Formal Evaluation of the Madigan Endoscopic Sinus Surgery Simulator

submitted to

Simulation and Training Systems
Lockheed Martin Tactical Defense Systems

Suzanne Weghorst
Christopher Airola
Peter Oppenheimer

Human Interface Technology Laboratory
University of Washington
Seattle, Washington

Lt. Col. Charles Edmond, M.D.
Troy Patience

Madigan Army Medical Center
Ft. Lewis, Washington

David Heskamp
Jeffrey Miller

Simulation and Training Systems
Lockheed Martin Tactical Defense Systems

Table of Contents

Abstract

[executive summary here]

Introduction

The use of computer simulation as a training tool for surgical procedures has been inspired by the success of computer simulation flight training as well as the high cost of medical education. In recent years advances in interactive graphics and virtual reality technology have greatly enhanced our arsenal of instructional tools (Caird, 1996) and are moving these systems toward more general commercial graphics platforms.

A recent survey of prior research in surgical simulation suggests that each surgical procedure has a unique set of simulation requirements (Edmonds et al., 1997). Research thus far include abdominal laparoscopy (Cover et al 1993; Hon, 1994; Merril, 1994; McGovern et al, 1994); limb surgery (Chen et al, 1992; Pieper et al, 1991); eye surgery (Peifer et al, 1994; Sagar et al, 1994); plastic surgery (Pieper, 1989); gastrointestinal endoscopy (Baillie et al, 1992; Bard, 1990; Gillies et al, 1992; Poon et al, 1988); anesthesiology (Good et al, 1993); epidural anesthesiology (Bostrom et al, 1993; Stredney et al, 1996); and interventional radiology procedures (Dawson et al, 1996).

Much of this recent work in medical simulation has received impetus from the leadership of both DARPA (Satava, 1996) and the Army Medical Command.

Wickham (1994) summarizes nicely the need for these novel and extensive training techniques for endoscopic surgery skills:

> Evaluation of new operative competence is urgently needed because of the rapidity of changes in interventional treatment. Training programmes must be established so that interventionists' training is similar to that of airline pilots. A surgeon or radiologist should not be allowed to treat patients with sophisticated and potentially dangerous instruments without the experience of simulated operations and closely supervised procedural training. Fully equipped training centers should be established with simulator laboratories where interventionists can develop the different hand-eye coordination required for the transition from open to endoscopic techniques...The need is urgent: the traditional methods of "see one, watch a video, do one" is completely inadequate preparation for minimally invasive techniques...A theoretical evaluation of competence by written or oral examination is totally insufficient to determine whether a clinician has gained the manual ability to carry out complex open or endoscopic surgery.

The motivation for medical training simulators is clear. It is important, however, to ask: Do they work? Is the simulator effective as a training device? What skills do students learn most effectively? Do the skills learned in the simulated training transfer to the operating room? How fast do students learn? How does an hour of simulator time compare to an hour of traditional training methods? Are these simulators currently cost effective? If not, when will they be cost effective for use in medical schools? And how can simulators best be integrated with medical school curriculum?

To answer these questions, formal evaluation of surgical simulation systems is essential. Without answers to these questions one remains uncertain as to whether students are receiving surgical training or merely learning to be good simulator users.

The evaluation study described here has been guided by the work of many others in the field. As Hoffman et al (1996) suggest, we have included the end-users during the formative stages of the simulator design, a vital step in establishing educational goals and curriculum design. In developing our evaluation criteria we have taken into account both objective and subjective considerations (Robb, 1997), have relied heavily on the basic surgical proficiency measures of time and accuracy (Johnson et al, 1996), and have incorporated much of our evaluation protocol into the simulator itself (Hon, 1996; Kaufmann, 1997).

Simulator Overview

The Madigan Endoscopic Sinus Surgery (ESS) Simulator uses virtual reality technology, a force feedback (haptic) display, and 3D computer-based anatomy models as a tool to teach a variety of skills needed to perform such surgery.

The heart of the system is a 3D model of the human nasal sinus anatomy derived from the National Library of Medicine's Visible Human Database. Researchers at the Ohio Supercomputer Center took the photographic cryosections of the male dataset and created segmented surface models of the sinus anatomy. The Lockheed Martin team then added stochastically generated surface textures to complete the anatomical model.

This model can be rendered in real time (15-30 frames per second, depending on primarily on anatomical position) on a Silicon Graphics Onyx System and viewed at NTSC resolution on a standard video monitor, thus simulating the view that a surgeon would see of a video endoscopic display.

To interact with the model and perform the simulated surgery, students in training use a pair of 6 degree of freedom input devices developed by Immersion Corporation. One such device represents the video endoscope and the other represents the surgical instrument, such as an injection needle or a forceps (Rosenberg and Stredney, 1996). An external view of an experienced ENT surgeon operating the system can be seen in Figure 1. Note that the position of the surgeon relative to the endoscopic monitor and simulated patient emulates a typical clinical configuration.

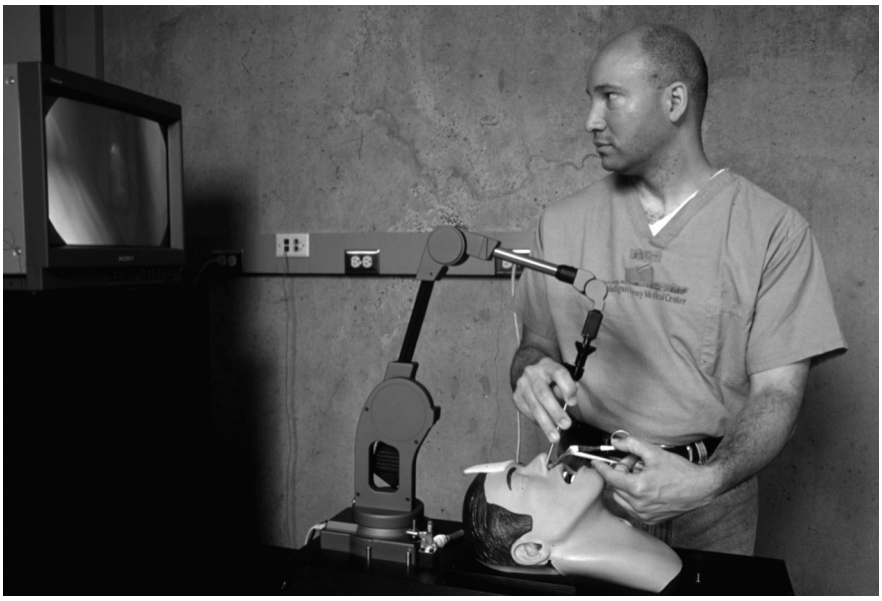


Figure 1. Operative configuration of the Madigan Endoscopic Sinus Surgery Simulator.

As the student manipulates the input devices, the simulator tracks the position and orientation of the devices, updates the positions and orientations of the virtual endoscope and virtual instrument, manipulates the virtual anatomical model accordingly, and displays the resulting virtual endoscopic view on the monitor. In addition, the system tracks the opening of the forceps handle of the instrument input device. All together, the system measures 13 degrees of freedom of the student's input.

The physical input devices are designed to resemble the feel of an endoscope and forceps and are assembled with a latex replica of a human head. The endoscope input device resides outside of the head, while the instrument input device is inserted into the nostril of the latex head and attached to the position tracker inside the head. In addition to displaying the simulated endoscopic view, the system computes the forces that the sinus tissue would apply to the tip of the instrument during surgery and applies the computed force to the instrument input device, via mechanical coupling within the mannequin head.

The proctor's console provides an interface to the run-time system parameters, along with optional radiographic views of the current anatomy and optional performance feedback summaries for the trainees.

Training Aids

Optional 3D graphical overlays are superimposed on the endoscopic view to provide performance aids for the student. These overlays include a path of circular hoops representing the desired endoscope trajectory, bullseye targets representing the desired injection sites, and text labels identifying anatomical feature landmarks.

The system provides voice audio feedback representing the current status of the surgical procedure, as well as a simulated heartbeat which responds to certain user actions.

Training Tasks

Students using the system are instructed to perform a simulated surgical procedure consisting of three subtasks: Navigation, Injection and Dissection. In addition to archiving the frame-by-frame position of the devices, the system measures the time required to perform these tasks, as well as the accuracy with which they are performed, to generate an evaluation of the student's overall performance.

Three versions of the procedure were developed corresponding to three skill levels: Novice, Intermediate, and Advanced. The primary differences among these procedures are the type of geometric model used and the presence of training aids.

The Novice procedure uses a simplified abstract model of the anatomy consisting of several spheres inside an open box. The Intermediate and Advanced procedures use the more complex surface sinus model generated by OSC. The Novice and Intermediate procedures use the training aids described above, whereas the Advanced procedure is performed without benefit of these aids (to more accurately simulate the target procedure).

Evaluation Overview

Evaluation efforts for this project fall into two general categories of activity:

- "formative" evaluation, which attempts to provide design specification input to the development team during the development process, and
- "summative" evaluation, which assesses the success of that effort by formally analyzing the effectiveness of the system.

Each of these activities is discussed in detail below.

Formative Evaluation

Throughout the development of the ESS simulator, the HITL team worked in close collaboration with Dr. Charles Edmond, staff otolaryngologist at Madigan Army Medical Center and principle investigator for the project, to perform ongoing formative evaluation and make design recommendations to the development teams at Lockheed-Martin, the Ohio Supercomputer Center, and Immersion Corporation.

The Methods of Approach for the formative evaluation phase included the following:

1) endoscopic video analysis to determine simulator performance requirements

2) geometric complexity requirements analysis

3) prototype anatomical modeling

4) development of spatial awareness aids, interface features and rendering approaches

5) development of a prototype simulator with an integrated expert system assistant

6) development of a surgical training curriculum to be embedded in simulator

7) survey of medical experts to determine feature and curriculum priorities.

Dr. Edmond provided the core of domain expertise in sinus surgery. After a few weeks of study, the rest of the HITL team had gained a basic familiarity with sinus anatomy and surgical issues. During this time Dr. Edmond gained familiarity with the computer graphics tools available at HITL for use in prototyping simulator design and computer-assisted surgery applications.

Design Requirements Analysis

Frame Rate Requirements

Method: Representative sequences of live videoendoscopic sinus surgery were digitized and manipulated. These sequences were

selected to include surgical interaction with tissue as well as to highlight early operative and late operative anatomy. Image degradation due to the scanning process itself was minimal.

We then re-recorded the scanned video at the following frame rates: 30 frames per second, 15 fps, 10 fps, 5 fps, 2 fps, and 1 fps. These recordings were then reviewed by expert surgeons to assess the minimum frame rates required for the sinus surgery simulator.

Results: To the extent possible 30 fps should be maintained. For surgical dissection a minimum of 15 fps (preferably 30 fps) should be maintained; temporary slow downs to 10 fps during deformation and dissection may be tolerable for certain instruments and maneuvers (e.g., side-biter, but not sickle knife). Frame rates lower than the 10fps minimum may cause disorientation in navigation.

Geometric Complexity Requirements

We applied the following guiding principles in assessing the geometric complexity requirements:

- Visual fidelity requirements are task driven.
- The task and sub-task relevant to each individual varies based on level of training and experience.
- Three categories of expertise were defined: novice, intermediate and expert. Tasks and sub-tasks were delineated for each category, and fidelity requirements proposed.
- Accurate anatomic representation (visual realism) and 3-dimensional spatial awareness are two of the most critical aspects for effective cognitive development. This was felt to be important across all levels of training and experience for endoscopic sinus surgery
- The development of psycho-motor skills necessary for successful endoscopic sinus surgery may require less visual realism and more 3-D spatial awareness.
- Low level visual realism might be less distracting to the novice, and therefore advantageous for both developing and assessing a user's psycho-motor skills.
- Varying degrees of visual fidelity are not only helpful, but necessary to develop the cognitive and psycho-motor skills for sinus surgery.

Methods: To help assess the geometric complexity requirements for the anatomical models we generated a prototype anatomical model using the following technique:

A 3D triangular mesh was generated by texture mapping an actual video endoscopic image onto a flat mesh surface and vertically displacing each vertex of the mesh in proportion to the brightness of texture at that vertex. We then varied the resolution of the mesh and evaluated the quality of the resulting images. This model is a fairly rough approximation, since video brightness does not precisely correspond to geometric distance from the endoscope. Specular highlights in the texture create spike artifacts in model. These were eliminated by painting over the highlights in the texture.

Geometric complexity requirements were determined by visual inspection of the resulting simulation mock-ups by the project domain experts. Evaluation criteria included:

- Ability to identify features
- Subjective evaluation of photo-realism
- Potential for simulator miscues

Results: We concluded that given the performance of the rendering system, frame rate was a critical performance requirement. When the initial versions of the anatomy model were delivered from OSC, the frame rate did not meet our specification. The geometric complexity of the model was then reduced until the specified frame rate was achieved. The geometric complexity requirement was essentially a function of the frame rate requirement and the rendering performance of the hardware and software engines.

The anatomical model developed for the ESS simulator adequately served its function. We noted considerable improvements in quality and efficiency of the model over the duration of our evaluation. Upon each release of a new model, our team would evaluate the model for rendering efficiency and fidelity. Dr. Edmond would suggest modifications to the model and these would be executed in subsequent releases. The addition of textural cues greatly enhanced the apparent geometric detail without sacrificing rendering speed.

Prototype Anatomical Modeling

Prototype anatomical virtual models were developed by Edmond and Oppenheimer using the Alias modeling package. The purpose of

these models was to provide guidelines in the design of the patient specific data-driven models produced by OSC, as well as to determine the necessary rendering and interface features for use in the simulator.

Methods: The geometry of these prototypes was based on cadaveral section photographs taken primarily from sinus anatomy atlases and surgery textbooks (Rohen and Yokochi, 1983; Rice and Schaefer, 1988). In one case, sections were scanned into the computer, and mapped onto parallel image planes. These image planes were then used as templates for drawing surface contours on the planes. These contours were then lofted into surfaces. Additional contour curves were added and edited as needed, based on Dr. Edmonds observations.

In other cases, freeform surface contours were drawn orthogonal to the scanned image plane. The scanned image was then projection texture mapped onto the resulting lofted surface. Although less geometrically accurate, this technique enabled photographic texture maps to be used in the final evaluation model.

Results: These prototype models provided the basis for experimenting with navigation aids, interface features and rendering approaches as described in the next section.

Interface Features and Rendering Approaches

The following design experiments using the prototype anatomical models resulted in recommendations and demonstrations of candidate features to be included in the development of the ESS Simulator:

- Use of texture
- Use of transparent or wireframe surfaces to reveal obscured anatomy
- Use of orientation cues to assist in navigation
- Use of tubular paths of hoops as a navigation aid
- Use of targets as an injection aid
- Use of cross hairs overlaid on instruments to represent orientation, and to assess distance from anatomy
- Use of patient face model
- Anatomical segmentation, and interactive selective segment display

- Displaying endoscope position on CT scans as a navigation aid
- Integration of expert system training aids

Texture mapping

The use of texture mapping offers several benefits. Texture maps can represent detail that would be too expensive to model as polygonal geometry. Photographic texture maps add to the realism of the model.

Texture also provides additional depth cues. If two parallel surfaces overlap from a given perspective view, than they may have the same shading coefficients in the lighting calculation. Without texture maps, the two surfaces would be rendered with the same pixel values and therefore the boundary between them would be indistinguishable. By placing texture maps on the surfaces, the discontinuity between the surfaces can be detected more easily, as a consequence of texture discontinuity.

In conditions of extreme ambient lighting, even non-parallel surfaces may have similar shading. Texture mapping can reveal subtle differences in orientation and distance from observer. In "real life" these shading ambiguities are also a noticible if the surfaces are of uniform texture. Real life objects with even subtle surface textural features are more readily discerned that smooth textureless objects.

Methods: Several texture mapping methods were applied to the models. In some cases an attempt was made to align the texture with the underlying model geometry; such textures included anatomical features. In other cases, the textures were closeup details of tissue type, such as mucosa, and did not include anatomical structures.

Results: On the basis of our experiments, we recommended the use of texture mapping in the target simulator. Since the polygonal resolution of the final models generated from the Visible Human database is as fine as the possible textures derived from the same data, detail texture had to be added algorithmically. Such texture, although not photo-realistic, does add to the realism of the simulation as well as improve the surface boundary visibility as described above.

Transparent and wireframe surfaces

By rendering outer surfaces transparently or in wireframe, the system can display otherwise obscured portions of the anatomy. This allows students to view anatomical landmarks, prior to revealing them

through dissection. These landmarks may include anatomy that should not be dissected (such as the lamina, optic nerve, or skullbase). Revealing them may enhance spatial awareness training, thereby preventing a severe surgical error. This feature is not currently implemented in the target simulator.

Orientation cues

By displaying a world-stabilized orthogonal grid or crosshairs, the user can determine the orientation of the endoscope with respect to the patient. This feature is not currently implemented in the target simulator.

Tubular path of hoops

By displaying a tubular surface whose axis is the desired trajectory of the endoscope, the system can provide a navigation tool to the user during surgical simulation. We experimented with two different rendering modes for the tube: (A) a partially transparent surface with alpha-mapped latitudinal rings, and (B) wireframe latitudinal hoops oriented orthogonal to the axial path. The target simulator uses the latter rendering mode (hoops).

Injection targets

Bullseye targets of concentric rings, rendered as flat shaded polygons can be placed on the anatomy at points of injection. These targets provide cues in the novice and intermediate modes of the simulator.

Crosshairs attached to the instrument

By attaching polygonal crosshairs to the tip of the virtual instruments, the system visually represents the relative orientation of the instrument with respect to the endoscope. If a grid texture is mapped onto the crosshairs, the user can assess the distance between the instrument and the anatomy by counting the grid lines between the instrument tip and the intersection between the crosshair and the tissue surface. Grid marks deeper than the tissue surface are obscured and will not be visible. This feature was not incorporated into the current implementation of the simulator.

Patient face model

A polygonal surface representation of the patient face was added to the virtual model. This face model provides the student with a position and orientation cue, as well as providing additional realism.

Anatomical segmentation, and interactive selective segment display

The prototype anatomical model built in Alias was constructed in segmented pieces, corresponding to recognizable anatomical features. This pre-segmentation allows one to selectively display certain parts of the anatomy as well as to highlight selected anatomical features at run time. In addition, virtual tissue segmentation permits collision detection with procedurally meaningful objects. The model based on the Visible Human dataset required an editing process to segment the surface anatomy.

Displaying endoscope position on CT scans as a navigation aid

As originally conceived the user interface to the system made use of an auxiliary CT view, in addition to the virtual endoscopic view of the anatomy. The system has the optional capability of updating the auxiliary CT image view with a crosshairs indicator of the current endoscope position. Although not used by the subjects in the current evaluation study, this feature is available and may be useful for future training protocols.

Prototype simulator with expert system assistant

In parallel with this development project, HIT Lab researcher Mark Billinghurst extended his work with Jesus Savage-Carmona on intelligent multi-modal environments to create a prototype sinus surgery simulator with an integrated expert system assistant (Billinghurst et al, 1996). Working with Oppenheimer and Dr. Edmond, Billinghurst developed a system which incorporates knowledge of the endoscopic procedure into a structured rule base, interprets the user's multi-modal inputs (currently voice and virtual endoscope position) and interacts with the user dyadically. While performing the simulated procedure, the user can query the system about anatomy and the specifics of the procedure, asking the system to identify features or demonstrate maneuvers. In turn the system recognizes the user's actions, and can provide vocal and visual feedback, as well as warnings when the user is about to execute a dangerous maneuver.

Although the system architecture is somewhat different, this prototype simulator provided ground work demonstration for certain features of the target surgical simulator, including navigation through the sinus cavity from an endoscopic viewpoint, use of abstract

graphical overlays as a navigation aid, and embedding of surgical task sequences into simulator

Iterative Testing

In addition to these design experiments based on prototype modeling, the HIT Lab served as a test site for successive versions of the anatomical model and the simulation system. Having the target hardware platform (SGI onyx) in house made iterative evaluation a viable and useful approach. In particular, the proximity of the lab to Madigan Army Medical Center made it relatively easy for the evaluation team to consult with Dr. Edmond regularly on simulator features and system performance.

Upon receiving each release of the ESS simulator, the HITL team would compile a list of feature enhancements, known bugs and other observations. This list would be proiritized by Dr. Edmond and returned to Lockheed Martin for inclusion in the subsequent release.

Haptic system emulator. Instrument tracking and force feedback displays were not integrated into the earliest test releases of the simulator. A menu driven "haptic system emulator" (HSE) provided a graphical user interface to navigate the position and orientation of the endoscope and instrument, as well as to select the instrument type and control the open/close parameter. The HSE consisted of up, down, left, and right buttons, sliders and radio buttons. Although not driven by a spatial input device, the HSE did allow for precisely controlled navigation, and also proved usful in positioning navigation aids within the model.

Although initially developed as a prototype testing tool , the HSE may have some commercial potential as well. Customers desiring a lower-cost system that does not require a complete 6 degree of freedom interface device could use a HSE to perform the surgical tasks (albeit in a way that does not capture many of the targeted clinical skills of the Madigan system). In this case the up, down, left, and right buttons could be replaced with a mouse-based interface or other commercial graphical input device.

Evolution of the Training Framework

During the early phases of the project we examined several other surgical simulator systems to look for places to improve the state of the art. What we noticed was that although these systems were making advances in anatomical modeling and user interaction, they

were missing a structured educational component. In general these systems simulated a specific surgical domain or task only and were therefore "orphaned" experiences in the educational process.

