Construct Validity Testing of a Laparoscopic Surgical Simulator

Elspeth M McDougall, MD, FRCSC, Federico A Corica, MD, John R Boker, PhD, Leandro G Sala, MD, Gabriela Stoliar, MD, James F Borin, MD, Frank T Chu, BSc, Ralph V Clayman, MD, FACS

BACKGROUND: We present initial data on the construct, content, and face validity of the LAPMentor (Simbionix), virtual reality laparoscopic surgical simulator.

STUDY DESIGN: Medical students (MS), residents and fellows (R/F), and experienced laparoscopic surgeons (ES), with < 30 laparoscopic cases per year (ES < 30) and those with > 30 laparoscopic cases per year (ES > 30), were tested on 9 basic skill tasks (SK) including manipulation of 0-degree and 30-degree cameras (SK1, SK2), eye-hand coordination (SK3), clipping (SK4), grasping and clipping (SK5), two-handed maneuvers (SK6), cutting (SK7), fulguration (SK8), and object-translocation (SK9).

RESULTS: Mean MS (n = 23), R/F (n = 24), ES < 30 (n = 26), and ES > 30 (n = 30) ages were 26 years (range 21 to 32 years), 31 years (range 27 to 39 years), 49 years (range 31 to 70 years) and 47 years (range 34 to 69 years), respectively. In the lower level skill tasks (SK3, SK4, SK5, and SK6) the ES > 30, ES < 30, and R/F had similar scores, but were all substantially better than the MS scores. In the higher level skill tasks (SK7, SK8, and SK9), the ES > 30 scores tended to be better than the R/F and ES < 30, which were similar, and these, in turn, were markedly better than the MS. The ES > 30 had notably higher SK8 scores than the R/F and ES < 30, who had similar scores, and these had notably better scores than the MS.

CONCLUSIONS: The noncamera skills (SK3 to 9) of the LAPMentor surgical simulator can distinguish between laparoscopically naive and ES. SK8 showed the highest level of construct validity, by accurately differentiating among the MS, R/F, ES < 30 and ES > 30. (J Am Coll Surg 2006;202: 779–787. © 2006 by the American College of Surgeons)

Laparoscopy is an integral component of today's urologic practice. The rapid evolution of this field at major medical centers, however, has been slow to disseminate into general urologic practice because most urologic applications involve a steep learning curve. The reduced depth perception and two-dimensional image, with more attenuation in tactile sensation and end force of the instruments that are manipulated in a totally counter-intuitive environment, make this a challenging learning process. But because of these same limitations, the laparoscopic environment, unlike the open surgical environment, can be replicated in a computer-generated virtual reality format. The need for repetitive practice of these challenging laparoscopic skills has made basic laparoscopy amenable to simulator-based training.[1] So surgical simulators may help neophyte surgeons master skills such as manipulation of laparoscopic instruments, compensation for the camera angle and the fulcrum effect of the instruments, and performance of ambidextrous tasks.[2]

Before a surgical simulator can be used to assess competency, it must be vigorously and objectively evaluated to determine both its scientific reliability and its validity.[3] Among the five recognized validities (content, face, construct, concurrent, and predictive), we sought to study the three most basic ones: content, face, and construct

Elspeth M McDougall, MD, FRCSC, is Director, Astellas Center for Urological Education. Astellas Pharma US, Inc, has received an unrestricted research grant from Endocare, Inc, and grants from Karl Storz America Endoscopy; Intuitive Surgical; Simbionix; and Ethicon Endo-surgery for equipment support in the laboratory.

Ralph V Clayman, MD, FACS, is a shareholder in Applied Medical and a consultant and speaker for Endocare, Inc and Intuitive Surgical. He has unrestricted educational grants from Baxter Corp, Endocare, Inc, and Boston Scientific. He receives royalties from Boston Scientific, Greenwald Surgical, Orthopedic Systems Inc, and Cook Urological.

Received November 18, 2005; Revised January 11, 2006; Accepted January 12, 2006.

From the Departments of Urology and Family Medicine, University of California, Irvine, Orange, CA.
Correspondence address: Elspeth M McDougall, MD, FRCSC, UCI Medical Center, Building 55, Room 304, Rt 81, 101 The City Drive, Orange, CA 92868.

Published by Elsevier Inc.

