

3.2 Evaluation Protocol for ALL-SAFE

Internal Evaluation/Validation Work:

Prior to implementation, internal and limited external evaluations of the ALL-SAFE program and components were completed. In some cases, modifications were made based on ratings and feedback. These changes are described when relevant. The summary below describes both the internal evaluation/validation process and validity evidence which resulted from this work.

Internal evaluation methods and results are detailed for ease of reproducibility for external evaluation. We also included all assessment and evaluation forms as appendices (Appendices 2-5) to facilitate GSTC external validation.

Completed internal evaluation/validation includes:

Part A. Ease of build and attributes of the low-cost ALL-SAFE box trainer and simulator, ability to perform relevant tasks [PAGES 1-6],

Part B. Quality of measures of ALL-SAFE cognitive testing tool [PAGES 6-8],

Part C. Quality of measures of psychomotor assessment tool (ALL-SAFE VOP) [PAGES 9-14],

Part D. Ease of navigation of the web-based ALL-SAFE program [PAGES 15-16]

External validation: To be performed at the discretion of GSTC [PAGE 17].

Part A. Internal Evaluation of Ease of Build, Attributes of ALL-SAFE Box Trainer and Simulator, and Ability to Perform Relevant Tasks

Validity evidence relevant to test content is supplied (Appendix 1. ALL-SAFE Validity Evidence Relevant to Test Content).

Internal evaluation of the ALL-SAFE components included the ALL-SAFE box trainer's and the ectopic pregnancy simulator's ease of build, attributes, and their perceived value as training and testing tools (Appendix 2 and link: https://umich.qualtrics.com/jfe/form/SV_3z1H6Gkg17uweTY)

Methods. The first week of April 2021 a total of 9 participants evaluated the ALL-SAFE box trainer, and/or the ectopic pregnancy simulator. The sample consisted of 4 General Surgery attendings, 3 Obstetrics/Gynecology attendings, and 2 general surgery residents (1-PGY1, 1-PGY3). The group reported a range of |0,50|, M=9.4 (SD=17.8) laparoscopic ectopic pregnancies, and |20,1000|, M=415 (SD=387.8) laparoscopic cases.

Four participants evaluated the box trainer's ease of build, while all 9 evaluated the box trainer's characteristics, and the ectopic pregnancy simulator.

Data capture methods followed best practices identified by Cook et al [1,2]. Reporting evidence relevant to content validity followed practices identified by Gao et al [3], Orringer et al [4], and Hsiung et al [5].

ALL-SAFE Box Trainer Ease of Build (N=4)

The ALL-SAFE box trainer's ease of build was measured using 6 items, rated on true Likert scales ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), and a seventh item to capture time required to build the box trainer. Means are reported in Table 1 below for each item.

Table 1. Participants means (*SD*) for ease of build

No.	Item	Mean (<i>SD</i>)	Meet criteria? (≥ 3.0)
1.	Build instructions for the trainer box were easy to understand	4.50 (.58)	YES
2.	Materials required to build trainer box were easy to acquire	4.50 (.58)	YES
3.	Reproduced trainer box matched the intended design	4.50 (.58)	YES
4.	Trainer box is of adequately stable construction, overall	4.33 (.58)	YES
5.	Estimated time to build (minutes)	90 (42.4)	—

RESULTING MODIFICATIONS: None. Mean ratings were well over 3.0 (Agree) criteria, and there were no suggestions for improvement on the build instruction, so no action taken.

Box Trainer Attributes (N=9)

The ALL-SAFE box trainer's attributes were measured using an additional 6 items, rated on true Likert scales ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), and a seventh item to capture if participants were required to stabilize the finished box trainer. Suggestions for improvements were captured using a fill-in-the-black option. Means are reported in table below for each item.

Table 2. Participants means (SD) for ALL-SAFE box-trainer attributes' ratings

No.	Item	Mean (SD)	Meet criteria? (≥3.0)
1.	The amount of light from window/cutouts allows for adequate visualization	3.25 (.50)	YES
2.	Trainer box dimensions provide a working space which adequately represents an average pelvic cavity	2.50 (.58)	NO
3.	Front camera placement provides a view adequately comparative to that of a 30° laparoscope	2.75 (.50)	NO
4.	Top camera placement provides a view adequately comparative to that of a 0° laparoscope	2.75 (.50)	NO
5.	The trocar site placement was flexible enough to allow comfortable instrument management	2.00 (.50)	NO
6.	Trainer box of adequately stable construction, overall	3.00 (.00)	YES
7.	During testing, were you required to take action to stabilize the trainer box on the table? (Scored Yes=1, No=0)	0.50 (.58)	—

Suggestions for improvement included “Need to readjust the port sites”, which targeted making the port sites bigger to better align with real surgical experience.

A final question which asked participants to rate current version of box trainer by selecting ONE statement below:

No.	Item	Raw (%)
1.	The ALL-SAFE Trainer Box requires <i>extensive</i> adjustments before it can be considered for use in salpingostomy training.	0 (0.0%)
2.	This simulation requires <i>minor</i> adjustments before it can be considered for use in salpingostomy training.	3 (33.3%)
3.	The ALL-SAFE Trainer Box can be used in training salpingostomy <i>as is</i> , but could be improved slightly	3 (33.3%)
4.	The ALL-SAFE Trainer Box can be used in salpingostomy training with no improvements made.	3 (33.3%)

RESULTING MODIFICATIONS:

- 1) Increasing the internal scaling to improve the view
- 2) Increase scale, and placement of the “port holes” to better align with authentic surgical experience

Ectopic Pregnancy Simulator

The ALL-SAFE ectopic pregnancy was evaluated by 9 participants across 2 domains— ease of build and attributes. The simulator’s ease of build was measured via 2 items, while its characteristics were measured using an additional 11 items, rated on true Likert scales ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), and a 14th item to capture time required to build the ectopic pregnancy simulator. Means are reported in table below for each item, broken out by entire group (ALL) and Obstetric/Gynecology participants (Ob/Gyne), specifically.

Table 3. Participants means (*SD*) for ALL-SAFE ectopic pregnancy simulator ease of build and attributes

No.	Item	ALL 9 (n=9) Mean (<i>SD</i>)	Ob/Gyne (n=2) (<i>SD</i>)	Meet criteria? (≥3.0)
Ease of Build				
1.	Build instructions for the ectopic simulation were easy to understand	4.50 (0.58)	4.00 (0.00)	YES
2.	Materials required to build ectopic simulation were easy to acquire	4.50 (0.58)	4.00 (0.00)	YES
Attributes				
3.	The proportions of the simulated pelvic organs were appropriate to each other	4.14 (0.69)	3.5 (0.71)	YES
4.	The proportions of the simulated pelvic organs were appropriate to the size of the trainer box	3.17 (0.76)	4.0 (0.00)	YES
5.	The uterus, fallopian tubes, and ectopic pregnancy were uniquely identifiable	4.63 (0.52)	5.0 0 (.00)	YES
6.	The uterus, fallopian tubes, and ectopic pregnancy were located realistically in relation to one another	4.00 (1.10)	3.5 (2.12)	YES
7.	The fallopian tube could be retracted with similar force to reality	4.20 (0.45)	4.0 (0.00)	YES
8.	The tactility of the uterus was realistic	3.43 (1.27)	1.0 (0.00)	NO
9.	The tactility of the fallopian tube was realistic	3.33 (.82)	3.0 (1.41)	NO
10.	The tactility of the ectopic pregnancy was realistic	3.40 (1.52)	2.5 (2.12)	NO
11.	Amount of force required to evacuate the ectopic contents was realistic	3.80 (1.10)	3.0 (1.41)	NO
12.	The feeling of the fallopian tube (penrose drain) during cutting of salpingostomy was realistic	3.60 (1.14)	3.0 (1.41)	NO
13.	Reproduced ectopic simulation matched the intended design	4.29 (0.49)	4.5 (0.71)	YES
14.	Estimated time to build (minutes)	45 (21.21)	—	—

Captured Comments:

- a) Ectopic pregnancy, with comments that included:

“Need to have something a bit more solid for the ectopic,” “Would suggest thicker substance such as playdoh to mimic clot/tissue of ectopic,”

- b) Fallopian tube, with comments that included:

“Typically fallopian tube has more resistance than penrose, easier to cut.”