We concluded that in order to be of significant educational value, one had to not only embed the simulator in the existing academic curriculum, but also develop curriculum within the simulation itself. This would not only serve the educational process but would also facilitate our evaluation of the educational effectiveness of the simulator.

The prevailing paradigm in surgical education is usually summarized as: "See One, Do One, Teach One." Our goal was to improve on this paradigm by taking the "See One, Simulate Many, Do One" approach that had proved effective in the domain of flight training. In order to achieve this we needed to develop a curriculum structure.

Task Analysis

Working with Dr. Edmond, we developed a taxonomy of ESS simulation objectives and simulator performance and interface features. Our objective, which proved far too demanding, was to elicit estimates from experienced ESS surgeons of the desirable system performance requirements for each training task. This approach did, however, provide us with a framework for approaching the issues of curriculum design in this domain.

ESS Domain Expert Survey

Twelve experienced ESS staff surgeons from several leading otolaryngology training programs were surveyed by Dr. Edmond to assess their judgments of the primary simulation requirements for physicians performing ESS procedures. The primary objective of this survey was to determine the curriculum needs and perceived importance of several of the candidate features of the system early in its development.

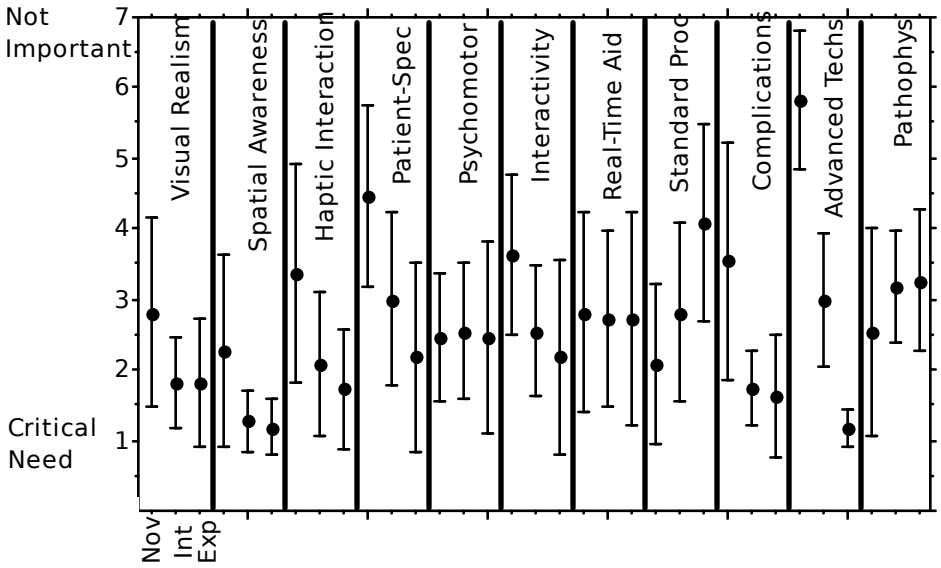


Figure 2. Mean survey response ratings (with 95% error bars) for each of eleven training and operational needs, for each of three hypothetical target groups (novice, intermediate, and experienced ENT physicians).

These domain experts were asked to rate the relative importance (from "critical need" to "not important") for 11 simulator characteristics: visual realism, spatial awareness training, haptic interaction, patient-specific modeling, psychomotor training, real-time interactivity, real-time (intra-operational) performance aids, standard surgical procedures, complications, advanced techniques, and pathophysiology.

In addition, they were asked to assess the value of these characteristics for target subjects at three levels of ENT experience: "novice" (i.e., junior residents), "intermediate" (senior residents), and "expert" (experienced ENT surgeons). Results of this survey are summarized in Figure 2, which indicates the mean rating for each characeristic for the three target groups, on a 1-7 scale.

As can be seen from this graph, these experts deemed "spatial awareness" the most crucial training need (of those presented in the survey) for all levels of subject experience, although "advanced techniques" were seen as equally critical for the experienced ESS target group. With the exception of "standard procedures" and perhaps "pathophysiology", all requirements were seen as equally important for all target levels or more important with increasing ESS experience.

It is interesting to note that almost all of these characteristics were rated on the "important" end of the scale. The characteristics deemed least important for simulation were "advanced techniques" and "patient-specific modeling" for the novice group, and "standard procedures" for the advanced group. In addition, "real-time interactivity", "haptic interaction" and training in "complications" were seen as of only moderate importance for ESS novices.

Finally, it should be noted that these survey respondents were overwhelmingly favorable toward the development of an ESS training simulator. While one or two expressed skepticism about any technological solutions, most respondents noted in their open-ended comments that this was an interesting and useful endeavor, and that the field had a need for such a system.

Simulator Integration

A considerable amount of effort was spent during the development phase on establishing a "curriculum wrapper" for the simulator. We initially envisioned the simulator as embedded within a multimedia training system which would provide a meaningful context and set the protocol for the trainee.

The development team elected instead to integrate relevant aspects of the emerging ENT curriculum into the simulator itself and to postpone further efforts to develop a total curriculum "package". The reasons for this included:

- development of a full-blown ESS curriculum was beyond the scope of this phase of the project
- it became apparent that a staged protocol approach was necessary to make the simulator effective as a training context
- useful techniques emerged (such as the use of navigation hoops and injection targets) which could be integrated relatively easily

The final product of the formative stage of the project was, in fact, more thoughtful with respect to an integrated ENT curriculum than had originally been anticipated and appears to be more useful as a training tool than other virtual reality medical simulators which have emerged in recent years. Its validity as a procedural simulator and its utility for training are evaluated in the following sections.

Summative Evaluation

After delivery and installation at the HIT Lab of Version 1.2 of the system by the Lockheed-Martin development team, the evaluation team began a “shakedown” of the system and testing protocols with the three targeted subject groups (described below). Midway through this phase the system was relocated to Madigan Army Medical Center for further analysis by staff and resident ENTs.

The primary goals of this phase of the evaluation effort were to

- validate its utility as an ESS training environment,
- assess the usability of the system, and
- provide additional iterative feedback to the development team.

The methods of investigation and results of this phase are described below, followed by a set of recommendations for further system development and evaluation derived from these findings. In a planned follow-on phase the primary focus of the evaluation effort will be on assessing the degree and nature of any transfer of this training to the real ESS operating environment (the ultimate objective of the simulator).

Methods

Subjects

Subjects were solicited from three distinct groups: (1) non-MDs with general intelligence and psychomotor abilities roughly comparable to the average otolaryngologist; (2) non-ENT physicians from a variety of specialties, and (3) ENTs with a wide range of ESS experience. We focused on these three groups in order to establish a baseline and asymptote for the evaluation of the efficacy of the simulator in training otolaryngology residents.

The first two groups also allowed us to "shake down" the system and research protocols without sacrificing valuable ENT resident subjects, and provided baseline scores for untrained/unfamiliar subjects. Finally, the non-ENT physician group could provide us with some valuable input about the extensibility of the simulator to non-ENT applications.

Non-MD Group: Twelve volunteers from the University of Washington College of Engineering comprised the non-MD group. These subjects ranged in age from 23 to 54, and included graduate students, professional staff and faculty. All had had some experience with simulation and virtual reality. None had previously used an endoscope or attended medical school.

Non-ENT Physician Group: Eight University of Washington MDs from specialties other than otolaryngology provided extensive feedback on the simulator design and utility for other medical and surgical tasks: 3 videoendoscopic surgeons, 2 radiologists, 1 neurosurgeon, 1 cardiologist, and 1 anesthesiologist. Four of these also provided us with complete trial performance data.

ENT Staff and Residents: Twelve staff and resident ENTs (1 female, 11 males) from MAMC served as subjects. The subjects in this group ranged in age from 28 to 46, with a mean age of 35.2 years, and a standard deviation of 6.15 years. One was left-handed, 10 right-handed, and the handedness of one was unknown.

ENT experience for this group broke down as follows:

- 4 staff (with an average of five years of training, six years of practice and more than 100 ESS procedures performed)
- 3 R2s (with an average of 1-5 ESS procedures performed or observed)
- 1 R3 (with 6-20 ESS procedures performed or observed)
- 2 R4s (with an average of 21-100 ESS procedures performed or observed)
- 2 R5s (with an average of 21-100 ESS procedures performed or observed).

Eight subjects had had occasional videogame experience, two reported playing videogames once, and two reported never playing videogames. Eleven subjects had had no virtual reality experience, while one reported occasional VR experience. Ten reported having no other simulator experience, while one reported one simulation experience and another reported "occasional" simulation experience. Because of the low incidence of prior experience with these systems, these factors were not evaluated further for this group.

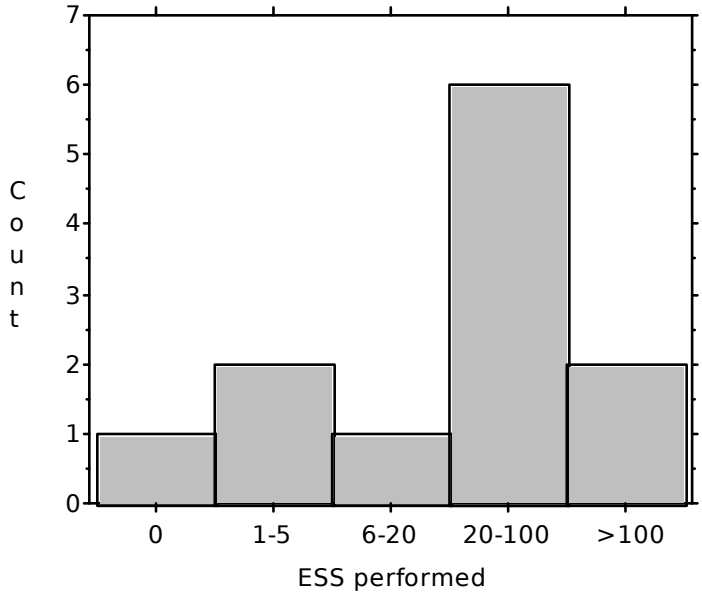


Figure 3. Frequency distribution of the reported number of prior ESS procedures performed by subjects in the ENT group.

Figure 3 shows the distribution of the number of prior ESS procedures performed by the subjects in this group. The two reporting "1-5" procedures were observational only.

Procedures

All subjects proceeded through a common protocol for one or more sessions, involving:

- general orientation and consent form
- pre-session background questionnaire (see Appendix)
- instructions and orientation to the simulator and tasks
- one or more proctored simulation trials using one of the three training models
- post-session debriefing and questionnaire (see Appendix)

All trials were videotaped for later analysis and "think-aloud" comments. Subjects were free to terminate the session at any time.

MODELS

Three models were constructed to provide for a sequentially more realistic ESS experience. For each model, there was both a right- and left-side version. The anatomy was built for the right nostril and was reflected to simulate the left nostril. The nostril (left or right) with which the subject started was determined by their handedness.

Model 1: The novice/abstract model consisted of only the skin of the face and the entrance to the nasal cavity. A 3D grid pattern replaced the sinus anatomy to provide depth of field during the three tasks: Navigation, Injection and Dissection. Training aids were used to guide the subjects through the task. Navigation training aids consisted of virtual hoops and Injection training aids consisted of virtual targets in space.

During Navigation the subjects maneuvered the endoscope through four sets of virtual hoops. The paths of the hoops represented three passes (sets two and three combined are one pass) commonly taken before the surgery begins to gain familiarity with the patient's anatomy and to allow cleaning of the areas of interest.

Injection consisted of maneuvering both the endoscope and instrumented forceps within the environment to inject five targets oriented obliquely in space. The instrumented forceps controlled a virtual needle for this task. The placement of the targets in space reflected the common areas of injection of a vasoconstrictor during a Maxillary Antrostomy.

During Dissection the subjects were also required to use both the endoscope and the instrumented forceps. The task consisted of dissecting each of a series of virtual spheres with pre-selected virtual tools. The instrumented forceps represented each of the tools most commonly used in the procedure: freer, needle, bent needle, sickle knife, microdebrider, suction, straight-biting forceps, up-biting forceps, left-biting forceps and right-biting forceps.

Navigation through the four sets of hoops, injection of the five targets and dissection of each of the spheres was required for a complete score. Digitized voice audio cues were given for each hoop negotiated in Navigation, for the percentage of each target Injected, for percentages completed of each sphere during Dissection, and for final completion of each subtask.

Model 2: The intermediate model was composed of the Navigation and Injection training aids from Model 1, overlaid within a virtual anatomical model of the sinus cavity. Injection and Dissection followed the protocol for a Maxillary Antrostomy: injection of the

inferior/anterior Middle Turbinate, superior root of the Middle Turbinate and the lateral nasal wall, followed by dissection of the Uncinate process, Bulla ethmoidalis and posterior Ethmoid cells. Labels were added for all anatomical structures with which the subject interacted.

Navigation through the four sets of hoops, injection of the five targets, medialization of the Middle Turbinate, dissection of the three anatomical structures, removal of two bone fragments placed in the Uncinate process and removal of three bone fragments placed in the Bulla ethmoidalis were required for a complete score. The widening of the Maxillary Ostium was not included in the procedure due to the inability to dissect enough of the lateral part of the Uncinate process for realistic viewing of the Ostium. Audio cues were given for each hoop negotiated in Navigation, for each target hit in Injection, for each bone fragment removed and for percentages completed for each anatomical structure in Dissection.

Model 3: The advanced model was composed of an anatomical model only. Subjects were expected to perform the three tasks without the training aids and to follow the protocol for a Maxillary Antrostomy. Three polyps were added superior/anterior to the Bulla ethmoidalis.

During Navigation the subject was required to perform the three passes in the same order as in Models 1 and 2: inferior pass along the floor of the nose to the Nasopharynx, followed by a more superior pass medial to the Middle Turbinate towards the upper aspect of the Nasopharynx and Sphenoid Ostium, then rolling under the Middle Turbinate to inspect the Ostial Meatal Complex, and finally the superior pass medial to the root of the Middle Turbinate towards the Sphenoethmoidal Recess.

During Injection, the subject was cued only by the amount of blanching (whitening) of the virtual tissue as to whether more vasoconstrictor was needed.

Dissection followed the protocol for a Maxillary Antrostomy. Navigation through the three passes, injection of the areas of interest, medialization of the Middle Turbinate, dissection of the three anatomical structures, dissection of the three polyps, removal of two bone fragments placed in the Uncinate process and removal of three bone fragments placed in the Bulla ethmoidalis were required for a complete score. Audio cues were given only at the end of each of the Navigation passes, for removal of bone fragments and for percentages completed in Injection and Dissection.

PROCTORING

During each trial, a proctor with knowledge of the procedure was present; proctoring for each of the groups varied according to their familiarity with the task. To assure that all records of the trial and all subject comments were noted, usually a second proctor was present. One proctor would be designated as an instructional proctor who would introduce the subject to the simulator and answer any questions during the trial. The second proctor would manage the forms, records and loading of the trials for the subject.

Non-MD Group

Each subject was initially introduced to the simulator and informed that we were evaluating the simulator as a possible trainer for residents in Otolaryngology. A summary of the reasons for this type of surgery was given along with a brief introduction to the anatomical structures and their locations in the sinus cavity to give the subjects a feel for the dimensions of the area in which they would be working.

Subjects were then introduced to the instrumented endoscope and informed of their ability to rotate the image axially by rotating the shaft of the endoscope. They were then introduced to the instrumented forceps and informed that the forceps would be simulating the virtual needle for Injection and all Dissection tools. The mechanics of the forceps (open and closed) were described as simulating the plunger for the syringe during Injection and opening and closing the jaws of the dissection tools. They were then shown how the instrumented forceps would be positioned in the opposite nostril until the beginning of Injection.

Subjects were given a brief verbal description of the three tasks for Model 1 (Navigation, Injection and Dissection) and what would be required of them during the trial. They were informed that their introduction to Model 1 would be broken up across the three subtasks. A videotape of Dr. Edmond performing the trial was then started, during which time the subject was allowed to step up to the mannequin and become familiar with the instrumentation, while the proctor described the task in more detail. Subjects were encouraged to ask any questions and "speak-aloud" during the entire procedure.

Following instructions and video for Navigation, the subject independently performed the Navigation task. The trial was then paused, proctoring and video for Injection was given, and the Injection task was performed. At that time the proctor would pull the instrumented forceps across the Columella Nasi into the nostril being

used and place a plug in the original nostril to inhibit re-crossing of the Columella Nasi by the instrumented forceps during the remainder of the trial. After Injection, the trial was then paused, instructions and video for Dissection were provided, and the Dissection task was performed.

Model 1 was the only model where this process was used and it was used only on their initial introduction to the simulator. On subsequent trials of Model 1, they were given verbal proctoring instructions only, and allowed to perform the task. Those who were introduced to Model 2 did not require breaking the trial into subtasks.

Progression to Model 2 was based on performance on Model 1. An average score of 54% was required of the subjects on Model 1 before progressing to Model 2. Since subjects in this group had no familiarity with the procedure or with paranasal sinus anatomy, Model 2 required extensive proctoring instructions on locations and anatomy to dissect during the procedure. No subjects from this group were run through Model 3 because of their inability to achieve adequate proficiency in Model 2 during the time course of the study.

Non-ENT Physician Group

Again, each subject was initially introduced to the simulator and informed that we were evaluating the simulator as a possible trainer for residents in Otolaryngology. They were encouraged to think of ways in which this type of simulator could be used in their own fields. They were then introduced to the virtual endoscope, informed of their ability to rotate the image axially by rotating the shaft of the endoscope, and of the availability of a 30 and 70 degree scope which could be swapped for the zero degree scope they would initially be given. The optics of the 30 and 70 degree scopes were explained where necessary.

Subjects were then introduced to the instrumented forceps and informed that the instrumented forceps would simulate the virtual needle for Injection and all Dissection tools. The mechanics of the instrumented forceps (open and closed) were described as simulating the plunger for the syringe during Injection and opening and closing the jaws of the dissection tools. They were then informed of how the instrumented forceps would be positioned in the nostril opposite the one used in the trial, until the beginning of Injection.

The subjects were then given a brief verbal description of the three tasks of Model 1 (Navigation, Injection and Dissection) and what would be required of them during the trial. A video of Dr. Edmond

performing the trial was then started, while the proctor continued to describe the subtasks in more detail and what was required for completion of the trial. During this time the subject was allowed to step up to the mannequin and become familiar with the instrumentation. The "blood effects scope" (scope becoming opaque within a set time interval) was shown in the video and reasons for it were described by the proctor, along with how to relieve the problem by wiping the "scope" on the foam pad located on the mannequin's forehead.

The subject was encouraged to ask any questions and "speak-aloud" during the procedure. This process was repeated before their first introduction to Model 2, with the appropriate, model-specific changes. On subsequent trials of Model 1 and Model 2 they were given verbal proctoring instructions only before performing the task.

The subject's familiarity with the procedure determined the need for further instruction by the proctor. On average this group had no familiarity with the procedure, but all had prior introduction to anatomy (during medical school) and had an understanding of the reasons for the procedure. Progression to Model 2 for this group was based on performance on Model 1; an average score of 69% was required on Model 1 before progressing to Model 2. Proctoring for Model 2 required instructions on locations and anatomy to dissect during the procedure. Training aids and video of the procedure provided adequate introduction to the task for Model 1's entirety and for Model 2's Navigation and Injection tasks. No subjects from this group were run through Model 3, due primarily to the limited availability of these MD subjects.

ENT Staff and Residents

Again, each subject was initially introduced to the simulator and informed that we were evaluating it as a possible trainer for residents in Otolaryngology. They were introduced to the virtual endoscope, informed of their ability to rotate the image axially by rotating the shaft of the endoscope, and of the availability of a 30 and 70 degree scope which could be swapped for the zero degree scope they would initially be given.

Subjects were then introduced to the instrumented forceps and informed that they would simulate the virtual needle for Injection and all Dissection tools. The mechanics of the instrumented forceps (open and closed) were described as simulating the plunger for the syringe during Injection and opening and closing the jaws of the dissection tools. They were then informed of how the instrumented forceps

would be positioned in the nostril opposite the one used in the trial until the beginning of Injection.

The subjects were then given a brief verbal description of the three tasks of Model 1 and what would be required of them during the trial. The video of Dr. Edmond performing the trial was then started, while the proctor continued to describe the subtasks in more detail and what was required for completion of the trial. During this time the subject was allowed to step up to the mannequin and become familiar with the instrumentation.

The "blood effects scope" was shown in the video and reasons for it were described by the proctor, along with how to relieve the problem by wiping the scope on the foam pad located on the mannequin's forehead. Subjects were encouraged to ask any questions and "speak-aloud" during the procedure. On subsequent trials they were given verbal proctoring instructions only (with the appropriate, model-specific changes), and allowed to perform the task.

The subject's familiarity with the procedure determined the need for further instruction by the proctor. Typically, the staff Otolaryngologists needed no further instruction on the procedure for the remainder of the session, except for the need to be shown the active areas for dissection of the Uncinate process, Bulla ethmoidalis and Posterior Ethmoid cells within the virtual model. In general, the staff ENT subjects ran through all three model levels.

Similarly, the ENT residents needed no further instruction for Model 1. For Model 2, however, more detailed instructions were sometimes needed during Dissection on what anatomy to dissect and where the active dissection areas were located within the anatomy. In Model 3 more detailed instructions were also needed during Navigation on the order of passes to perform and where the active dissection areas were located within the anatomy. All residents were run through Model 1-right, Model 2-right, Model 3-right, Model 3-left and Model 2-left.

