The use of a virtual reality surgical simulator for cataract surgical skill assessment with 6 months of intervening operating room experience

Shameema Sikder¹
Jia Luo²
P Pat Banerjee²
Cristian Luciano²
Patrick Kania²
Jonathan C Song¹
Eman S Kahtani³
Deepak P Edward¹,³
Abdul-Elah Al Towerkı³

¹Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA;
²College of Engineering, University of Illinois at Chicago, Chicago, IL, USA;
³King Khaled Eye Specialist Hospital, Riyadh, Saudi Arabia

Purpose: To evaluate a haptic-based simulator, MicroVisTouch™, as an assessment tool for capsulorhexis performance in cataract surgery. The study is a prospective, unmasked, nonrandomized dual academic institution study conducted at the Wilmer Eye Institute at Johns Hopkins Medical Center (Baltimore, MD, USA) and King Khaled Eye Specialist Hospital (Riyadh, Saudi Arabia).

Methods: This prospective study evaluated capsulorhexis simulator performance in 78 ophthalmology residents in the US and Saudi Arabia in the first round of testing and 40 residents in a second round for follow-up.

Results: Four variables (circularity, accuracy, fluency, and overall) were tested by the simulator and graded on a 0–100 scale.Circularity (42%), accuracy (55%), and fluency (3%) were compiled to give an overall score. Capsulorhexis performance was retested in the original cohort 6 months after baseline assessment. Average scores in all measured metrics demonstrated statistically significant improvement (except for circularity, which trended toward improvement) after baseline assessment. A reduction in standard deviation and improvement in process capability indices over the 6-month period was also observed.

Conclusion: An interval objective improvement in capsulorhexis skill on a haptic-enabled cataract surgery simulator was associated with intervening operating room experience. Further work investigating the role of formalized simulator training programs requiring independent simulator use must be studied to determine its usefulness as an evaluation tool.

Keywords: medical education, computer simulation, educational assessment, technology assessment, cataract surgery

Introduction

Surgical training using simulators has been adopted by many specialties to provide training in a controlled environment and generate objective assessment of skills to track progress.¹⁻⁶ The need for surgical simulators in cataract surgery education is driven by many forces, including pressure to optimize surgical training with a reduced work week for trainees.⁷ The Accreditation Council for Graduate Medical Education (ACGME) is in the process of mandating milestones that will monitor the development of surgical competencies in training of residents in surgical specialties, and objective assessment tools are vitally important in this process.⁸

The ACGME deems that ophthalmology residents must perform a minimum of 86 cases before they practice as primary cataract surgeons.⁹ From ACGME published reports, current ophthalmology residents graduate after performing an average of 150 cases,⁹ one report states that full surgical maturity is not achieved until the surgeon has performed between 400 and 1,000 cases.¹⁰ This finding suggests that freshly
graduated surgeons likely have not reached their full surgical maturity. While it is understandable that a resident cannot be expected to perform over 400 cases – about three times the average resident caseload – it can be suggested that surgical simulators might help the ophthalmology resident gain additional experience and log practices that would otherwise be challenging to perform in the operating room (OR).

Three cataract surgery simulators are currently available: Eyesi (VRmagic Holding AG, Mannheim, Germany), PhacoVision® (Melerit Medical, Linkoping, Sweden), and MicroVisTouch™ (ImmersiveTouch, Chicago, IL, USA).

The number of published articles describing the use of ImmersiveTouch and PhacoVision simulators is limited, while the Eyesi simulator has been described in peer-reviewed publications. The ImmersiveTouch simulator is the only device that has incorporated a haptic (tactile) feedback interface. In a published survey, a majority of the participating ophthalmologists agreed that integration of tactile feedback could provide more realistic operative experience.18

A previous study has shown that simulators retain realism or “face validity” and present a parallel that can be used to further enhance the resident’s skill set.19 With their realism and accuracy in depicting ocular surgery, it is assumed that improvement on the simulator may result in concurrent improvement in OR performance. This paper attempts to assess the reverse, i.e., whether OR experience can enhance performance on the surgical simulator and be used as an assessment tool for microsurgical skill. While current educational efforts are geared toward reducing the number of surgical cases needed to achieve competency with simulator use, this paper seeks to correlate real-life operating experience with improvements in simulator performance. In the future, the authors plan prospective masked training on the simulator to evaluate the simulator’s influence on cataract surgery capsulorhexis performance using the same metrics and statistical analysis.

