

## **Evaluating the assessment metrics of the Integrated Cognitive Simulator's laparoscopic appendicectomy module**

Background: Simtics Integrated Cognitive Simulator (ICS) is a software learning environment, previously demonstrated to be of use to surgical trainees in learning the cognitive aspects of procedural skills. It also includes a Test-Mode, which has not previously been assessed for its ability to discriminate between experienced and non-experienced operators. Methods: 10 experienced and 10 non-experienced participants performed a laparoscopic appendicectomy using the ICS laparoscopic appendicectomy (LA) module. Total score, total time, and four further metrics (incorrect hand, incorrect instrument, incorrect location, time > 15s per step) were recorded for each individual, across the four sections and nine steps of the simulation. Results: Median total score was greater amongst surgically experienced participants than surgically inexperienced participants (26.5 %,  $p = 0.023$ ). Instrument use was a significant discriminator between these groups in 2 of 4 of the ICS LA sections.

# Evaluating the assessment metrics of the Integrated Cognitive Simulator's laparoscopic appendicectomy module

---

Authors:

Dylan A Mordaunt      d.a.mordaunt@gmail.com

John A Windsor      j.windsor@auckland.ac.nz

[Ben Loveday]

Corresponding author:

Professor John A Windsor

Department of Surgery

Faculty of Medical and Health Sciences

University of Auckland

Private Bag 92019

Auckland 1023

New Zealand

Affiliations:

1. Department of Surgery, University of Auckland, New Zealand
2. Department of Genetic and Molecular Pathology, SA Pathology, South Australia

Type of manuscript: Original Article

## Contents

Abstract .....	2
Introduction.....	4
Methods.....	4
Materials.....	4
Procedure.....	4
Statistical analysis.....	5
Results.....	5
Demographics.....	5
Metrics.....	5
Discussion .....	5
Discriminant and non-discriminant metrics.....	5
Validity as an evidence-based argument .....	6
References .....	7
Tables and Figures .....	8
Table 1. Demographics, experience and training level.....	8
Table 2. Comparison of metrics between participant groups.....	9

## Abstract

**Background:** Simtics Integrated Cognitive Simulator (ICS) is a software learning environment, previously demonstrated to be of use to surgical trainees in learning the cognitive aspects of procedural skills. It also includes a Test-Mode, which has not previously been assessed for its ability to discriminate between experienced and non-experienced operators.

**Methods:** 10 experienced and 10 non-experienced participants performed a laparoscopic appendectomy using the ICS laparoscopic appendectomy (LA) module. Total score, total time, and four further metrics (incorrect hand, incorrect instrument, incorrect location, time > 15s per step) were recorded for each individual, across the four sections and nine steps of the simulation.

**Results:** Median total score was greater amongst surgically experienced participants than surgically inexperienced participants (26.5 %,  $p = 0.023$ ). Instrument use was a significant discriminator between these groups in 2 of 4 of the ICS LA sections.

Conclusions: ICS LA metrics are able to discriminate between experienced and inexperienced operators. Differences in the discriminant ability of these metrics could be used to optimise the assessment metrics for assessment metrics of procedural simulations.

## Introduction

Simulation technologies are garnering increased attention as a tool for medical and surgical education. These simulators focus on skills in different learning domains, including psychomotor, visuospatial perception, and cognitive skills (1). Simulators are used for both learning and assessing skills in these domains, though little attention has been given to the metrics used in simulation-based assessments.

Simtics Integrated Cognitive Simulator module Test-Mode provides an assessment environment designed to assess the executive routine involved in performing the laparoscopic appendicectomy procedure. This study aimed to determine which of the module's testing metrics are able to distinguish between individuals with different levels of procedural knowledge and experience.