USER PERFORMANCE CRITERIA

The scoring algorithm for Version 1.2 takes into consideration three major performance measures for Endoscopic Sinus Surgery: Accuracy, Completeness and Time. The overall trial score is calculated as:

(navigation score + injection score + dissection score - hazard score)/3

Subtask scores are calculated as:

*subtask score = accuracy*optimal-time/completed-time.*

The scoring algorithm used for the evaluation of this phase of the project took into account the most important skills needed to perform sinus surgery, as well as the level of difficulty of the model. In an attempt to normalize across model conditions, the following optimal times (in seconds) were used in this scoring equation:

Navigation: advanced=97, intermediate=97, novice=97
Injection: advanced=90, intermediate=165, novice=90
Dissection: advanced=815, intermediate=635, novice=165

These times were derived from the approximate performance times for the project's lead otolaryngology advisers. If a subject performed the subtask in less than the optimal time, then their score was equal to their accuracy on that subtask.

Navigation accuracy was based on the percentage of hoops negotiated (with the hoops in the advanced model rendered invisibly). Injection accuracy scores were based on the ratio of the percentage of each target injected to a criterion percentage for that object: 100% for the graphical targets, 25% for the middle turbinate, and 10% for the nasal wall. Similarly, dissection accuracy was based on a criterion dissection percentage for each dissectable object (markers, polyps, uncinate, ethmoid cells, ethmoid bulla, maxillary ostium, and bone fragments).

The score for hazards is the sum of the collision scores for each of the hazards in the patient model, which range from 0 to 100, based on the proximity of the intrument. Given the above algorithm, hitting a hazard will have a huge impact on the overall score; hitting just one target with a collision score of 100 limits the overall score to no more than 66%. This relatively large penalty is considered appropriate here because of the potentially devastating consequences of hitting a hazard (e.g., the skull base or optic nerve) in an actual surgery.

Results

Non-MD Evaluation

Although not the target domain audience for this simulator, the non-MD group gave us an opportunity to look at the baseline difficulty of the simulator for domain-naive users and to further work out requirements for instructional presentation, proctoring, and pacing of the "curriculum" (after initial "shake down" and protocol development by the project staff). Since the novice model did not require interaction with the complex paranasal sinus anatomy or detailed knowledge of the surgical procedure, it was felt that this group should be able to complete those trials.

The inherent difficulty of the task for inexperienced users is evidenced by subject attrition and by their initial trial scores. Indeed, complete trial data are available for only nine of the 12 subjects in this group. Of the three lost to attrition, one terminated the session due to disorientation and discomfort with visual interface, and two were unable to complete even the navigation task during the time period available for testing. The primary difficulty observed with these three subjects was an inability to adapt to the psychomotor demands of the interface; that is, they were not able to acquire the basic skill necessary to control the positioning of the virtual endoscope within the model.

Of the nine remaining subjects in this group, five had had considerable experience with videogames and three with commercial flight simulators. Thus the trials scores for this group may represent the high end of non-MD users and should not be taken as a sample from the general non-MD population.

Novice model performance

Still, despite attrition of the low-end performers and the unusual prior experience of the remaining subjects, the non-MD group performed significantly worse than the ENT group on their initial novice level trials (as can be seen graphically in a later section).

Unpaired Student's t comparisons with the ENT group revealed significantly

- lower overall trial scores (nonMD mean = 45.9, ENT mean = 65.1, t = 2.87, p = .0098),
- longer overall trial times (nonMD mean = 1012 sec, ENT mean = 692 sec, t = 2.14, p = .0458),
- lower navigation scores (nonMD mean = 47.1, ENT mean = 69.4, t = 2.39, p = .0276),

longer navigation times (nonMD mean = 239 sec, ENT mean = 146 sec, t = 2.52, p = .0209), and

lower injection scores (nonMD mean = 55.3, ENT mean = 81.5, t = 2.46, p = .0235).

Injection times, while not significantly longer, did approach significance (nonMD mean = 257 sec, ENT mean = 117 sec, t = 1.96, p = .0649). Surprisingly, while dissection scores and dissection times were slightly worse for the non-MD group, these differences were not significant. This relatively poor performance by the ENT group on their initial dissection trial is discussed further below.

Practice effect

Six of the 12 subjects in this group had complete trial scores for at least two novice trials. Student's t comparisons of their first and second novice trials showed that their second trial resulted in significantly

higher overall trial scores (trial 1 mean = 46.7, trial 2 mean = 69.5, t = 2.33, p = .0423),

higher injection scores (trial 1 mean = 53.8, trial 2 mean = 86.7, t = 2.22, p = .0500),

shorter dissection times (trial 1 mean = 492 sec, trial 2 mean = 328 sec, t = 2.39, p = .0382), and

higher dissection scores (trial 1 mean = 35.8, trial 2 mean = 51.8, t = 2.55, p = .0287).

Total trial times, navigation times and scores, and injection times all changed in the expected direction (that is, they improved across trials), but none of these improvements was statistically significant. It appears that, for this non-MD group, the primary challenge may have been mastering the endoscope positioning.

Intermediate model performance

First-time intermediate level scores were acquired for only 3 of the subjects in the non-MD group. Despite this small N, unpaired Student's t comparisons with the ENT group revealed significantly

lower overall trial scores (nonMD mean = 53.7, ENT mean = 76.7, t = 3.53, p = .0041),

longer overall injection times (nonMD mean = 359 sec, ENT mean = 132 sec, t = 6.91, p = .0001), and

lower injection scores (nonMD mean = 47.1, ENT mean = 69.4, t = 2.39, p = .0276).

Overall trial times, while not significantly longer, did approach significance (nonMD mean = 1532 sec, ENT mean = 1215 sec, t = 1.79, p = .098), as did navigation times (nonMD mean = 199 sec, ENT mean = 138 sec, t = 2.08, p = .0592) and navigation scores (nonMD mean = 48.0, ENT mean = 73.7, t = 1.99, p = .0698). Again, no difference was found between the two groups in initial dissection times and dissection scores on the intermediate model.

The experiential advantage of the ENTs is perhaps most telling in their superior injection task performance. As will be seen below in examining trends within the ENT group, ability to perform efficient injections may by the hallmark of the accomplished surgeon, much more so than the ability to perform efficient dissection (at least as indicated by simulator performance).

Non-ENT Physician Evaluation

The primary objectives of the non-ENT physician evaluations were to provide a medically-trained comparison and "shakedown" group and to explore the perspectives of other medical specialties regarding additional potential applications for this sort of simulation training. In particular, we were interested in applications which might make use of the integrated approach and specific components incorporated into the Madigan ESS simulator.

Novice model performance

Novice performance measures for the non-ENT MD group were very similar to the those for the experienced ENT group (as summarized graphically below in the ENT analysis section). While there were no statistically significant differences between these two groups, it should be noted that the mean differences for all of the performance measures were in the direction we had expected (that is, the means performance measures were consistently better for the ENT group).

This finding may be due to the considerably broad prior medical and interface experience for all of the subjects in this group. Most of the subjects in this group have been involved over the years in developing or testing novel interface devices for medical tasks, and several had previous experience test-driving similar medical simulators.

Perhaps the greatest mitigating factor, however, is the extensive experience that two of these four subjects had acquired as videoendoscopic surgeons. This experience provided them with a background comparable to the ENT group in basic endoscopic psychomotor skills, as well as in general procedural knowledge and confidence.

Although there were no statistically significant differences in any performance measures between the non-ENT MDs and the non-MD subjects, it was apparent that they were much more comfortable with the experimental task requirements. This finding (and the nature of the questions asked during the session) suggests the value of a general medical perspective in assuring confidence with this sort of procedural simulator.

Intermediate model performance

Three of the non-ENT MD subjects also tried the intermediate model. Comparison of the non-ENT MDs with the ENTs on their first trial on the intermediate model reveals an interesting finding. While there was no significant difference between the groups in overall score and trial time, the non-ENTs did notably better than the ENTs on the dissection subtask. Dissection scores were significantly better for the non-ENTs (nonENT mean = 84.0, ENT mean = 63.1, $t = 2.439$, $p = .0312$), which appears to be due primarily to their faster times on that task (nonENT mean = 662 sec, ENT mean = 880 sec, $t = 1.99$, $p = .0692$), a non-significant but clearly suggestive finding.

Looking at individual scores, we note that this effect is due to the relatively fast dissection times for the two videoendoscopic surgeons. While we would expect that the ENT group would be superior to all other groups on this task, the extensive background in similar procedures appears to have prepared the general surgeons well for this task.

Related Medical and Surgical Applications

Response by the non-ENT physician group was uniformly positive. Each of them suggested additional applications, both within and outside of their particular specialties, that would be amenable to a similar simulation approach. These suggested applications included:

Neurosurgery: approaching the pituitary gland through the sphenoid sinus

Anesthesiology: intubation, bronchioscopy

Videoendoscopic surgery: cholecystectomy, bladder, joints, throat, chest procedures

Cardiology: cannulization, thoracoscopic surgery

Gastroenterology: endoscopic procedures

Radiology: flouroscopy, intrusive ultrasound procedures (e.g., transesophageal echocardiography)

In general, procedures that use endoscopes or probes to explore internal structure and to perform manual procedures should be amenable to this simulation approach. Several of the physicians in this group indicated that they would like to see work proceed in that direction and that there would be support for this development from the leaders within their specialties.

Staff and Resident ENT Evaluations

Simulator performance measures were acquired on 53 separate trials for the 12 ENT subjects between 5/13/97 and 8/12/97 (23 trials run at HITL, and 30 trials run at MAMC). The number of trials acquired for each subject ranged from 2 to 12, with all subjects being tested (minimally) on both the novice and the intermediate model.

Initial Novice Model Performance

Figure 4 shows the distribution of trial scores on the initial novice trials for the 12 ENT subjects.

To examine the degree to which prior OR experience might contribute to these initial score differences, we calculated the Pearson product-moment correlation coefficients ("r") across the primary performance measures (trial time and trial score) and the primary experience measures (age, years of ESS training, and approximate number of actual ESS procedures performed. A matrix of these correlations appears in Table 1 (critical r = .6021 for probability $p < .05$, degrees of freedom df = 9).

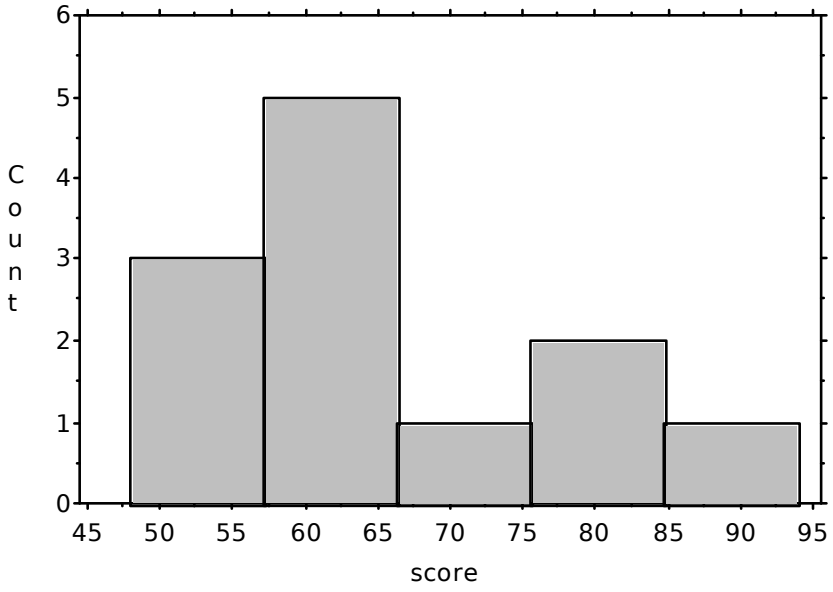


Figure 4. Frequency distribution of overall scores on initial trial on novice model by subjects in the ENT group.

As can be seen, the best predictor of novice trial time from among the surgical experience measures is the number of actual ESS cases performed (r = -.652), with trial time decreasing significantly with surgical experience. For trial score, the findings are less conclusive, although the best experiential predictor is again the number of ESS cases performed (r = .526). While age is naturally correlated with ESS experience and training, age alone does not appear to be a significant predictor of simulator performance for these ENT subjects.

	score	ESS perf	years_tr	age	trialtime
score	1				
ESS performed	.526	1			
years_training	.38	.923	1		
age	.236	.662	.713	1	
trialtime	-.966	-.652	-.514	-.33	1

Note: 1 case deleted with missing values.

Table 1. Correlation matrix showing relationships among ESS experience measures and overall performance measures on the novice model for the 12 ENT subjects.

In general we see that trial score on the initial novice simulator trial is indeed positively correlated with prior ESS experience. This finding suggests that the simulator provides a valid reflection of the skills acquired in ESS procedures.

Subtask Performance on Initial Novice Trial

Not surprisingly, overall trial scores are, with the sole exception of injection time, significantly correlated with subtask scores and times, as shown in Table 2 (critical r = .5760 for $p < .05$, df 10). Overall trial times are similarly related to subtask performance measures, although by far the strongest predictor of overall trial time is performance on the dissection task (r = .897 for dissection time and -.918 for dissection score).

	score	trialtime	navtime	navscore	injtime	injscore	disstime	disscore
score	1							
trialtime	-.965	1						
navtime	-.587	.469	1					
navscore	.68	-.573	-.97	1				
injtime	-.531	.508	-.109	.092	1			
injscore	.684	-.667	.016	.033	-.957	1		
disstime	-.784	.897	.189	-.342	.3	-.476	1	
disscore	.856	-.918	-.275	.413	-.352	.524	-.964	1
ESS perf...	.504	-.614	-.245	.227	-.425	.428	-.526	.471

Table 2. Correlation matrix showing relationships among ESS procedures previously performed and subtask performance measures on the novice model for the 12 ENT subjects.

As before, prior ESS experience is predictive of overall trial time (r = -.614), and, although its relationship to each of the subtask measures is consistent with simulator validity, none of these correlations are statistically significant.

Novice Model Performance Across ENT Groups

Figure 5 shows the breakdown of first-time novice (abstract) model scores for this group. The linear relationship between overall novice trial score and year of residency is apparent from this mapping. It is interesting to note that the R5 residents appear to have higher scores on the novice model than do the experienced ENT staff subjects. This could be due a number of factors, including the relatively high number of procedures being performed routinely by the R5s.

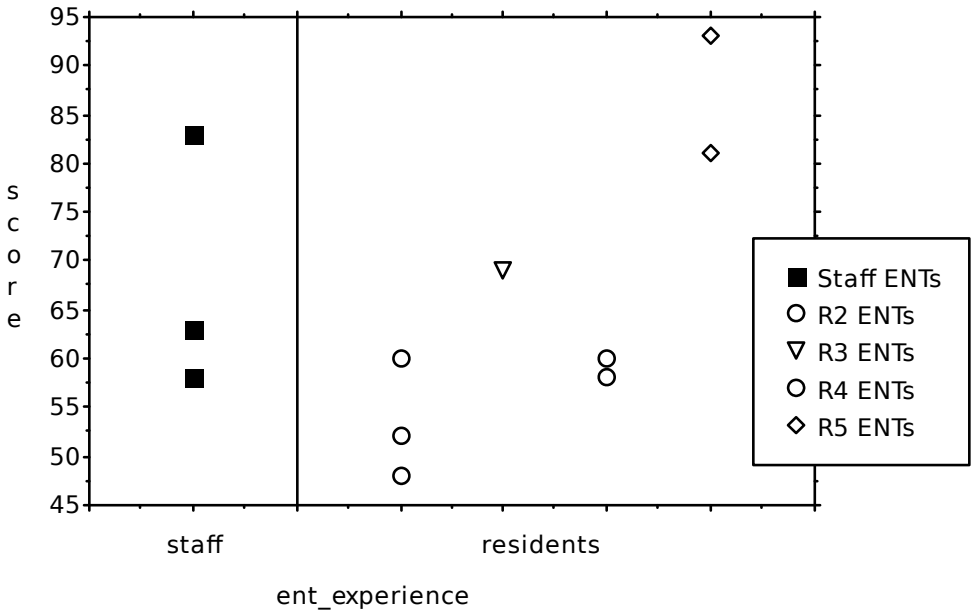


Figure 5. Overall scores for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

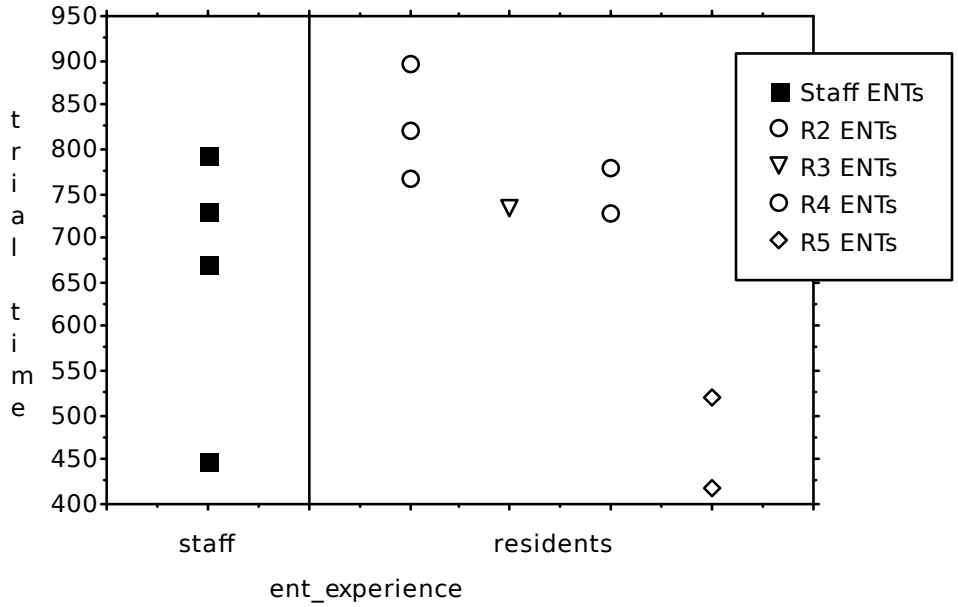


Figure 6. Overall trial times for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

Similarly, it can be seen in Figure 6 that overall trial times for the initial novice trials fall off as expected with year of residency. Both of these findings provide strong evidence for the validity of the simulator for the ESS task, even when the model is abstract.

Subtask Times By Year of Residency

A breakdown of trial times by subtask is shown for the range of ENT subjects in Figures 7-9.

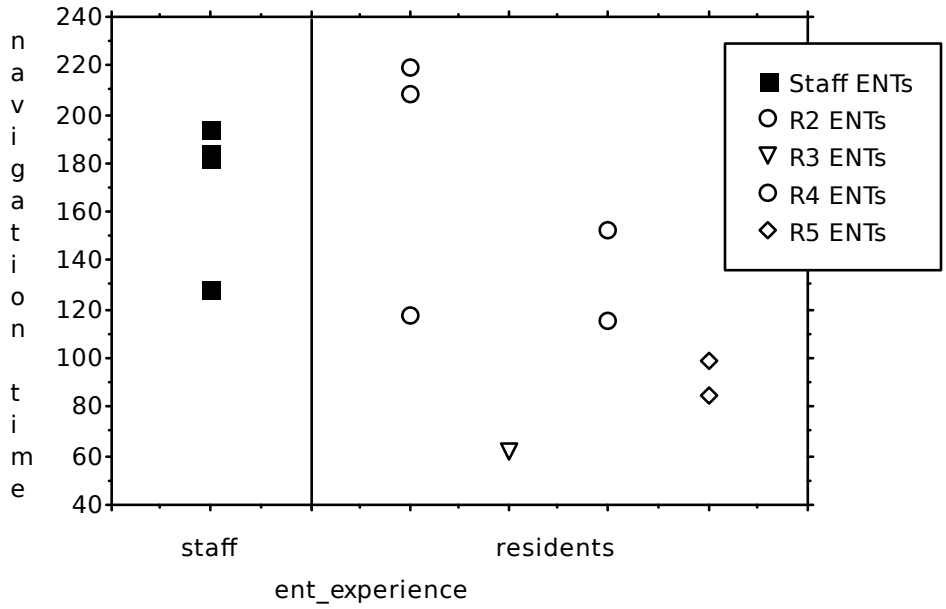


Figure 7. Navigation times for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

While resident subtask times generally decrease with year of residency, dissection times indicate a nonlinear trend, with residents in the middle years of training taking longer to dissect than the R5 residents (who have had considerably more experience). This effect may perhaps be due to the degree of caution observed by the R3s and R4s as they become aware of the procedural risks during dissection but do not yet have their instrument control skills well-honed.