ISSN 1072-7515/06/$32.00
doi:10.1016/j.jamcollsurg.2006.01.004

Abbreviations and Acronyms

ES = experienced laparoscopic surgeons
MS = medical students
R/F = residents and fellows
SK = skill tasks

validity. Construct validity is one of the most valuable and mandatory assessments because it confirms that the simulator can distinguish the experienced from the inexperienced surgeon based on the performance score. Content validity is the assessment of the appropriateness of the simulator as a teaching modality and involves formal evaluation by experts knowledgeable about the device. This determines whether the simulator can realistically teach what it is supposed to teach. Face validity is usually assessed informally by nonexperts and is used to determine the realism of the simulator, or whether the simulator represents what it is supposed to represent.[3] The purpose of this study was to present the initial construct, content, and face validity testing results of the basic laparoscopic skills training tasks of the LAPMentor (Simbionix) surgical simulator. Determination of concurrent validity, the extent to which the simulator correlates with the gold standard for teaching the technique, and predictive validity, the demonstration that surgical performance scores measured on the simulator during training can predict the trainee's future skill level when transferred to the clinical operating room, is only reasonable after these three basic validities have been corroborated.

METHODS

After Institutional Review Board approval at the University of California, Irvine, medical students (MS), general surgery and obstetrics and gynecology residents and fellows (R/F), and experienced laparoscopic surgeons (ES) were invited to participate in the study, from August 2004 to January 2005. ES were defined as practicing urologists, general surgeons, or gynecologists performing laparoscopic surgery in their practice. The ES group was subdivided into surgeons performing < 30 laparoscopic cases per year (ES < 30) and those performing > 30 laparoscopic cases per year (ES > 30). The majority of the ES > 30 were entered into the study voluntarily during the 13th International Congress and Endo Expo annual meeting of the Society of Laparoendoscopic Surgeons in New York, NY.

Table 1. The Nine Skill Tasks of the LAPMentor

LAPMentor skill task	Description of skill task
SK 1	Manipulation of 0-degree camera
SK 2	Manipulation of 30-degree camera
SK 3	Eye-hand coordination
SK 4	Clipping leaking hoses
SK 5	Grasping and clipping leaking hoses
SK 6	Two-handed maneuvers
SK 7	Cutting
SK 8	Fulguration
SK 9	Object translocation

The LAPMentor is a computer-based virtual reality simulator for learning basic laparoscopic skills, featuring two mock working instruments and a camera. Instrument and camera movements were translated into a virtual surgical environment, including haptic feedback, and displayed through a 17-inch flat liquid crystal display. The basic skill tasks (SK) included manipulation of a 0-degree camera (SK1), manipulation of a 30-degree camera (SK2), eye-hand coordination (SK3), clipping leaking hoses (SK4), grasping and clipping leaking hoses (SK5), two-handed maneuvers (SK6), cutting (SK7), fulguration (SK8), and object translocation (SK9) (Table 1). Each basic SK was divided into component measures that varied according to each task. Total time for each exercise completion was assessed. Accuracy rate, as a percentage, was defined as the number of red balls hit or total of shots (SK1) and (SK2), or both, the number of balls correctly touched or total of touched balls (SK3), or both, the number of correctly applied clips or total number of clips used (SK4), or both, the number of injury-free cuts or total number of cuts (SK7), or both, and the number of burned highlighted bands or total number of burned bands (SK8) or both. Each basic SK was automatically scored according to an algorithm predetermined by the manufacturer and stored in a password protected computerized spreadsheet folder. A final, overall LAPMentor performance score was generated by summing the nine laparoscopy basic SK scores. Basic SK scores and their component measures were compared among the three groups. The manufacturer's predetermined skill task scores, which took into consideration both speed and accuracy, were then compared among themselves and correlated with each one of their own component measures.

