Repeatability and Reproducibility of Performance on the Eyesi Ophthalmic Surgery Simulator

B Hussain, GM Saleh, K Theodoraki, SN Gillan, PM Sullivan, F O'Sullivan, C Bunce, I Athanasiadis Moorfields Eye Hospital, London, United Kingdom

Purpose

To evaluate the variability of performance among novice ophthalmic trainees in a range of repeated tasks using the Eyesi virtual reality (VR) ۲ ophthalmic surgery simulator.

<u>Methods</u>

Prospective study conducted at Moorfields Eye Hospital, London. A consultant trainer gave 18 novice ophthalmic trainees (<2 hours of</p>



simulation and intraocular surgical experience) a standardised induction to the Eyesi ophthalmic surgery simulator (VR Magic, Mannheim, Germany), figure 1.

Five modules were selected for completion: one cataract specific task (capsulorhexis level 1, figures a and b) and four generic 3-D tasks (cracking and chopping level 2, figure c; cataract navigation level 3, figure d; cataract bimanual training level 1, figure e; and anti-tremor level 2, figure e).

Each of the tasks was repeated three times to assess repeatability and reliability for each subject – scores, with a maximum of 100 for each attempt at each task were given.

Data were analysed using non-parametric tests because of evidence of non-normaliity. The signed-rank test was used to asess differences in scores for each of the three attempts, whilst the Kruskal-Wallis test was used to assess whether the overall and the range of scores differed between individuals and/or between tasks.

For all tests, a P-value <0.05 was considered significant.</p>



| Scores in all tasks by trainee | | | | Difference between highest and lowest score in a single task | | | | |
|--------------------------------|----------------|--------|----------------|--|----------------|--------|----------------|--|
| Trainee | Lower limit | Median | Upper limit | Trainee | Lower limit | Median | Upper limit | |
| 1 | 0 | 32 | 76 | 1 | 35 | 51 | 57 | |
| 2 | 0 | 69 | 70 | 2 | 10 | 18 | 22 | |
| 3 | 0 | 78 | 94 | 3 | 4 | 6 | 61 | |
| 4 | 6 | 64 | 85 | 4 | 21 | 23 | 32 | |
| 5 | 0 | 19.5 | 81 | 5 | 0 | 4 | 21.5 | |
| 6 | 0 | 72 | 84 | 6 | 6 | 12 | 24 | |
| 7 | 17 | 59 | 90 | 7 | 30 | 46 | 59 | |
| 8 | 11 | 65 | 93 | 8 | 37 | 39 | 65 | |
| 9 | 0 | 34 | 76 | 9 | 8 | 24 | 39 | |
| 10 | 0 | 57 | 86 | 10 | 21 | 28 | 29 | |
| 11 | 0 | 53.5 | 75 | 11 | 16 | 47.5 | 78 | |
| 12 | 37 | 74 | 88 | 12 | 28 | 36 | 36 | |
| 13 | 0 | 53 | 69 | 13 | 3 | 24 | 42 | |
| 14 | 49 | 77 | 95 | 14 | 30 | 39 | 66 | |
| 15 | 0 | 35 | 73 | 15 | 0 | 21 | 42 | |
| 16 | 0 | 79 | 92 | 16 | 4 | 5 | 6 | |
| 17 | 0 | 62 | 79 | | | | | |
| 18 | 0 | 0 | 58 | | | | | |
| P-value | | 0.1104 | | P-va | lue | 0.3878 | | |

| | Tab | s across attempts | s (intra-novice per | rformance) | | | | | |
|-----------------------------|---|--------------------|---------------------|-------------------------|-----------------|-------------|--------|-------------|--|
| | | | 2-Scores of all t | | | | | | |
| | Atte | empt | Lower limit | Median | Upper limit | | | | |
| | 1st | | 0 | 36.5 | | | | | |
| | 2nd | | 0 | 68 | 83.5 | | | | |
| | 3rd | | 13 | 69 | 83 | | | | |
| | | 1st vs 2nd attempt | | <i>P</i> -value <0.0001 | | | | | |
| | | 1st vs 3rd attempt | | | P-value <0.0001 | | | | |
| | | 2nd vs 3rd attempt | | P-value 0.65 | | | | | |
| | | | | | | | | | |
| Table 3 Scores across tasks | 8 | | | | | | | | |
| Overall p | Difference in highest and lowest score in a single task | | | | | | | | |
| Task | Lower limit | Median | Upper limit | Task | | Lower limit | Median | Upper Limit | |
| Capsulorhexis 1 | 0 | 0 | 45 | Capsulorhex | is 1 | 21 | 51 | 61.5 | |
| Cracking and Chopping 2 | 65 | 89.5 | 94 | Cracking and chopping 2 | | 7 | 23 | 38 | |
| Cataract navigation 3 | 42 | 73.5 | 78 | Cataract navigation 3 | | 23 | 40.5 | 47.5 | |
| Bimanual training 1 | 68.5 | 81 | 91 | Bimanual training 1 | | 6 | 19.5 | 28 | |
| Anti-tremor 2 | 0 | 0 | 0 | Anti-tremor | 2 | 0 | 1 | 34 | |
| P-value | | 0.0001 | | P-value | | | (| 0.0030 | |

- Table1: No significant differences in the scores were demonstrated between the trainees using different tasks (P=0.1104). This indicates that in this group of first year trainees there was no significant difference in their ability to execute the tasks. There was also no difference between their highest and lowest scores (P=0.3878).
- Table 2: Highly significant differences were found between the scores achieved in the first attempt and that during the second (P<0.0001) and third (P<0.0001), but not between the second and third attempt (P=0.65). This indicates initial poor reproducibility for the high-fidelity tasks by this group of trainees, while a certain level of consistency in scores is achieved between the 2nd and the 3rd attempt.

• Table 3: Highly significant differences between tasks were shown both in the overall score (P=0.0001) and in the difference between highest and lowest score (P=0.003); i.e. performance varies significantly with the complexity of the task.

<u>Conclusions</u>

• This study, which is the first to quantify reproducibility of performance in entry level ophthalmic trainees using a VR tool, demonstrated significant intra-novice variability.

The cohort of subjects performed equally overall in the range of tasks (no inter-novice variability) but each showed that performance varies significantly with the complexity of the task when using this high-fidelity instrument.

• The data presented show that the simulator would be more useful to monitor performance (formative assessment) rather than to evaluate and quantify overall skills (summative assessment).

The authors gratefully acknowledge support from the London School of Ophthalmology, STeLI (Simulation and Technology-enhanced Learning Initiative) and the IFOS (International Forum of Ophthalmic Simulation)