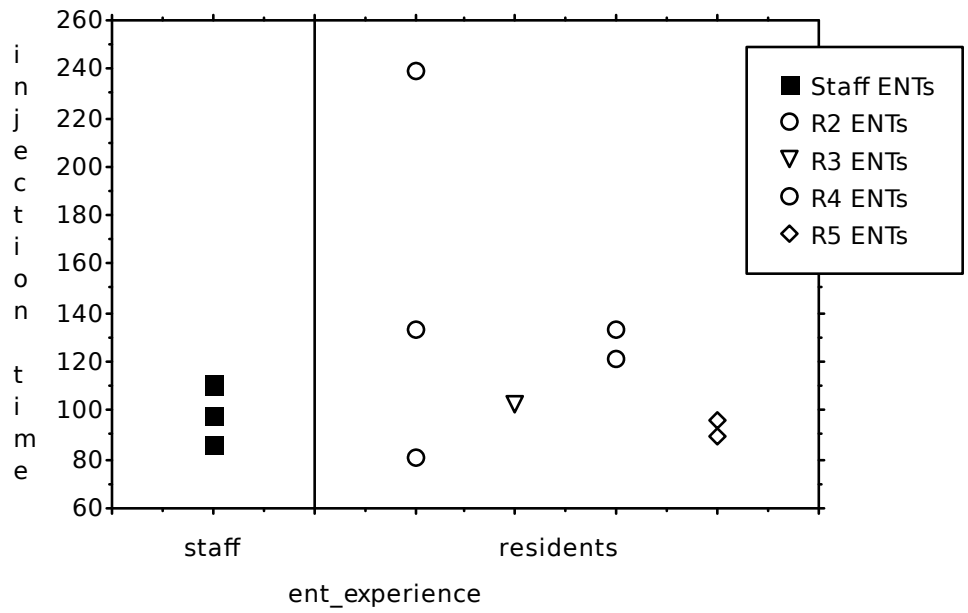


Figure 8. Injection times for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

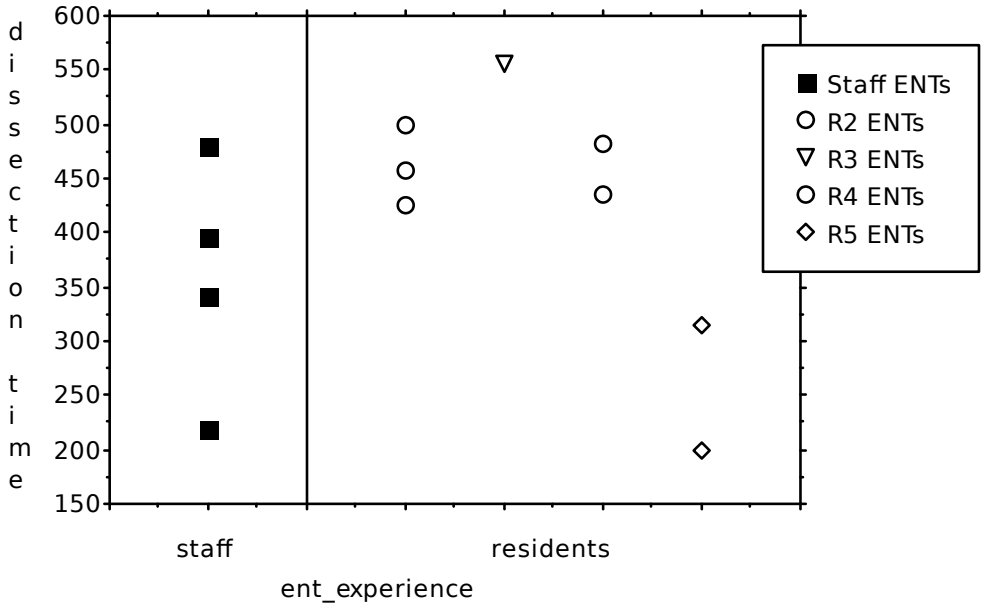


Figure 9. Dissection times for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

These subtask times for the staff ENTs suggest that injection skill is well-learned by this time, while the finding of relatively more variability in navigation and dissection times may reflect differences in style, with some experienced surgeons remaining (appropriately) more cautious and perhaps more attentive to the potential hazards of making a mistake.

Subtask Scores By Year of Residency

A similar breakdown for subtask scores (shown in Figures 10-12) mimics the trends for subtask time.

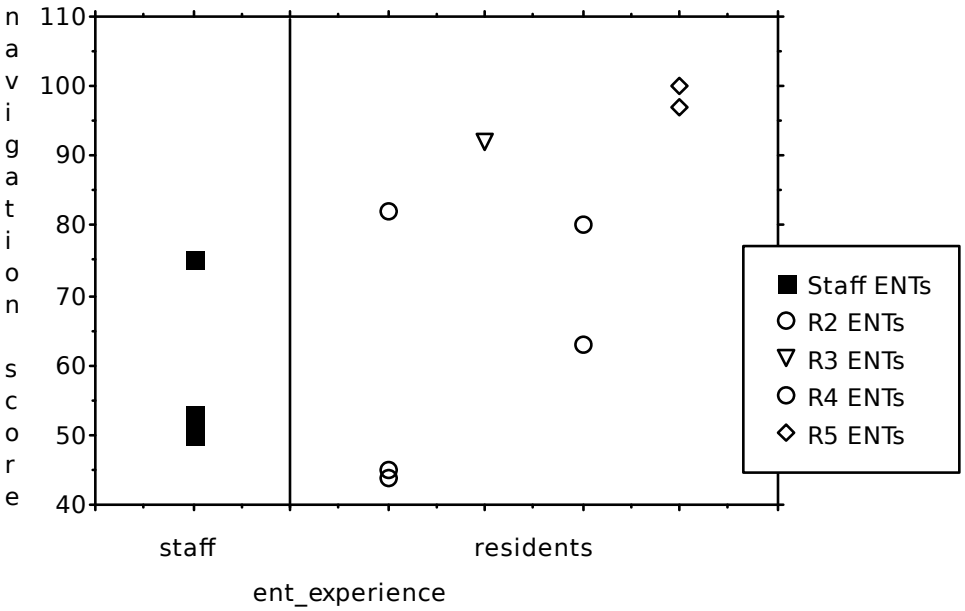


Figure 10. Navigation scores for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

Scores on each subtask generally increase with year of residency, with the exception of R3s and R4s achieving slightly lower scores for dissection than their junior and senior colleagues. Injection scores for R5s are comparable with staff ENTs, while R5 navigation and dissection scores appear to be superior.

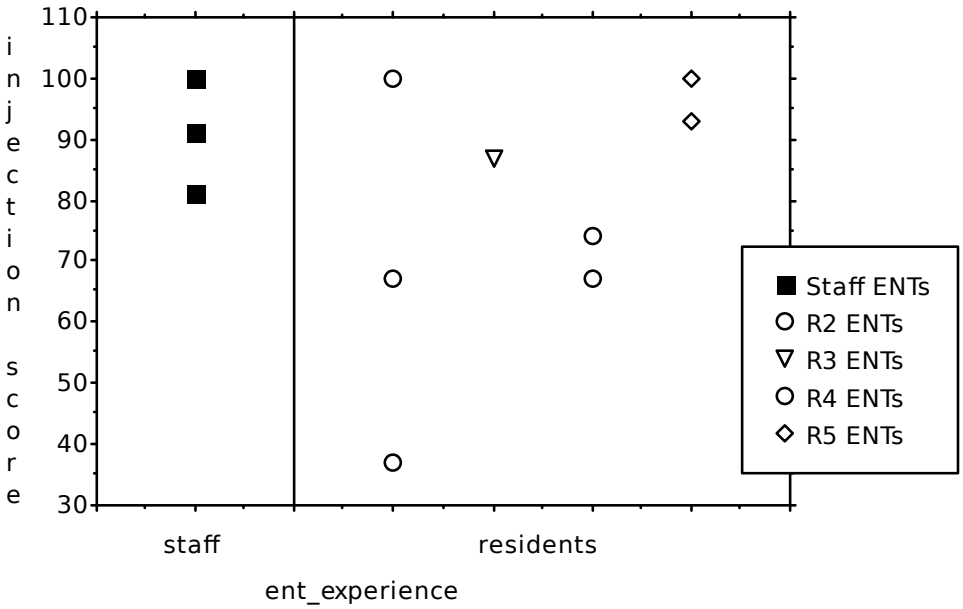


Figure 11. Injection scores for the first trial on the novice model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

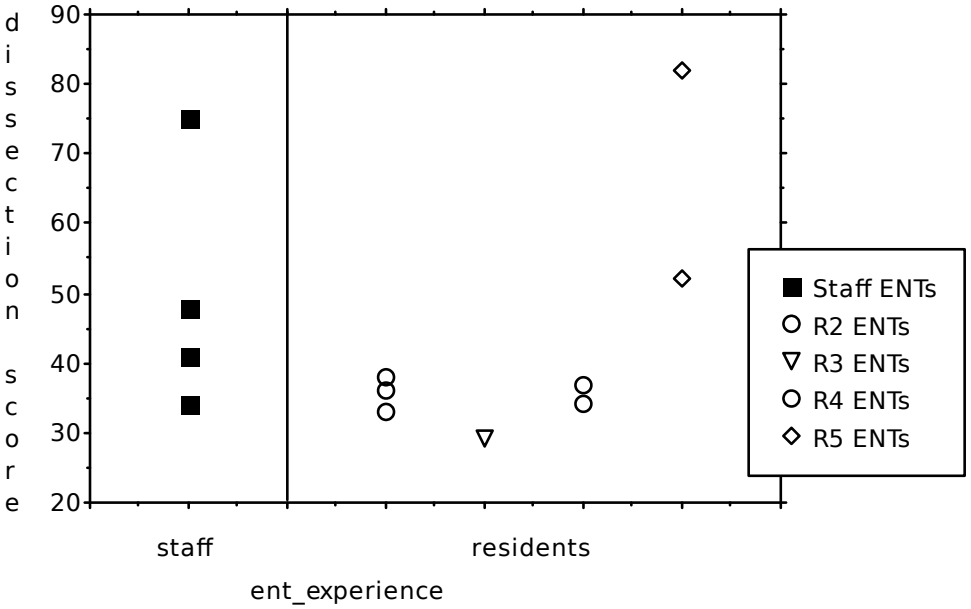


Figure 12. Dissection scores for the first trial on the novice model for ENT staff (left) and ENT residents (right).

Novice Model Performance Across Subject Groups

Figure 13 shows the distribution on overall scores on the first novice trial for all three subject groups, while Figure 14 shows the same thing for overall trial times.

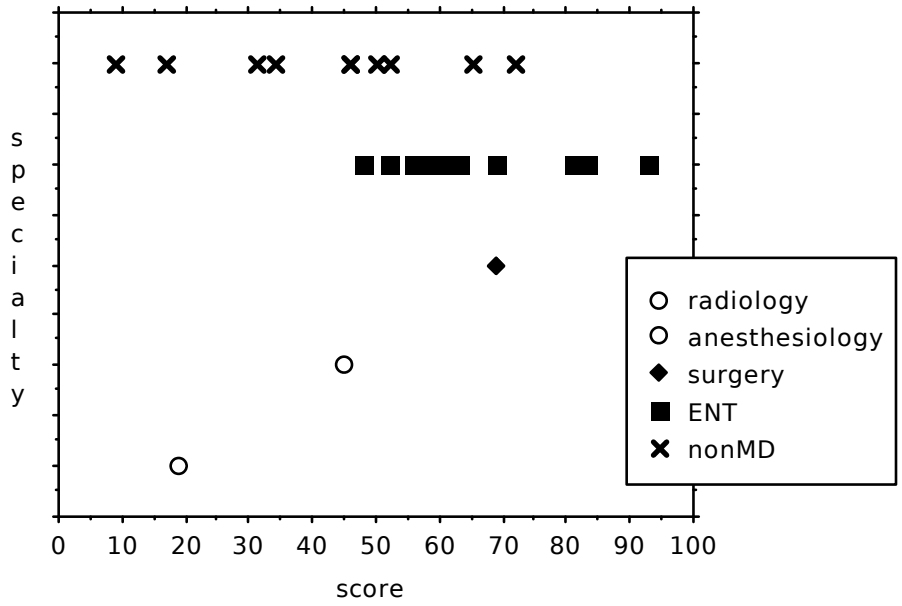


Figure 13. Overall scores for the first trial on the novice model for the nonMD subject group, the nonENT MD subjects (representing three specialties), and the ENT group.

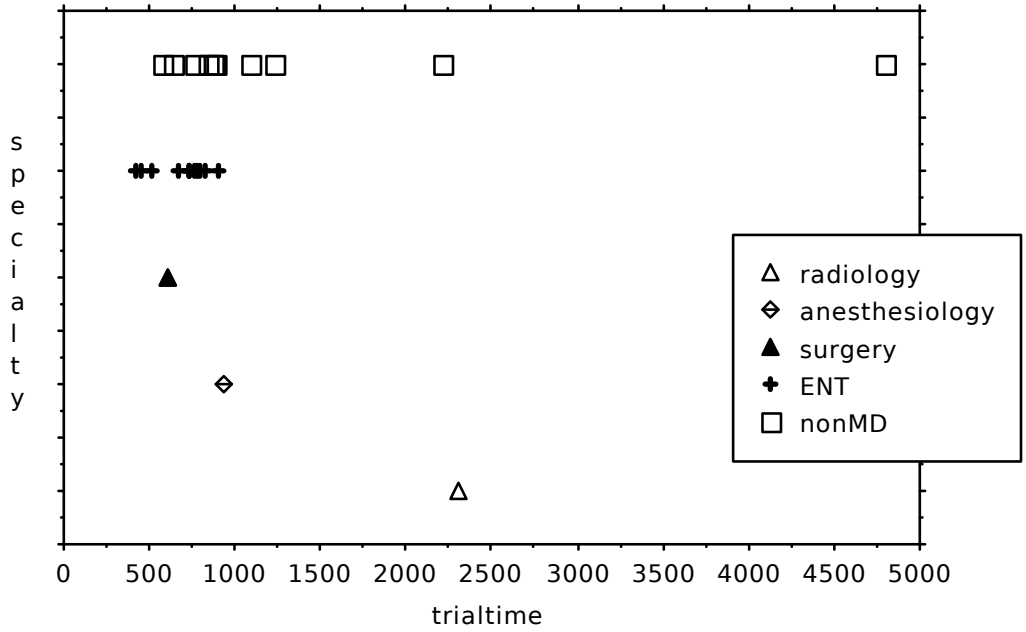


Figure 14. Overall trial times for the first trial on the novice model for the nonMD subject group, the

nonENT MD subjects (representing three specialties), and the ENT group.

The nonMD group is clearly less skilled than the ENT group (as discussed above), while the nonENT physicians demonstrated more variable performance. Note the two videoendoscopic surgeons had identical overall scores on their novice trials (and overall times that were essentially the saem, as well), and that their performance was comparable to the average ENT subject. The anesthesiologist also did surprisingly well, due perhaps to that subject's extensive clinical experience performing image-guided and vital sign-guided procedures which require new psychomotor skills.

Initial Intermediate Model Performance

Figure 15 shows the distribution of trial scores on the initial intermediate trials for the 12 ENT subjects (Note that the first intermediate trial for one of these subject was terminated after the navigation phase and is consequently not included here).

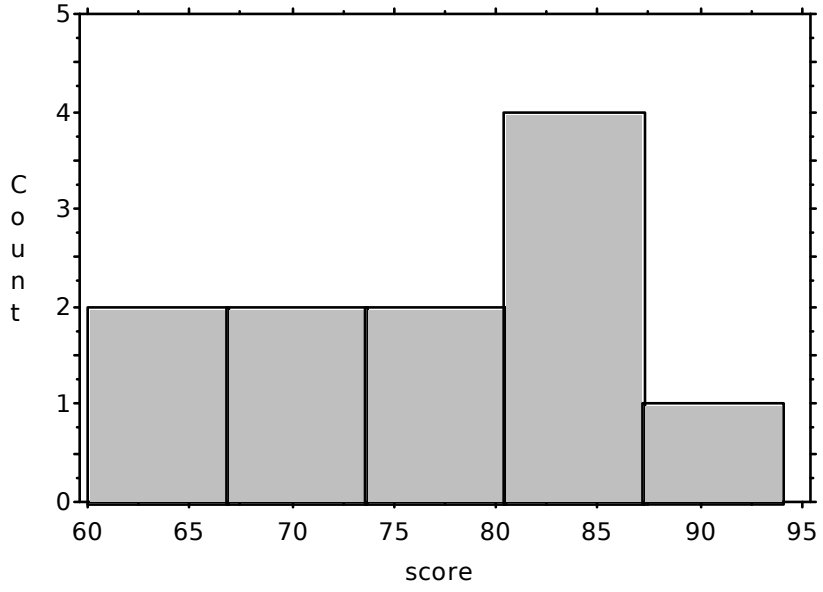


Figure 15. Frequency distribution of overall scores on initial trial on the intermediate model by subjects in the ENT group.

Although we might expect a positive relationship between ESS experience and novice model performance, we would expect that OR experience would be especially predictive of performance when the

sinus anatomy serves as the model. A matrix of these correlations appears in Table 3 (critical r = .6319 for p < .05, df 8).

In contrast to the initial novice scores, these results indicate that the best predictor of initial intermediate trial time (from among the surgical experience measures) is subject age (r = -.475), but this correlation is not statistically significant. For trial score, the findings are also inconclusive, with the best experiential predictors being subject age (.581) and the number of ESS cases performed (r = .555).

	score	ESS perf	years_tr	age	trialtime
score	1				
ESS performed	.555	1			
years_training	.356	.931	1		
age	.581	.667	.68	1	
trialtime	-.795	-.144	-.014	-.475	1

Note: 1 case deleted with missing values.

Table 3. Correlation matrix showing relationships among ESS experience measures and overall performance measures on the intermediate model for the 12 ENT subjects.

In general we see that initial performance on the intermediate model is indeed positively correlated with prior ESS experience, but the relationship is not as strong as for initial novice model performance. Table 4 attempts to shed light on this unexpected finding by looking more closely at experiential correlates of performance on procedural subtasks for the initial intermediate trial (critical r = .6319 for p < .05, df 8).

	ESS perf...	years_t...	age	navtime	navscore	injtime	injscore	disstime
ESS perf...	1							
years_tr...	.931	1						
age	.667	.68	1					
navtime	-.569	-.314	-.386	1				
navscore	.509	.288	.36	-.971	1			
injtime	-.869	-.847	-.624	.435	-.413	1		
injscore	.859	.798	.525	-.476	.446	-.967	1	
disstime	.199	.229	-.428	.191	-.224	-.275	.362	1
disscore	-.059	-.144	.462	-.354	.381	.186	-.226	-.967

Note: 1 case deleted with missing values.

Table 4. Correlation matrix showing relationships among ESS experience measures and subtask performance measures on the intermediate model for the 12 ENT subjects.

It appears from Table 4 that prior ESS experience is indeed a strong predictor of performance on the initial intermediate injection task (r = -.859 for injection time and .859 for injection score) and a slightly weaker predictor of performance on the initial intermediate navigation task (r = -.569 for navigation time and .509 for navigation score).

Performance on the initial intermediate *dissection* task, however, is actually the reverse of our predicted effect: dissection times are slightly longer for more experienced surgeons on their first encounter with the intermediate model, and dissection scores are slightly lower.

These findings are deserving of further study. One possible explanation might be that there is significant negative transfer to the simulated dissection task due to more extensive experience with the real instruments and real anatomical dissection. Comments from experienced ESS surgeons regarding the difficulty of the dissection task support this notion.

Intermediate Model Performance Across ENT Groups

As shown in Figures 16 and 17 scores on the initial intermediate trial generally improved with year of residency. Overall trial times, on the other hand, appear to have remained relatively constant, but perhaps less variable, over residency training year.

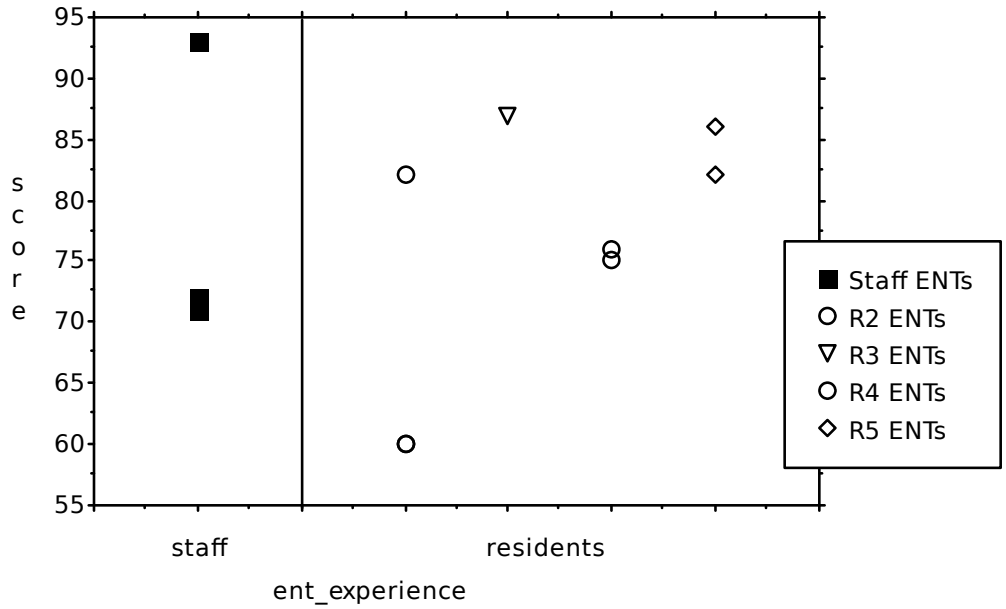


Figure 16. Overall scores for the first trial on the intermediate model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

While intermediate trial times might be expected to be lower than the those for the residents, their real world experience with the procedure may actually make them more cautious, as noted above. The experienced sinus surgeon knows, for example, that if he traumatizes the septum or middle turbinate during the navigation phase, it will create bleeding that will effect the rest of the surgical procedure.