Favorable feedback included: “None - the feel of the trainer felt very conducive to educational simulation”

RESULTING MODIFICATIONS:

- 1) Assessment. Removal of item 8. The tactility of the uterus was realistic, for irrelevance (confirmed by Ob/Gyne)
- 2) Change ectopic pregnancy contents to play dough recipe
- 3) No changes to fallopian tubes as no viable alternative to penrose drain that is low cost.

Ability to Perform Tasks

Six participants rated their personal ability to perform tasks associated with the laparoscopic management of ectopic pregnancy on the ALL-SAFE simulator and trainer box. Ability was measured via 5 items (tasks), rated on 5-point rating scales, scored from 1 (Too difficult to perform), 2 (Difficult to perform), 3 (Reasonably difficult to perform), 4 (Easy to perform), to 5 (Too easy to perform), with a “Don’t know” option provided. Means are reported in table below for each item.

Table 4. Participants means (*SD*) for ability to perform tasks using ALL-SAFE box trainer and simulator

No.	Item (Task)	All (n=6) Mean (<i>SD</i>)	Ob/Gyne (n=2) Mean (<i>SD</i>)	Meet criteria? (≥2.5*)
1.	Evaluate both fallopian tubes	3.86 (0.38)	3.50 (0.71)	YES
2.	Identify ectopic pregnancy site	3.75 (0.46)	3.50 (0.71)	YES
3.	Place suture ligature on fallopian tube	2.86 (0.90)	3.00 (1.41)	YES
4.	Complete 1-2cm longitudinal incision along ectopic pregnancy site opposite to mesosalpinx	3.50 (0.53)	3.50 (0.71)	YES
5.	Complete salpingostomy without injury to ovary	2.67 (0.82)	2.00 (0.00)	YES

*Criteria for appropriateness does not follow criteria for other evidence. Variability across items (tasks) is expected and could be later used for expert/novice discrimination for items (tasks) considered more difficult by less experienced surgeons.

Value

Value of the simulation was rated using a 4-point rating scale, scored as 1 (No value/relevance), 2 (Little value/relevance), 3 (Some value/relevance, and 4 (High value/relevance) with a “Don’t know” option provided. Means are reported in table below for each item.

Table 5. Participants means (SD) for perceived value

No.	Item (Task)	All (n=8) Mean (SD)	Ob/Gyne (n=2) Mean (SD)	Meet criteria? (≥3.0)
1.	Value as a training tool	3.88 (0.35)	4.00 (0.00)	YES
2.	Value as a testing tool	3.57 (0.53)	3.00 (0.00)	YES
3.	Relevance to practice	3.00 (1.07)	4.00 (0.00)	YES

Part B. Internal Evaluation of Measures of ALL-SAFE Cognitive Testing Tool

Methods:

Study. 20 participants from 3 sites completed the web-based module. Participants included 2 Ob/Gyne attendings, 2 General Surgery attendings, 3 first-year residents, 1 second-year resident, 1 third-year resident, and 7 medical students. All participating sites were represented (Mbingo, n=3; Soddo, n=6; SIU, n=3; UM, n=8).

Scoring and Statistical analyses.

The identical (but shuffled in presentation) 10-item pre- and post-module quizzes were scored dichotomously (1=correct, 0=incorrect) and summed for each participant. Pre- and post-module summed scores were compared using paired student-test with SPSS Statistics for Windows v.25 (IBM, Armonk, NY) while differences in scores across participants' experience and site were analyzed using a many-facet Rasch model using Facets software v. 3.50 (Winsteps.com, Beaverton, OR) following anchoring on subjects to accommodate for nested design across sites.

RESULTS

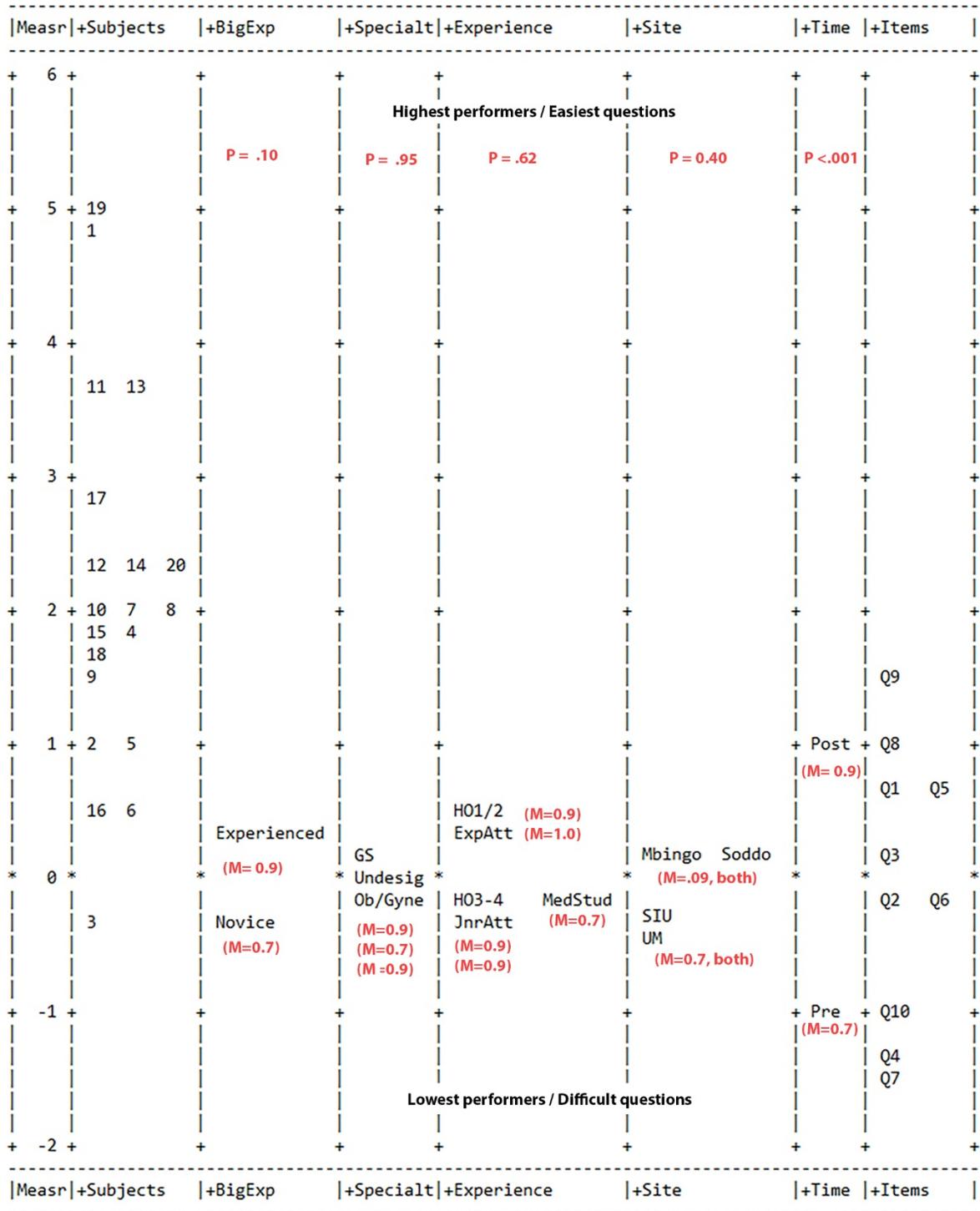
Paired Student T-Test

Comparison of pre- and post-intervention Quiz (Appendix 3) summed scores from all 20 participants indicated that there was a statistically significant improvement in mean summed scores from Pre (M=6.7, SD=2.2) to Post (M=9.0, SD = 1.5), $t(20) = -4.76$, $P < .000$.

Analysis using 7-facet Rasch model

Rasch analysis confirmed statistical improvement across Pre- and Post-module Quiz scores, $P < .001$. Analysis also indicated higher scores for experienced (HO3 and Attendings; M=0.9) over novice (M=0.7) participants, but statistical significance was not reached, $P = 0.10$ (Figure 1. Rasch Variable Map). Further, there were no difference in performance across Specialty ($P = .95$), or Site ($P = 0.40$).

