Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery

Jeffrey H. Peters, MD, Gerald M. Fried, MD, Lee L. Swanstrom, MD, Nathaniel J. Soper, MD, Lelan F. Sillin, MD, Bruce Schirmer, MD, Kaaren Hoffman, PhD, *and* **the SAGES FLS Committee,** *Los Angeles, Calif*

From The Society of American Gastrointestinal Endoscopic Surgeons, Los Angeles, Calif

In the late 1990s the Society of American Gastrointestinal Endoscopic Surgery (SAGES) formed a committee (Fundamentals of Laparoscopic Surgery or FLS) and charged it to develop educational materials covering the basic fundamentals of laparoscopic surgery. Four major principles guided the committee's developmental process. First, comprehensive coverage of the domain of basic laparoscopy was seen as involving two components: one cognitive (declarative knowledge); and the other psychomotor (procedural skill). Second, the focus of the program was to be on the educational material considered unique to laparoscopy and not on material normally encountered during open surgical training. Third, in accordance with the idea of basic fundamentals, any content specific to a particular anatomic location or to a specific laparoscopic procedure was to be avoided. And fourth, the program was to contain mechanisms for assessment as well as for didactic instruction.

The overall goal of the FLS program was to "teach a standard set of cognitive and psychomotor skills to practitioners of laparoscopic surgery" in the belief that knowledge and application of these fundamentals would help "ensure a minimal standard of care for all patients undergoing laparoscopic surgery." The didactic learning modules are CD-ROM based and teach the underlying physiology, fundamental concepts, and component manual skills involved in laparoscopic surgery. The assessment instruments were designed to be in accord with the competency movement and the American Board of Medical Specialties recommendation for maintenance of practice and practice-based learning and improvement,[1] and to that end they were rigorously developed.

It was not SAGES intent to develop a certifying examination but rather to provide tools for the teaching and assessment of the cognitive knowledge, technical skills, and clinical judgment related specifically to basic laparoscopic surgery. These tools could then be used in any number of ways as determined by the intent and purpose of the organization desiring to use them. The assessment tools consist of both a paper-and-pencil examination (cognitive in orientation) and a hands-on, performance-based, technical skills examination. The processes involved in designing both these measurement tools relied heavily of the Test Standards prepared by the Joint Committee on Standards of the American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education.

Throughout the development process, the FLS committee and other subject matter experts continually reviewed the learning modules for appropriateness. Several studies were performed to ensure the effectiveness of the manual skills model, based on the McGill Inanimate Systems for Training and Evaluation of Laparoscopic Skills (MISTELS) system. An initial pilot test of the assessment and learning materials was conducted in the spring of 2000, and several modifications were undertaken.

Accepted for publication April 6, 2003.

Reprint requests: Jeffrey H. Peters, MD, Professor of Surgery, 1510 San Pablo St, Los Angeles, CA 90033.

Surgery 2004;135:21-7.

0039-6060/$ - see front matter

doi:10.1016/S0039-6060(03)00156-9

Table I. Content areas for basic laparoscopic examination

1. Equipment: tools of the trade
2. Energy sources: electrosurgery, ultrasound, and laser
3. Patient considerations: patient selection, contraindications, preparation
4. Anesthesia: types, complications
5. Patient positioning
6. Establishment and physiology of pneumoperitoneum: gas biologic characteristics, pressure and flow characteristics
7. Abdominal access and trocar placement: techniques, complications
8. Tissue handling, exposure, and examination of abdomen and pelvis
9. Biopsy techniques
10. Hemorrhage and hemostasis
11. Tissue approximation: indications, techniques
12. Exiting the abdomen: drains, site closures
13. Postoperative care

Most recently a beta or field-testing of the didactic and manual skills assessment tools was undertaken in eight specially designated testing centers across the United States and Canada. Each center tested from 10-15 randomly selected participants from the defined geographic area. Participants were stratified across defined levels of training and experience. Pertinent data from the previous studies and the beta-test investigation are presented in the discussion of the FLS program that follows.