## Methods

### Materials

The Integrated Cognitive Simulator (ICS) produced by SIMTICS ([www.simtics.com](http://www.simtics.com)) is a web-based learning environment, with training modules for a wide range of procedures. This study assessed the laparoscopic appendicectomy (LA) module, previously described by Loveday et al. (2010). The procedure is divided into four sections, with a total of nine steps. These four sections are: 1) Separation of the mesoappendix, 2) Application of Endoloop, 3) Division of appendix, 4) Removal of the appendix.

Two modes exist for learning "Learn-Mode" and formative assessment "Test-Mode". "Learn mode" provides feedback through visual cues, allowing operators to correct an incorrect action. "Test mode" does not provide these visual cues, and records a number of metrics in the logbook, including a proprietary measure of total score (represented as a percentage of total), and total time taken for the procedure.

### Procedure

Institutional ethics approval was obtained. A convenience sample was obtained by emailing surgical trainees and supervisors working in the departments of general and paediatric surgery at Auckland District Health Board. No pre-existing was available with which to develop power calculations, and therefore an arbitrary number of individuals were recruited. Individuals were classified into two groups- those with basic medical and anatomical knowledge but without surgical experience (group 1), and those actively working and training in an area of surgery (group 2). Individuals were excluded if they had previously used the Simtics ICS LA module.

Participants were asked to complete a pre-test questionnaire on three areas- demographics, level of training, and of surgical experience. Testing was completed on a single laptop equipped with the minimum specifications required to run the Simtics module (2), and executed by a single individual. Participants were introduced to the interface in a standardised way, and were provided no additional assistance beyond this introduction. A clinical scenario introduced the participants to the module, and the participants were

advised to use the hook diathermy and forceps retrieval method of appendicectomy, using the ICS LA module in Test-Mode.

In-simulation metrics were retrieved in addition to total score (%) and total time (seconds). These included the recording of incorrect hand, incorrect instrument, incorrect location, time greater than 15s (per step) for each of the four sub-sections of the procedure. The following were calculated from the recorded data- total number of errors across the simulation and for each section, and total number of errors recorded in each of the above categories (hand, instrument, location and time).

## Statistical analysis

Data were analysed using SPSS 22 (3). Continuous variables were assessed using the Mann-Whitney U test, with an alpha level of 0.05.

## Results

### Demographics

10 final year medical students from the University of Auckland were recruited into the first group and 10 surgical registrars from Auckland District Health Board (Auckland City Hospital and Starship Children's Hospital) were recruited into the second group. In general, the second group were older, more experienced, and at a higher level of training than group 1. Demographic characteristics, level-of-training and surgical experience are outlined in Table 1.

### Metrics

ICS LA total score discriminated between groups 1 and 2 ( $p = 0.023$ ). Median time taken to complete the procedure was different between groups 1 and 2 (by 79.5 s), but this difference was not significant ( $p = 0.179$ ). Choice of instrument was a significant discriminator between groups 1 and 2 in section 1 ( $p = 0.005$ ) and section 4 ( $p = 0.070$ ), and overall ( $p = 0.001$ ), but not in sections 2 or 3.

Section score discriminated between the groups in sections 2 ( $p = 0.023$ ) and 4 ( $p = 0.020$ ), but not sections 1 and 3. Time >15 seconds to complete a step and choice of hand, were not significant discriminators between groups in any of the sub-sections, nor overall. Full details are outlined in Table 2.

## Discussion

This study evaluated the ability of internal and external ICS LA metrics to discriminate between surgically experienced and non-experienced participants. We demonstrated that some of the ICS LA metrics are able to discriminate between experienced and inexperienced participants. This provides the basis with which to improve the simulator metrics, and a focus for the future design of surgical assessments.

### Discriminant and non-discriminant metrics

Of the four metrics tested, only incorrect instrument choice showed a significant difference in three of the four test sub-sections; no other metric discriminated between the groups. Sub-section score was

discriminant in two of the four sections. This suggests a need to include other metrics or optimise the existing methods, in order to optimise the discriminant ability of the ICS LA in Test-Mode.