Figure 1. Rasch Variable Map



Item discrimination

Review of item discrimination showed reasonable distribution of item difficulty for items, with item 7 as the most difficult item (item discrimination=.68), and Q9 as the easiest (item discrimination =1.03) (Table 6).

Table 6. Item discrimination values for ALL-SAFE cognitive test items, listed highest to lowest.

Item No.	Difficulty	Estimated Discrimination*	Discrimination Power	Suggested Action
Q4	Most difficult	1.40	Highest	Dblcheck language for ambiguity, too difficult?
Q5	Relatively easy	1.36	High	–
Q1	Relatively easy	1.29	High	–
Q6	Moderate	1.04	Good	–
Q9	Easy	1.03	Good	–
Q3	Moderate	0.99	Good	–
Q8	Easy	.92	Good	–
Q10	Moderate	.87	Acceptable	–
Q7	Most difficult	.68	Low	Consider removal/rewriting-too difficult to discriminate
Q2	Moderate	.52	Lowest	Consider rewriting-too difficult to discriminate across participants. Review of response patterns (not shown here) indicate possible guessing from lower performers (higher OutfitMnSq=1.73)

*Values over 1 indicate this item has more discrimination power than expected for its difficulty while values under 1 indicate less discrimination power for its difficulty.

Item modifications included:

- a) Modification of question 2 to avoid ambiguity
- b) No changes to questions 4/7. Discussion with team surgeon educators, indicates topic is difficult, but critical and relevant to all participating sites. Modify language to "Hasson (open technique)" and "Veress (needle insufflation)

Part C. Internal Evaluation of the Quality of Measures from Psychomotor Assessment Tool (ALL-SAFE VOP)

Validity Evidence for use of ALL-SAFE VOP with Checklist (11 items) and Global (5 items) ALL-SAFE Verification of Proficiency (Appendix 4.)

Internal Structure: Comparison of Novice v. Expert Performance Ratings

Methods: Comparison of performance scores across 14 performances (Novice, n=9; Expert, n=5) performances using Kruskal-Wallis (non-normal distribution confirmed) x 12 judges. Done in two ways; 1) *all raters* (n=12; including novices) and 2) *only experts* (n=5; only attending) raters. **Scoring:** Checklist items 1,2,4,5,8,9,10, 11 rescored max score=2. Items 3,6,7 max score=3. Global items all max score=5. Max comb sum=50.

Table 7. Comparison of mean novice v. expert performance ratings

item	Checklist item	Novice Mean (SD) Combined All Raters	Experts Mean (SD) All Raters	P-value	Novice Mean (SD) Combined Expert Raters Only	Experts Mean (SD) All Raters Expert Raters Only	P-value
1	Evaluates both fallopian tubes by pointing to both with an instrument	1.67 (.75)	1.90 (.44)	.055	1.57 (.84)	1.91 (.42)	.094
2	Stabilizes involved fallopian tube by grasping adjacent to ectopic pregnancy site	1.82 (.58)	1.93 (.36)	.225	1.91 (.42)	1.91 (.42)	.975
3	Avoids excessive grasping of fallopian tubes	2.68 (.94)	2.80 (.75)	.473	2.74 (.86)	3.00 (.00)	0.62
4	Creates a longitudinal salpingostomy	2.00 (.00)	2.00 (.00)	1.0	2.00 (.00)	2.00 (.00)	1.00
5	Extends salpingostomy to encompass length of ectopic pregnancy	1.61 (.80)	1.90 (.44)	.019	1.39 (.94)	1.91 (.42)	.025
6	Avoids transecting involved fallopian tube	3.0 (.00)	3.0 (.00)	1.0	3.00 (.00)	3.00 (.00)	1.00
7	Avoids damaging mesosalpinx when performing the salpingostomy	2.79 (.78)	2.90 (.54)	.379	2.87 (.63)	3.00 (.00)	.328
8	Evacuates at least 80% of ectopic contents from tube	1.11 (1.00)	1.43 (.91)	.091	1.04 (1.02)	1.48 (.90)	.160
9	Retrieves specimen from abdomen with laparoscopic instrument	1.71 (.71)	1.93 (.36)	.042	1.65 (.78)	2.00 (.00)	.043
10	Places single suture at marked edge of fallopian tube	1.75 (.67)	1.73 (.69)	.893	1.83 (.58)	1.83 (.58)	.963
11	Performs intracorporeal knot with a surgeon's knot followed by two additional throws	1.68 (.74)	1.97 (.26)	.007	1.65 (.78)	2.0 (.00)	.043
–	SUMMED	21.79 (2.81)	23.50 (1.88)	.001	21.65 (2.84)	24.04 (1.33)	.002

item	Global Domain	Novice Mean (SD) Combined All Raters	Experts Mean (SD) All Raters	P-value	Novice Mean (SD) Combined Expert Raters Only	Experts Mean (SD) Expert Raters Only	P-value
1	Respect for Tissue	3.22 (1.01)	3.89 (.96)	<.01	3.39 (1.16)	4.09 (.79)	.039
2	Economy of Time and Motion	2.78 (1.18)	3.63 (1.03)	<.01	2.65 (1.27)	3.48 (1.04)	.041
3	Instrument Handling	2.79 (1.21)	3.72 (1.06)	<.01	2.65 (1.27)	3.74 (1.01)	.004
4	Flow of Operation	3.37 (1.04)	4.14 (.92)	<.01	3.26 (1.14)	4.09 (.95)	.016
5	Overall Performance	2.59 (1.01)	3.42 (1.04)	<.01	2.61 (1.20)	3.57 (.99)	.006
–	GLOBAL SUMMED	14.75 (4.83)	18.79 (4.51)	<.01	14.57 (5.27)	18.96 (4.16)	.006
–	TOTAL SUMMED	36.79 (6.67)	42.13 (5.37)	<.01	36.22 (7.47)	43.00 (4.46)	.001
	Final Rating	2.30 (.78)	2.75 (.51)	<.01	2.22 (.85)	2.83 (.49)	.006

Summary. Findings suggest that comparison of novice /expert performance ratings at the item-level was not helpful to discriminate performance levels. In spite of this, summed total of the checklist did discriminate novice and expert performances, regardless of judge expertise. Further, global ratings at the domain-level were able to discriminate novice versus expert performances, as were the global summed and total (checklist and global summed). This suggests that the summed checklist and global ratings could be used to discriminate novice and expert performances.

Supplemental Analyses:

Many-facet Rasch model which examined ratings differences using a 6-facet Rasch model (ID x Subject Expertise x Institution X Judge Expertise x Final Rating x Item) indicated significant ratings differences across *Subject Expertise* facet, with Novice subjects' performance ratings (M=3.00) statistically lower than Expert subjects' performance ratings (M=3.7), $\chi^2 (2,114)= 96.9, p=.001$, suggesting ratings were able to discriminate between Novice and Expert performances.

The same Many-facet Rasch Model was used to examine ratings differences across *Final Rating* (e.g. Competent, Borderline, and Not Competent) response options. This analysis indicated statistical ratings differences across *Final Ratings*, shown below;

Competent (M=4.0) → Borderline (M=2.5) → Not Competent (M=2.1)

$\chi^2 (2,114)= 455.5, p=.001$, suggesting that these three response options could adequately discriminate subjects.

Internal Structure: Rater agreement across novice and expert judges

Description: Review of inter-rater agreement of 10 performances* across Novice (n=7) and Expert (n=5) raters, measured by averaged two-way mixed Intraclass correlation

(*selected for completeness of data, represented at least 3 from each site)

Table 8. Inter-rater reliability by domain

item	Domain	ICC	95% Confidence Interval
Checklist			
–	Checklist Summed	.96	.85 - .95
Global			
1	Respect for Tissue	.90	.70 - .90
2	Economy of Time and Motion	.90	.83 - .95
3	Instrument Handling	.90	.82 - .94
4	Flow of Operation	.89	.80 - .94
5	Overall Performance	.77	.58 - .87
–	GLOBAL SUMMED	.93	.88 - .96
–	TOTAL SUMMED	.95	.91 - .97
–	Final Rating	.88	.79 - .94

Summary: There were a good amount of rater agreement across novice and expert judges The lowest ICC value was estimated to be .77, for *Global-Overall Performance* item, suggesting moderate agreement between novice and expert judges for that item. Remaining ICC values ranged between .88 and .96, suggesting excellent interrater agreement for those items.¹

Supplemental Analyses: Many-facet Rasch model which examined ratings differences using a 6-facet Rasch model (ID x Subject Expertise x Institution X Judge Expertise x Final Rating x Item) indicated no statistical ratings differences across *Judge Expertise* facet, with Trainee ratings (M=3.4) not statistically different than Expert ratings (M=3.4), $X^2(2,114) = 1.9, p = .16$.

1. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016; 15(2):155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.

Internal Structure: Comparison of performance ratings across 3 sites

Description: Comparison of ranked ratings of same 10 performances across 3 sites (UMichigan, Mbingo, and Soddo) using Kruskal-Wallis test to ensure generalizability of scoring across sites

Table 9. Comparison of median performance ratings across 3 participating sites

item	Domain	Mbingo (n=2) Mdn, Range	Soddo (n=7) Mdn, Range	UMichigan (n=5) Mdn, Range	P- Value
Checklist					
–	Checklist Summed	23.00, 19,25	23.00, 16,25	24.00, 14,25	.375
Global					
1	Respect for Tissue	4.0, 2,5	3.0, 2,5	4.0, 1,5	.128
2	Economy of Time and Motion	3.5, 1,5	3.0, 1,5	3.5, 1,5	.917
3	Instrument Handling	4.0, 1,5	3.0, 2,5	3.0, 1,5	.675
4	Flow of Operation	5.0, 4,5	4.0, 2,5	3.5, 1,5	.002
5	Overall Performance	3.5, 2,5	3.0, 1,5	3.0, 1,5	.162
–	GLOBAL SUMMED	20.0, 10,25	16.0, 7,25	17, 5,25	.268
–	TOTAL SUMMED	43.0, 29,50	39.0, 24,50	39.0, 20,50	.199
–	Final Rating	3.0, 1,3	3.0, 1,3	3.0, 1,3	.568

Summary: This supported preliminary evidence of generalizability of scoring across participating sites.

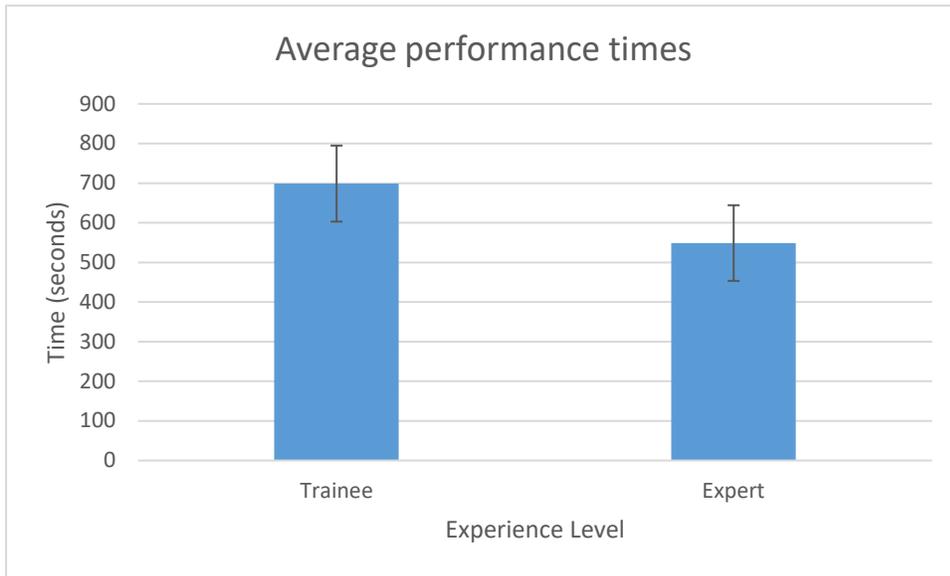
A Kruskal-Wallis test was conducted to evaluate scoring differences among the three sites (Mbingo, Soddo, and University of Michigan) for the Summed Checklist, each of the 5 Global items (domains), the Global summer score, total summed score (combines the checklist summed score and the Global summed score), and finally, the final overall rating. For all but one item (Global rating #4 Flow of operation), the test, which was corrected for tied ranks, was not significant, $p = |.128, .568|$, suggesting the proportion of variability in the ranked dependent variable (score) was not accounted for by participating sites, indicating a little relationship between site and the test scores. This supported preliminary evidence of generalizability of scoring across participating sites.

Supplemental Analyses: Many-facet Rasch model which examined ratings differences using a 6-facet Rasch model (ID x Subject Expertise x Institution X Judge Expertise x Final Rating x Item) indicated statistical ratings differences across *Institutions* with Mbingo (M=3.6) having statistically higher ratings than U Michigan (M=3.2) and Soddo (M=3.3), $\chi^2(2,114) = 74.8, p=0.001$, suggesting rater training will be important to ensure rating calibration across institutions.

Internal Structure: Comparison of Novice v. Expert Performance Times

Description: Comparison of performance times across Novice (n=5) and Expert (n=5) performances using independent t-test (normality confirmed using both Kolmogorov-Smirnov and Shapiro-Wilk tests, $p \geq 0.20$).

Figure 2. Comparison of performance times across trainees and experts



Summary. Although average Trainee times were higher than Expert times, findings suggest that comparison of novice /expert performance times for this cohort were not helpful at discriminating performance levels.

$M_{\text{Novice}} = 698.80$ seconds ($SD=214.6$)

$M_{\text{Expert}} = 548.80$ seconds ($SD=213.5$)

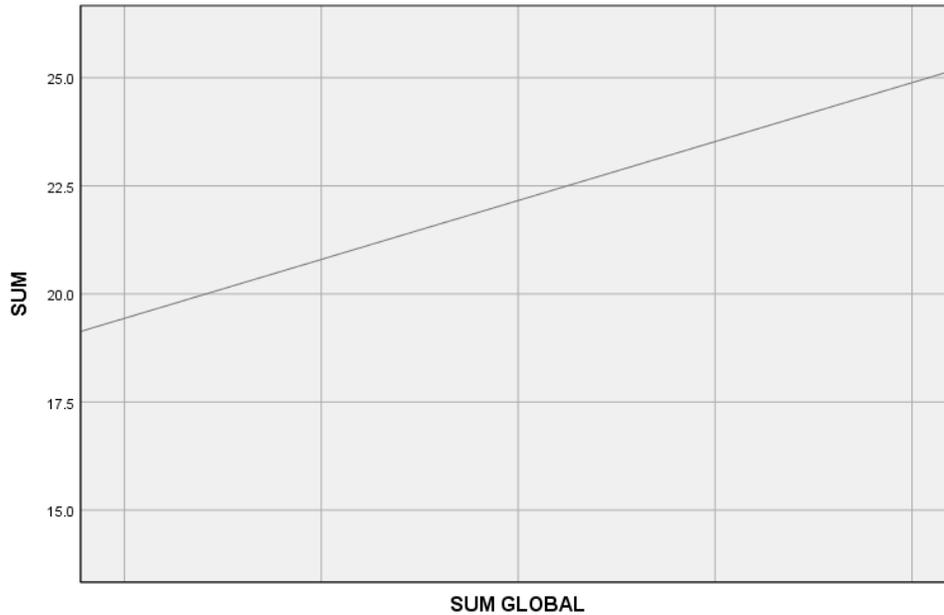
$P = .30$ (NOT STATISTICALLY SIGNIFICANT)

This exercise pointed out an issue with the checklist, along with a few considerations re: timing. These are listed below, with potential solutions.

