Validation study of a computer-based open surgical trainer: SimPraxis® simulation platform

Linh N Tran1
Priyanka Gupta2
Lauren H Poniatowski2
Shaheen Alanee3
Marc A Dall’Era4
Robert M Sweet2

1Department of Internal Medicine, Loma Linda University, Loma Linda, CA, 2Department of Urology, University of Minnesota, Minneapolis, MN, 3Urology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, 4Department of Urology, University of California, Davis, CA, USA

Background: Technological advances have dramatically changed medical education, particularly in the era of work-hour restrictions, which increasingly highlights a need for novel methods to teach surgical skills. The purpose of this study was to evaluate the validity of a novel, computer-based, interactive, cognitive simulator for training surgeons to perform pelvic lymph node dissection (PLND).

Methods: Eight prostate cancer experts evaluated the content of the simulator. Contextual aspects of the simulator were rated on a five-point Likert scale. The experts and nine first-year residents completed a simulated PLND. Time and deviations were logged, and the results were compared between experts and novices using the Mann–Whitney test.

Results: Before training, 88% of the experts felt that a validated simulator would be useful for PLND training. After testing, 100% of the experts felt that it would be more useful than standard video training. Eighty-eight percent stated that they would like to see the simulator in the curriculum of residency programs and 56% thought it would be useful for accreditation purposes. The experts felt that the simulator aided in overall understanding, training indications, concepts and steps of the procedure, training how to use an assistant, and enhanced the knowledge of anatomy. Median performance times taken by experts and interns to complete a PLND procedure on the simulator were 12.62 and 23.97 minutes, respectively. Median deviation from the incorporated procedure pathway for experts was 24.5 and was 89 for novices.

Conclusion: We describe an interactive, computer-based simulator designed to assist in mastery of the cognitive steps of an open surgical procedure. This platform is intuitive and flexible, and could be applied to any stepwise medical procedure. Overall, experts outperformed novices in their performance on the trainer. Experts agreed that the content was acceptable, accurate, and representative.

Keywords: simulation, surgical education, training, simulator, video

Introduction

Traditionally, surgical training has been based on a master-apprentice model in which trainees gain experience through repetitive, supervised training with increasing levels of clinical responsibility over time.1,2 This model creates challenges for novice surgeons, who face shortened work weeks and expanding skill requirements. There are fewer opportunities to perform and refine open surgical techniques with the emergence of minimally invasive techniques.3 Additionally, subjective assessment of skill by the faculty lacks reliability and may introduce bias based on resident qualities not related to technical performance4 as well as interpersonal inconsistencies among the faculty.5 Simulation addresses these issues by providing a risk-free, learner-centered environment in which education of trainees and assessment of their performance can take place.
without time constraints. Simulators are also able to provide objective feedback and assessment of performance.6

High-fidelity simulators have been designed to train in minimally invasive skills and procedures such as tendon transplant, cholecystectomy, limb trauma, hysteroscopy, liver surgery, anastomosis, intravenous insertion,7 and urologic surgery.8–11 These trainers focus on developing the psychomotor and visual-spatial skills necessary to perform these procedures. However, they do not emphasize the cognitive aspects of performing surgery, including judgment and decision-making, which may constitute as much as 75% of what is learned during a medical procedure.12 In order to master a surgical procedure, a novice surgeon must learn to make decisions related to patient positioning, choosing medical instruments, recognizing important anatomical features, determining the next logical step, and interacting effectively with an assistant. A simulator that addresses these decision-making issues, allows for interaction between the user and the virtual operating room, gives the user informative and summative feedback, and has the potential to increase the efficiency and success of a procedure when performed on an actual patient.

In this study, we evaluated the acceptability, content validity, and certain aspects of construct validity of a simulation engine prototype called the SimPraxis8 (Red Llama, Inc, Seattle, WA, USA). The SimPraxis is a novel, computer-based cognitive trainer for performing complex medical procedures. Our research questions were developed with the purpose of the trainer in mind, ie, to prepare surgeons and medical students in training for the cognitive aspects needed to perform surgery in the operating room. More specifically, it attempts to improve the cognitive preparation process taken from an atlas standard by providing the opportunity for the user to interact with videos of real structures and instruments and to practice decision-making under the supervision of a virtual mentor. The prototype was designed to teach a complex open procedure known as pelvic lymph node dissection (PLND), which is the first series of steps during a retropubic radical prostatectomy.