Figure 17. Overall times for the first trial on the intermediate model for ENT staff (left) and ENT residents (right).

Intermediate Model Performance Across Subject Groups

Overall times and scores for initial intermediate model trials are broken down by specialty in Figures 18 and 19. As expected, the ENT group performed better on the anatomical model than did the non-MD group.

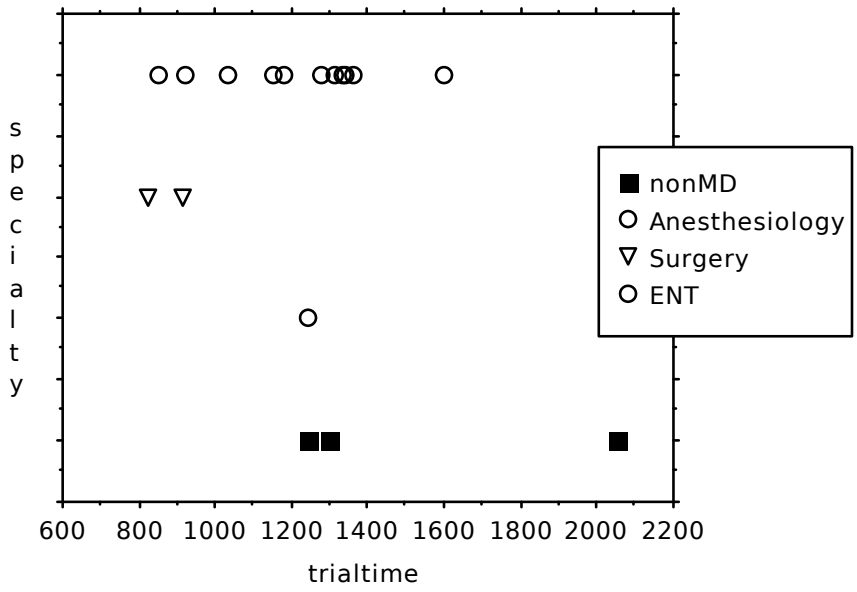


Figure 18. Overall trial times for the first trial on the intermediate model for the nonMD group, the nonENT MD subjects (representing three specialties), and the ENT group.

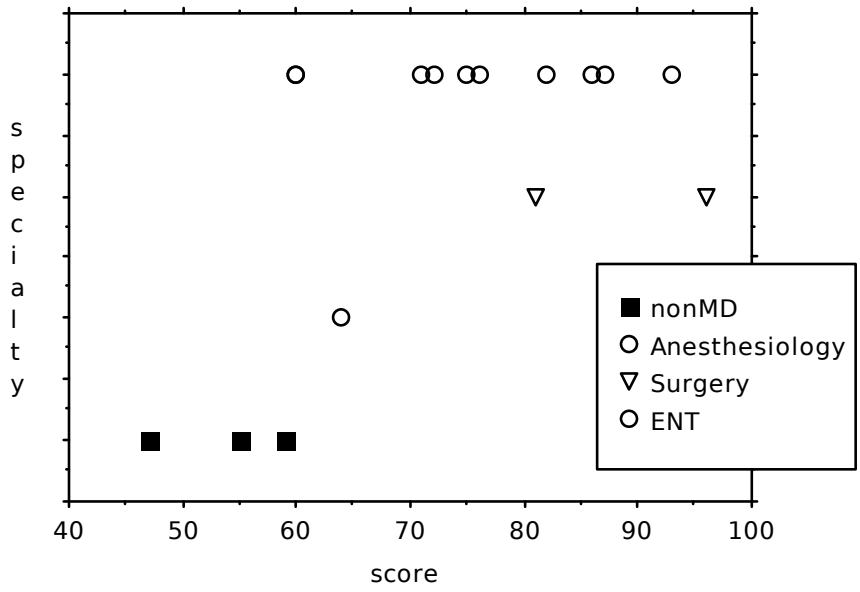


Figure 19. Overall scores for the first trial on the intermediate model for the nonMD group, the nonENT MD subjects (representing three specialties), and the ENT group.

While the trial time difference between the ENT and non-MD groups was not statistically significant, the trial score difference was (as summarized in the table below). This finding again suggests that ESS experience is predictive of simulator performance, thus providing evidence for the validity of the system for ESS training.

Unpaired t-Test X $_1$: specialty Y $_1$: score

DF:	Unpaired t Value:	Prob. (2-tail):
12	-3.533	.0041

Group:	Count:	Mean:	Std. Dev.:	Std. Error:
nonMD	3	53.667	6.11	3.528
ENT	11	76.727	10.631	3.205

Figures 18 and 19 also show that the nonENT physicians in this sample had scores that were comparable to the ENT group (Note that the radiologist from the novice trial comparisons above was not tested on the intermediate or advanced models). This is not surprising, given that these subjects were videoendoscopic surgeons who typically perform more endoscopic procedures on a regular basis than do ENT surgeons.

General videoendoscopic procedures may also be more challenging from a psychomotor perspective. In addition to navigation, injection and dissection, their procedures may require retracting, cautery, cutting, suturing and true dissection, with the reflection of tissues away from the area of interest. Furthermore, they perform these task in a more bi-dexterous fashion; the ENT surgeon uses both hands, but usually one at a time. Finally, the videoendoscopic surgeon's psychomotor abilities may be challenged more during their procedures, because the endoscope and instruments are usually not coaxial.

Initial Advanced Model Performance

Advanced Model Performance Across ENT Groups

Figures 20 and 21 show the overall scores and trial times for the first advanced trial by subjects in the ENT group. Note that two of the staff subjects did not attempt the advanced model.

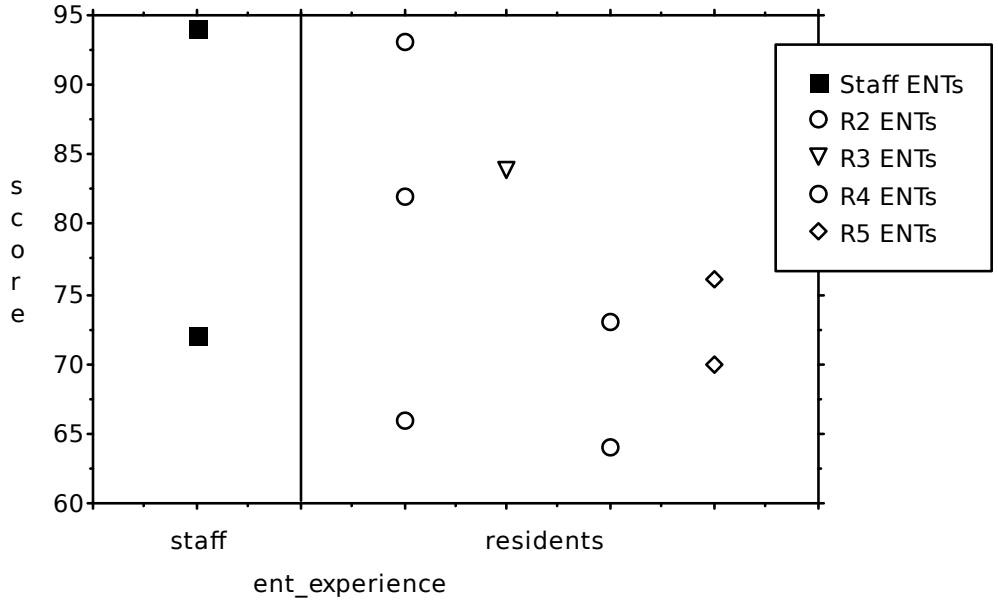


Figure 20. Overall trial scores for the first trial on the advanced model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

Surprisingly, initial performance on the advanced model revealed a tendency for the more senior residents to perform at about the same speed as the staff ENT subjects but with slightly lower overall scores. This may reflect the ability of the more experienced surgeons to

successfully avoid hazards and perform maneuvers with a higher degree of accuracy.

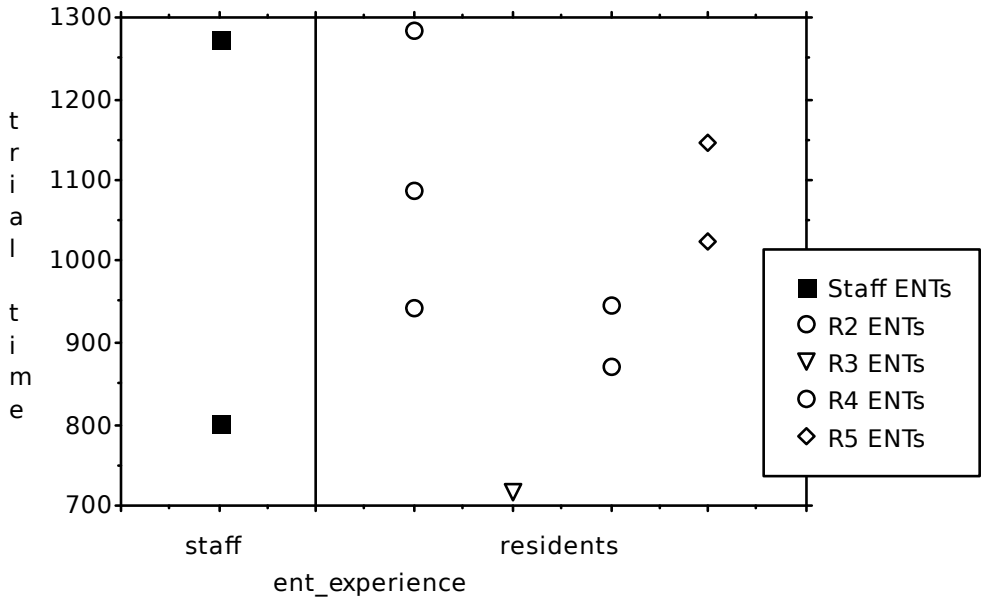


Figure 21. Overall trial times for the first trial on the advanced model for ENT staff (left) and ENT residents (right).

While not a strictly linear relationship across years, the R3 and R4s had shorter trial times on their first advanced trial, but the R4s also had lower scores on those trials, indicating that their accuracy or hazard avoidance was lower.

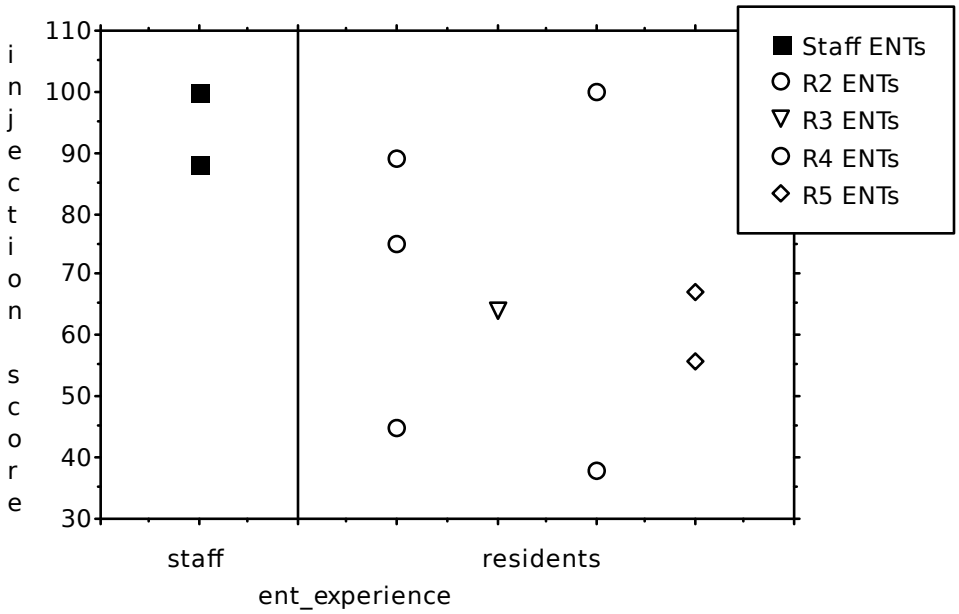


Figure 22. Injection scores for the first trial on the advanced model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

While overall time and score for staff ENTs did not appear to differ systematically from the resident times and scores on initial advanced model trials, their performance on the *injection* subtask appeared to be superior to the resident group (as shown in Figures 22 and 23). Similar differential performance was not seen for the navigation and dissection subtasks.

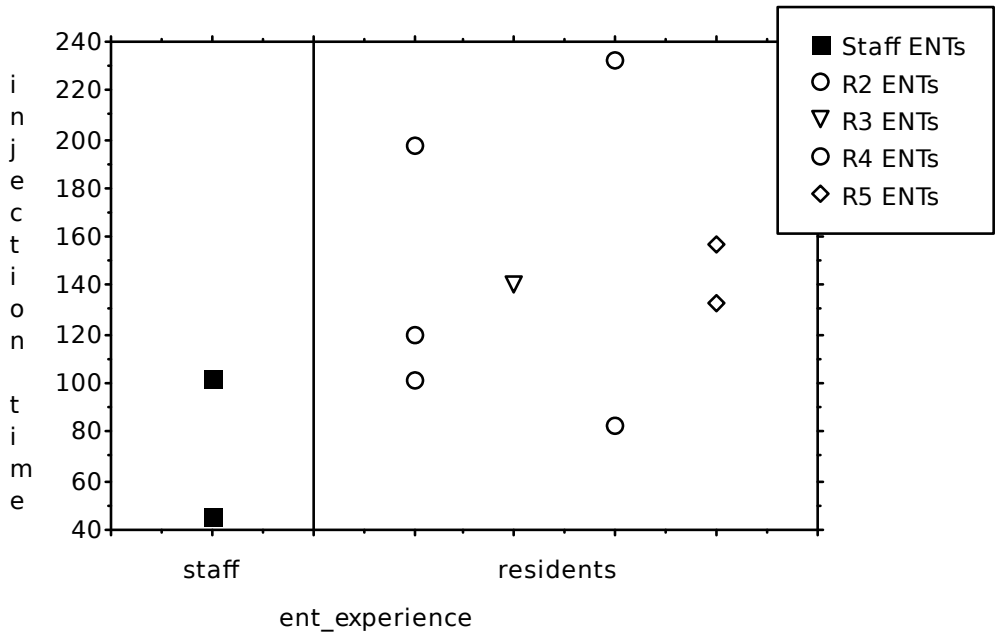


Figure 23. Injection times for the first trial on the advanced model for ENT staff (on the left) and ENT residents (on the right, plotted by year of residency).

Although navigation and dissection may vary with individual care and priorities, the data suggest that injection efficiency may be a hallmark of the experienced sinus surgeon. This finding should be followed up and verified in future evaluation studies.

Asymptotic Performance

While extensive learning curve data are not available, repeated measures on the steady-state system (version 1.2) are available for our two primary evaluation proctors (CE, an ESS surgeon, and CA, an engineering student).

Given that these two subjects had each tested the simulator in various degrees of development approximately 30-40 times over a 6-8 month period, their later scores may provide an estimate of asymptotic performance on the simulator.

ENT steady-state performance

Simulator performance scores were analyzed for the last eleven completed steady-state trials by the experienced ENT proctor. As can be seen in Figure 24, steady state trial times for the novice model (mean = 259 seconds, s.d. = 15.9) were lower than those for the intermediate and advanced models (combined mean = 512 seconds, s.d. = 57.1). The similarity between trials times on the intermediate and advanced models may suggest that the two tasks are comparable for the highly experienced subject and, further, that the intermediate model training aids are no longer necessary for this subject.

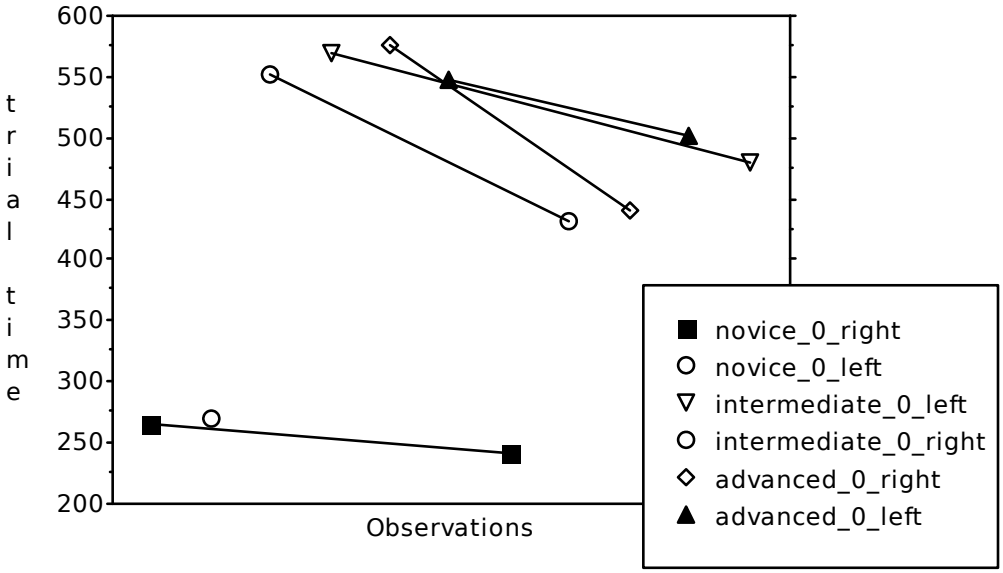


Figure 24. Asymptotic overall trial times on each model for the experienced ENT proctor. Times are plotted in order of trial presentation.

It also appears from this figure that trial times for each model are consistently lower in the last trial than in the previous trial, suggesting that these trials times may, in fact, *not* represent steady state performance. An examination of overall scores for these trials (see Figure 25) reveals, however, that for at least some of these conditions the decrease in overall trial time for these later trials came at the cost of trial score. It should be noted that, since this subject

achieved the minimum required time on all subtask trials, these trial scores precisely reflect a decrement in performance *accuracy*.

Given the better trial score for the first samples for each model condition, we may tentatively treat the trial times for those first samples as more representative of optimal performance.

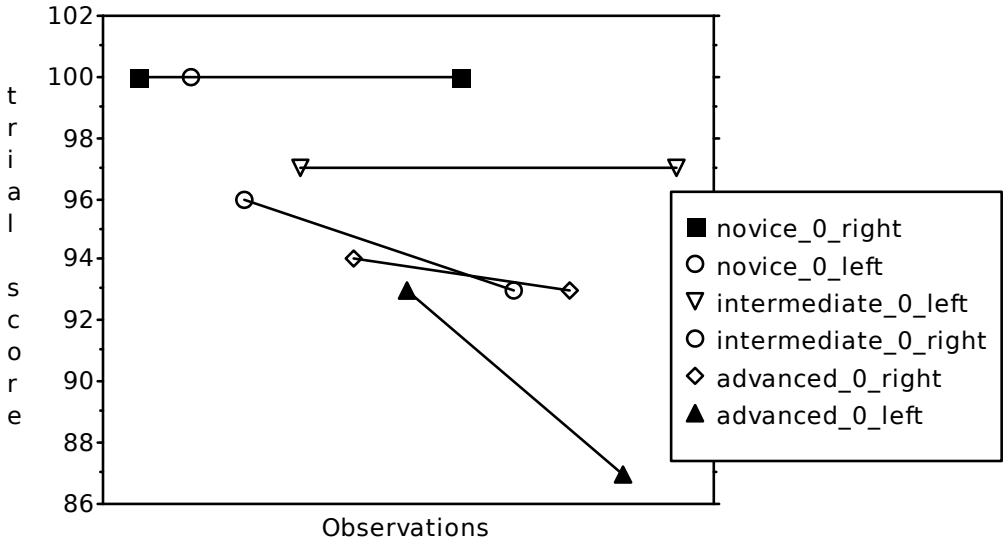


Figure 25. Asymptotic overall trial scores on each model for the experienced ENT proctor. Scores are plotted in order of trial presentation.

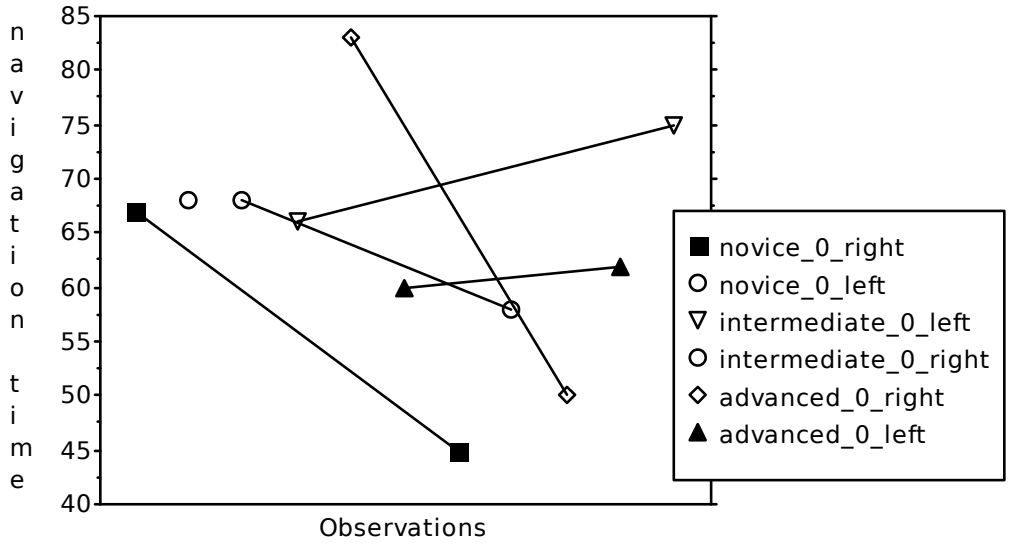


Figure 26. Asymptotic navigation times on each model for the experienced ENT proctor. Times are plotted in order of trial presentation.