1. Many examinees did not cut suture at end of task, making it difficult to tell when task was done.
Solution: Add “Cut suture tail on competing knot” to checklist
Justification: Allows hard time stamp to be considered (from first touch of fallopian tube to cutting of suture)
*For the sake of this exercise, I timed all from first touch to cinching of knot for consistency.
Many examinees did not cut suture at end of task, making it difficult to tell when task was done.
2. Many examinees left substantial amount of ectopic contents.
Could not adequately visualize to estimate amount (%) left, compromising associated item on checklist, and made performance times shorter for those examinees that chose not fully empty contents (biased benefit)
Solution: Should consider adding time penalties for this (do not have good example form this sample, so can't estimate time penalty)
justification: FLS standard does this.
3. Some examinees did not examine fallopian tube.
Not having a clear “start time” makes it difficult to measure, and erroneously shortens time for those that did not perform this task (biased benefit)
Solution: Add time penalty (average time was 5-7 seconds in sample)

Relationships to Other Variables: Correlation of summed ALL-SAFE checklist scores with OSATS summed scores

Figure 3. Graph illustrating Correlation of summed checklist scores with OSATS scores (n=10) estimated by Pearson's r.



Summary. Findings suggest a positive correlation between summed ALL-SAFE Checklist score (SUM) and established OSATS summed score (SUM GLOBAL), $r(114) = .534$, $p=.0001$, supporting use of ALL-SAFE checklist summed score to measure performance skill.

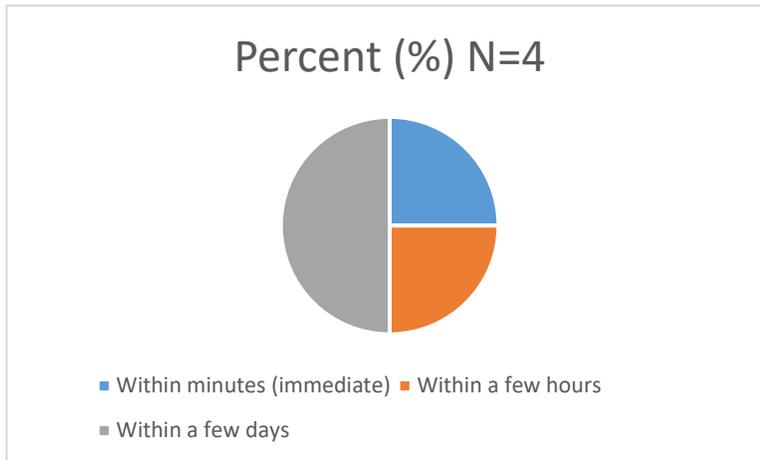
Similarly, these summed scores (SUM) correlated with the Combined summed score of the Checklist and OSATS, $r(114) = .96$, $p=.0001$), as well as the Final Rating, scored on 3-point scale (1 = Does not demonstrate competence, 2 = Borderline, 3 = Does demonstrate competence), $r(114)=.76$, $p = .0001$.

Part D. Internal Evaluation of Ease of Navigation of the Web-Based ALL-SAFE Program

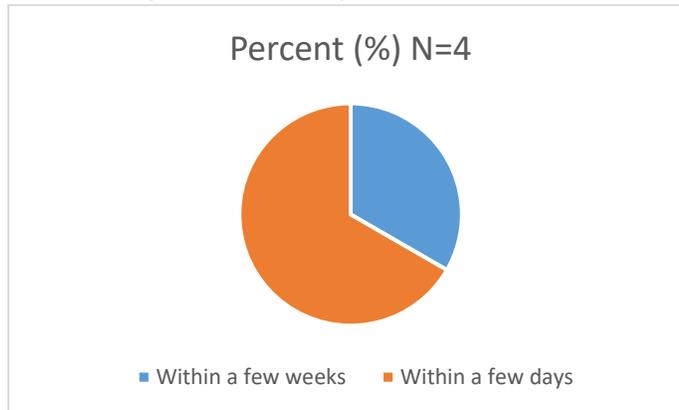
Methods. On October 11, 2021 a recruitment email was sent to a University of Michigan medical student surgery interest group and Mbingo residents who have not previously participated in the development of the ALL-SAFE program, with a request to evaluate the ALL-SAFE computer-based program. Four participants evaluated the site using an 11-item survey, with mixed format items (Appendix 4). The sample consisted of 4 medical students unfamiliar with the ALL-SAFE program, two from US and 2 from Mbingo participant sites. Summary of reporting follows each item of evaluation.

1. **Was this exercise a good use of your time?**
100% selected Yes.
2. **How likely are you to recommend ALL-SAFE to your colleagues? Please rate your agreement using the scale 0 (Not at all) to 10 (Extremely likely).: Reported M=9.5 (SD=1.0)**
3. **Is the site navigation clear and easy-to-follow? Yes/No (please select best response):** 100% selected Yes.
4. **Please write down where and how site navigation can be improved using the space provided:**
Comments included: "Nothing comes to mind," "The videos take a lot of time to play even on the best of internet network, this should be looked into," "Please you have to work the video uploading section," "would recommend the ability to click through each section before doing the one prior to it. Also the timing estimates were not accurate from my experience completing it."
5. **Are the instructions and language clear and easy-to-follow? Yes/No:** 100% selected Yes.
6. **Please write down where and how the instructions and language can be improved using the space provided:**
Comments included: "For the build: create printable templates that can be printed on 8.5x11" sheets," "For the uterus: recommend cling wrap over saran wrap, as well as taping the plate to the table," "They are very okay as at now,"
"In the construction of materials. Include specifics on exact measurements/types of materials and a print-version of materials list would be helpful so a participant could just take the list to their local supply area and get them all at once without clicking through the actual module. Including this in the beginning as some kind of "you will need these materials before beginning" would be helpful."
7. **Does this module seem culturally appropriate for the country in which you practice medicine? Please explain any improvements that could be made.**
Comments included: "Yes, it is" "The module is appropriate. The only difficulty is the availability of instruments for laparoscopic surgery. I don't know there is way to make those available in order to facilitate practice," "Yes, from USA. However have worked in Liberia and I don't believe all the construction and module materials like laproscopic tools, penrose drain, cardboard, are all available."

8. What is your preferred timeframe for receiving feedback? (Please select best choice):



9. What is the longest timeframe you'd feel comfortable with for receiving feedback?



10. Please provide any other feedback you may have in the space provided below.

"Great job!" and "None for now"

11. If you'd like to be available for additional questions, please provide your email:

all 4 participants supplied, not included in this report

ALL-SAFE Documents for External Evaluation:

Internal evaluation methods and results are detailed for ease of reproducibility for external evaluation. To facilitate GSTC external validation, we also provided all assessment and evaluation forms that were used in internal evaluation. These include:

Appendix 2 Evaluation form. Ease of build and attributes of the low-cost ALL-SAFE box trainer and simulator, ability to perform relevant tasks for evaluation of evidence of test content and feasibility of use/build of ALL-SAFE components found within the web-based ALL-SAFE program (Support documents available via the Appropedia site)

Appendix 3 Cognitive test. 10-item multiple choice knowledge test for evaluators' review of relevance to training and alignment with ALL-SAFE Assessment blueprint (Appendix 6)

Appendix 4 Technical skills assessment form. 17-item checklist/global rating ALL-SAFE Laparoscopic Salpingostomy: Psychomotor Skills Assessment (VOP) for external evaluation of psychomotor assessment tool (ALL-SAFE VOP)

and

Appendix 5. Evaluation form. 11-item ALL-SAFE web-based program for external evaluation of ease of navigation and clarity of language of the web-based ALL-SAFE program (<https://www.allsafe.education/>). **External evaluators please select "3rd Party Test Course-Fully functional course for GSTC Testers"**

REFERENCES

1. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*, 2014 May;19(2):233-50.
2. Cook DA, R, Zendejas B, SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013 Jun;88(6):872-83.
3. Gao RW, Rooney D, Harvey R, Malloy KM, VanKoeveering KK. To Pack a Nose: High-Fidelity Epistaxis Simulation Using 3D Printing Technology. *Laryngoscope*, 2021 Jul;00:1-7.
4. Orringer MB, Hennigar D, Lin J, Rooney DM. A novel cervical esophagogastric anastomosis simulator. *J Thorac Cardiovasc Surg* 2020 Dec;160(6):1598-1607.
5. Hsiung GE, Schwab B, O'Brien EK, Gause CD, Hebal F, Barsness KA, Rooney DM. Preliminary Evaluation of a Novel Rigid Bronchoscopy Simulator. *J Laparoendosc Adv Surg Tech A*. 2017 Jul;27(7):737-743.