COGNITIVE KNOWLEDGE—INSTRUCTION AND ASSESSMENT

The overall educational component was developed as two CD-ROM based study guides. The first CD contains a multimedia didactic presentation of cognitive material organized into 10 content areas.

Chapters

1. Tools	6. Abdominal access
2. Energy sources	7. Abdominal examination
3. Patient considerations	8. Lap suturing
4. Anesthesia and positioning	9. Biopsy and hemostasis
5. Pneumoperitoneum	10. Exiting and postop considerations

Each chapter includes printable textual review, references, practice questions, and a practice clinical scenario. The intent is to provide verbal and visual instruction and interactive patient scenarios.

The 10 chapters cover 13 major content areas (Table I). The content outline was formulated collaboratively by the FLS committee and consists of the basic principles, procedures, and management considerations deemed important when undertaking any laparoscopic operation. The committee defined the 13 major content categories and identified 66 subject areas/concepts that fell within these categories.

The content outline also served as a framework for the development of the cognitive assessment. To validate that the didactic outline and test blueprint effectively covered the domain of basic laparoscopy, samples of practicing laparoscopic surgeons attending the 2001 SAGES Annual Meeting were asked to rate the importance of the 66 content areas. Participants were asked: "In order to provide quality patient care, how essential is it for the laparoscopic surgeon to have knowledge and/or clinical judgment in the listed content area?" Ratings were made on a scale of 1 to 5, where 1 was "essential for laparoscopic surgery," 3 was "desirable but not essential," and 5 was "not applicable to laparoscopic surgery." The relative importance of content areas was determined both by examining average content area ratings and by having the participants directly assign percentage weights to the 13 major content areas. Finally, participants were asked to add any content areas or concepts they felt were essential but were not included in the survey.

One hundred and seventeen surveys were evaluated, involving 77% from SAGES members and 21% from non-SAGES fellows and practicing surgeons attending the conference. Seventy-two of respondents practiced basic, intermediate, and advanced laparoscopic surgery; the remaining (28%) were involved in either basic or intermediate laparoscopy or both, but not advanced. In general, rating variations between respondents with differing backgrounds were more those of degree rather than kind. Data were analyzed both by separate groups and as a whole. No significant differences that would influence conclusions were found. The vast majority of content areas were judged as essential with most receiving ratings less than 2.0. Nine content areas were judged as more "desirable" than "essential" (ie, their rating was greater than 2.0). These areas were reviewed and revised by the FLS committee as appropriate. In addition, participants listed some 14 areas as being not present but appropriate for the examination. The examination committee reviewed these areas and either integrated them into the blueprint or eliminated them based on their appropriateness to the examination's definition and purpose (ie, their uniqueness to laparoscopic surgery and their basic, procedure-free nature).

Weights for each of the major categories were obtained both by averaging the ratings in a given category and by direct estimates of weights by the respondents. For both methods of obtaining relative weights, the differences among the categories were minor. Data indicated all areas seemed equally important with slightly less weight falling to anesthesia, biopsy techniques, and site closures. The final test blueprint outlines the major content areas and the weights applied among them (Table II).

Cognitive item/question development. The cognitive assessment was designed to be computer-based and highly visual, and to test the understanding and application of the basic fundamentals of laparoscopy placing as much emphasis as possible on clinical judgment or intraoperative decision making. Two question formats were used: single multiple choice, and a newer format called scenarios involving sets of multiple choice questions integrated within the context of a case scenario. Guidelines for item writing were developed for both formats, and emphasis was placed on writing questions that required examinees to problem solve and apply or use (rather than simply recall) their knowledge of the fundamentals of laparoscopic surgery. The scenario-type question format, in particular, was designed to capture the type of sequential decision making or judgment found in the operating room. As an aid for creating scenarios, a generic surgical case was broken down into 4 sections that summarized the order of events during a laparoscopic procedure. The guidelines for writing scenarios then listed the types of essential steps, considerations, or decisions ordinarily made within each of the sections. Each of these decision points was seen as a potential question within a case scenario format. These problems are based on a visually documented case history to create a simulated, "real-life" scenario.