We assessed the acceptability, content validity, and construct validity of the trainer using definitions described by Gallagher et al, with an overall focus on construct validity as reported by Sweet et al.13,14 Acceptability addresses whether potential users like the trainer and would recommend and/or use it. Content validity is a subjective measure of a simulator’s cohesiveness and completeness in teaching and testing criteria. The content aspects of validity typically rely on an examination and evaluation for completeness of the content of the simulator by a subject matter expert. Construct validity is the extent of the ability of the simulator to “identify the quality, ability or trait that it was designed to measure.”13 The construct we examined was whether previous experience in the real operating room environment predicted performance on the simulator. In order to test this construct, we compared the performance of first-year surgical residents rotating through the urology department with that of board-certified urologists.

Materials and methods
SimPraxis PLND simulator
A modified PLND procedure was recorded on video, edited, and divided into subtask sections. These sections were then integrated under the direction of the subject matter expert (RS) into the SimPraxis interactive software engine using C-sharp programming language. The procedure was modeled after the gold standard surgical atlas in urology, ie, Hinman’s Atlas of Urologic Surgery.15 The entire simulator runs on a standard laptop with downloadable software.

In the opening interface of the simulator, users can choose from a potential library of procedures. The user begins by viewing the patient history link which contains individual patient histories with relevant laboratory, imaging, and/or pathology results. Next, the user is introduced to his or her “virtual mentor,” who provides feedback throughout the experience. The introduction to the trainer includes indications and contraindications for surgery, technical tips, and potential complications.

Users learn to choose appropriate instruments by clicking buttons on the right side of the display (representing the “back table”) and dragging them onto a Mayo stand. The content includes all steps, beginning with the initial exposure of the abdominal cavity to the removal of the lymph node packet (Table 1). To begin the surgery, the user is prompted to place the patient in the correct position for optimal exposure. This may include rotating the patient toward or away from the surgeon and/or in and out of the Trendelenburg position. After positioning the patient correctly, the user begins the procedure by selecting an actor (surgeon or assistant), choosing an instrument for that actor to use, and then placing the instrument in the appropriate anatomical location on the paused surgical field. Choice of actor, choice of instrument, and instrument placement can be changed at each substep. If the user selections are correct, he/she is rewarded with verbal feedback from the virtual mentor, a video clip of the moves contained within that portion of the procedure, and permission to proceed to the next step. Following completion of the exercise, the user is asked questions regarding postoperative management related to prostate cancer.
After finishing the practice exercise or test, the user receives summative feedback, including time taken to complete the task and total number of deviations from the pathway of the simulated procedure. Deviations from simulator standards include choice of actor, choice of surgical instrument, and placement of instrument within the video scene. These performance metrics can be broken down and displayed during individual steps of the operation.

The user has the choice of performing the procedure in two different modes, ie, a practice mode or a testing mode (Figure 1). In the practice mode, the virtual mentor makes comments to help guide the user regarding who should perform each action, which instruments should be used, and when and why certain moves are appropriate at any time during the virtual procedure. The practice mode allows the trainee to become familiar with the simulator, practice the steps of the procedure, and request mentorship in the form of hints and suggestions. In the test mode, the embedded virtual mentor comments are turned off and the user’s selections, including requests for assistance, are monitored and logged. This log is available for later review by residency directors or continuing medical education administrators.

**Validation study design**

This study was approved by the internal review board committee at the University of Washington. All rules and regulations were followed with regard to the responsible conduct of research as established by the university.

The validation study was completed at the 2004 Western sectional meeting of the American Urological Association in San Diego, CA, USA. Experts (board-certified urologists who perform PLND in their practice) and residents were eligible to participate. Following signing of informed consent, participants were introduced to the simulator in a standard fashion with an explanation of the function of each component by a single investigator (LT). Next, each participant was given an introductory procedure unrelated to PLND to allow the user to become familiar with the simulator’s user interface. When each participant indicated that they were comfortable with using the simulator, he or she was asked to perform a monitored PLND procedure in test mode.