Previous work specifically addressed the validity of the MicroVisTouch simulator.19,20 The haptic-based simulator MicroVisTouch was evaluated over a 1-year period at two primary institutions: Wilmer Eye Institute at Johns Hopkins Medical Center (Baltimore, MD, USA) and King Khaled Eye Specialist Hospital (Riyadh, Saudi Arabia [SA]) to assess whether simulator performance improved over the course of a 6-month interval of OR experience.

Methods
Study design
This prospective study was carried out at the Wilmer Eye Institute and the King Khaled Eye Specialist Hospital, affiliated with the Wilmer Eye Institute with its own independent residency training program. Institutional review board approval was obtained at both sites, and all participants provided written consent. An invitation was made to residents at both locations to participate in the study (n=25 US; n=65 SA). Those residents who were interested contacted the research coordinators to participate. Those who participated in the first round were contacted to participate at the 6-month interval. Participation was completely voluntary.

With regards to simulator experience, capsulorhexis and wound creation were specifically assessed. Performing a capsulorhexis on the MicroVisTouch begins with using a virtual keratome to make a corneal incision (Figure 1). When the keratome moves through the cornea, haptic resistance is felt. The user is notified (visually and haptically) if the instrument is in contact with the lens and cornea, or the wound is distorted. After the corneal incision is made, the user withdraws the keratome and changes to a virtual cystotome. The cystotome is used to create a radial slit in the capsule. After the slit is made, the cystotome changes to virtual forceps. A pressure sensor, attached to the haptic stylus, opens and closes the virtual forceps to tear the capsule (Figure 1). Once the capsule is completely torn, it is removed through the corneal wound.

Inclusion criteria
Ophthalmology residents from the Baltimore region were invited to participate in the study at the Wilmer Eye Institute (n=25) as well as residents in Riyadh (n=65). Participants had not previously used the simulator and were acclimated before the first training session by completing a dexterity program. The dexterity program allowed the user to trace around a set circular path using the haptic device, allowing them to orient to the simulator (Figure 1). Participants were given instructions once in a demonstration of the capsulorhexis module prior to performing capsulorhexis. The participants then performed the capsulorhexis simulation six times. Computer generated scores following the procedure were recorded with their lowest score dropped. The data was saved for future comparison.

Training program
The study was conducted on the MicroVisTouch surgical simulator, with the following specifications: Dell® Precision T7500 with Intel® Xeon® X5690 6-core CPU, 12 GB RAM, NVIDIA® Quadro® FX 6000 graphics card, SensAble Phantom® Omni haptic device, Microsoft® Windows® 7 64-bit OS. A 10-day study was conducted at the King
Khaled Eye Specialist Hospital in Riyadh, followed by a 7-day study in Baltimore at the Wilmer Eye Institute. Six months later, this study was repeated at the King Khaled Eye Specialist Hospital and Wilmer Eye Institute. In the intervening period, participants self-reported surgical and simulator experience. Surgical cases reported were those as the resident as a primary or assistant surgeon on cataract and non-cataract cases.

**Variables**

The simulator tested four variables on a 0–100 scale (100 being a perfect score). The sample size for baseline testing was (n=49 SA; n=29 US). Ten of the residents from the US were non-Wilmer Maryland residents. In the 6-month testing, the same residents were re-evaluated, although decreased participation was noted (n=25 SA; n=15 US). Decreased participation from the first round of testing to the second round was a result of increased OR requirements and student availability.

The four variables measured were circularity, accuracy, fluency, and overall score. Circularity was a measure of how perfectly circular the capsulorhexis was made. To determine how perfect of a circle was made, the average radius from the performance was used along with how smooth and circular the capsulorhexis was. Accuracy was a measure of contact frequency of the cornea with the virtual instrumentation (keratome, cystotome, or forceps). When the cornea was touched for a third time, a 4% penalty was applied, as well as for each subsequent touching of the cornea. While the virtual instrument was in the anterior chamber, if the instrument put an unacceptable amount of pressure on the corneal incision, a 5% penalty was applied each time this happened. Fluency was a measure of how many times the capsule was grasped and torn with the virtual forceps during capsulorhexis. The ideal number of grasps and tear was set to four, with each subsequent grasp and tear resulting in a 3% penalty. The overall score metric was calculated by assigning a weight to...
each of the above-mentioned variables – circularity (42%), accuracy (55%), and fluency (3%).