Three such scenarios (involving 45 questions in all) were pilot tested using 86 residents and practitioners. Several scoring systems were employed. Based on data from the pilot test, the questions within a scenario were modified to a multiple choice format in which the examinee selects the one best (or two best, etc.) alternatives. The current format is a cross between "context-dependent" questions and "patient management problems." A scenario-type format begins with the presentation of a brief case vignette followed by 3 to 7 questions. The computerized administration requires test takers to answer a question before the next question is presented. Examinees cannot return to previous questions but are able to access previously given information. Dependency among questions within a scenario is often reduced by informing the examinee which of the possible alternative decisions or actions were actually undertaken on previous questions.

Table II. Fundamentals of laparoscopic surgery; subject area weights

Subject area	*Weight (%)*
Equipment: tools of the trade	11
Energy sources; electrosurgery, ultrasound, and laser	7
Patient considerations: patient selection, contraindications, and preparation	11
Anesthesia: types and complications	4
Patient postioning	7
Establishment and physiology of the pneumoperitoneum	11
Abdominal access and trocar placement	11
Tissue handling, exposure and examination of the abdomen and pelvis	11
Biopsy techniques	5
Hemorrhage and hemostasis	7
Tissue approximation: indication and techniques	4
Exiting the abdomen: drains, site closure	4
Postoperative care	4

Question review and preliminary standard setting. After development, questions were reviewed by small groups of laparoscopic surgeons. To avoid bias, these review groups were selected to be geographically diverse and to represent a variety of practice patterns: academic; private practice urban; or rural. Seven review sessions were held over the course of a year. In addition to answering the questions, participants were asked to designate the subject area of each question, rate its relevance to the area of laparoscopic surgery, and ascertain if the question required problem solving or the application of basic knowledge of laparoscopic surgery. Data from such judgments provided an independent review of the clarity and relevance/importance of each proposed question to the field of laparoscopic surgery and verified that the questions fulfilled the objectives of the blueprint for the examination.

Preliminary standard setting judgments were also collected during the review process. Participants were introduced to the concept of a "just qualified candidate" and engaged in a discussion of its definition. Using a modified Angoff method, participants estimated the likelihood that the just qualified candidates would answer each question correctly. Participants had the opportunity to practice this methodology and engage in a group discussion on a series of questions that had statistical data derived from previous pilot testing.

The data from each review session were summarized and integrated with the participants' suggested question modifications from the review sessions. Poorly performing questions were inspected and either eliminated or revised in accordance with the data and participants' suggestions. The revised questions were then re-reviewed at a later session. In all, the subject-specific experts reviewed approximately 203 multiple choice questions and 138 scenario-type questions, some of these being revised questions from earlier sessions. In general, a question was considered appropriate for beta testing if at least 60% of the reviewers answered it correctly, at least 70% agreed it was highly relevant to laparoscopic surgery, and at least 70% felt it required clinical problem solving rather than simple recall of information.

Psychometric properties (metrics). Questions that passed the review process were finalized and entered into the computer system for field testing. To accommodate limited testing time, questions ready for beta testing were assigned to two separate assessments—Test A and B. A common set of 21 questions with an "equating" section that included both multiple-choice and scenario formats was embedded into both Test A and B. The field testing process involved 8 testing centers in the United States and Canada, so that the appropriateness of both question content and the administration mode could be investigated. Between October and December 2002, each center randomly selected 10-15 participants from lists of residents and practicing laparoscopic surgeons in the defined geographic area. Demographic and biographic data were gathered on all participants and used to investigate the relationships between the examination scores and variables of interest to establish the construct validity of the examination scores.