It may also be instructive to look at asymptotic performance for the three subtasks for this highly experienced ENT subject. In Figures 26-28 we present the subtask times for this same set of trials. As can be seen from these figures, the systematic decrease in overall trial time for intermediate and advanced model (shown in Figure 24) appears to be due primarily to a systematic decrease in completion time for the dissection subtask, while navigation and injection times appear to vary more randomly across these temporal samples.

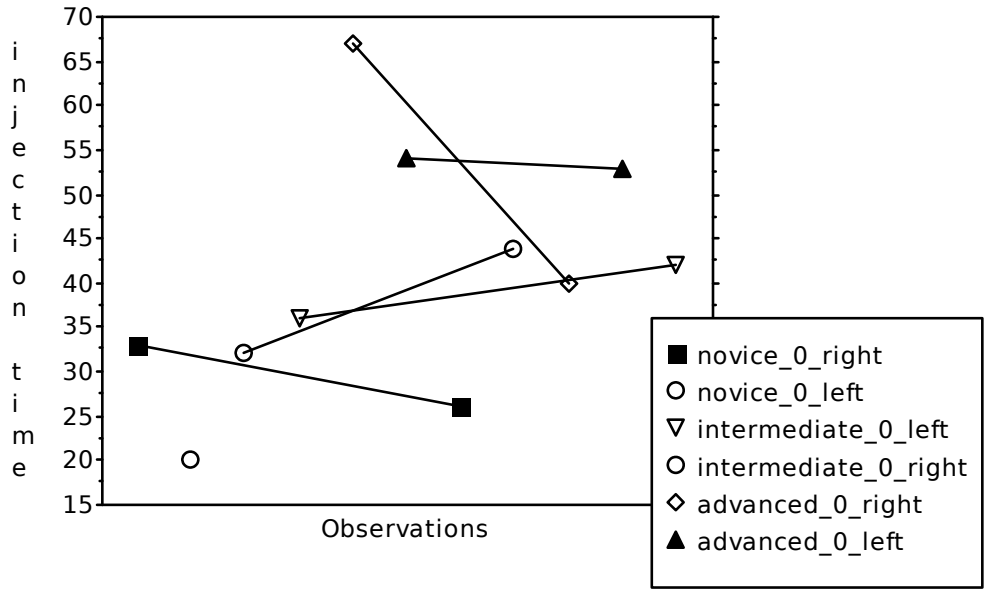


Figure 27. Asymptotic injection times on each model for the experienced ENT proctor. Times are plotted in order of trial presentation.

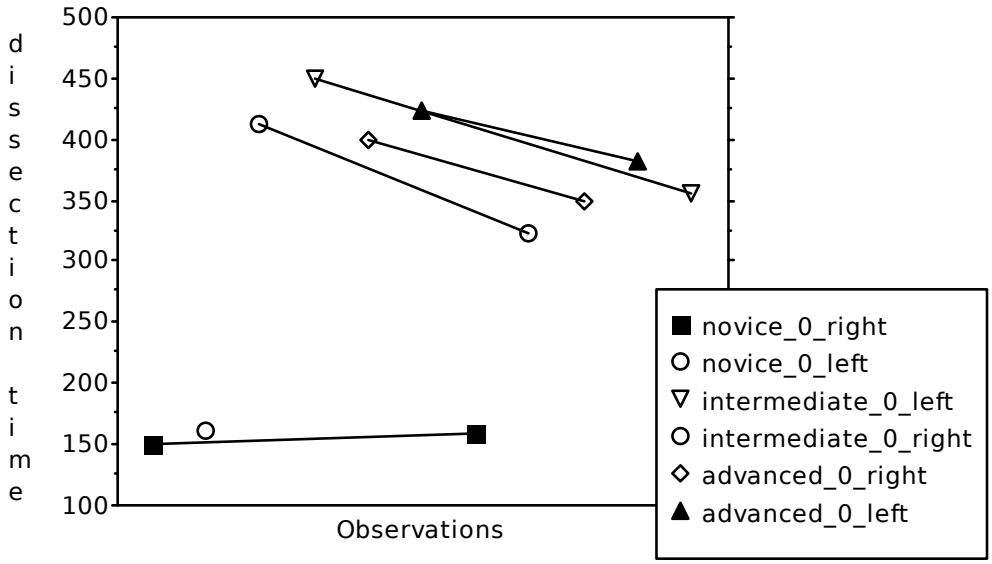


Figure 28. Asymptotic dissection times on each model for the experienced ENT proctor. Times are plotted in order of trial presentation.

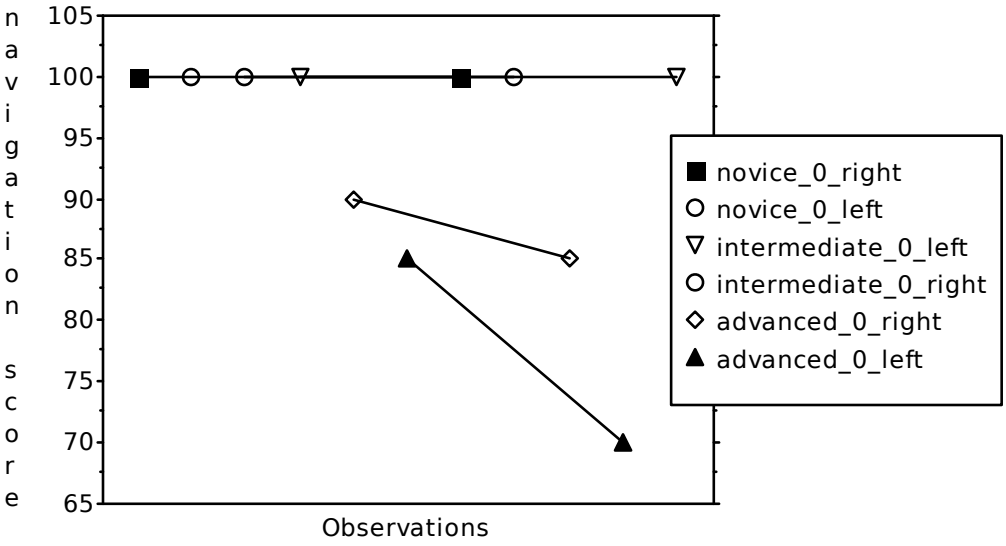


Figure 29. Asymptotic navigation scores on each model for the experienced ENT proctor. Scores are plotted in order of trial presentation.

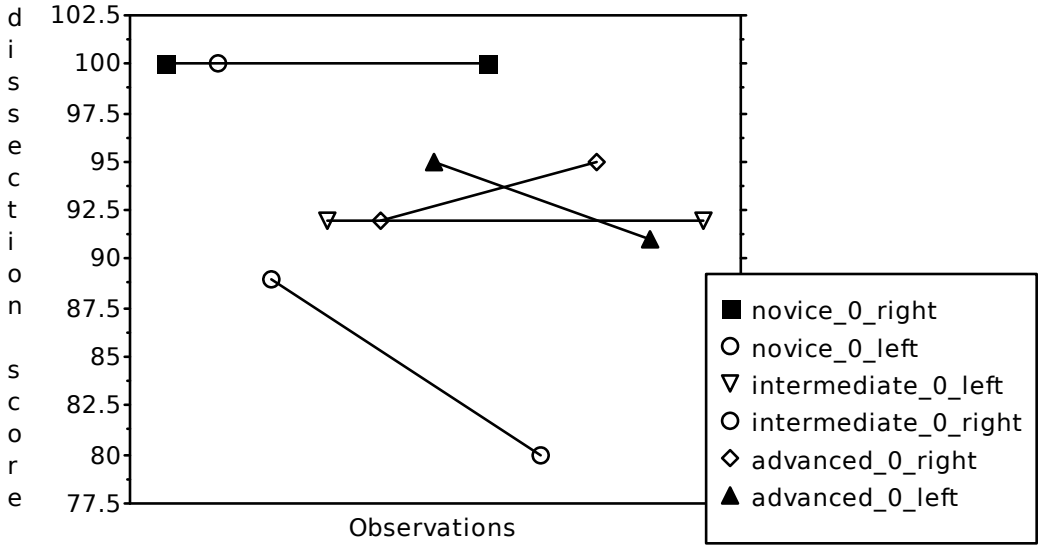


Figure 30. Asymptotic dissection scores on each model for the experienced ENT proctor. Scores are plotted in order of trial presentation.

Similarly, examination of Figures 29 and 30 suggests that the overall decrease in trial score for the intermediate and advanced models for this experienced subject are due to decreases in navigation score for the advanced model condition and a decrease in dissection score for one of the intermediate models (intermediate-0-right). Injection performance and performance on the navigation and dissection subtasks for the other models appear to indeed be at steady-state asymptotic levels. (Note that injection subtask scores were at 100 percent for each of these eleven trials, and are consequently not charted here.)

Collapsing across model over these 11 trials for the expert ENT subject, we observe the Pearson product-moment correlation coefficients ("r") shown in Table 5 among these performance variables (critical r = .6021 for $p < .05$, df 9).

	trialscore	navtime	navscore	injtime	disstime	disscore	trial time
trialscore	1						
navtime	.037	1					
navscore	.856	.119	1				
injtime	-.718	.384	-.608	1			
disstime	-.643	.29	-.387	.642	1		
disscore	.557	-.136	.049	-.416	-.593	1	
trial time	-.609	.424	-.357	.689	.987	-.581	1

Table 5. Correlation matrix showing relationships among subtask and overall performance measures across the final eleven trials by the experienced ENT proctor.

These findings suggest that (for this subject, at least) the best asymptotic predictor of overall trial score is navigation score (r = .856), and that the best predictor of overall trial time is dissection time (r = .987), as suggested above.

Non-ENT proctor steady-state performance

When we look at the correlations among these same performance variables for the last six trials of the other highly experience proctor (CA), we see the same result, as shown in Table 6 (critical r = .8114 for $p < .05$, df 4). Again, the best predictor of overall trial score is navigation score (r = .924), and the best predictor of overall trial time is dissection time (r = .657), although the latter relationship is not statistically significant.

Indeed, if we plot the trial times and trial scores for this proctor side-by-side with the values for the expert ENT proctor, we see that they are remarkably similar. In Figures 31 and 32 we see the trial times and scores, respectively, for the non-ENT proctor on the left-hand side of each graph and the expert ENT proctor on the right-hand side of the each graph. Again, trial times on the novice model are in the 250-second range, while trial times for the other model condition are in the 550-second range for the non-ENT proctor.

	trialscore	navtime	navscore	injtime	disstime	disscore	trial time
trialscore	1						
navtime	-.459	1					
navscore	.924	-.193	1				
injtime	-.686	.795	-.592	1			
disstime	-.889	.627	-.664	.657	1		
disscore	.694	-.805	.37	-.602	-.916	1	
trial time	-.471	.151	-.27	-.079	.657	-.626	1

Table 6. Correlation matrix showing relationships among subtask and overall performance measures across the final six trials by the nonENT proctor.

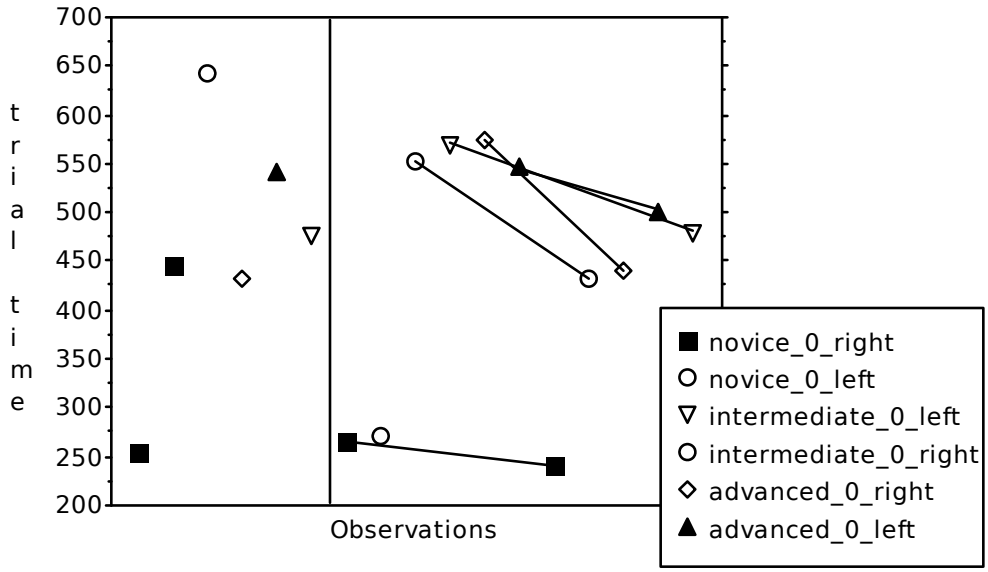


Figure 31. Comparison of the asymptotic trial times on each model for both the nonENT proctor (on the left) and the experienced ENT proctor (on the right). Trial times are plotted for each proctor in order of trial presentation.

These findings suggest that extensive experience with the simulator may afford a non-ENT subject performance values which are comparable with an experienced ENT proctor. Assuming the validation findings discussed above, this bodes well for the training effectiveness of the simulator.

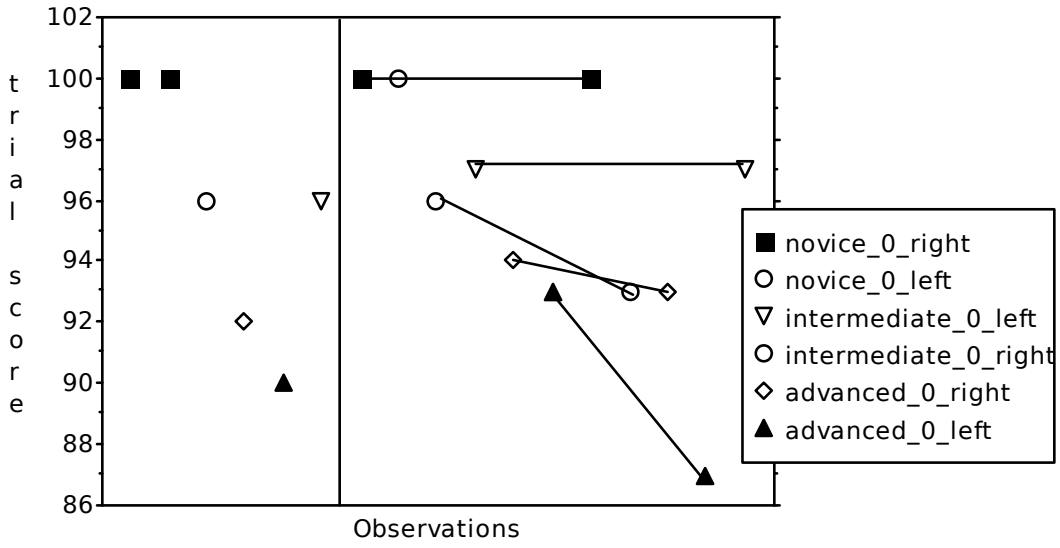


Figure 32. Comparison of the asymptotic trial scores on each model for both the nonENT proctor (on the left) and the experienced ENT proctor (on the right). Scores are plotted for each proctor in order of trial presentation.

Post-Session Questionnaire

All 12 ENT subjects provide us with post-session questionnaire data. All scale responses indicated here are for 10-point scales anchored at the end points.

Proctor's Instructions: Most subjects found the proctor's instructions "very useful" (6/12) or "adequate" (4/12). The one area singled out for improvement was "endoscope and tool handling".

Simulator Layout: Simulator layout and interaction with the model and instruments was rated as moderately realistic (mean = 6.5 and 6.2, respectively).

Abstract Model: The assessed training benefit of the abstract model was relatively variable (mean = 5.3, s.d. = 2.6), perhaps reflecting the variability in baseline experience of the subjects. The assessed training value of each of the five "training aids" (hoops, targets, dissecting spheres, voice feedback, and heart rate) is shown in Figure 33.

As can be seen in these notched box plots (which plot the response percentiles around the median, with the center of the notch being the median and the ends of the box representing the 25th and 75th percentile score), heart rate was viewed as the least beneficial training aid for this model.

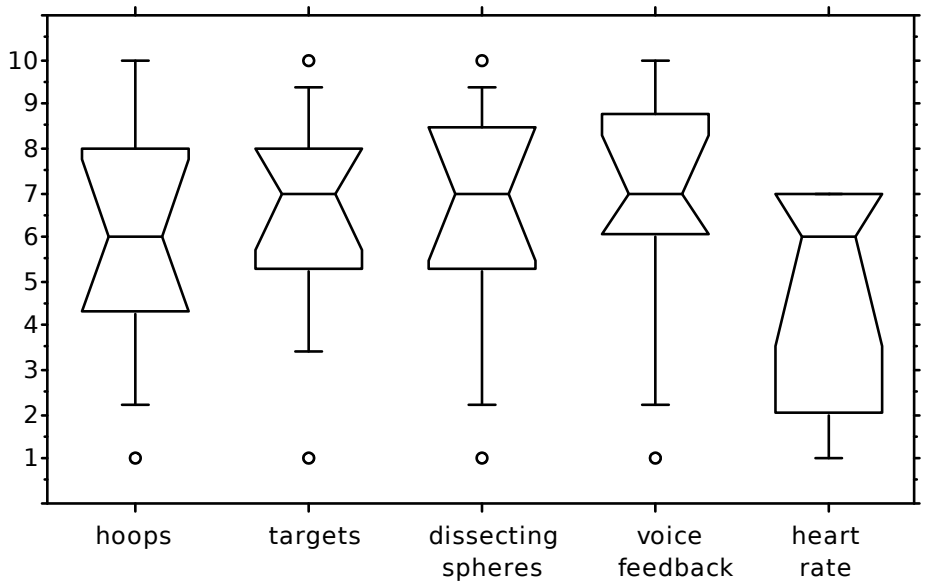


Figure 33. Subject ratings of the training value of each of the training aids in the abstract ("novice") model (1 = "none", 10 = "significant").

In addition, most subjects indicated that the level of difficulty of the abstract model should remain the same (9 subjects out of 12) or increase (3 out of 12).

Intermediate Model: Subjects overwhelmingly indicated that the experience with the Novice Model prepared them for Model 2 (mean = 8.3, s.d. = 1.5). Dissection was judged as more difficult in Model 2 than in Model 1 (median = 7.1), while navigation and injection were seen as about the same level of difficulty (median = 5.5 and 5.0, respectively).

Subject ratings of the benefits of the Model 2 training aids are shown in Figure 34. As is evident from the non-overlapping notches in the box plots, the heart rate cue was judged to be significantly less beneficial than the hoops, targets or voice feedback. Furthermore, note that the ratings for these training aids are considerably higher for the intermediate model than for the abstract (novice) model.

One possible explanation for this surprising finding is that in the novice model the cues were useful for training psychomotor skills and instrument control, while in the intermediate model they were *also* useful for training the specific surgical procedure.

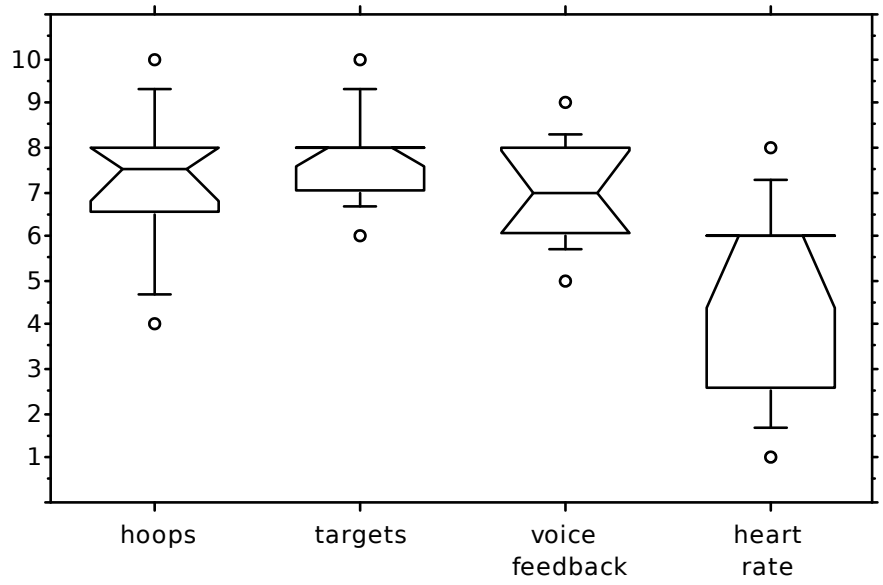


Figure 34. Subject ratings of the training value of each of the training aids in the intermediate model (1 = "none", 10 = "significant").

Ratings of realism of the anatomical model were also extremely high, with a median across respondents of 8.0 and no ratings below "6" on the 10-point scale. It is also interesting to note (in examining the correlations among survey responses) that the more realistic the anatomical model was rated by these subject, the easier the Model 2 subtasks were perceived to be (r = -.503, -.477, and -.499, for navigation, injection and dissection, respectively).

Advanced Model: Figure 35 presents a summary of the distributions of responses by the ENT subjects to all questionnaire items regarding the advanced model. Responses to question 12 ("model2 benefit") indicate that the vast majority of these subjects felt that the experience with Model 2 was highly beneficial for performing on Model 3.

