambeck AE (2008) Analysis of a computer based simulator as an educational tool for cystoscopy: subjective and objective results. J Urol 179(1):267–271
39. Shah J, Darzi A (2002) Virtual reality flexible cystoscopy: a validation study. BJU Int 90(9):828–832
40. Davis DA, Mazmanian PE, Fordis M, Van HR, Thorpe KE, Perrier L (2006) Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 296(9):1094–1102

41. Farmer et al (1999) Handbook of simulator-based training. Ashgate, England

42. Peyton JWR (1998) Teaching and learning in medical practice. Manticore Europe Ltd, UK

43. Verdaasdonk EG, Stassen LP, Schijven MP, Dankelman J (2007) Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Surg Endosc 21(8):1406–1412

44. Satava RM (2001) Surgical education and surgical simulation. World J Surg 25(11):1484–1489

45. Spencer FC (1978) Teaching and measuring surgical techniques—the technical evaluation of competence. Am Coll Surg 63:9–12

46. Dankelman J, Chmarra MK, Verdaasdonk EG, Stassen LP, Grimbergen CA (2005) Fundamental aspects of learning minimally invasive surgical skills. Minim Invasive Ther Allied Technol 14(4):247–256

47. Enochsson L, Isaksson B, Tour R, Kjellin A, Hedman L, Wredmark T et al (2004) Visuospatial skills and computer game experience influence the performance of virtual endoscopy. J Gastrointest Surg 8(7):876–882

48. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J (2003) Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc 17(7):1082–1085

49. Hutchinson L (2003) Educational environment. Br Med J 326(7393):810–812

50. Lacorte MA, Risucci DA (1993) Personality, clinical performance and knowledge in paediatric residents. Med Educ 27(2):165–169

51. Mitchell M, Srinivasan M, West DC, Franks P, Keenan C, Henderson M et al (2005) Factors affecting resident performance: development of a theoretical model and a focused literature review. Acad Med 80(4):376–389

52. Reich DL, Uysal S, Bodian CA, Gabriele S, Hibbard M, Gordon W et al (1999) The relationship of cognitive, personality, and academic measures to anesthesiology resident clinical performance. Anesth Analg 88(5):1092–1100

53. Ringsted C, Skaarup AM, Henriksen AH, Davis D (2006) Person-task-context: a model for designing curriculum and in-training assessment in postgraduate education. Med Teach 28(1):70–76

54. Schwind CJ, Boehler ML, Rogers DA, Williams RG, Dunnington G, Folse R et al (2004) Variables influencing medical student learning in the operating room. Am J Surg 187(2):198–200

55. Williams S, Dale J, Glucksman E, Wellesley A (1997) Senior house officers' work related stressors, psychological distress, and confidence in performing clinical tasks in accident and emergency: a questionnaire study. Br Med J 314(7082):713–718

56. Kneebone R (2005) Evaluating clinical simulations for learning procedural skills: a theory-based approach. Acad Med 80(6):549–553

57. Teunissen PW, Stapel DA, Scheele F, Scherpbier AJ, Boor K, van Diemen-Steenvoorde JA et al. (2007) The influence of context on residents' evaluations: effects of priming on clinical judgment and affect. Adv Health Sci Educ Theory Pract. doi: 10.1007/s10459-007-9082-2, Oct 17, 2007

58. Vincent C (2003) Understanding and responding to adverse events. N Engl J Med 348(11):1051–1056

59. Kneebone R, Nestel D, Wetzel C, Black S, Jacklin R, Aggarwal R et al (2006) The human face of simulation: patient-focused simulation training. Acad Med 81(10):919–924

60. Kneebone R, Kidd J, Nestel D, Asvall S, Paraskeva P, Darzi A (2002) An innovative model for teaching and learning clinical procedures. Med Educ 36(7):628–634

61. Satava RM, Cuschieri A, Hamdorf J (2003) Metrics for objective assessment. Surg Endosc 17(2):220–226

62. Brigden D, Dangerfield P (2008) The role of simulation in medical education. Clin Teach 5:167–170

63. Chmarra MK, Grimbergen CA, Dankelman J (2007) Systems for tracking minimally invasive surgical instruments. Minim Invasive Ther Allied Technol 16(6):328–340

64. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Andersen DK, Satava RM (2004) Analysis of errors in laparoscopic surgical procedures. Surg Endosc 18(4):592–595

65. Van Sickle KR, Baghai M, Huang IP, Goldenberg A, Smith CD, Ritter EM (2008) Construct validity of an objective assessment method for laparoscopic intracorporeal suturing and knot tying. Am J Surg 196(1):74–80

66. Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S et al (1999) An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 189(6):566–574

67. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C et al (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84(2): 273–278

68. Miller GE (1990) The assessment of clinical skills/competence/performance. Acad Med 65(9 Suppl):S63–S67

69. Shah J, Montgomery B, Langley S, Darzi A (2002) Validation of a flexible cystoscopy course. BJU Int 90(9):833–835

70. Reich O, Noll M, Gratzke C, Bachmann A, Waidelich R, Seitz M et al (2006) High-level virtual reality simulator for endourologic procedures of lower urinary tract. Urology 67(6):1144–1148

71. Watterson JD, Beiko DT, Kuan JK, Denstedt JD (2002) Randomized prospective blinded study validating acquistion of ureteroscopy skills using computer based virtual reality endourological simulator. J Urol 168(5):1928–1932

72. Jacomides L, Ogan K, Cadeddu JA, Pearle MS (2004) Use of a virtual reality simulator for ureteroscopy training. J Urol 171(1):320–323

73. Wilhelm DM, Ogan K, Roehrborn CG, Cadeddu JA, Pearle MS (2002) Assessment of basic endoscopic performance using a virtual reality simulator. J Am Coll Surg 195(5):675–681

74. Matsumoto ED, Pace KT, D'A Honey RJ (2006) Virtual reality ureteroscopy simulator as a valid tool for assessing endourological skills. Int J Urol 13(7):896–901

75. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD (2001) A novel approach to endourological training: training at the Surgical Skills Center. J Urol 166(4):1261–1266

76. Rashid HH, Kowalewski T, Oppenheimer P, Ooms A, Krieger JN, Sweet RM (2007) The virtual reality transurethral prostatic resection trainer: evaluation of discriminate validity. J Urol 177(6):2283–2286

77. Kallstrom R, Hjertberg H, Kjolhede H, Svanvik J (2005) Use of a virtual reality, real-time, simulation model for the training of urologists in transurethral resection of the prostate. Scand J Urol Nephrol 39(4):313–320

78. Knudsen BE, Matsumoto ED, Chew BH, Johnson B, Margulis V, Cadeddu JA, Pearle MS, Pautler SE, Denstedt JD (2006) A randomized, controlled, prospective study validating the acquisition of percutaneous renal collecting system access skills using a computer based hybrid virtual reality surgical simulator: phase I. J Urol 176(5):2173–2178