After one-on-one, hands-on instruction by a trained individual on each one of the basic SK, including the

Table 2. LAPMentor Scores of the Four Study Groups on the Nine Basic Laparoscopy Skill Tasks and Overall Performance Score

Skill task #	Surgeons > 30 cases per year	Surgeons < 30 cases per year	Residents and fellows	Medical students
1	86 ± 14	85 ± 7	91 ± 8	82 ± 14
2	86 ± 15	82 ± 16	89 ± 17	81 ± 17
3	87 ± 8	79±16	88 ± 7	69 ± 22
4	96 ± 8	90 ± 19	97 ± 4	86 ± 14
5	81 ± 7	77 ± 13	82 ± 12	66 ± 22
6	91 ± 9	87 ± 13	92 ± 13	78 ± 23
7	92 ± 14	73 ± 19	87 ± 17	76 ± 22
8	75 ± 25	54 ± 17	59 ± 22	43 ± 15
9	60 ± 26	40 ± 25	62 ± 26	27 ± 20
Overall	756 ± 79	680 ± 86	754 ± 72	615 ± 99

Scores represent mean ± SD.

expected performance, recorded variables, and penalized errors, participants were allowed a single practice trial, after which their performance was recorded. Participants were tested on the nine basic SK of the LAPMentor, moving sequentially through skill tasks 1 to 9. A survey questionnaire was administered to each of the participants immediately after their simulator testing inquiring about demographic data, previous surgical experience, and opinions of the simulator. Survey data relevant to the determination of content validity, by ES, and face validity, by the MS responses, were analyzed.

Data were examined by computing descriptive statistics, correlation coefficients, and comparison of study group mean scores. Statistical significance of differences among study group mean scores was determined by either one-way ANOVA or independent-group *t*-tests. Significant ANOVA was followed by pair-wise multiple comparisons of group mean scores using either the Student-Newman-Keuls test (for homogeneous variances) or the Games-Howell test (for heterogeneous group variances). The nominal alpha level of < 0.05 was the criterion for statistical significance, and no correction was made to control Type I error rates associated with computing multiple tests. All statistical tests were performed with SPSS version 12.0 (SPSS Inc).

RESULTS

The total number of participants in the study was 103; this included 23 MS, 24 (R/F), 26 ES with < 30 laparoscopic cases per year (ES < 30), and 30 ES with > 30 laparoscopic cases per year (ES > 30). The mean ages of the groups were MS, 26 years (range 21 to 32 years), R/F, 31 years (range 27 to 39 years), ES < 30, 49 years (range 31 to 70 years), and ES > 30, 47 years (range 34 to 69 years). The two ES groups had similar ages, but were notably older than the R/F, and these, in turn, were notably older than the MS. Thirty-five percent of MS, 21% of R/F, and 4% of ES were women. All of the R/F had assisted in < 30 laparoscopic cases by the time of testing, except for one general surgery chief resident and one urology fellow, who had assisted in > 50 cases. None of the MS had earlier laparoscopic surgical experience.

There was no notable difference in the scores in the four different study groups for the camera manipulation skill tasks (SK1 and SK2) (Tables 2, 3). The ES > 30, ES < 30 and R/F had similar scores in SK3, SK4, SK5, and SK6. But the scores for these three groups were all substantially better than those for the MS in these same four skill tasks. The SK7, laparoscopic cutting task, and SK9, object translocation task, demonstrated similar scores for the ES > 30 and R/F, but these two groups had considerably higher scores ($p < 0.01$ for SK7, $p < 0.001$ for SK9) than the ES < 30 and the MS groups, which also had similar scores. Interestingly, SK8 laparoscopic fulguration task revealed that the ES < 30 and the R/F had similar scores. But the ES > 30 had notably higher SK8 scores than the ES < 30 and R/F groups ($p < 0.01$), and these in turn, had notably higher scores than the MS group ($p < 0.001$). In analysis of the overall or cumulative skill task scores, the ES > 30 and R/F had similar total scores ($p = 0.95$), and these two groups had notably higher total scores compared with the ES < 30 ($p = 0.0001$). The ES < 30 likewise had considerably higher total scores compared with the MS group ($p = 0.0001$).

Table 3. Statistical Comparison of LAPMentor Scores for the Basic Laparoscopy Skill Tasks in the Four Study Groups

LAPMentor skill task	Comparison of study group scores
SK 1	ES > 30 = ES < 30 = R/F = MS (p = 0.07)
SK 2	ES > 30 = ES < 30 = R/F = MS (p = 0.30)
SK 3	ES > 30 = ES < 30 = R/F > MS (p < 0.001)
SK 4	ES > 30 = ES < 30 = R/F > MS (p < 0.006)
SK 5	ES > 30 = ES < 30 = R/F > MS (p < 0.01)
SK 6	ES > 30 = ES < 30 = R/F > MS (p < 0.009)
SK 7	ES > 30 = R/F > ES < 30 = MS (p < 0.01)
SK 8	ES > 30 > R/F (p < 0.01) = ES < 30 > MS (p < 0.001)
SK 9	ES > 30 = R/F > ES < 30 = MS (p < 0.001)
Overall score	ES > 30 = R/F (p = 0.95) > ES < 30 (p = 0.0001) > MS (p = 0.0001)

ES < 30, experienced surgeons performing < 30 laparoscopic cases per year; ES > 30, experienced surgeons performing > 30 laparoscopic cases per year; MS, medical students; R/F, residents and fellows; SK, skill task.

