Assessment of Construct Validity of the Endoscopic Sinus Surgery Simulator

Harman Arora, MD; Jose Uribe, MD; Walter Ralph, MD, PhD; Michael Zeltsan, MS; Hernando Cuellar, MD; Anthony Gallagher, PhD; Marvin P. Fried, MD

Objective: To study the relationship between performance on an endoscopic sinus surgery simulator (ES3) and fundamental perceptual, visuospatial, and psychomotor abilities.

Design: Validation study.

Setting: Tertiary care medical center.

Participants: Thirty-four medical students and 4 otolaryngology residents voluntarily enrolled.

Interventions: Subjects performed tasks on the ES3, minimally invasive surgical trainer virtual reality (MIST-VR), pictorial surface orientation (PicSOr), and 3 visuospatial tests (cube comparison, card rotation, and map planning).

Main Outcome Measures: The MIST-VR was scored for time, task error, economy of hand movement, economy of diathermy, and total score. Scores were generated for the PicSOr task and visuospatial tests. Scores were correlated with time, accuracy, and total subscore on navigation, injection, and dissection tasks, as well as hazard score and total trial score on the ES3.

Results: The PicSOr score was statistically significantly correlated with the hazard score on the ES3 ($r=0.50, P<.001$). Cube comparison ($r=0.43, P<.01$) and card rotation ($r=0.45, P<.01$) scores correlated significantly with the ES3 trial score, as did the MIST-VR total score and the ES3 trial score ($r=0.57, P<.001$). In a multiple regression model, the PicSOr, cube comparison, and MIST-VR total scores were statistically significant predictors of ES3 performance ($r=0.63, P<.01$).

Conclusions: Scores on the ES3 correlate strongly with scores on previously validated measures of perceptual, visuospatial, and psychomotor performance. The ES3 provides a reliable assessment of factors that are important to the acquisition of minimally invasive surgical skills, demonstrating construct validity.

practice of new techniques, which is of particular importance in the realm of minimally invasive surgery (MIS). Minimally invasive surgery requires a skill set that is different from that required for open surgery. Procedures require coordination of surgical instruments with an endoscope in 3-dimensional space. Surgeons must become acquainted with the fulcrum effect associated with instrument handling, as well as the psychomotor constraints imposed by the endoscopic interface. These tasks require complex ambidextrous perceptual, visuospatial, and psychomotor performance. Virtual reality simulators are gaining acceptance as a means of training in the skills necessary for MIS.

Endoscopic sinus surgery (ESS) is considered the standard of care for the operative treatment of many disorders of the sinuses and nasal cavity. Manipulation and instrumentation during the procedure are challenging because of the complex anatomy and the close proximity of important structures such as the brain, orbital contents, and carotid artery. Overall rates for complications of ESS vary, with most studies reporting a rate between 5% and 10%. Complications include cerebrospinal fluid leaks, orbital injury, anosmia, septal perforation, and bleeding. Endoscopic sinus surgery is a well-suited procedure for computer-based medical simulation. It is associated with minimal deformation of anatomy because of the rigid structure of the nasal cavity and sinuses. Furthermore, a solid grasp of intranasal anatomic relationships and attentiveness to spatial boundaries are considered more important to the task at hand than the actual psychomotor task of endoscopic tissue removal.

The endoscopic sinus surgery simulator (ES3) was developed by Lockheed Martin Corporation (Akron, Ohio) based on flight simulation technology. It is under investigation as an adjunct to training in otolaryngology residency training programs. An important aspect of performance on the simulator is having the skill set necessary to conduct ESS. As already mentioned, these skills differ considerably from those associated with open surgery. It is important to ensure that any technology that proposes to simulate such a procedure will use the same skill set. In an effort to validate the ES3 in this manner, we used standardized tasks of perceptual, visuospatial, and psychomotor skills to investigate whether performance on these fundamental tasks is correlated with performance on the ES3. Independent evaluation of such abilities is an important indicator of the ability of the ES3 to measure these human factors that are believed to be essential for the learning and practice of minimally invasive surgical skills.