Data from field testing showed that skilled laparoscopic surgeons performed similarly on the questions assigned to each test with an expected average score of 81%. The average relevancy rating was 2.5 (on a scale of 1 = low to 3 = high) for questions in Test A and 2.6 for those in Test B. Questions in both tests were judged as requiring problem solving rather than simple recall (an average of 73% for Test A and 77% for Test B). Finally, the projected score for a "minimally qualified" laparoscopic surgeon (obtained from the preliminary standard setting judgments) was estimated to be 67% for Test A and 68% for Test B.

Validity analyses. Data from the field testing verified that the questions in the tests were relevant to the domain of laparoscopic surgery and appropriate to the purpose of the test and the test blueprint. Although this is a necessary aspect of any assessment, it is not sufficient evidence for judging overall validity. Evidence is needed to verify that the decisions and inferences made from the assessment are reasonable and valid in terms of their consequences. In other words, does the assessment really measure what you want it to measure (ie, knowledge of the fundamentals of laparoscopic surgery that has an influence on competence in this area)? Such "construct" evidence is often indirect and must be gathered over time from a variety of analyses. The beta testing process was designed to provide some of this evidence.

Demographic and biographic data were gathered on all beta testing participants and, along with the performance data, were used to investigate the following 4 construct validity questions:

1. Are there expected differences in performance score between groups of participants with varying training levels or credentials?
2. Are the assessment components related to external data such as supervisory ratings, experience variables, or self-estimates of ability in an expected manner?
3. Are the test scores free from any geographic or administrative location bias?
4. Are the assessment components (multiple choice, scenarios, and performance-based scores) related in an expected or theoretically sound manner?

As expected, preliminary results from the beta testing showed significant differences in cognitive performance for participants with different levels of training, specifically among second-year residents, fifth-year residents, and a combined group of fellows and practicing surgeons who are SAGES members. These differences were noted for both Test A and B and for the overlapping sections.

Significant differences in performance were also noted for different experience levels as determined by a self-report of number of various laparoscopic procedures performed. The correlation between the two variables was around 0.81 in the beta testing sample. From a validity standpoint, one would hope that performance/differences in training level would be less than those noted for laparoscopic experience levels, as differences in training could be due to a progressive increase in general surgical knowledge rather than an increase in the specific knowledge of laparoscopic surgery. To examine this point, differences in laparoscopic experience was controlled statistically (ie, entered as a covariate), and another analysis of cognitive performance across training levels was performed. When experience level was controlled, the differences in cognitive performance across training levels were no longer statistically significant. The reverse situation, which is controlling for training level and then examining differences among experience levels, still showed statistically significant differences. These findings help ensure that the cognitive assessment is sensitive to changes in laparoscopic knowledge and not just to increases in general surgical knowledge.

Further evidence that the cognitive assessment is measuring laparoscopic knowledge was sought by examining the relationship between assessment scores and self-ratings of competence. Self-ratings of competence in performing basic and advanced laparoscopic procedures independently and in performing general laparoscopic technical skills were all significantly correlated with performance on the cognitive assessment (.62, .63, and .56, respectively). Participants who rated themselves above the median in competence scored significantly higher on the cognitive assessment.

Finally, the components of the assessments relate to each other in an appropriate manner. Within the cognitive assessment, the multiple-choice questions are correlated with the scenario questions, but not so much that the two formats can be seen as duplicative ($r = \sim.60$).

TECHNICAL SKILLS—INSTRUCTION AND ASSESSMENT

As noted previously, the FLS committee felt that a component that addresses the training and evaluation of technical skill is essential to any curriculum on the fundamentals of laparoscopic surgery. Basic manual skills must be acquired for a surgeon to be competent in laparoscopy. There are several domains in which the skills required for laparoscopy are quite distinct from those that must be attained to perform open surgery,[2,3] including differences in visual spatial perception and the use of very long instruments constrained through a fixed trocar in the abdominal wall. Using a monocular optical system, a surgeon must be able to develop new cues to enable depth perception, manipulate objects within the abdomen or thorax, and use both hands in a complementary manner to cut, dissect, control tubular structures, and suture.