The SK8, laparoscopic fulguration task, was reanalyzed. The mean and median scores for the ES > 30 were 86 and 93, respectively (Fig. 1). Although there were a few R/F with SK8 scores greater than the ES > 30 mean score, none of the ES < 30 had SK8 scores greater than the mean score in the ES > 30 group. Similarly, a few of the MS had SK8 scores greater than the mean scores of R/F and ES < 30, but none of the MS scores were greater than the ES > 30 scores. The mean scores of the MS, R/F, and ES < 30 were all less than the lowest score of the ES > 30 group.

It was noted that one of the fellows in the R/F group had extensive laparoscopic general surgery experience in her native South American practice before undertaking a surgical subspecialty fellowship. So the data for the SK8 was reanalyzed after eliminating this participant's scores from the R/F group and reassigning her scores to the ES < 30 group, to which she more appropriately belonged. With this modification of the data, the mean SK8 score for the ES > 30 (86 ± 14) was higher than that for the ES < 30 (79 ± 15), although not statistically significant. These two groups did have statistically significant higher SK8 scores ($p < 0.0005$) compared with the revised R/F group (52 ± 17) and the MS group (43 ± 16). Although the revised R/F group had higher SK8 scores than the MS group, this was not statistically significant. Also interestingly, when the top three SK8 scoring R/Fs were compared with the ES > 30, there was no notable difference in the scores on this task (91 ± 2.5 and 86 ± 14, respectively). In these three R/Fs, one had extensive experience with laparoscopy and two were participating in a dedicated laparoscopic fellowship. The SK8 score of these three R/F members (91 ± 2.5) was substantially higher than that of the R/F group when analyzed without these four high scoring R/Fs (52 ± 17), ($p < 0.0005$).

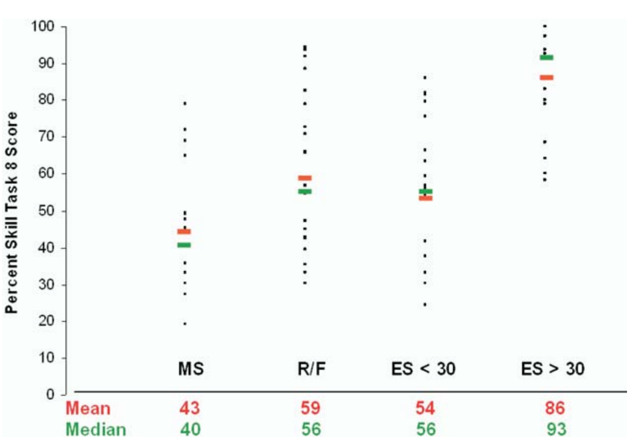

Figure 1. Comparison of the study group performance scores of skill task 8 for the four study groups with means and medians for each group. ES, experienced surgeons; MS, medical students; R/F, residents and fellows.

Also, scores were analyzed as percent score quartiles: 76% to 100% quartile, 51% to 75% quartile, 26% to 50% quartile, and 0% to 25% quartile (Fig. 2). Eighty-five percent of the ES > 30 had SK8 scores in the 76% to 100% quartile, and 15% had scores in the 51% to 75%. None of the ES > 30 participants had SK8 scores < 50%. In the ES < 30 group, 25% of participants scored in the 76% to 100% quartile, 70% scored in the 51% to 75% and 26% to 50% quartiles (35% in each of these quartile score groups), and 5% in the 0% to 25% quartile. The R/F group had 27.5% of the group scoring in the 76% to 100% quartile, 27.5% in the 51% to 75% quartile, 45% in the 26% to 50% quartile, and none in the 0% to 25% quartile. The MS group had 77% scor-