A six sigma-based statistical quality control measure was used.14 Process capability indices were used to compute the training process improvement from baseline simulator test (T1) to simulator tests performed after 6 months of clinical training (T2). Process capability indices assessed the distance of a lower specification limit from the mean by comparing it to three sigma. Similarly the distance of higher specification limit from the mean was also compared to three sigma.

Average and standard deviation were calculated for the two groups of residents. A two-sample t-test with a significance level of 0.05 was conducted for data comparison after the normality of the data was checked.

Results

Study participants for baseline data collection were postgraduates who were in the first (PGY-1; n=11 SA; n=0 US), second (PGY-2; n=16 SA; n=11 US), third (PGY-3; n=8 SA; 11 US), or fourth (PGY-4; n=14 SA; n=7 US) year of residency.

Study participants for the 6-month data collection were in PGY-1 (n=8 SA; n=0 US), PGY-2 (n=10 SA; n=6 US), PGY-3 (n=4 SA; n=5 US), or PGY-4 (n=3 SA; n=4 US). The ophthalmology training program in SA is a 4-year program (excluding internship), whereas the duration of the training in the US is 3 years.

The first round of testing (baseline) from hereafter will be referred to as T1 (Fall of 2012) and second round of testing (after 6 months) is referred to as T2 (Spring of 2013). For T1, there were 49 residents from SA and 29 residents from the US. For T2, there were 25 residents from SA and 15 residents from the US. The average and standard deviation of resident performance in each subgroup were computed for comparison (Tables 1–2). The two-sample t-test was conducted for comparing the data. A significance value of 0.05 was used for the hypothesis test (Table 3).

Comparing the 6-month data to the baseline data, with the exception of circularity, the improvement in all test variables was statistically significant for both SA and US ophthalmology residents (P<0.05). Reduction in standard deviation in the performance score was also calculated (Figures 2–5). The results are summarized in Tables 1 and 2. The intervening OR and simulator experience for the participants are described in Tables 4 and 5. As expected, more senior trainees had increased surgical experience. Of the 40 residents who were retested in both centers, 28 (n=7 PGY-1; n=10 PGY-2; n=5 PGY-3; n=6 PGY-4) performed better the second time. Five of the six residents that practiced on the simulator between the two rounds of testing showed improvements from the first time being tested.

Discussion

Using a surgical simulator as an assessment tool to teach surgical concepts in conjunction with training in the OR may streamline the teaching of ophthalmic surgery and address criticisms aimed at wet laboratory training and microsurgical skills courses.14 Ideally, the use of a simulator should be a part of a specific curriculum that assesses the knowledge and skill of the user and adapts to the individual learner.

Using the concept of face validity, it can be hypothesized that resident performance on the simulator will indicate how he or she will perform in the OR. The face validity of the MicroVisTouch has previously been presented.19,20 Such face validity studies on a simulator would be ideal but can be challenging to design and execute for cataract surgery with

**Table 1** Experimental results from all residents of Saudi Arabia in the first (n=49) and second (n=25) round of testing

<table>
<thead>
<tr>
<th></th>
<th>Circularity</th>
<th>Accuracy</th>
<th>Fluency</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SAT2</td>
<td>SAT1</td>
<td>SAT2</td>
<td>SAT1</td>
</tr>
<tr>
<td>Mean, %</td>
<td>68.29</td>
<td>65.32</td>
<td>84.54</td>
<td>71.41</td>
</tr>
<tr>
<td>Standard deviation, %</td>
<td>14.50</td>
<td>15.69</td>
<td>10.46</td>
<td>25.97</td>
</tr>
</tbody>
</table>

**Abbreviations:** SAT1, residents from Saudi Arabia in the first round of testing; SAT2, residents from Saudi Arabia in the second round of testing.

**Table 2** Experimental results from all residents from the US in the first (n=29) and second (n=15) round of testing

<table>
<thead>
<tr>
<th></th>
<th>Circularity</th>
<th>Accuracy</th>
<th>Fluency</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UST2</td>
<td>UST1</td>
<td>UST2</td>
<td>UST1</td>
</tr>
<tr>
<td>Mean, %</td>
<td>82.50</td>
<td>80.84</td>
<td>96.57</td>
<td>89.33</td>
</tr>
<tr>
<td>Standard deviation, %</td>
<td>12.97</td>
<td>9.09</td>
<td>2.85</td>
<td>9.23</td>
</tr>
</tbody>
</table>

**Abbreviations:** UST1, residents from the US in the first round of testing; UST2, residents from the US in the second round of testing.
Table 3  P-values of the t-tests from residents present in both rounds of testing