METHODS

Medical student subjects were recruited via e-mail. Thirty-four of approximately 360 students e-mailed were included in the study. Otolaryngology residents were recruited based on level of training. Four of 8 eligible residents were included. Appropriate institutional review board approval was obtained for the study, and consent forms were filed for all subjects. Each participant was asked to complete a novice trial on the ES3, a set of 35 trials on the pictorial surface orientation (PicSOr), 3 visuospatial tasks (including cube comparison, card rotation, and map planning), and 6 tasks on the minimally invasive surgical trainer virtual reality (MIST-VR). All tasks are described herein. Before beginning the study, the subjects were instructed on how to perform all tasks and had demonstrations of each task being performed. No subjects had prior formal training in any of the tasks.

The ES3 consists of 3 linked computer systems. These included one containing the virtual patient model responsible for the endoscopic image, the surgical interactions, and the user interface; another dedicated to the control of the electro-mechanical hardware; and the third responsible for voice recognition and virtual instruction. The ES3 depicts images based on the Visible Human project conducted by the National Library of Medicine. A haptic (force feedback) system is incorporated into the simulator to allow for tracking the position and orientation of the endoscope, as well as those of the surgical instrument. This system allows force to be applied to the tip of the surgical instrument as would be experienced in surgery, although there is no force applied to the endoscope. Three levels of instruction are available for training. The novice mode introduces the subject to an abstract environment in which the student can become accustomed to the use of the endoscope and the forceps instrument. The student performs basic skills at this level with the help of training aids. The student navigates through virtual space via 4 sets of hoops, injects local anesthesia into 3 targets, and dissects 12 three-dimensional structures, while avoiding adjacent placed structures acting as hazards. The intermediate mode introduces the anatomic structures of the sinuses and nasal cavity. The student performs navigation through nasal anatomy using hoops as training aids, injection of epinephrine with targets placed on anatomic structures, and dissection of the uncinate, ethmoid cells, agger nasi cells, and maxillary antrum with the help of anatomic labels. The advanced mode is a replica of the intermediate mode without the training aids so that the student must use knowledge gained in the intermediate mode to perform the same tasks. Hazards in the intermediate and advanced modes include real anatomic structures such as the lamina papyracea, periorbital fat, optic nerve, and anterior ethmoid artery. Each trial is assigned an overall score based on 4 subscores, including a navigation score, an injection score, a dissection score, and a hazard score.

The PicSOr is a test developed to assess normal human depth perception. It uses a subset of the techniques described by Cowie. Each item is a picture on a computer monitor, showing a spinning arrowhead with its point touching the surface of a geometric object (a cube or a sphere). The subject’s task is to maneuver the arrowhead (using cursor keys) until its shaft is perpendicular to the object surface at the point where they touch. The motor element of the task is deliberately trivial: the shaft changes its orientation only when a cursor key is pressed, and it can be nudged backward and forward repeatedly until the user is satisfied that it is correctly aligned with the cube. Hence, the task is a pure test of the ability to recover the pictorial cues that specify how structures are oriented in (virtual) pictorial space and to compare the implied orientations.

Following pilot studies by the Northern Ireland Center for Endoscopic Training and Research at the Queen’s University of Belfast, 3 main simplifications were made to produce a tractable test. First, although a cube or a sphere can be used in this task, the cube, representing the perceptual problem in its simplest form, was chosen. Second, the research group identified the most important measure of performance in this task as the correlation r between theoretically correct arrowhead orientation and the results chosen by the subject. Third, as a compromise between brevity and reliability, the number of trials per person was reduced to 35.