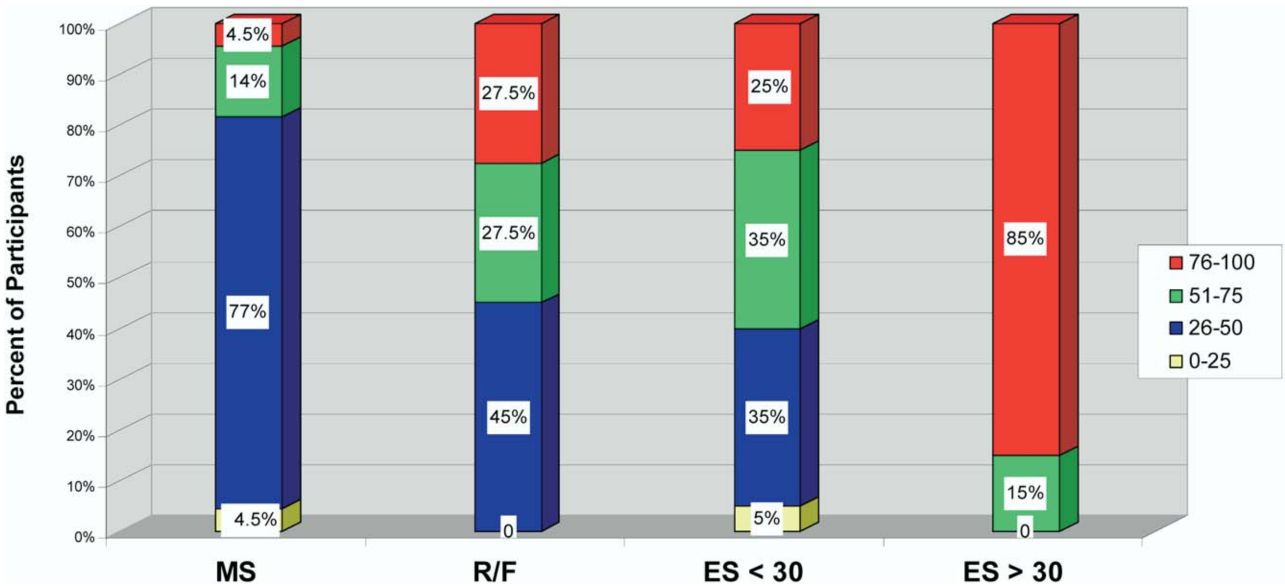

Figure 2. Performance score quartile comparison of skill task 8 for the four study groups. ES, experienced surgeons; MS, medical students; R/F, residents and fellows.

ing in the 26% to 50% quartile, 14% in the 51% to 75% quartile and 4.5% scoring in the 76% to 100% and 0% to 25% quartiles each.

The skill task scores for each study group are depicted in Table 2. MS were the least accurate of all the groups in the completion of SK3, SK5, SK6, and SK9. Importantly, MS did not outperform R/F on any of the skill measures assessed.

The LAPMentor was evaluated for its face validity or realism by the MS and the R/F. Ninety-four percent of these study participants assessed the simulator to be average to easy to use and to master the various skill tasks. Also, 89% considered the LAPMentor to be a realistic practice format (Table 4).

The ES > 30 were asked several questions to assess the content validity of the LAPMentor (Table 5). The majority (91%) considered the LAPMentor to be a useful training device, and 87% considered it as good as, or better than, the conventional pelvic trainer for teaching basic laparoscopic skills. In addition, 91% of ES > 30 would recommend acquisition of the LAPMentor for a laparoscopic training program (Table 5). Seventy-four percent of the ES > 30 considered the LAPMentor to be a good testing format for residents before their operating room experience in laparoscopy. Thirty-five percent of the ES > 30 considered the LAPMentor appropriate as a privileging and certifying tool, and 39% were unsure of its use in this capacity (Table 5).