A series of human learning experiments on both practicing surgeons and nonsurgeons was performed in the early 1990s by Tendick et al.[4] These studies demonstrated that 10% of practitioners had an innate ability to perform complex maneuvers endoscopically, 80% of practitioners had a slow, but steady learning curve to gain mastery, and some 10% of the population had a fundamental inability to work in the 2-dimensional "video" environment. The endoscopic abilities were not associated with open surgical skill or experience—it was more a factor of neuromotor and spacial/perception functions. Although these skills can be learned, learning is best achieved in an environment outside the operating room.

Development of the FLS manual skills exercises. The development process involved identifying a set of simulations that incorporated a large number of the psychomotor skills required during laparoscopic surgery. These exercises could then be used for training as well as for assessment. After a skill-oriented review of a series of videotapes of laparoscopic operations, seven exercises representing common elements of laparoscopic procedures were developed using a physical simulator.[5] A physical simulator was chosen over a computer-based or virtual reality simulator because of the advantages of being inexpensive, portable, and reproducible, and because it included an optical system and instrumentation similar to those used in the operating room (Fig 1). Previous studies[6,7] showed that 2 of the exercises (clipping a tubular structure and placing and securing a mesh over a defect) failed to contribute any additional discriminatory value to the training or assessment, and so for simplicity they were eliminated from the FLS program. The 5 remaining exercises included peg transfer, pattern cutting, ligating loop, suturing with an intracorporeal knot, and suturing with an extracorporeal knot.

An instructional CD in a "watch and do" manner demonstrates the 5 exercises being carried out in a simple laparoscopy box. The exercises are non-procedure-specific and designed to improve the technical facility of a basic laparoscopic skill set.

Manual skills test metrics. Intrinsic to any educational system is feedback. To provide useful feedback to the trainee, each exercise was accompanied with metrics, which provided a score for performing the exercise based on time and accuracy. Thus, the trainee is rewarded by performing an exercise efficiently and without error. To ensure that the measurements attained were objective and reproducible, the reliability of the measuring system was tested by measuring inter-rater reliability and test-retest reliability. All tasks proved to be highly reliable with interclass correlation coefficients of 0.99 and 0.89, respectively, for the entire manual skills test.

Fig 1. Laparoscopic skills simulator used for practice and for testing.

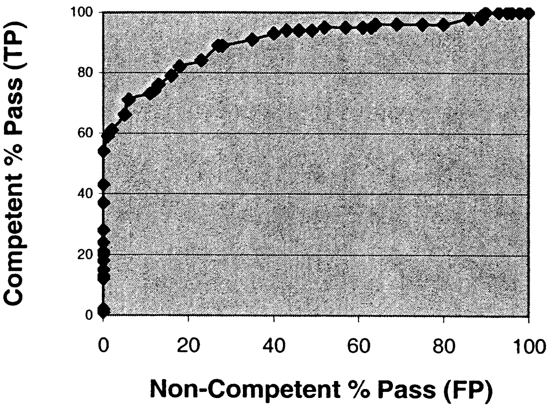

Fig 2. Receiver operator curve; prevalence of a passing score for competent versus non-competent surgeons on MISTELS skills test. (Reprinted with permission from Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass-fail score for the MISTELS system. Surg Endosc 2003;17:964-7. Copyright Springer-Verlag.)

Test validity. The use of the exercises for training or assessment requires evidence that they are representative of the domain of manual skills in basic laparoscopy. Experienced laparoscopic surgeons assembled a list of 14 skills commonly required in basic laparoscopic procedures. Participants in the beta testing (n = 44) were asked after completing manual skills assessment to indicate which of the listed skills was required for successfully completing each of the FLS exercises. The majority of participants felt that 11 of the listed skills were incorporated in 1 or more of the exercises. Four of the skills were viewed as required by all the exercises, and 3 of the skills seemed to be only required by 1 of the exercises. Global ratings on the exercises revealed that beta test respondents felt they would be useful for training and that the skills required were similar to those required by actual laparoscopic surgery.