Performance data were collected from both the expert and resident groups. Key metrics collected for each substep included testing time, instrument choice, instrument placement, choice of actor, and number of hints given. Time taken to complete the task is the time (in minutes) taken by the user.
to complete all steps of the procedure. The total number of deviations is the sum of all deviations made during each substep of the procedure. The minimum number of deviations for the procedure is 0, which corresponds to error-free execution of the procedure, and the maximum number of deviations is unlimited. Choice of actor requires the user to choose the correct actor for each substep and to use his or her assistant effectively to help with the procedure when necessary. Deviations from correct patient positioning were collected for step 1.

After completing the test, each expert participant was asked to complete a questionnaire. In the questionnaire, each expert provided his or her demographic information and rated certain aspects of the simulator on a five-point Likert scale, with a rating of 1 corresponding to “totally disagree;” 2 as “somewhat disagree;” 3 as “neutral;” 4 as “somewhat agree;” and 5 as “totally agree.” They were also asked to provide qualitative feedback on how to improve the content of the simulator. Residents did not complete these questionnaires because of their relative inexperience in performing PLND.

Time taken to complete the task and total number of deviations were compared between experts and novices using a Mann–Whitney test because of the data set not conforming to a normal distribution. Internal reliability of the SimPraxis PLND simulator was analyzed using a multiple comparison test for proportions in cross-tabulation for each step in order to assess if residents and experts are spending equal proportions of time with the same level of accuracy for the following domains: choice of instruments, choice of actor, and choice of anatomic location. Statistical significance was defined as $P < 0.05$ and a trend was considered at $P < 0.10$.

**Results**

**Demographics**

A total of 17 subjects comprising eight experts and nine residents participated in the study. The experts were from five institutions, with ages ranging from 34 to 45 years. All experts were male, and seven of the eight were from academic practice and one was from a large group practice. All experts had completed their residency in the United States within the previous 20 years and had performed more than 20 PLND procedures in the previous 2 years. The nine residents had rotations scheduled with the Department of Urology, University of Washington. Eight of the nine residents had not performed any PLND procedures prior to the study; and the ninth had viewed one PLND before. The number of general surgical procedures performed by the residents ranged from seven to 70 cases. Eighty-nine percent of residents were familiar with the name, appearance, and function of different surgical instruments.
Acceptability
The expert participants were asked to answer pre-task and post-task questions related to the acceptability of simulators for PLND training. Upon entry into the study, 88% of experts felt that a validated simulator would be useful for PLND training. After using the simulator, 100% of experts felt that the SimPraxis PLND simulator would be useful for training, and in addition felt that it was more useful than standard video training. Using the five-point Likert scale, experts rated the simulator an average value of 2.5 when asked if they would “use the simulator to maintain their skills” (Figure 2). Experts found the simulator was acceptable and rated it at an average of 3.5 when asked if they would “rather have their assistant train using the SimPraxis PLND simulator” prior to working with him or her in the operating room (Figure 2). The experts rated the simulator at an average score of 3.6 when asked how well the simulator “trained the user to work with an assistant” (Figure 2).

Content validity
The SimPraxis PLND simulator was shown to have good content validity with regard to its appropriateness for PLND training. Using the five-point Likert scale, experts gave the simulator an average score of 3.9 when asked whether the simulator “supplements anatomical knowledge, aids in understanding of the procedure, and trains indications and steps for PLND” (Figure 2). The simulator was given average scores of 4.4 and 3.9 when asked if it “aids in understanding” and “enhances knowledge of anatomy,” respectively (Figure 2). There was weaker evidence of content validity as related to certification purposes. Fifty-six percent of experts stated that the simulator would be useful for recertification of established practitioners.

Overall, the experts thought the SimPraxis PLND simulator was a useful tool for training residents. Eighty-eight percent of the experts stated that they would like to see the simulator integrated into their residency program curriculum, and the simulator was given an average rating of 4 when asked if they would “recommend the simulator for training residents” (Figure 2).

Construct validity
There was a clear difference in performance on the simulator between experts and residents. Overall, experts needed less time ($P = 0.0023$) and showed fewer total deviations ($P = 0.0062$) than residents when performing the simulated PLND procedure (Figure 3). The total number of deviations was defined as the sum of all deviations for choice of actor, choice of instrument, and instrument placement for all substeps. Table 2 shows the median deviations from the entire incorporated procedure in subcategories of instrument choice, instrument placement, and choice of actor. Experts showed significantly fewer deviations than novices for instrument choice ($P = 0.0006$) in the five individual steps for which this decision was relevant (Table 2). The trend was similar for deviations in instrument placement, but there was no significant difference between experts and novices ($P = 0.015$, Table 2). Although experts had fewer deviations for choice of actor, there was no significant difference between experts and residents ($P = 0.3669$, Table 2). There was a significant difference between experts and novices on the first step involving patient positioning, with experts making no deviations on this step ($P = 0.0062$).