DISCUSSION

The camera skill tasks (SK1 and SK2) of the LAPMentor surgical simulator were unable to distinguish between any of the study groups, as would be expected for tasks representing fundamental abilities. The higher level, but simple skill tasks, such as eye-hand coordination (SK3), clipping leaking hoses (SK4), grasping and clipping leaking hoses (SK5), and two-handed maneuvers (SK6) could distinguish between surgeons who were laparoscopically naïve and the surgeons with laparoscopic experience. But these skill tasks were unable to differenti-

Table 4. Face Validity of LAPMentor as Evaluated by Medical Students and Residents/Fellows (n = 36)

	Responses				
Evaluating question	**1**	**2**	**3**	**4**	**5**
How was the LAPMentor to use? n (%) (1 = difficult, 3 = average, 5 = easy)	0	2 (6)	13 (36)	15 (42)	6 (17)
Is the LAPMentor a realistic practice format? n (%) (1 = not at all, 3 = somewhat, 5 = extremely good)	0	4 (11)	16 (46)	12 (34)	3 (9)

Table 5. Content Validity of LAPMentor as Evaluated by Experienced Surgeons > 30 (n = 23)

	Content validity scores					
Evaluating question	**1**	**2**	**3**	**4**	**5**	**6**
Is LAPMentor useful for training residents? n (%) (1 = no, 3 = somewhat, 5 = yes)	0	2 (9)	3 (13)	11 (48)	7 (30)	
Is LAPMentor better or worse than pelvic trainer? n (%) (1 = worse, 3 = equivalent, 5 = better)	3 (13)	—	11 (48)	—	9 (39)	
Do you recommend acquiring the LAPMentor for training? n (%) (1 = no, 5 = yes)	2 (9)	—	—	—	21 (91)	—
Is LAPMentor appropriate testing format before operating room experience for residents? n (%) (1 = no, 5 = yes, 6 = don't know)	2 (8)				17 (74)	4 (17)
Could LAPMentor be used as a privileging or certifying device? n (%) (1 = no, 5 = yes, 6 = don't know)	6 (26)				8 (35)	9 (39)

ate surgeons according to their degree of laparoscopic surgical experience.

The higher level, but moderately complex, skill tasks including laparoscopic cutting (SK7) and object translocation (SK9), were also able to distinguish the laparoscopic naïve and the laparoscopic experienced surgeons. Interestingly, in these skill tasks, the ES > 30 and the R/F achieved similar scores and the ES < 30 and the MS achieved similar scores, but the ES > 30 and R/F scores were notably better than those for the ES < 30 and MS.

The fulguration (SK8) skill task is a very high level performance skill. It requires simultaneous two-handed manipulation of instruments, maneuvers similar to dissecting tissues and foot-control of the fulgurating instrument. In addition, the right foot pedal must be used in conjunction with the right-handed instrument and the left foot pedal must be used in conjunction with the left-handed instrument. This skill task, as such, may afford a greater distinction of surgical experience because of its degree of complexity.

In fact, the ES > 30 scored considerably better than all the other groups in this one skill task. Interestingly, the ES < 30 and R/F had similar scores in this SK8, but they had considerably higher scores than the MS in SK8.

The maximum score on SK8 for the R/F group did exceed the mean score of the ES > 30 in a few subjects. But the lowest SK8 score in the ES > 30 group was not less than the mean score of the R/F group. Interestingly, the maximum score in the ES < 30 did not exceed the mean score in the ES > 30 group, and the ES < 30 scores were essentially the same as those for the R/F group. Similarly, the maximum score on SK8 for the MS group did not exceed the mean score of the ES > 30 group. So although there were a few MS and several R/F who performed very well on this high level task, none of them scored better than the mean performance score of the ES > 30 group. This finding supports the construct validity of SK8 on the LAPMentor.

Interestingly, surgeons with some laparoscopic experience (R/F, ES < 30, and ES > 30) had similar scores compared with those of the MS in the camera manipulation skills. This may reflect the fact that ES rarely perform camera manipulation, which is more likely to be assigned to R/F or less ES. Also, camera skills actually represent tasks of fundamental abilities, so would not be expected to show a high degree of discrimination between the novice and the ES. So the experience with this skill task was reflected in the performance scores, supporting the hypothesis that camera skills would lack valuable construct validity as tested with this task of the LAPMentor. R/F consistently performed better than MS and similar to the ES. MS took the longest to perform most of the basic SK.