<table>
<thead>
<tr>
<th></th>
<th>SAT2 &gt; SAT1</th>
<th>UST2 &gt; UST1</th>
<th>T2 &gt; T1 (SA and US combined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circularity</td>
<td>0.16</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.031</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>Fluency</td>
<td>0.0002</td>
<td>0.002</td>
<td>2.941E-06</td>
</tr>
<tr>
<td>Overall</td>
<td>0.002</td>
<td>0.03</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Abbreviations: SA, Saudi Arabia; SAT1, residents from Saudi Arabia in the first round of testing; SAT2, residents from Saudi Arabia in the second round of testing; T1, first round of testing; T2, second round of testing; UST1, residents from the US in the first round of testing; UST2, residents from the US in the second round of testing.

a new simulator without pilot data. As a corollary, it can be proposed for a pilot study that a performance improvement on the simulator can be correlated to a performance improvement in the OR. The purpose of this study was to demonstrate performance changes on the simulator in two groups of residents after a 6-month interval with exposure to surgery.

Previous retrospective studies using the Eyesi simulator have been used to examine the outcomes of implementing the simulator in ophthalmology training programs and its effect on clinical skills. Belyea et al\textsuperscript{21} retrospectively compared the performance of the non-simulator experienced residents trained by a single attending with a newer group of residents who have been trained by the same attending ophthalmologist with the Eyesi simulator. For all the measured parameters, the simulator group showed significantly better results, with the exception of the complication rate and complication grade in which there were no significant differences. Complications were graded as follows: 1) Descemet’s membrane tear or detachment, 2) anterior chamber or posterior chamber capsule tear without vitreous loss, 3) vitreous loss and a sulcus intraocular lens, 4) vitreous loss, vitrectomy, and anterior chamber intraocular lens placement. Pokroy et al\textsuperscript{22} also investigated retrospectively the incidence of posterior capsule rupture and operation time for the first 50 phacoemulsification procedures of the non-simulator trainees (before 2007–2008) and simulator trainees (after 2009–2010). A total of 20 trainees (ten in each group) performed 1,000 cases. Both groups showed no significant difference in the number of posterior capsular rupture and the overall operation time for the first 50 cases. However, the simulator group showed a shorter median operating time in cases 11 to 50. A recent study by McCannel et al demonstrates a reduction in errant capsulorhexis with simulator training.\textsuperscript{23} Daly et al conducted a prospective study demonstrating efficacy of using simulators to train when evaluating clinical skills (capsulorhexis) in the OR.\textsuperscript{24} The international cohort in this prospective study comprised residents from differing training levels. The simulation protocol acquired several scores for comparison, including individual variables evaluating components of capsulorhexis performance and an overall performance metric. Reductions in standard deviation of performance scores and process capability indices over the 6-month period were also calculated.

The study showed improved performance by the residents in both institutions on the simulator and was statistically significant for all parameters except circularity. Improved

Figure 2  Mean and standard deviation of circularity scores for residents present in both rounds of testing.

Abbreviations: PGY, postgraduate year; SA, Saudi Arabia; T1, first round of testing; T2, second round of testing.
As Figures 2–5 show, the average of circularity, accuracy, fluency, and overall score for US PGY-2 residents were lower than the PGY-3 and PGY-4 residents. With the exception that the PGY-2 residents scored a slightly higher average on accuracy than PGY-3 and PGY-4, the correlation was valid for all other parameters for residents in all experience levels ($P<0.05$). The trend of slightly decreased overall performance of the SA PGY-2 residents was related to a decrease in the measured circularity, although the accuracy and fluency were noted to increase.

The correlation between experience in the OR and improved performance on the simulator is further substantiated by the finding that simulator performance correlated with cataract surgery experience in the OR for 28 residents at all levels of training. These findings suggest that the simulator may be used as a measure to index the actual surgical proficiency and OR performance of an individual at two time intervals. An alternative explanation for improved simulator performance might be improved familiarity with the simulator system during T2, resulting in improved performance independent of OR experience. However, it is to be noted that in the 6-month period between T1 and T2 the residents had no access to the simulator. On average, the more advanced training level the individual had, the higher their simulator score.
by stratification of data by levels of resident surgical experience. In general, residents who were more experienced in cataract cases demonstrated an increase in their performance.