Table 3 demonstrates a comparison of median deviations for the individual steps (steps 2–8) of the procedure. For step 1, involving patient positioning (which did not involve instrument

![Figure 2](https://www.dovepress.com/)

**Figure 2** Expert acceptance and evaluation of content of the SimPraxis® simulator (Red Llama Inc, Seattle, WA, USA).
choice, instrument placement, or choice of actor), there was a significant difference between experts and novices, with median deviations of 0 and 4, respectively ($P = 0.0062$).

While performing PLND, the experts completed the procedure in the same amount of time in 92% of the steps, while residents used the same amount of time in only 40% of steps. Mistakes in choosing instruments for the different steps were more similar for residents (80%) than for experts (48%). Mistakes in choosing the actor and location of the instruments were highly similar for both experts and residents, with no significant $P$-values in 88%–100% of the tested steps. A summary of these results is shown in Table 4.

**Discussion**

The importance of the cognitive domain and lack of virtual reality tools to train in open skills were the motivations behind creation of the SimPraxis cognitive training platform. The engine was built as an interactive atlas standard to accommodate any procedure that has well defined steps and decision points that reflect the cognitive ability of users. The increasingly challenging cognitive aspects of learning, as identified by Bloom’s taxonomy,16 are essential for performing at a high level. The SimPraxis trainer will potentially allow users to move more quickly to higher levels of psychomotor skill, as described by Dave’s taxonomy,16 by reducing cognitive load.

One benefit of the platform is its flexibility for expansion to many medical procedures. Most current simulators contain hardware and software specifically for one type of procedure. Therefore, incorporating many procedures into one simulator is a difficult task which requires significant changes in hardware and interface. In contrast, the SimPraxis simulator is able to expand more easily to other procedures, without the need for changes in the hardware or user interface.

Another benefit of the SimPraxis is its feasibility as compared with virtual reality trainers. Dexterity trainers often require expensive hardware for motion tracking, force measurement, and haptic feedback. In addition, software for these trainers is costly and difficult to design due to the need for developers to create the entire virtual environment of the surgical procedure.17–20 The SimPraxis interface includes a video of actual surgical procedures which provides the user with a realistic training environment and prevents the costly need for reprogramming the virtual environment for each procedure. One particular learning objective, ie, reconstruction of tissue planes, is especially challenging for assessment and training. The use of high-fidelity video has advantages over virtual reality for accomplishing this goal.

**Table 2** Comparison of number of deviations between experts and novices for key metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Experts</th>
<th>Novices</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median deviations (n)</td>
<td>Interquartile range (25%–75%)</td>
<td>Median deviations (n)</td>
</tr>
<tr>
<td>Instrument choice</td>
<td>6</td>
<td>3.75–10.25</td>
<td>37</td>
</tr>
<tr>
<td>Instrument placement</td>
<td>11</td>
<td>4.00–27.75</td>
<td>32</td>
</tr>
<tr>
<td>Choice of actor</td>
<td>7</td>
<td>2.75–17.75</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 3  Median values of deviation from steps 2 to 8 of the simulated procedure

<table>
<thead>
<tr>
<th>Instrument choice</th>
<th>Placement location of instrument</th>
<th>Choosing actor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experts</strong></td>
<td><strong>Novices</strong></td>
<td><strong>P-values</strong></td>
</tr>
<tr>
<td>Step 2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Step 3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Step 4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Step 5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Step 6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Step 7</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Step 8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Experts</strong></td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Novices</strong></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><strong>P-values</strong></td>
<td>0.1922</td>
<td>0.0559</td>
</tr>
<tr>
<td><strong>Novices</strong></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>P-values</strong></td>
<td>0.0192*</td>
<td>0.2514</td>
</tr>
<tr>
<td><strong>Novices</strong></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>P-values</strong></td>
<td>0.2514</td>
<td>0.2981</td>
</tr>
<tr>
<td><strong>Novices</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>P-values</strong></td>
<td>0.5000</td>
<td></td>
</tr>
</tbody>
</table>

Note: *Statistically significant P-values using Mann–Whitney test.