Appendix 1. ALL-SAFE Validity Evidence Relevant to Test Content

Analyses and summary by DMRooney, PhD

Full description of development process of associated curriculum materials and assessment tools to ensure transparency, and initial validation processes used with summary of findings

Materials: ALL-SAFE Box Trainer

1. Development of ALL-SAFE box trainer was done by team of 3 Engineer students (3DI Lab, UMich) following parameters of PI
(material = cardboard, functions and approximate dimensions of avg lap. pelvis/abdomen field)
2. Initial prototype trialed by PI/CO-Is (Kim, Barnard, Jeffcoach, Snell) for ease of build, stability, lighting.
3. Expanded trial: 9 laparoscopic surgeons (2 Gen Surgery residents, 4 General Surgery attendings, 3 Ob/Gyne attendings) trialed the box. Four participants evaluated the box trainer's ease of build, while all 9 evaluated the box trainer's characteristics, and the ectopic pregnancy simulator, using 2 checklists;
 - a. Ease of build (5-item 5-point rating scales)
- Ease of Build M ratings range = |4.33, 4.50|, with no suggestions for improvements
 - b. Box Trainer Attributes (6-item 5-point rating scales)
- Box Trainer Attributes ratings range = |2.50, 3.25|, with suggestions for improvements including
"Need to readjust the port sites", which targeted making the port sites bigger to better align with real surgical experience

ACTIONS TAKEN:

- 1) Increasing the internal scaling to improve the view
- 2) Increase scale, and placement of the "port holes" to better align with authentic surgical experience

ALL-SAFE Uterus/Ectopic Pregnancy Simulator

1. Development of simulated uterus/ectopic pregnancy was done CO-I (Barnard)
(parameters = representative of relevant anatomy, low-cost, accessible materials)
2. Initial prototype trialed by PI/CO-Is (Kim, Barnard, Jeffcoach, Snell)
3. Expanded trial: 9 participants (2 Gen Surgery residents, 4 General Surgery attendings, 3 Ob/Gyn attendings) trialed the simulated ectopic pregnancy. All participants evaluated the associated build directions (ease of build) and the simulator's characteristics, using 3 checklists;
 - a. Ease of build (5-item 5-point rating scales)
- Ease of Build M ratings range = |4.00, 4.50|, with no suggestions for improvements
 - b. Ability to Perform Task (5-item 5-point rating scales)
- Ability M rating range = |2.00, 3.86|

ALL-SAFE Uterus/Ectopic Pregnancy Simulator, continued.

- c. Simulator Attributes (13-item 5-point rating scales)
- Box Trainer Attributes ratings range = |1.00, 5.00|, with suggestions for improvements targeted 2 primary areas;

- a) Ectopic pregnancy, with comments that included:
 “Need to have something a bit more solid for the ectopic,”
 “Would suggest thicker substance such as playdoh to mimic clot/tissue of ectopic”
- b) Fallopian tube, with comments that included:
 “Typically fallopian tube has more resistance than penrose, so penrose was easier to cut”

ACTIONS TAKEN:

- 1) Change ectopic pregnancy contents to play dough recipe
- 2) No changes to fallopian tubes as no viable alternative to penrose drain that is low cost was identified

Cognitive Assessment Tool: Pre/post Multiple Choice Quiz

- 1. Originally drafted by PGY1 resident (Hsu, UMich), based on the case scenario
- 2. Reviewed and edited for relevance/content by PI (Kim)
- 3. Reviewed by 2 M4 students-research assistants (Mott/ Ryder) for clarity
- 4. Reviewed by psychometrician and PGY2 resident (Rooney, Obayemi) for stem clarity, response option bias, for content clarity, relevance, and flow.
- 5. Disseminated to entire research team for review
- 6. Final review and approval by PI (Kim)

Case Scenario and Associated Questions

- 1. Originally drafted by a PGY2 resident (Freneh, Soddo Ethiopia) to ensure relevance to targeted learners
- 2. Reviewed by 2 Co-Is (Jeffcoach, GenSx; Marzano, OBGyn; UM) for content and relevance
- 3. Reviewed by 2 M4 students-research assistants (Mott/ Yoonhee, UM) for clarity, content and flow
- 4. Reviewed by a native English speaker copy-editor for clarity and grammar
- 5. Reviewed by psychometrician and (Rooney, Obayemi, UM) for stem clarity, response option bias, for content clarity, relevance, alignment of questions with scenario, and flow, based on 7/17 Self-assessment workshop (O’Keeffe; Royal College of Surgeons in Ireland)
- 6. Disseminated to entire research team for review
- 7. Final review and approval by PI/Co-I (Kim, UM; Barnard, SIU)

Psychomotor Assessment Tool: ALL-SAFE Performance Assessment (Verification of Proficiency)

- 1. Drafted by CO-I (Barnard)
- 2. Reviewed by entire research team for content and relevance
- 3. Dissemination to sites for trial of practical use with performance videos (change language form salpingectomy to salpinostomy)
- 4. Disseminated to entire research team for review
- 5. Review with Global Surgical Training Challenge assessment expert (Dara O’Keeffe, Royal College of Surgeons in Ireland), 6/17/21
- 5. Review with ALL-SAFE research team to discuss proposed changes based on 6/17 suggestions

6. Edited by PI/Co-I to split 1 item (item 3), and add 3 additional “error-based items,” and split of final designation to “Competent, Borderline, Not competent “
7. Review by psychometrician (Rooney) for clarity, relevance, alignment of questions with skills
8. Captured data with 10 performances (2N/2E from 3 sites (Soddo, Mbingo, UM) x judges/raters, indicated need for additional item “Cut suture” to allow an actual end time to be observable.

Appendix 2. ALL-SAFE Box Trainer and Simulator Evaluation

Student : _____

Level (please circle): Student PGY- 1 2 3 4 5 6 7 Fellow Attending

Specialty (please circle): N/A General Surgery Ob/Gyn

Your institution: _____

Approximately how many laparoscopic salpingostomy procedures have you performed? _____

Approximately how many laparoscopic cases have you performed? _____

This trainer box and associated task trainer are designed to provide surgical learners with new safe opportunities for laparoscopic salpingostomy training.

Please take a moment to complete the evaluation below. Your feedback on both the trainer box and the simulator will guide their improvements. Please answer all questions you are able to answer and are applicable to your participation.

Trainer Box and Ectopic Build	Don't Know (.)	Strongly Disagree (1)	Disagree (2)	Not Disagree, nor Agree (3)	Agree (4)	Strongly Agree (5)	Relative to Reality
Build instructions for the trainer box were easy to understand							N/A
Materials required to build trainer box were easy to acquire							N/A
Reproduced trainer box matched the intended design							
Trainer box is of adequately stable construction, overall							
Build instructions for the ectopic simulator were easy to understand							
Materials required to build ectopic simulation were easy to acquire							
Please leave comments/suggestions for improvements to the <i>Box trainer and simulator build</i> :							
What method did you use to stabilize the trainer box on the table (please select all that apply) TAPE WEIGHTS LAPTOP OTHER:							
Time to build ectopic simulator did not exceed 1 hour (please provide estimated time to build ectopic simulation): _____)							

Box Trainer Attributes	Don't Know (.)	Strongly Disagree (1)	Disagree (2)	Not Disagree, Nor Agree (3)	Agree (4)	Strongly Agree (5)	
The amount of light from window/cutouts allows for adequate visualization							
Box trainer dimensions provides a working space which adequately represents an average pelvic cavity							
Front camera placement provides a view adequately comparative to that of a 30° laparoscope							
Top camera placement provides a view adequately comparative to that of a 0° laparoscope							
The trocar site placement was flexible enough to allow comfortable instrument management							
Trainer box is of adequately stable construction, overall							
During testing, were you required to take action to stabilize the trainer box on the table (please circle)		YES	NO				
CONDITIONAL: What did you use to stabilize the box trainer?							
Please leave comments/suggestions for improvements to the <i>Box trainer attributes</i> :							