In rating the difficulty of Model 3 on the three essential subtasks subjects indicated that Model 3 was moderately more difficult than Model 2. Finally, the level of procedural realism for Model 3 was rated quite high, with a median of 8.0 on the 10-point scale ranging from "far from reality" to "close approximation".

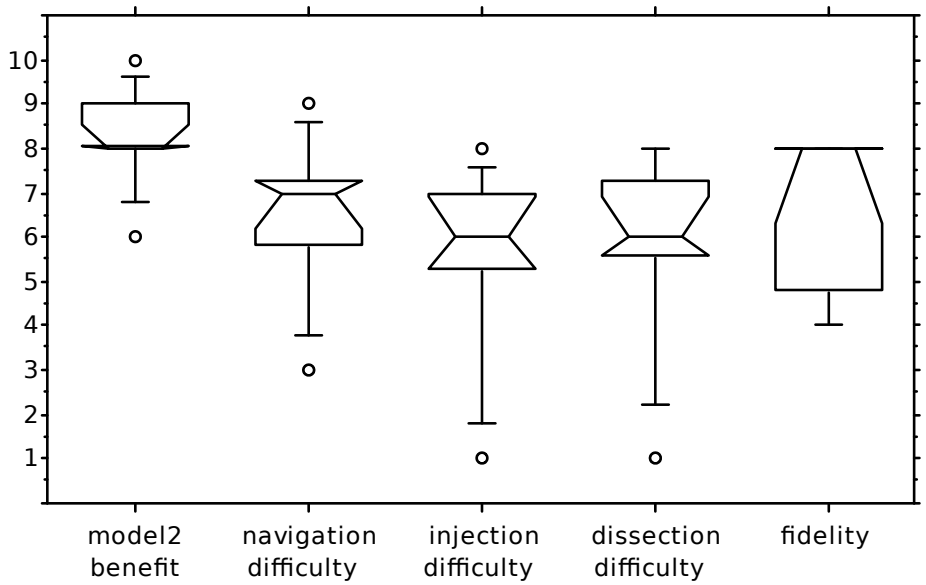


Figure 35. Subject ratings of the training benefit of Model 2 for subsequent performance on Model 3, the difficulty of the three subtasks (1 = "Model 3 easier", 10 = "Model 3 more difficult"), and the fidelity of the simulation in the advanced model (1 = "far from reality", 10 = "close approximation").

Force Feedback: Seven of the 12 ENT subjects rated the use of the haptic subsystem. When asked whether they preferred the force feedback on or off, 4 preferred it "on", one preferred it "off", and one initially preferred it "off" but later preferred it "on". When asked if they preferred the force display on the forceps or on the endoscope, 6 indicated a preference for forces on the forceps, while 1 indicated a preference for forces on the endoscope. (Note that they only experienced the forces on the forceps, not on the endoscope, so these repsonses were partially speculative.

Ratings of the realism of the force feedback were moderately low, with a mean of 5.1 and a standard deviation of 1.7 on the 10-point scale from "far from reality" to "close approximation", although one of the 7 subjects rated it an "8" and none of the subjects gave it a realism rating below "3".

Overall Evaluation: Responses to items regarding the overall evaluation of the simulator are summarized in Figures 36 through 38. Level of difficulty of the simulator for the three subtasks is assessed in Figure 36; the results suggest that the tasks were perceived as

roughly the same level of difficulty as an actual procedure, with the exception of the dissection task, which is perceived as somewhat more difficult on the simulator.

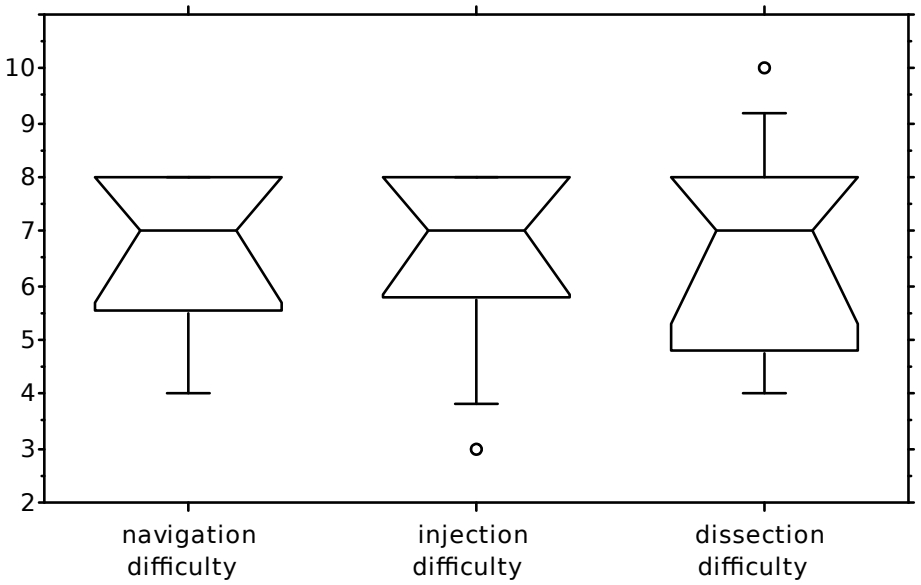


Figure 36. Subject ratings of the overall difficulty of the three subtasks compared with an actual procedure (1 = "simulator easier", 10 = "simulator more difficult").

In assessing the training value of each model for themselves (Figure 37), there appears to be a linear trend from the novice model to the advanced model, with all seen as valuable but the Advanced model especially so.

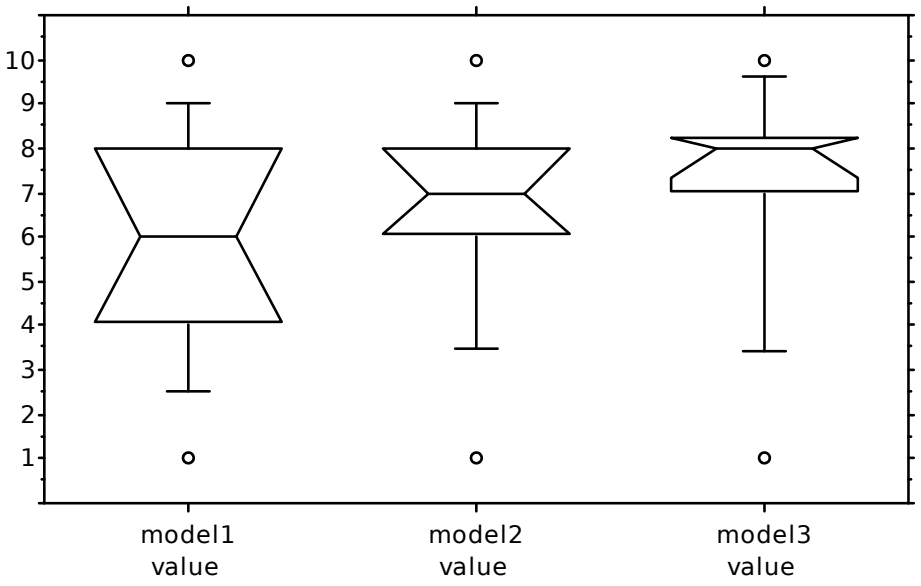


Figure 37. Subject ratings of the overall training value of each of the three models (1 = "none", 10 = "significant").

In assessing the adequacy of the virtual anatomical model for developing proficiency at the common tasks (Figure 38), these subjects indicate that the model is quite adequate for training navigation and injection skills, but may be less than adequate for training dissection skills.

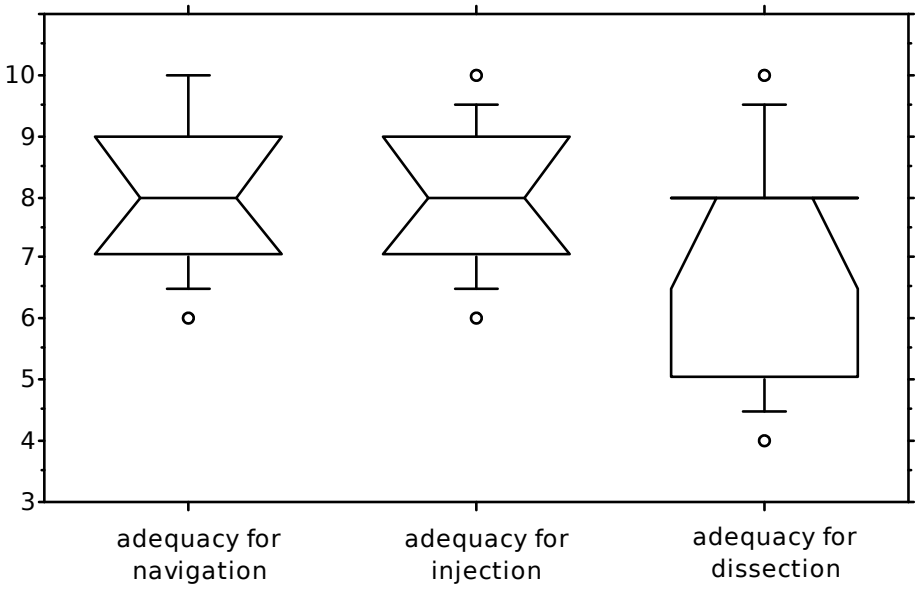


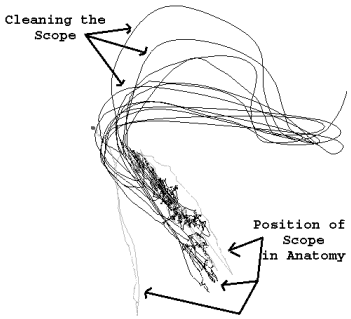
Figure 38. Subject ratings of the adequacy of the anatomical model for developing proficiency at the three common subtasks (1 = "inadequate", 10 = "more than adequate").

When asked to indicate what levels of ENT training would benefit from exposure to the simulator, the 10 ENT respondents offered the following distribution of responses, suggesting that in its current implementation, the simulator is best suited for junior residents (as designed):

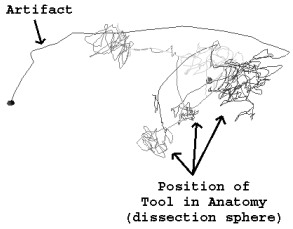
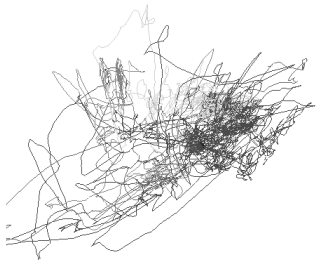
med student = 2
R1 = 6
R2 = 6
R3 = 4
R4 = 2
R5 = 2
R6 = 0
ALL = 4

Finally, would further realism enhance training on the common tasks? All 10 respondents indicated "yes", as we would expect.

ENDOSCOPE POSITION DATA:

<table>
<tr><td>A. Staff Otolaryntgologist
 
</td><td>B. Non-ENT Subject
  </td></tr>
<tr><td>C. Jr. Resident Otolaryngologist
  </td><td>D. Sr. Resident Otolaryngologist
  </td></tr>
</table>

TOOL POSTION DATA:

<table>
<tr><td>E. Staff Otolaryngologist
 
</td><td>F. Non-ENT Subject
  </td></tr>
<tr><td>G. Jr. Resident Otolaryngologist</td><td>Sr. Resident Otolaryngologist</td></tr>
</table>

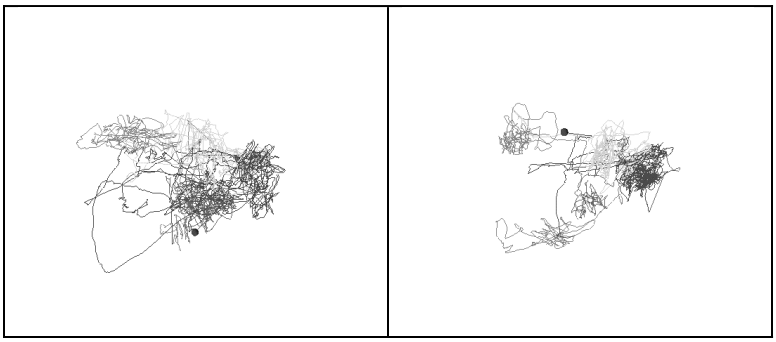


Figure 39. Greyscale path renderings for endoscope and tool traversals during a procedure by a representative sample of subjects. Endoscope paths (A, B, C, D) are for trials using the intermediate model and include "cleaning" of the "bloody scope". Tool paths (E, F, G, H) are for trials using the novice models, in which tool selection was done automatically by the system for each dissection target ("sphere").

Optimal Path Analysis

One promising technique for evaluating user performance on these surgical simulation tasks is to look at the paths of the virtual endoscope and virtual
tools over the course of a procedure. A sample of such paths is shown in Figure 39 in greyscale and in the attached color sheet in color. We have selected a sample of trials to illustrate the variability in user performance.

The colors (shades of grey) in the endoscope path renderings (figures A, B, C,and D) correspond to the three subtasks of the procedure (navigation, injection, and dissection) during a trial on the intermediate model. Note that there are a number of large sweeps outside the main area of activity; these represent the periodic removal of the "bloody" scope to wipe it on the pad on the forehead of the mannequin.

For the tool paths (E, F, G, and H) the colors represent the dissection of each sphere in the novice model. Since the tool for each sphere was automatically chosen by the system in this model (to provide subjects with a systematic exposure to the instrument set), the colors in these paths also represent performance with the various tools.

Tool path scores. In addition to drawing the paths for visual analysis, we calculated a set of objective scores based on the tool path

data. The total time taken for each sphere was determined by the number of frames for that segment. These measures effectively translate into overall time, since the frame rate is determined primarily by position of the renderer in the anatomy, which was essentially constant across subjects for each (automatically selected) tool.

A second measure was the mean deviation of the tool position (in each frame) from the centroid of the dissecting sphere for that tool. A third (composite) measure was then calculated for each sphere as

*sphere path score = total frames * mean tool deviation from centroid/1000*

Mean frames/sphere, mean deviation from sphere centroid, and mean sphere path score were also calculated for each subject's path, yielding an overall "path score". Note that for each of these measures (frames, deviation, and path score), the lower the score, the better the performance.

First-order correlations among these new path measures and the other subject performance measures are presented in Table 6 (critical r = .4060 for p <.05, df 22). As can be seen, the mean frames/sphere and tool path scores were correlated significantly with all other performance measures, and most importantly, dissection score.

In addition, the mean deviation/sphere measure correlated significantly with dissection score, overall score, trial time and navigation score. Note further that all of these significant correlations were in the direction expected. These results may provide validation evidence for the system-generated scoring algorithm.

	path score	mean dev	frames/tool
path score	1		
mean dev...	.625	1	
frames/t...	.951	.381	1
disscore	-.74	-.41	-.76
disstime	.936	.398	.973
injscore	-.708	-.269	-.748
injtime	.814	.364	.837
navscore	-.559	-.44	-.526
navtime	.596	.374	.592
score	-.798	-.439	-.809
trialtime	.922	.416	.95

Table 6. Correlation matrix showing relationships between the tool path measures and the system-generated subtask and overall performance measures for the initial novice trial.

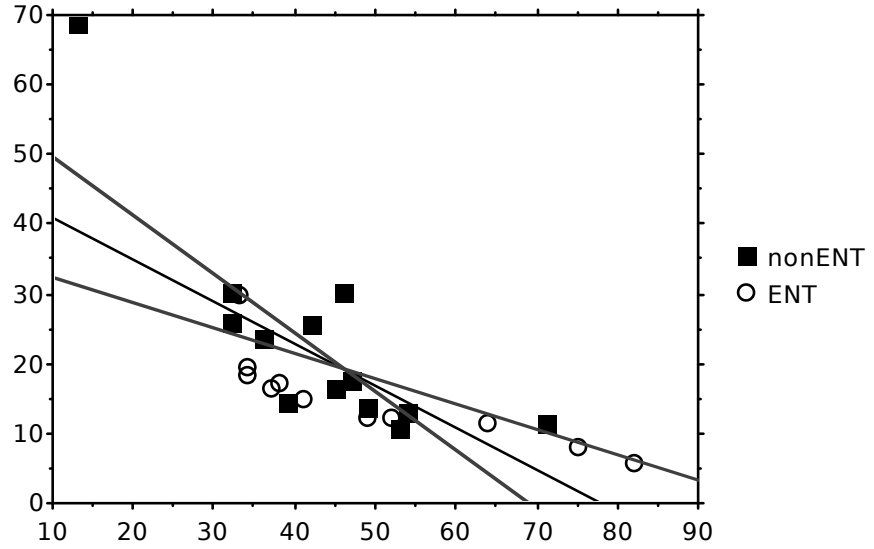


Figure 39. Linear regression for novice tool path score by system-generated dissection score, with 95% confidence limits for the slope of the regression line, broken down by subject group.

The strong (inverse) relationship between the tool path scores and the system-generated dissection scores can be seen clearly in Figure 39. It is interesting to note that the relationship appears to be even stronger for the ENT group alone.

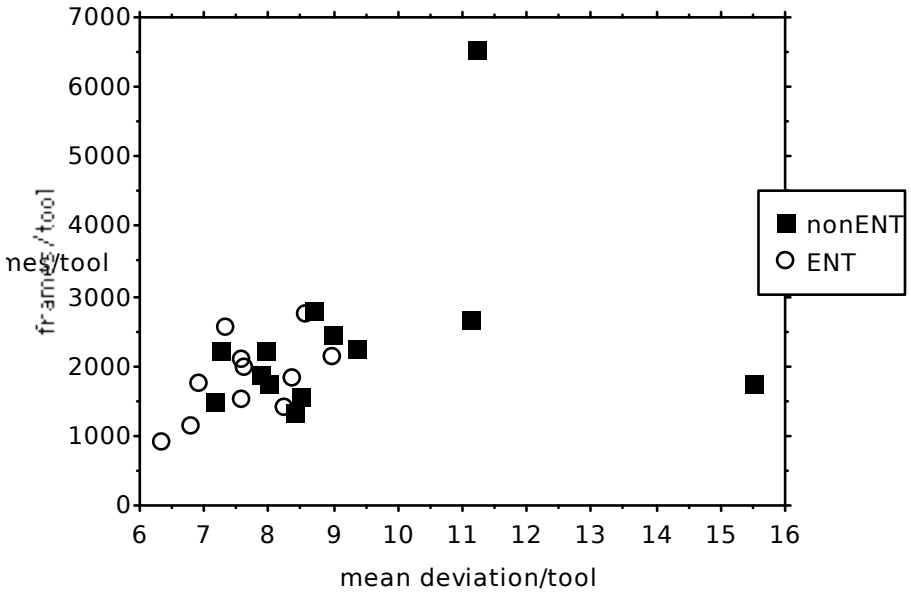


Figure 40. Scatterplot showing the relationship between the two components of the tool path score (mean frames per tool, or sphere, and mean deviation from the sphere centroids) for all subjects on initial novice trial, broken down by subject group.

Figure 40 illustrates the relationship between the two components of the tool path score (essentially, speed and accuracy). Points closest to the origin on both axis reflect better performance. It can be seen that the ENT scores cluster much closer to the origin.

In examining these differences further using Student's t, we find that the nonENT subjects did indeed deviate significantly more from the sphere centroids (ENT mean = 7.66, s.d. = .807, nonENT mean = 9.22, s.d. = 2.27, p = .0422). However, their time scores (ENT mean = 1883 frames/sphere, s.d. = 562.25, nonENT mean = 2381 frames/sphere, s.d. = 1327.07, p = .2162) and overall tool path scores (ENT mean = 15.06, s.d. = 6.49, nonENT mean = 23.24, s.d. = 15.33, p = .1142), while worse, were not significantly so.

Proctor Observations

A number of significant insights can be gleaned from subject's open-ended comments, informal observations by the proctor and problems experienced during trials. Following are some observations and recommendations derived from the proctors' experiences.

Haptics

Subjects (most notably the residents and ENT staff) had difficulty with the need to steady their thumb to control the instrumented forceps during injection and dissection. This stabilization while opening and closing the forceps was required to keep the tip of the virtual tool from moving away from the area of interaction. This need for stabilization has been attributed to a lack of realistic haptic feedback.

While the version of the haptics subsytem tested here was somewhat rough, Version 1.3 of the software (not yet formally evaluated) has made significant improvements to the haptics, also providing a much more realistic representation of grasping, tearing and injecting the virtual tissue. Grasping of the virtual tissue is represented by a "hold" on the tip of the instrumented forceps, after closing the jaw on the tissue, as if holding onto a static object. Tearing of virtual tissue is represented by a resistive force on the tip of the instrumented forceps as it is being pulled away from the "hold" position, with a final "release" after a predetermined distance. Injecting is represented by a "pop" when initially passing the virtual needle through the tissue and a "hold", keeping the instrumented forceps static in the X,Y,Z position, but not the heading, pitch and roll.

In addition to these enhancements for stabilization, the suction tool is "pulled" toward the virtual tissue based on its proximity, an initial "jolt" is placed on all tools when initially interacting with the virtual tissue and the feel of the sickle knife simulates cutting paper with a straight razor. These improvements have added a tremendous amount of realism, not only to the injection and dissection tasks, but also to the realism of navigating the instrumented forceps through the sinus cavity. These improvements will be implemented and tested in the next phase of the simulator's development.