To assess visuospatial skills, we chose 3 tasks from a battery of tests generated by the Educational Testing Service (Manual for Kit of Factor-Referenced Cognitive Tasks).\cite{10} The chosen visuospatial tasks ask a subject to appreciate the spatial representation of objects drawn on paper that are arranged in different manners. The specific tasks used in our study (cube comparison, card rotation, and map planning) were found in a recent study\cite{11} to have the strongest relationship to endoscopic performance.

Psychomotor ability is essential for the adaptation, consolidation, and development of skills in MIS. The visuomotor discordance of the fulcrum effect on instrument navigation and handling flattens the learning curve in terms of overall surgical performance.\cite{12} The MIST-VR is a virtual reality system developed by Mentice Corporation (Gothenburg, Sweden) in which a subject is able to execute specific tasks that are functionally related to tasks performed in laparoscopic surgery and receive feedback about his or her performance.

In task 1, the subject simulates grasping tissue by gripping a sphere and transferring it to a 3-dimensional location. In task 2, the task is repeated with the addition of a transfer of the sphere from one gripper to the other before transfer to the 3-dimensional box. Task 3 simulates moving along the bowel by using hand-over-hand transfer of a cylinder. In task 4, the subject is directed to remove a tool from the operating field and reinsert it accurately. Task 5 prompts the subject to cauterize 3 subtargets placed consecutively on a sphere using a foot pedal. Finally, in task 6, the subject is directed to maintain the sphere within the target box while cauterizing 3 consecutive subtargets with the foot pedal.\cite{13} The MIST-VR has been independently validated as a good measure of endoscopic psychomotor performance.\cite{14,15}

Data were analyzed using univariate Pearson product moment correlation coefficient and multiple regression to assess the strength of the relationship between the different measures of fundamental abilities (i.e., perceptual, visuospatial, and psychomotor) and performance on the ES3. Statistical significance was set at \( P < .05 \).

### RESULTS

The *Table* summarizes the main findings from the analysis of the ES3 scores and the scores on the tests of perceptual, visuospatial, and psychomotor ability. Performance on these tests was correlated with time, accuracy, and total subscore on the navigation, injection, and dissection tasks on the ES3.

The only score with which the PicSOr score was statistically significantly correlated was the hazard score for the entire trial \( (r=0.50, P < .001) \). Hazards are a measure of error for the trial as a whole. The cube comparison \( (r=0.43, P < .01) \) and card rotation \( (r=0.45, P < .01) \) visuospatial test scores correlated strongly and statistically significantly with the ES3 scores, particularly the overall score. The correlations observed between the ES3 scores and the map planning task scores were much weaker than those of the other 2 tests and were not found to be statistically significant.

The MIST-VR total score varied in the strength of the correlation between it and the ES3 scores. Five of these correlations were statistically significant. However, one of the strongest correlations observed was between the MIST-VR total score for all 6 tasks and the ES3 (total) trial score \( (r=0.57, P < .001) \). When the perceptual task PicSOr score, the visuospatial task cube comparison score, and the psychomotor task MIST-VR total score were included in a multiple regression model, they were found to be strong and statistically significant predictors of ES3 performance \( (r=0.63, r^2=0.39, F_{3,22}=4.77; P < .01) \).

### COMMENT

Although there are many other factors involved in surgical training, mastery of technical skill is critical to becoming a competent and safe surgeon. There is an overwhelming consensus in the medical community toward improving patient safety. The field of surgery is progressing rapidly to the forefront with the use of simulation technology, not only as a training tool but also as an assessment tool for skills of established surgeons.\cite{16} The benefits are numerous. Performance can be objectively measured without the need for an experienced surgeon to be present during instruction. Simulators can be used at any time, rather than having training hindered by the lack of availability of procedures during a specified time frame. In addition, the use of simulators has the potential of decreasing cost. This can be achieved by reducing errors and the subsequent expense of prolonged hospitalization, revision procedures, and legal consequences that might ensue.