Whether the simulated exercises are used for training or assessment, the measurements obtained must be a valid reflection of performance in the operating room. Because there is no gold standard for technical performance in the operating room, validity of the skills training had to be demonstrated in other ways. Performance scores were assessed of residents at different levels of training, assuming that the scores would show progressive improvement, which was indeed confirmed.[6] The data showed that the variance in performance scores is large early in the residency program, but at the chief resident level, not only were their scores higher, but they also showed less variance, providing further evidence for the validity of the measurement.

Preliminary data from the recent beta testing confirm that performance on the manual skills assessment increases significantly with training. Second-year residents had greater variance than did fifth-year residents or fellows, but the variance of practicing surgeons in the beta test sample was nearly as large as that of the second-year residents. Beta test sample scores were also positively related to self-reports of laparoscopic surgical experience and to self-ratings of competence in technical skills and in handling an advanced case. And similar to the cognitive assessment, differences in training level were no longer present once laparoscopic experience was controlled.

Investigators from McGill University compared novice surgeons (surgical residents in their first 2 years) with a "competent" group (chief residents, fellows, and practicing laparoscopic surgeons). A highly significant difference in performance scores was found for each of the tasks and the sum of the tasks.[8] Frequency distribution curves were developed for performance of these 2 groups. From these data, receiver operator curves were generated, allowing determination of a cutoff score to maximize the sensitivity and specificity of the FLS skills test in separating competent from non-competent surgeons (Fig 2).[8] Using a passing score generated by this means, the test provided sensitivity, specificity, and positive and negative predictive values all in excess of .80.

The final step in validation of the FLS skills program was provided by showing a correlation between performance scores in the FLS manual skills component and in-training evaluations performed at the end of clinical rotations, which address technical skills performance. These data also showed very good separation in FLS scores between resident surgeons that were more highly rated in their technical skill performance in the operating room versus poorer performers.[9] Given that the same exercises could be used for both training and assessment, it is important to establish that practice on the exercises leads to an increase in performance in the operating room as well as in the simulated tasks.

SUMMARY

In summary, the FLS program was developed because of an identified need to educate surgeons in the underlying principles and basic skills of laparoscopic surgery and because of the growing demand to document competency in surgical practice. The program was designed by a committee of experts who first defined relevant content and then created an educational program. Examinations of both cognitive and manual skill were developed to assess learning, and metrics were used to value technical efficiency and accuracy. The inclusion of a testing component ensures that competence in laparoscopy is both taught and evaluated. The program was shown to be reliable and valid by a rigorous metrics process and from the results of recently completed multi-institutional beta testing. Ideally, measurements attained in the FLS program can be shown to be a valid reflection of performance in the operating room. SAGES hopes to further document the validity and impact of FLS by collection of longitudinal data and performance comparisons. We believe the process used to develop this program can be followed in other areas where new procedures or new technologies are introduced.

REFERENCES

1. Leach DC. The ACGME competencies: substance or form? J Am Coll Surg 2001;192:396-8.
2. Hanna GB, Shimi SM, Cuschicri A. Task performance in endoscopic surgery is influenced by location of the image display. Ann Surg 1998;227:481-4.
3. Alarcom A, Berguer R. A comparison of operating room crowding between open and laparoscopic operations. Surg Endosc 1996;10:916-9.
4. Tendick F, Bhoyrul S, Way LW. Comparison of laparoscopic imaging systems and conditions using a knot tying task. Comp Aided Surg 1997;2:24-33.
5. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg 1998;175:482-7.
6. Derossis AM, Antoniuk M, Fried GM. Evaluation of laparoscopic skills: a 2-year follow-up during residency training. Can J Surg 1999;42:293-6.
7. Fried GM, Derossis AM, Bothwell J, Sigman HH. Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. Surg Endosc 1999;13:1077-81.
8. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. Surg Endosc 2003;17:964-7.
9. Feldman LS, Hagarty SE, Ghitulescu GA, Stanbridge D, Fried GM. Relationship between objective and subjective assessment of technical skill in surgical residents. J Am Coll Surg. In Press.