The SimPraxis simulator was determined to be an acceptable training tool for the PLND procedure. This platform is best suited to training medical students and residents to help them quickly acquire the cognitive knowledge required to perform the procedure in order to prepare them better for learning in the operating room. Although experts did not think the simulator would be useful for maintenance of their own skills, they did think it would be useful for training with their assistants in the procedure prior to doing cases in the operating room. This is somewhat expected, because experts who perform PLND frequently may not need to review the steps of the procedure, but may benefit from working with an assistant who was trained on the procedure using this tool. Experts showed their acceptance of the simulator as a new tool of training by ranking the usefulness of the SimPraxis PLND platform higher than that of the gold standard independent study medium for surgery, ie, video. This is in agreement with the educational literature that provides evidence for the superior training effect of interactive over passive learning tools.21 Given the economic and technological feasibility of interactive engines such as this, and considering the large volume of resources devoted by medical governing bodies towards passive video media, consideration should be given to directing resources toward interactive training engines.

All aspects of content validity were rated positively by the experts. The weakest area was in use of the tool for accreditation purposes. Further studies and improvements should be made to gather more evidence that the SimPraxis simulator can be used for recertification of surgeons performing PLND.

To determine the construct validity of the SimPraxis simulator, differences in performance between novices and experts were analyzed. The results of this preliminary study show that there was a significant difference between the expert and novice groups for both time taken to complete the task and total number of deviations. However, for choice of actor, although experts outperformed novices, there was no significant difference in total deviations between novices and experts. Specifically, within each step, there were a small number of deviations in the choice of instrument, instrument placement, and choice of actor for the expert group, as shown in Table 4. However, in step 6, the experts had a greater amount of deviations as compared with other steps for instrument placement. Step 6 involves distal dissection and division of the lymph node packet, which could be performed differently by surgeons depending on their training and how they utilized their assistant. The variation in performance of this step also highlights the benefit of such a tool in identifying steps in a procedure that should allow for variation and needs a component of experiential learning.

Novices showed variation in all three areas throughout the steps when compared with experts, which again highlights the construct validity of this platform.

The data provided in Table 3 support the validity and reliability of the instrument because one would expect individuals on the learning curve (interns) to show a larger range of time taken to complete the task. In contrast, one would expect experts to show a wider range of variability in choice of instruments because there are many ways to complete the steps of a procedure. Therefore, we suggest that the SimPraxis simulator can determine variability in technique among groups of users.

There are numerous methods for performing a surgical procedure such as PLND. One limitation of the SimPraxis simulator is its inability to distinguish whether deviations from the incorporated procedure are due to user incompetence or discrepancies among practicing styles. The intent of this
trainer was to provide training in the procedure as defined in Hinman’s Atlas of Urologic Surgery, so the poor reliability amongst the experts on “instrument choice” may reflect differing opinions among experts rather than error. Future versions of the simulator should incorporate techniques from more than one expert to address this limitation and distinguish deviations from errors.

Conclusion
Overall, the SimPraxis PLND cognitive trainer has evidence of validity and demonstrates a novel, economically feasible engine that allows for high-fidelity interactive learning and assessment for surgical procedures. Attractive features of this engine are its relatively low developmental cost when compared with the immersive virtual reality model, its uniform useable medium (a compact disc which can run on any laptop), and its flexible platform that can be applied to any surgical technique, including open surgery. In addition to supplementing the learning of students and residents, the simulator engine has the potential to be a useful tool for practicing physicians to rehearse rarely performed or complex procedures. Another potential application is to train and demonstrate cognitive competency in new procedures. Future versions of the PLND trainer should include input from more experts and include other less common methods for performing a particular procedure. Overall, the data presented in this study show that the simulator is a promising tool for training in an open surgical procedure and has the potential to be incorporated into the residency training curriculum. Future studies may include a more global population of participants who are at many different points along the learning curve. Additionally, studies should be completed to examine validity to use for high-stakes assessment.

Acknowledgment
The authors acknowledge the west central section of the American Urological Association for enabling collection of data on site.

Disclosure
RS is a former chief medical officer of Red Llama Inc, but has no financial ties with this company. All other authors report no conflicts of interest in this work. This research was sponsored by Astra Zeneca.

References