As types of procedures gain in complexity with the increasing use of minimally invasive techniques, training must be more thoroughly aimed at these procedures.
to ensure satisfactory outcomes. Because such techniques have become more commonly practiced, associated complications have inevitably increased among surgeons' early experience with these procedures. Therefore, extensive skills training and further assessment of an individual's fine motor skills and hand-eye coordination must be core components of any surgical training program. Fundamental abilities such as those evaluated in our study are important to consider, but they do not necessarily translate directly into operative situations. Simulation devices such as the ES3 have moved a step beyond current measures of surgical ability by providing ongoing assessment of the user's performance. The ES3 is a suitable prototype that answers the need for directed training. The additional value provided by such training and the integrated evaluation of performance would be a novel adjunct to current teaching programs.

In this study, we sought to determine whether the ES3 actually captures aspects of the skills that it purports to measure, thus providing evidence for construct validity, defined as the evaluation of a testing instrument based on the degree to which the test items identify the quality, ability, or trait it was designed to measure. Endoscopic sinus surgery is associated with a skill set that must be acquired by first becoming accustomed to the psychomotor and perceptual aspects of the procedure. Skills required for endoscopic surgery differ markedly from those used in conventional surgery. Training must address the use of videoscopic interface, as increasing the level of this skill will directly relate to the clinical situation. Virtual reality training that replicates such a skill will help trainees progress in a manner that is objectively quantified without inherent risk to patients. The MIST-VR has been previously validated as a measure of endoscopic psychomotor performance. In the study reported herein, MIST-VR performance scores were found to be strongly correlated with performance on the ES3, suggesting that the ES3 captures important aspects of psychomotor performance necessary for the practice of endoscopic surgery.

The PicSOr was specifically developed to assess perceptual skill by testing an individual's ability to recover pictorial cues. When performing the dissection task on the ES3 in novice mode, the geometric objects representing the target and the hazards can be differentiated by recovering such pictorial cues. One would expect the hazard score on the ES3 to correlate with the PicSOr score, which it did significantly. Spatial orientation and navigation require that a subject be able to determine the relative position of an object with respect to other objects in the environment. The ability to process spatial representations of one's environment and navigate through this environment is important for performance in MIS. Two of the tasks used in the study for assessing visuospatial performance, cube comparison and card rotation, correlated significantly with performance on the ES3, indicating that the ES3 adequately assesses an individual's spatial orientation ability. It is unclear why scores on the map planning visuospatial task did not correlate significantly with ES3 performance. However, cube comparison and card rotation assess spatial orientation, while map planning assesses spatial navigation, and this may be why the relationship is weaker. Our understanding of the relationship between visuospatial ability and MIS is still evolving, and we have taken special note of this finding.

Our objective in completing this study was to investigate whether the ES3 demonstrates validity in measuring what it proposes to measure, thus providing evidence for the construct validity of this instrument. Several observations from this study confirm a relationship between ES3 performance and performance on fundamental perceptual, visuospatial, and psychomotor tasks.

Validation of simulation devices is imperative, as the surgical community plans to use them as an adjunct to train future surgeons. This study suggests that the ES3 is a reliable assessment tool. It also lays the groundwork for future validation studies, especially to confirm ES3 predictive validity, defined as the extent to which scores on a test are predictive of actual operating room performance. Ongoing assessment of predicting validity will determine whether this virtual reality simulator will translate to improved performance in the operating room. It is anticipated that as this and other simulation devices are validated, they will be widely used in surgical training programs so that their benefits may be assumed by all surgeons and surgeons in training.

Submitted for Publication: August 10, 2004; accepted November 11, 2004.

Correspondence: Marvin P. Fried, MD, Department of Otolaryngology, Albert Einstein College of Medicine and Montefiore Medical Center, 3400 Bainbridge Ave, Third Floor, Bronx, NY 10467 (mfried@montefiore.org).

Funding/Support: This work was supported by grant R18 HS11866 from the Agency for Healthcare Research and Quality, Rockville, Md.

REFERENCES