The limitation of this study that could support criticism of the results is that the participant subjects were allowed only one practice session before the test evaluation of the skill task performances. Gallagher and colleagues[4] demonstrated that, in a small group of surgeons (six experienced and six inexperienced laparoscopic surgeons), performance scores improved considerably in individuals and in the group with successive practice trials, up to approximately three trials. But other researchers have suggested that individuals with more virtual reality game playing experience performed better on virtual reality and model-based laparoscopic surgery simulators than those who do not play these types of games on a regular basis.[5] One might expect this to bias our study against the older, more experienced group of laparoscopic surgeons, whose median age was 45 years compared with 26 years for the MS and 32 years for the R/F. These two younger groups would be expected to have more experience in video-type game playing, which may theoretically improve their scores on the virtual reality simulator.

The LAPMentor was deemed a realistic simulator format and was considered relatively easy to learn to use. So this simulator possesses face validity. Content validity was also confirmed with the LAPMentor. The majority of the ES, with > 30 laparoscopic cases per year, considered the simulator to be a valuable teaching and training device and would recommend its incorporation into a laparoscopic training program. It is most interesting to note that, despite the fact that the LAPMentor lacks predictive and concurrent validity testing at this time, a third of the ES considered it appropriate as a privileging and certifying tool. But this may merely reflect a lack of understanding of the importance of scientific validity in assessing surgical simulators. Alternatively, the value of simulation, through analogy to pilot flight training, may be so intuitive that some surgeons recognize the need for this type of evaluation process in the privileging and certification of surgeons.

Computer-based surgical simulation is a rapidly evolving technology that may prove to be a useful surgical training and credentialing tool. Albeit not conclusively superior to pelvic trainers, computer-based surgical laparoscopic simulators assess surgical skills objectively and relatively free of nonrandom or systematic error, allowing for lower intra- and interparticipant variation in skill assessment, and higher reliability.[2,6,7] Performance of an individual is recorded impartially by the machine, eliminating the time-consuming and costly use of a human preceptor to observe and score an individual, which may be inconsistent because of personal bias or inattentiveness of the preceptor. This modality, as such, may become a valuable one for tracking individual trainee progress over time and providing formative assessments. Additionally, objective assessment of accuracy and efficiency of skill performance are unique to computer-based simulation. But virtual reality surgical simulators, as they currently exist, are less flexible than pelvic trainers at incorporating new tasks and are expensive to add to the educational curriculum.[7]

With more attention to virtual reality surgical simulation development, the potential for infinite flexibility in both abstract and realistic simulations exists. Additional programming and modules will allow these training devices to have multidisciplinary applicability for multiple disease states. Admittedly, the initial cost of this technology may seem prohibitive, but the potential for improving surgical clinical performance may provide a strong argument in favor of investing in this technology at specific teaching centers. Seymour and colleagues[1] reported a 29% faster gallbladder dissection and a sixfold reduction in errors among surgical residents randomized to the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) simulator training compared with those not trained on a simulator. The simulator also eliminates interobserver variation, and intraobserver variation, which may justify the increased expense of the computerized simulator. Another advantage is that simulators reduce the need for dedicated faculty teaching time. Once the initial basic training on the simulator is complete, with a comprehensive curriculum built into the simulator, the student can receive constructive feedback and training by the simulator alone.

Several other computer-based laparoscopic simulators with specific skill tasks, anatomic representation, or presence or absence of haptic feedback, have been proved to achieve construct validity. These include the MIST-VR (Mentice AB),[8,9,10] the LapSim (Immersion Medical),[11] the MISTELS (SAGES FLS program),[12] and the Xitact LS500 (Xitact SA).[13] It should be recognized that the currently validated surgical simulators are related primarily to basic laparoscopic skills training. Simulators related to training higher level skills and specific laparoscopic procedures still lack complete validation. It will also be important to compare the available surgical simulators with each other with the same validation process. These studies are currently underway through the Society of Laparoendoscopic Surgeons' Surgical Simulation Committee.

Although construct validity is important and is necessary before a simulator can be used as an assessment tool, this is only the beginning of the evaluation of this laparoscopic surgical training device. More longterm evaluations will now be necessary to determine the other two very important validities: predictive and concurrent validity; only when these are documented can this simulator be endorsed as an appropriate training tool and assessment device. Predictive validity testing with the LAPMentor is currently being undertaken.