	Don't Know (.)	Strongly Disagree (1)	Disagree (2)	Not Disagree, nor Agree (3)	Agree (4)	Strongly Agree (5)	
Simulator Physical attributes							
The proportions of the simulated pelvic organs were appropriate to each other							
The proportions of the simulated pelvic organs were appropriate to the size of the box trainer							
The uterus, fallopian tubes, and ectopic pregnancy were uniquely identifiable							
The uterus, fallopian tubes, and ectopic pregnancy were located realistically in relation to one another							
The fallopian tube could be retracted with similar force to reality							

	Don't Know (.)	Strongly Disagree (1)	Disagree (2)	Not Disagree, nor Agree (3)	Agree (4)	Strongly Agree (5)	
The tactility of the fallopian tube was realistic							
The tactility of the ectopic pregnancy was realistic							
Amount of force required to evacuate the ectopic contents was realistic							
The feeling of "fallopian tube" (penrose drain) during cutting of salpingostomy was realistic							
Reproduced ectopic simulator matched the intended design							
Please leave comments/suggestions for improvements to the <i>Ectopic pregnancy simulator</i> attributes:							
Please rate your ability to perform the below tasks on the simulator	Don't Know (.)	Too difficult to perform (1)	Very difficult to perform (2)	Reasonably difficult to perform (3)	Very easy to perform (4)	Too easy to perform (5)	
Evaluate both fallopian tubes							
Identify ectopic pregnancy site							
Places suture ligature on mesosalpinx laparoscopically							
Completes 1-2 cm longitudinal incision along ectopic pregnancy site opposite to mesosalpinx							
Completes salpingostomy without injury to ovary							
Retrieves specimen from "abdomen" with grasper							
Please leave comments/suggestions for improvements of the simulator that might improve your ability to perform tasks:							
		Don't Know (.)	No value (1)	Little value (2)	Some value (3)	High value (4)	
Please rate the value of the simulation as a <i>training</i> tool							
Please rate the value of the simulation as a <i>testing</i> tool							

		Don't Know (.)	No relevance (1)	Little relevance (2)	Some relevance (3)	Very relevant (4)	
Please rate the relevance of this simulation to your practice							

TRAINER BOX:

Global- Please check the *one* statement below with which you most agree.

-
- The ALL-SAFE Trainer Box requires **extensive improvements** before it can be considered for use in salpingostomy training.
 - The ALL-SAFE Trainer Box requires **minor adjustments** before it can be considered for use in salpingostomy training.
 - The ALL-SAFE Trainer Box can be used in training salpingostomy as is, but **could be improved slightly**.
 - This simulator can be used in salpingostomy training with no improvements made.
-

ECTOPIC PREGNANCY SIMULATION:

Global- Please check the *one* statement below with which you most agree.

- This simulator requires **extensive improvements** before it can be considered for use in salpingostomy training.
- This simulator requires **minor adjustments** before it can be considered for use in salpingostomy training.
- This simulator can be used in training salpingostomy as is but **could be improved slightly**.
- This simulator can be used in salpingostomy training with no improvements made.

Please leave comments in space provided below:

MODIFIED FROM:

Seagull FJ, Rooney DM. Filling a void: developing a standard subjective assessment tool for surgical simulation through focused review of current practices. *Surgery*. 2014 Sep;156(3):718-22. doi: 10.1016/j.surg.2014.04.048.

Barsness KA, Rooney DM, Davis LM, O'Brien E. Preliminary evaluation of a novel thoracoscopic infant lobectomy simulator. *J Laparoendosc Adv Surg Tech A*. 2015 May;25(5):429-34. doi: 10.1089/lap.2014.0364. Epub 2014 Dec 23.

Appendix 3. ALL-SAFE Cognitive Test Items

*Correct response. Items and response options are randomly presented across pre- and post-treatment

1. Which of the following patient criteria would require surgical management over medical management in the treatment of ectopic pregnancy?

- A. Patient is hemodynamically stable
- B. Patient has a serum hCG level of 4000 IU/L
- *C. Patient has a tubal diameter of 5 cm
- D. Patient is able to follow up reliably for continued monitoring and observation

2. Which of the following is true of laparoscopic salpingostomy compared to open salpingostomy?

- *A. Lower overall cost for the patient
- B. Higher operative blood loss
- C. Longer hospital stay
- D. Lower anesthetic risk
- E. Longer operative time

3. In which of the following situations is open salpingostomy most strongly preferred over laparoscopic salpingectomy?

- A. Younger patient
- B. Ruptured ectopic pregnancy
- *C. Hemorrhagic shock
- D. History of pelvic surgery

4. An advantage of using the Veress (needle insufflation) technique over the Hasson (open) technique to enter the abdomen is:

- *A. Shorter operative time
- B. Direct visualization of all layers of the abdominal wall
- C. Lower risk of major vascular and viscus injury
- D. Ease of use compared to Hasson technique in patients with significant adhesions from prior surgery

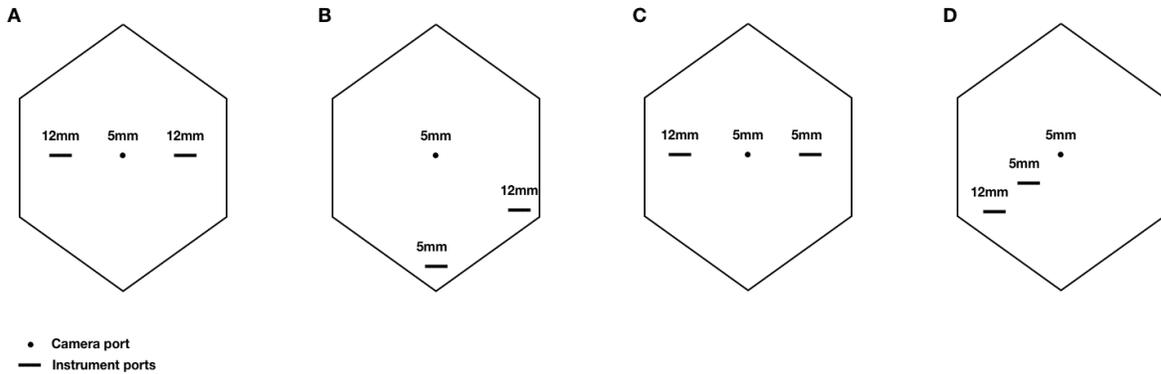
5. What is the standard pressure of pneumoperitoneum for a standard transabdominal laparoscopic procedure?

- A. 5 mmHg
- B. 10 mmHg
- *C. 15 mmHg
- D. 20 mmHg

6. During entry of the abdomen with a trocar, the patient sustains a large vessel laceration. The best next step is to:

- A. Insert a second trocar to attempt to control the bleeding laparoscopically
- B. Request angioembolization of the vessel
- C. Increase pneumoperitoneum to 30 mmHg
- *D. Immediately convert to open surgery

7. Which of the following options is NOT an effective port configuration for laparoscopic treatment of an ectopic pregnancy in the right tube?



- A
- B
- C
- *D

8. During the surgery, you notice mild bleeding from the mesosalpinx. Of the following, the best next step is to:

- A. Administer vitamin K
- B. Observe the bleeding, and if it continues, convert to open surgery
- *C. Apply an atraumatic laparoscopic grasper to the site of the bleeding for hemostasis
- D. Increase pneumoperitoneum pressure to 20 mmHg

9. After evacuating the ectopic pregnancy and inspecting for hemostasis, the best next step to remove trocars and instruments from the abdomen is:

- A. Evacuate pneumoperitoneum by direct suction on the trocar
- B. Remove all instruments, followed by the camera, followed by the trocars
- *C. Remove all instruments, followed by the trocars, followed by the camera

10. Which of the following is the correct method of closing 5 mm ports?

- A. Closure of skin with tape
- B. Closure of skin with suture
- *C. Either method is permissible

Appendix 4. ALL-SAFE Laparoscopic Salpingostomy: Psychomotor Skills Assessment (VOP)