### **Metric validity as an evidence-based argument**

Though *choice* of surgical instrument would seem an obvious source of error, existing assessments appear to concentrate on the physical *use*, rather than *choice* of instrument. This is reflected by the absence of correct instrument choice as a metric in both the Objective Structured Assessment of Technical Skills (OSATS), and Global Operative Assessment of Laparoscopic Skills (GOALS) assessment tools (4, 5). It would seem reasonable that the addition and increased complexity of metrics would add to the cognitive challenge/fidelity of a procedural assessment tool, by incorporation in the unified metric (total score).

### **Repetition**

This study assessed the discriminant ability of the ICS LA test-mode in individuals naïve to the module. The effect of repetition on the metrics should be assessed, in particular the rate at which (number of repeated attempts until) a perfect score is achieved, should be assessed.

### **Conclusion**

ICS LA Test-Mode are able to discriminate between experienced and inexperienced individuals. Differences in the discriminant ability of these metrics, provide an opportunity to optimise the discriminant ability of the ICS LA test-mode, and also to assist in the development of metrics for other laparoscopic ICS modules.

## References

1. Miller RD. Miller's anesthesia. Philadelphia, PA: Churchill Livingstone/Elsevier; 2010.
2. Loveday BP, Oosthuizen GV, Diener BS, Windsor JA. A randomized trial evaluating a cognitive simulator for laparoscopic appendectomy. ANZ journal of surgery. 2010;80(9):588-94.
3. IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp. Released 2013.
4. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. The British journal of surgery. 1997;84(2):273-8.
5. Niitsu H, Takiyama W, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. Surgery Today. 2013;43(3):271-5.

## Tables and Figures

Table 1. Demographics, experience and training level.

Demographics	Group 1	Group 2
Number	10	10
Age in years - median (range)	23.5 (23-26)	31.5 (26-43)
Sex - male : female	4:6	5:5
Handedness - right : left : ambidextrous	9:1:0	8:2:0
Level of surgical training		
Training level - non-training : SET 1 : >=SET 2	10 : 0 : 0	0 : 2 : 8
PGY - median (range)	0 (0)	3 (3-12)
ASSET - yes : no	0 : 10	7 : 3
Previous LA operative experience		
Number of procedures performed as primary surgeon		
0	9	1
1-4	1	1
5-29	0	0
>30	0	8
Number of procedures performed as assistant		
0	3	0
1-4	6	0
5-29	1	1
>30	0	9
Number of procedures observed		
0	1	1
1-4	0	0
5-29	8	1
>30	1	8
PGY, postgraduate year; LA, laparoscopic appendicectomy		

Table 2. Comparison of metrics between participant groups.

	Group 1 Median	Group 2 Median	Difference	P value
Total Score (%)	50	76.5	26.5	<b>0.023</b>
Total Time (sec)	451.5	372	79.5	0.179
Total Hand	0	0	0	-
Total Instrument	15	6	9	<b>0.001</b>
Total Placement	10	6	4	0.656
Total >15s	0	0	0	-
Total Score .	26	9.5	16.5	0.179
S1 Hand	0	0	0	-
S1 Instrument	3	0	3	<b>0.005</b>
S1 Placement	3.5	1	2.5	0.370
S1 >15s	0	0	0	-
S1 Sub-Score	7	1.5	5.5	0.070
S2 Hand	0	0	0	-
S2 Instrument	9.5	4.5	5	0.070
S2 Placement	4	2	2	0.656
S2 >15s	0	0	0	-
S2 Sub-Score	13	7	6	<b>0.023</b>
S3 Hand	0	0	0	-
S3 Instrument	0	0	0	-
S3 Placement	0	0	0	-
S3 >15s	0	0	0	-
S3 Sub-Score	0.5	0	0.5	0.350
S4 Hand	0	0	0	-
S4 Instrument	2.5	0.5	2	<b>0.020</b>
S4 Placement	0	0	0	-
S4 >15s	0	0	0	-
S4 Sub-Score	2.5	0.5	2	<b>0.020</b>