Another important trend that surfaced when examining the longitudinal data for both SA and US ophthalmology residents was the reduction in standard deviation for all variables, which is measured by process capability indices\textsuperscript{14} based on six sigma quality improvement metrics. This trend was particularly strong in the SA data (Tables 1 and 2). In the US data, the 2013 population had slightly higher standard deviation in the circularity and overall score parameters despite higher averages, but the trend of standard deviation decreasing was still sound for the accuracy and fluency parameters. Similar trends were noticed for the SA data, but to a lesser extent. When relating the differences in standard deviation using process capability indices,\textsuperscript{14} results show that process improvement from 2012 to 2013 as a result of simulation-based training combined with more practice. With the lower performance limit set to 70% (as is commonly defined as “pass” in medical education) for the baseline T1 data, an improvement from 70% in T1 to 87.44% in T2 was noted on average for accuracy, to 84.29% in T2 on average for fluency, and to 78% in T2 on average for overall scores for all residents.

Increased standard deviations in performance score likely represents increased variability in performance, which can be observed in T1. A more uniform performance as observed in T2, characterized by a low standard deviation, is characteristic of data with few outliers.\textsuperscript{16} A similar trend was noted in a previous publication using the MicroVisTouch.\textsuperscript{19} The authors suggest that decreasing standard deviation in the various components of assessment over time may be a sensitive indicator to evaluate surgical maturity and serve as a reasonable evaluation tool. Whether this measure was true for other simulators on the market is yet to be determined.

Intervening OR experience and simulator use may play a role in the trends. The data indicates that the performance of residents with low to moderate OR experience improved their performance more substantially than the residents with

**Table 4** Range of surgical caseload reported by postgraduate year level in Saudi Arabian residents in 2012–2013

<table>
<thead>
<tr>
<th>SA resident surgical caseload range (cataract and non-cataract, primary and assistant surgeon) in 2012–2013</th>
<th>Simulator performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-1 (7 residents): 17–75 cases</td>
<td>Improved by 13.02%</td>
</tr>
<tr>
<td>PGY-2 (9 residents): 46–171 cases</td>
<td>Declined by 0.74%</td>
</tr>
<tr>
<td>PGY-3 (3 residents): 124–315 cases</td>
<td>Improved by 10.1%</td>
</tr>
<tr>
<td>PGY-4 (3 residents): 95–150 cases</td>
<td>Improved by 7.84%</td>
</tr>
</tbody>
</table>

**Table 5** Range of surgical caseload reported by postgraduate year level in US residents in 2012–2013

<table>
<thead>
<tr>
<th>US resident surgical caseload range (cataract and non-cataract, primary and assistant surgeon) in 2012–2013</th>
<th>Simulator performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-2 (6 residents): 20–64 cases</td>
<td>Improved by 3.71%</td>
</tr>
<tr>
<td>PGY-3 (5 residents): 58–139 cases</td>
<td>Improved by 1.52%</td>
</tr>
<tr>
<td>PGY-4 (4 residents): 85–140 cases</td>
<td>Improved by 4.2%</td>
</tr>
</tbody>
</table>

**Abbreviations:** PGY, postgraduate year; SA, Saudi Arabia.
more OR experience, indicating that the simulator may be more useful earlier on in their training. Limitations of this study include the limited sample size, due to the number of residents available. Furthermore, the attrition in follow-up testing may also lead to sample bias as it is unclear if those who did not follow-up had significant improvement in their surgical skills. Additionally, there may be biases due to a lack of exposure to simulator technology. Given the constraints of testing subjects within a live surgical training residency, a control group of subjects with no interval live training was not included as part of the study. Although not ideal, those trainees with lower surgical experience served as a surrogate for this group. A lower standard deviation, higher average score, and statistically significant improvement all seem to point to the fact that the users did improve from their initial attempts on the simulator. An improvement in the simulator performance may signify an improvement in the OR as well, but needs to be tested in more rigorous future experiments designed to confirm this hypothesis.

Acknowledgments
This study was funded by King Khaled Eye Specialist Hospital (Riyadh, Saudi Arabia), National Eye Institute: 5R42EY018965-03 (Bethesda, MD, USA), and Research to Prevent Blindness (New York, NY, USA). The authors would like to acknowledge University of Illinois at Chicago pre-med undergraduate student Naga Dharmavaram for his assistance with this paper.

Disclosure
Mr Kania, Dr Banerjee, Mr Luo, and Dr Luciano’s work was supported in part by ImmersiveTouch, Inc (Chicago, IL, USA). The other authors report no conflicts of interest in this work.

References