Retraction of instrumented forceps

During pilot studies, when subjects wanted to swap tools they were required to retract the instrumented forceps from the sinus cavity (retracting to just posterior to the Columella Nasi) before acquiring the new virtual instrument. This requirement was the major contributor to two problems with the encoders on the instrumented forceps. The first problem took place during the pilot studies; the retraction of the instrumented forceps caused the endoscope shaft to collide with the encoders measuring the heading and pitch of the instrumented forceps, which would then cause either fraying or displacement of the cables on the sectors of the encoders.

The second problem was interaction of the sectors of the encoders and the hard palate of the mannequin during retraction. This interaction forced rotation of the sectors, introducing an offset in the initial calibration of the instrumented forceps and therefore an offset in the vector of the virtual tool. The problem was repaired by application of LocTite to the sectors and eliminating the requirement that the subjects retract the forceps during instrument swapping. Immersion Corporation is currently upgrading and improving the design of the instrumented forceps encoders for the next phase of the evaluation.

Hummer speed

The typical microdebrider in the OR operates at 3000 RPM. Experienced surgeons develop a certain rhythm to dissection, and expect and anticipate a certain volume of tissue to be removed over time. In version 1.2 of the simulator, the dissection speed is at best "quarter speed". The more experienced surgeons may have become impatient with this rate of dissection, and may have had a tendency to move on to another area of the anatomy to dissect, to the detriment of their dissection scores. It also appeared frustrating to have areas in version 1.2 that should have been dissectable but would not dissect.

Ambiguity in the posterior ethmoid

In the posterior ethmoid the model takes on an amorphous look, making it somewhat difficult to determine what to dissect next. The inexperienced sinus surgeon would just dissect anything that would disappear. For the experienced surgeon, knowing that you cannot dissect with reckless abandon for fear of creating a major complication, dissection naturally slows down in this area close to the carotid artery, optic nerve and skullbase. In addition, the surgeon normally relies heavily on haptic cues on dissecting posterior ethmoid cells. The absence of haptic cues in an amorphous environment creates an anxious situation for the experienced sinus surgeon. The improved haptic system in version 1.3 is likely have significant impact on dissection performance in this region.

Grabbing bone fragments with jawed tools

Removal of two bone fragments from the Uncinate process and three bone fragments from the Bulla ethmoidalis, which were required for a complete score, was unrealistic, due to the difficulty in grasping with the jaw of the virtual tool; the fragments could only be grasped at their center. Regardless of skill, subjects were required to learn how

to grasp the bone fragments in the simulator and required instruction by the proctor on how to do so. Many of the staff and resident ENT surgeons would initially try to grasp the bone fragments on their edge, as they would in a real surgical environement, but without success. Version 1.3 has corrected this problem by allowing a jawed virtual tool to grasp the bone fragments anywhere along its circumference as well as at its center.

Improvement of the virtual suction tool

During testing, the suction tool's only function was to reduce the volume of the blood spheres inside the anatomy after dissection of part of the tissue. It was brought to our attention that in a surgical environment the ENT surgeon would use the real suction tool not only to remove blood, but also to remove small amounts of tissue and mucous in the Posterior Ethmoid cells, and bone fragments throughout. Version 1.3 allows a small amount of dissection by the virtual suction tool, as well as the ability to grasp and slowly remove bone fragments.

Need to break up Model 1 into separate tasks for the non-MD group

During pilot testing of the non-MD group it was discovered that an initial verbal summary of the task to be performed, without giving a visual example of the task, was overwhelming to the untrained/unfamiliar subject. The need for the subject to memorize each subtask and also to gain an understanding of what was needed to complete the task proved to be too demanding. The original presentation of the material consisted of a videotaped introduction to the simulator (described below) followed by verbal instructions by the proctor, describing each subtask (Navigation, Injection and Dissection) in order and in totality.

The subject was then introduced to the endoscope and instrumented forceps and instructed in their functions for each task. The subject was then asked if there was anything which needed further explanation. We found that the proctoring instructions needed to be repeated multiple times throughout the trial to reinforce what the task entailed. By trial and error, it was finally decided to split each subtask into three independent tasks for their initial trial on the simulator. This breakdown was required for only their initial trial.

Ability to "push through" the virtual anatomy during trials

While performing the three tasks subjects had the ability to push through the virtual anatomy with the shaft of both the virtual endoscope and the virtual tool. The staff ENTs expressed greatest concern with the virtual *tool* passing through the virtual anatomical structures during Injection and Dissection. A potential, but expensive and cumbersome, solution to this problem would be implementation of haptics with 6 degrees of freedom on the instrumented tool. This solution would be too cumbersome to implement in the current design of the mannequin box, but should be considered for future designs.

The unrealistic ability of the virtual *endoscope* to pass through the anatomy, however, was welcomed by the majority of the staff ENTs. When the virtual scope passes through the anatomy, the image on the monitor currently disappears and the screen turns black until the position of the instrumented endoscope is maneuvered back inside the anatomical model. This solution was seen as helpful because the residents were taught to concentrate on staying within the anatomy for the duration of the procedure without potentially traumatizing real tissue in the operating room. This was seen as one of the many advantages of using the simulator to gain the hand-eye coordination necessary to maneuver the endoscope through the anatomy.

Ability to see Maxillary Ostium for Antrostomy

Although the subjects were told to perform a Maxillary Antrostomy, the antrostomy of the Maxillary Ostium was removed from the requirements during the procedure on the simulator because of the inability to realistically view the Ostium. After dissection of the Uncinate process in an actual procedure, the Maxillary Ostium would be in view with a zero or 30 degree scope and the antrostomy would easily be conducted. After dissecting the Uncinate process in the simulation, however, to view the Ostium required a 70 degree scope, with unrealistic positioning of the scope inside the sinus cavity. To correct this problem, Version 1.3 allows dissection of the Uncinate process more superior/laterally and inferior/laterally than Version 1.2. This added realism will be tested and evaluated in the next phase of the evaluation of the simulator.

Instructional video

An instructional video was made to introduce the subjects to the simulator in order to standardize instructions across subjects. The video was designed to instruct them on the procedure, the anatomical structures inside the sinus cavity, and all available tools at their disposal.

The content structure of the original videotape was as follows:

- Endoscope and Forceps
    - Introduction to the instrumented endoscope
        - How to handle it
        - The need and process for stabilization using both hands
        - Hints on guiding the scope, by use of angles, through the anatomy
        - How to eliminate the effects of your natural tremor
        - Introduction to transferring the tool across the Columella Nasi
        - Instructions on cleaning the scope when it becomes "bloody"
        - Hints on alignment of the scope within the plane of the mannequin to keep track of your heading and orientation
        - The scope's ability to rotate the image, by axially rotating the shaft of the scope
        - Explanation of the 30 and 70 degree scopes and their uses to view around corners
    - Introduction to the instrumented forceps
        - Limitations of the subject's ability to fully retract the forceps
        - Rationale for passing the forceps across the Columella Nasi
            - Their rigid connection to the mechanics of the haptics inside the mannequin
        - How to swap tools in models 2 and 3
            - Calling out the desired virtual instrument
            - Retracting the forceps to just posterior to the Columella Nasi to receive the desired virtual instrument
- Body Positioning
    - Keeping your body parallel with the mannequin
        - Turning head to see the monitor instead of turning body
- Anatomy
    - Introduction to major anatomical structures on physical "pull away" model
        - Nasal Passage, Septum and Nasopharynx
    - Introduction to sinus cavity anatomical structures on physical "pull away" model
        - Superior, Middle and Inferior Turbinates, Uncinate Process and Bulla ethmoidalis
    - Description of anatomical structures which will be interacted with and removed during procedure
        - Definition and procedure for a Maxillary Antrostomy

- Description and demonstration of sites for injection during procedure in models 2 and 3
    - Medial Middle Turbinate, root of Uncinate Process and the lateral Nasal wall
- Description and "pull away" of Dissection tasks for models 2 and 3
    - Medialization of Middle Turbinate and dissection of Uncinate Process, Bulla ethmoidalis and widening of the Maxillary Ostium

- Procedures
    - Playback of video of Novice model, performed by Dr. Edmond, no voice over to allow the proctor to discuss the trial with the subject
    - Examples of all virtual instruments available to the subject
    - Playback of video of Intermediate model, performed by Dr. Edmond, no voice over to allow the proctor to discuss the trial with the subject
    - Playback of video of Model 3, performed by Dr. Edmond, no voice over to allow the proctor to discuss the trial with the subject

The video proved to be too advanced for the non-MD group, presenting an excessive amount of information for the untrained or unfamiliar subject. Pilot subjects from this group who were shown the video reported being confused as to what was required of them. Specifically, these subjects did not have enough time to comprehend both the simulator interface and the purpose of the procedure being simulated.

Future work is planned in this area to create group-specific videos for each subject group. Shortening the duration of the instructional video will make the testing protocol more efficient, and automating more of the proctoring instructions will permit a single proctor to complete all necessary tasks without difficulty.

Conclusions and Recommendations

Procedural validity

The validity of the simulator for the ESS domain is suggested by a number findings:

- ENT subjects performed better than non-MD subjects on both the novice (abstract) and intermediate (anatomical with aids) models

- initial performance on the novice model was correlated with residency level and degree of prior ESS experience
- patterns of difficulty for asymptotic performance on the simulator seem to match the typical pattern of subtask difficulty in the OR
- subject ratings of the realism of the virtual anatomical model were consistently high on the post-session questionnaire and in open-ended comments
- post-training questionnaire responses confirm that the simulator was generally perceived as valid and useful for ESS training by the ENT subjects

Curriculum design

The ESS simulator takes medical simulation several major steps forward in its evolution. Aside from its technical accomplishments, the integration of a well-thought-out curricular framework allows it to take advantage of virtual reality without sacrificing the benefits of more traditional computer-aided instruction.

Model 1 introduces the student into an abstract environment allowing the student to gain the required hand-eye coordination with the endoscope and the special skills needed to maneuver the instrumented forceps, without requiring them to concentrate on anatomy.

Model 2 introduces the student to the anatomy, but still utilizes the training aids from Model 1. This model gives the student the help of hoops for the initial passes through the anatomy, targets for injection areas and labels on the anatomical structures with which interaction is necessary. The educational advantages of simulation can best be achieved with a model of this kind.

Model 3 introduces the student to a more realistic environment. There are no longer any training aids to guide the student through the procedure. For Navigation of the scope, the student must rely on what was learned when navigating through the hoops in Model 2. For Injection the student must remember where injection of the vasoconstrictor is useful. For dissection, the student has no labels to indicate what anatomy to interact with, and so must rely on what was learned in Model 2 to perform the procedure.

Future Directions for Evaluation

Scoring Algorithm Validation

Considerable attention was given to the scoring algorithm currently implemented in the system, in an attempt to capture the primary clinical performance factors. Still, it is somewhat complex and ad hoc, and deserves more detailed analysis and validaton.

In the next evaluation phase we will attempt to correlate the algorithmic scores with independent measures of surgical skill, and to adjust the optimal values appropriately. A further refinement will be to normalize the algorithm across models, to facilitate comparisons and to arrive at a common criterion for advancement.

Optimal Instrument/Scope Path Analysis

The optimal path techniques introduced above are compelling because of their "face validity" but will require further development.

Tool path analysis. Preliminary pilot data indicate that these tool path efficiency measures closely reflect observers' subjective rankings of path "goodness". The validity of these measures will be studied more formally in the next evaluation phase. In addition to calculating overall scope and tool path scores, we will look more closely at user performance with the individual tools in the virtual instrument set and at the effects of anatomical position on performance with each tool.

Determination of the clinically acceptable range of path scores is desirable for several reasons, including the development of training protocols that reinforce appropriate skills. The range of path scores is bounded on the low end by the radius of the dissecting spheres and the minimal time to dissect (which is not known, but can be approximated by the asymptotic performance of an experienced ENT surgeon.

The constraints on the high end are not as clear cut (at least for the time component), and should perhaps be determined clinically. While theoretically a surgeon could take infinitely long to perform a procedure, there are several practical and clinical constraints, including the increased risk to the patient during protracted procedures.

Scope path analysis. We have not yet attempted a quantitative analysis of the endoscope path data. In the next evaluation phase we will experiment with various methods of determining mean deviation from the optimal scope path. Initially the most direct path between

the training aids (hoops, targets and dissecting spheres) may serve as an optimal path, but this approach does not take into account the surgeon's knowledge and skill in navigating around critical anatomical features.

A more appropriate approach might be to calculate the optimal path using a robust lowess regression technique on the datasets from a group of experienced ENT surgeons, and then determine the mean deviation from that optimal path for each subject.

Evaluation of Improved Haptics

The value of the haptic feedback subsystem to the simulator was assessed only subjectively in the current phase. In general, it was perceived to be only moderately useful, and we suspect that the relatively poor performance on the dissection task, in particular, may have been due largely to inadequate feedback about the tissue forces.

As described above, preliminary evaluation of Version 1.3 suggests that the haptic subsystem is greatly improved. The improved system will allow us to evaluate more formally its impact on training trial performance and its contribution to OR preformance. In addition, we will investigate more formally the relative merits of force feedback on the endoscope versus the intrument.

Transfer of Training

Transfer of training to the real operative environment is the primary objective of this procedural simulator, and will be the primary focus of the next phase of our research.

In preparation for that phase, we have collected initial OR videotape data from several first-time ESS procedures by residents for whom we have previous simulator performance data. These tapes will be formally analyzed for subtask proficiency, as rated by their attending staff instructors, to derive a set of systematic measures of performance on the target task. These will then be correlated with prior simulator experience for current and entering residents to determine the training effectiveness of the simulator and to provide guidance for its further development.

Model Enhancements

As noted earlier, the development team elected to focus the first iteration of the ESS simulator (appropriately) on the needs of junior residents. Several enhancements to the library of available anatomical models are currently being integrated which are intended to extend the utility of the system to senior residents and more experienced staff physicians.

Appropriate enhancement implementation will require iterative formative evaluation, as before, and the training utility of each enhancement will need to be assessed. Transfer of training may need to be assessed indirectly in some cases, since real-world incidence of certain conditions of interest may be infrequent under normal conditions.

Criterion Performance Standards

One longer-term objective of this evaluation activity will be to determine the clinically acceptable range of user performance scores on the simulator, including those generated by the basic scoring algorithm and those calulated from the optimal path data. Furthermore, more detailed analysis of these scores (e.g., time vs accuracy for specific tools) will also provide us with useful information for customizing training protocols for individual residents.

Ideally, subject performance on the simulator will be reliably predictive of OR performance. Achievement of this correpondence will enable us to better establish performance criteria for advancement through the protocol. In a sense, the scoring algorithm represents a theoretical model of surgical proficiency; additional validation studies of the simulator will help to correct and refine those theoretical models.

In addition to examining the learning curve for simulator performance over time, the distribution of trials over time is also of interest. For our "steady-state" proctors, for example, the temporal distribution of training trials was highly variable, ranging from several trials per day to one trial every 2-3 weeks. Routinely collecting repeated measurements on resident subjects will enable us to explore the optimal number and timing of training trials in the resident's curriculum. Undoubtedly some degree of trial spacing (as opposed to massed practice) will prove most effective, as has been shown in many other domains; of interest is just what that spacing strategy should be for ESS procedure training.

Ultimately, we may also be able to establish an equivalence between time on the simulator and time in the OR. Such a correspondence would be useful for residency curriculum development, and perhaps eventually for professional credentialling. Such a system would be of significant benefit for patients, and could greatly extend accessibility to procedural training opportunities.

Summary

This study presents the results of our systematic evaluation of a high-end virtual reality simulator aimed at training a set of skills essential to endoscopic sinus surgery. Our findings suggest that the simulator represents a valid and useful implementation of the target ESS tasks. In addition, the thoughtful integration of an organized curriculum perspective makes this system uniquely valuable among emerging medical simulation systems.

This study also suggests a framework for incorporating systematic evaluation into the process of developing procedural training simulators in the medical domain. Incorporating both formative and summative aspects of evaluation has greatly enhanced the development process and assures the continuing evolution of a usable and effective system.

Bibliography

Baillie J.,et al, The Future of Endoscopy Simulation: A Duke Perspective, Endoscopy, 24, 1992.

Bard, C., Simulation Modeling of the Colon; First International Symposium on Endoscopy Simulation, World Congresses of Gastroenterology, Sydney, 1990.

Bartlett, Worley, Simpson, Aylward, Bailey, Fish. (1995) Report out to the Modeling and Simulation Working group. Alexandria, VA. Modeling and Simulation Benefits Task Force, Defense Modeling and Simulation Office.

Billinghurst, M., Savage, J., Edmond, C., Oppenheimer, P. The Expert Surgical Assistant: An Intelligent Virtual Environment with Multimodal Input. Medicine Meets Virtual Reality IV, 1996.

Bostrom, M., Singh, S.K., Wiley, C.W., Design of an Interactive Lumbar Puncture Simulator with Tactile Feedback, IEEE Annual Virtual Reality Symposium, 429-435, 1993.

Burnside, Billy, U.S Army Research Institute. Assessing the Capabilities of Training Simulations: A Method and Simulation Networking (SIMNET) Application, 1990.

Caird, J. (1996) Persistent Issues in the Application of Virtual Environment Systems to Training. IEEE Computer Society: Human Interaction with Complex Systems. Vol 3. 124-132

Chen, D., Rosen, J., Zeltzer, D., Surgical Simulation Models: From Body Parts to Artificial Person, Proc. Image VI Conference, July 1992.

Cohen, MacRae, Jamieson. (June 1996) Teaching Effectiveness of surgeons. The American Journal of Surgery. Vol 171. 612-614

Cover, S. et al, Interactively Deformable Models for Surgery Simulation; IEEE CG&A, 13,6,68-75, 1993.

Dawson et al, An Interactive Training/Simulator for Interventional Radiology, J. Vascular and Interventional Radiology, 7(1, suppl 2): 132-136, 1996.

Geis, Kim, McAfee, Kang, Brennan. (March 1996) Synergistic benefits of combined technologies in complex, minimally invasive surgical procedures. Surgical Endoscopy. Vol 10. 1025-1028

Gillies, D. et al, Computer Simulation for Teaching Endoscopic Procedures, Endoscopy, 24, 1992.

Good, M.L., Gravenstein, J.S., Training for Safety in an Anesthesia Simulator, Seminars in Anesthesia, 12, 4, 1993.

Gupta S. et al, Introduction of New Technologies in the Medical Undergraduate Curriculum, Medicine Meets Virtual Reality 4, 1996.

Hoffman, H.M., Murray, M., Irwin, A., McCracken, T., Developing a Virtual Reality-Multimedia System for Anatomy Training, Medicine Meets Virtual Reality 4, 1996.

Hon, D., Ixion's Laproscopic Surgical Surgical Skills Simulator, Medicine Meets Virtual Reality II, 1994.

Hon, D., Medical Reality and Virtual Reality, Medicine Meets Virtual Reality 4, 1996.

Johnson R., Bhoyrul, S., Way, L., Satava, R., McGovern, K., Fletcher, J.D., Rangel, S., Loftin, R.B., Assessing a Virtual Reality Surgical Skills Simulator, Medicine Meets Virtual Reality II, 1994.

Kaufmann, D.M., Bell,W. Teaching and Assessing Clinical Skills Using Virtual Reality Medicine Meets Virtual Reality: Global Healthcare Grid, 1997.

McGovern, K.T. and McGovern, L.T., The Virtual Clinic, A Virtual Reality Surgical Simulator, Medicine Meets Virtual Reality II, 1994.

Merril, J. Virtual Reality in Surgery and Medical Education, Medicine Meets Virtual Reality II, 1994.

Peifer J., et al, Virtual Environment for Eye Surgery Simulation, Medicine Meets Virtual Reality II, 1994.

Pieper, S., CAPS: Computer-Aided Plastic Surgery, PhD Thesis, MIT, 1989.

Pieper, S., et al, A virtual environment system for simulation of leg surgery, Stereoscopic Displays and Applications II, 1991.

Poon A., Williams C., Gillies D., The Use of Three-Dimensional Dynamic and Kinematic Modeling in the Design of a Colonoscopy Simulator, New Trends in Computer Graphics, Springer Verlag, 1988.

Rice, D., and Schaefer, S. Endoscopic Paranasal Sinus Surgery, Raven Press, NY. 1988.

Robb, R.A.; Virtual Endoscopy: Evaluation Using the Visible Human Datasets and Comparison with Real Endoscopy in Patients, Medicine Meets Virtual Reality: Global Healthcare Grid, 1997.

Rohen, J., Yokochi, C., Color Atlas of Anatomy, Igaku-Shoin, New York, 1983.

Rosenberg, L., Stredney, D. A Haptic Interface for Virtual Reality Simulation of Endoscopic Surgery. Medicine Meets Virtual Reality IV, 1996.

Rosse, C. The American Journal of Surgery. Vol 171. 612-614

Rosser J, Rosser L, Savalgi. (Feb 1997) Skill Acquisition and Assessment for Laparoscopic Surgery. Archives of Surgery. Vol. 132., 200-204

Sagar, M.A. et al, Virtual Environment and Model of the Eye for Surgical Simulation, Computer Graphics Proceedings, 205-212, 1994.

Satava, S., Doorway to the Future, Medicine Meets Virtual Reality: Global Healthcare Grid, 1997.

Stredney, D. et al, An Virtual Simulation Environment for Learning Epidural Anesthesia, Medicine Meets Virtual Reality 4, 1996.

Wickham, J.E.A., (Jan 1994), Minimally Invasive Surgery: Future Developments, The British Medical Journal, Vol 308(6922), pp 193-196.