Teaching and testing of technical skills in surgery is the least systematic or standardized component of the classic surgical curriculum. Recent limitations in resident work hours, increasing costs of operating room time, the public's attention to medical errors, and the ethics of learning basic surgical skills in the operating room have encouraged development of curricula geared to teaching fundamental laparoscopic skills in a risk-free setting. Incorporation of surgical simulators into training curricula would seem to provide an opportunity for repetitive skills training with exploration of possible outcomes in a risk-free environment to maximize the educational experience and reduce the time of training for surgeons in complex surgical techniques, while limiting patient morbidity. Mastering laparoscopic surgical skills on a simulator has been shown to improve patient safety and minimize operating room time.[1]

Although inanimate trainer practice results in tangible acquisition of laparoscopic skills, objective quantification of these skills is difficult, requiring human supervision and scoring.[2] Computer-based simulators allow for objective assessment of laparoscopic skill performance based on the quality of the performance, efficiency of instrument movement, and consideration of errors, rather than on time to task completion alone. This computer-based format also allows individual progress to be tracked and compared with the individual's performance over time. This represents the individual's learning curve for a specific skill or procedure. In addition, the individual's performance can be compared with that of trained peers and expert surgeons. With proved predictive validity, criteria can be established for proficiency levels in specific skills and procedure training. This creates the potential for determining when it may be appropriate for surgeons to advance to the clinical operating room based on the simulation proficiency score on the skill or procedure. Being in the operating room would no longer be a rite of passage for surgical residents, but rather a privilege learned and earned. In addition, simulators such as the LAPMentor record not only what is being done, but also have a built-in program that notifies the surgeon of errors in technique when they occur, allowing the surgeon to immediately correct an error and proceed to practice the correct (ie, safe) method for accomplishing a given task, procedure or both.

In conclusion, the LAPMentor surgical simulator has appropriate content and face validity, and it has documented construct validity. Specifically, SK 8, two-handed, foot-controlled electrosurgical coagulation, was the single performance outcomes that could accurately differentiate among MS, R/F, and ES performing $<$ 30 laparoscopic cases per year and ES performing $>$ 30 laparoscopic cases per year. More testing will be necessary to determine whether this laparoscopic simulator also has predictive and concurrent validity before it is incorporated into an educational curriculum as a truly valid and reliable training and assessment device.

Author Contributions

Study conception and design: McDougall, Boker, Clayman
Acquisition of data: McDougall, Corica, Sala, Stoliar, Borin, Chu
Analysis and interpretation of data: McDougall, Corica, Boker, Clayman
Drafting of manuscript: McDougall, Corica, Boker, Clayman
Critical revision: Boker, Clayman
Statistical expertise: Boker
Obtaining funding: McDougall, Clayman
Supervision: McDougall, Clayman

REFERENCES

1. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance. Results of a randomized, double-blind study. Ann Surg 2002;236:458–464.
2. Madan AK, Frantzides CT, Shervin N, Tebbit CL. Assessment of individual hand performance in box trainers compared to virtual reality trainers. Ann Surg 2003;69:1112–1114.
3. McDougall EM. Surgical simulation. AUA Updates (Accepted for publication).
4. Gallagher AG, Lederman AB, McGlade K, et al. Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on experience. Surg Endosc 2004;18:660–665.
5. Enochsson L, Isaksson B, Tour R, et al. Visuospatial skills and computer game experience influence the performance of virtual endoscopy. J Gastrointest Surg 2004;8:876–882.
6. Munz Y, Kumar BD, Moorthy K, et al. Laparoscopic virtual reality and box trainers: is one superior to the other? Surg Endosc 2004;18:485–494.
7. Kothari SN, Kaplan BJ, DeMaria EJ, et al. Training in laparoscopic suturing skills using a new computer-based virtual reality

simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. J Laparoendosc & Adv Surg Tech A 2002;12:167–173.

8. Taffinder N, Sutton C, Fishwick RJ, et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomized controlled studies using the MIST-VR laparoscopic simulator. Stud Health Technol Inform 1998;50:124–130.
9. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001;25:1478–1483.
10. McNatt SS, Smith CD. A computer-based laparoscopic skill assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 2001;15:1085–1089.
11. Duffy AJ, Hogle NJ, McCarthy H, et al. Construct validity for the LapSim laparoscopic surgical simulator. Surg Endosc 2005; 19:401–405.
12. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the values of simulation in laparoscopic surgery. Ann Surg 2004;240:518–528.
13. Schijven M, Jakimowicz J. Construct validity. Expert and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 2003;17:803–810.