Tasks		Y	N	Comments	
Evaluates both fallopian tubes by pointing to both with an instrument					
Stabilizes involved fallopian tube by grasping adjacent to ectopic pregnancy site					
Avoids excessive grasping of fallopian tubes					
Performs longitudinal salpingostomy					
Extends salpingostomy to encompass length of ectopic pregnancy					
Avoids transecting involved fallopian tube					
Avoids damaging mesosalpinx when performing the salpingostomy					
Evacuates at least 80% of ectopic contents from tube					
Retrieves specimen from abdomen with laparoscopic instrument					
Places single suture at marked edge of fallopian tube					
Performs intracorporeal knot with a surgeon's knot followed by two additional throws					
Global					
Respect for Tissue	1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue/occasionally caused inadvertent damage	4	5 Consistently handled tissue appropriately with minimal damage
Economy of Time and Motion	1 Many unnecessary / disorganized movements	2	3 Organized time / motion, but some unnecessary movements	4	5 Economy of movements and maximum efficiency
Instrument Handling	1 Repeatedly made tentative or awkward moves with instruments	2	3 Competent use of instruments/occasionally appeared stiff or awkward	4	5 Fluid moves with instruments and no awkwardness
Flow of Operation	1 Frequently stopped operating and seemed unsure of next move	2	3 Demonstrated ability for forward planning with steady progression of operative procedure	4	5 Obviously planned course of operation with effortless flow from one move to next
Overall Performance	1 Undeveloped	2 Developing	3 Competent	4 Skilled	5 Master
Final Rating: Does this participant demonstrate competency (<i>choose only one</i>)? Yes Borderline No		Other Summative Comments:			

Appendix 5. ALL-SAFE Web-based Program Evaluation

We would like you to complete the entire ALL-SAFE program found via the link: (<https://www.allsafe.education/>). Following completion, please answer the questions below. We appreciate your feedback.

1. Was this exercise a good use of your time? (please select one response): Yes No
2. How likely are you to recommend ALL-SAFE to your colleagues? Please rate your agreement using the scale 0 (Not at all) to 10 (Extremely likely).: _____
3. Is the site navigation clear and easy-to-follow? (please select best response): Yes No
4. Please write down where and how site navigation can be improved using the space provided:
5. Are the instructions and language clear and easy-to-follow? Yes No
6. Please write down where and how the instructions and language can be improved using the space provided:
7. Does this module seem culturally appropriate for the country in which you practice medicine? Please explain any improvements that could be made in this regard using space provided.
8. What is your preferred timeframe for receiving feedback? (Please select best choice):
 - a. Within minutes (immediate)
 - b. Within a few hours
 - c. Within a few days
 - d. Within a few weeks
9. What is the longest timeframe you'd feel comfortable with for receiving feedback?
 - a. Within minutes (immediate)
 - b. Within a few hours
 - c. Within a few days
 - d. Within a few weeks
10. Please provide any other feedback you may have in the space provided below.
11. If you'd like to be available for additional questions, please provide your email:

Appendix 6. ALL-SAFE BLUEPRINT

Instructional Objectives/Goals:

- To improve the surgeons' confidence and competence in laparoscopic management of ectopic pregnancy
- To support the development of basic cognitive and technical skills required for the safe laparoscopic management of ectopic pregnancy
- To provide surgeons with standardized educational content on the laparoscopic management of ectopic pregnancy to assist in development of relevant clinical judgment
- To provide a formative assessment framework and objective quantifiable measures to assess knowledge, judgment, and manual skills in the laparoscopic management of ectopic pregnancy
- To promote collaboration between countries of diverse resource levels to provide low cost, portable educational and assessment framework to improve quality of surgical care

To meet the targeted instructional objectives we have developed the ALL-SAFE training program with the following specific learning objectives across three domains. We will evaluate the impact of the ALL-SAFE training program using Kirkpatrick's training evaluation model.¹ The Kirkpatrick Model is an internationally recognized method of evaluating the outcomes of training programs on four distinct levels of criteria: reaction, learning, behavior, and results. In this case, each level presents an order of potential impact the educational program may have on learners' abilities and the system in which that learner performs. In our setting, the *Reaction* level (Level 1) evaluates the attitude of the learner towards the ALL-SAFE program; *Learning* level (Level 2) evaluates the knowledge and skills achieved by the learners; the *Behavior* level (Level 3) measures the adoption of learners' new knowledge and skill in the operating room; and the *Results* level (Level 4) measures the large-scale impact of the training program on patient care.

The ALL-SAFE Blueprint on the following 3 pages identifies the specific measures which will be used to evaluate the program's impact on the specific learning outcomes, which correlate with Levels 1 and 2 of Kirkpatrick's Model. The Blueprint follows Miller's Pyramid of Assessment^{a-c}, which provides a framework for assessing clinical competence in medical education by matching learning outcomes (observable behaviors) with the expected level of learner skill along the continuum of their training (e.g. Low, Mid, High).

¹Kirkpatrick, D. L. (1998). Evaluating Training Programs: the Kirkpatrick model and the Four Levels (2nd Ed.). San Francisco, Berrett-Koehler.

ALL-SAFE Blueprint

Knowledge – Cognitive Domain ^a			
Domain (KSA) Miller's Level (relative)	Learning Objective	Outcome	Content Source
Low (Comprehension)	Identify diagnostic ectopic pregnancy	Case Scenario items 1-3	Case Scenario
Low (Comprehension)	Compare and contrast different treatment options for ectopic pregnancy	Correct Quiz items 1-3	Case Scenario, Quiz
Low (Comprehension)	Compare differences in common laparoscopic techniques used to access the abdomen	Correct Quiz item 4	Case Scenario, Quiz
Mid (Application)	Employ relevant knowledge of MIS technologies to optimize best visual field during laparoscopic management of ectopic pregnancy	Correct Quiz items 5,7	Case Scenario, Quiz
High (Synthesis)	Initiate appropriate clinical judgement when presented with a common complication during a clinical scenario	Correct Quiz items 6,8	Case Scenario, Quiz
High (Synthesis)	Initiate appropriate management of equipment and instruments on completion of laparoscopic management of ectopic pregnancy	Correct Quiz item 9	Case Scenario, Quiz
Mid (Application)	Select appropriate methods of port closure	Correct Quiz item 10	Case Scenario, Quiz
Mid (Strategic thinking) High (Extended thinking)	Using the provided ALL-SAFE assessment checklist and global ratings, accurately critique and assess other learners' technical competency at managing laparoscopic ectopic pregnancy in simulated setting	Completed checklist and global ratings for at least 3 number of peer learners	On-line assessment (checklist, Global)
Mid (Strategic thinking) High (Extended thinking)	When indicated, using the All-SAFE checklist and global ratings, provide written feedback and recommended suggestions for performance improvements to other learners regarding their technical competency at managing laparoscopic ectopic pregnancy	Clearly-written feedback that includes directives for demonstrable changes in learner action	On-line assessment (checklist, Global)

^a Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook: The Cognitive Domain*. New York: David McKay Co Inc.

Skill – Psychomotor ^b			
Domain (KSA) Miller’s Level (relative)	Learning Objective	Outcome	Content Source
Low (Imitation)	Carefully examine both fallopian tube and identifies ectopic pregnancy (pathology)	“Yes” selected on checklist item 1, and minimum rating of “3” on global item 2	Skills Lesson video 1, On-line assessment (checklist, Global)
Mid (Precision) Mid-High (Articulation)	With coordinated actions, safely performs salpingostomy by making full length incision (no more) and only grasping surrounding tissue as necessary to stabilize fallopian tube and pathology	“Yes” selected on checklist items 1-4, and minimum rating of “3” on global items 1,3.	Skills Lesson video 1, On-line assessment (checklist, Global)
Mid (Precision) Mid-High (Articulation)	With coordinated actions, avoids damaging critical structures (i.e. fallopian tube, mesosalpinx), during laparoscopic management of ectopic pregnancy	“Yes” selected on checklist items 6-7, and minimum rating of “3” on global item 1	Skills Lesson videos 1-2, On-line assessment (checklist, Global)
Low-Mid (Manipulation)	Effectively and safely removes specimen (pathology) in simulated setting	“Yes” selected on checklist items 8-9, and minimum rating of “3” on global items 3-4	Skills Lesson videos 1-2, On-line assessment (checklist, Global)
High (Naturalization)	Combining skills, demonstrate basic laparoscopic suturing proficiency by carefully and efficiently throwing a single intracorporeal knot with a surgeon’s knot followed by two throws	“Yes” selected on checklist items 10-11, and minimum rating of “3” on global items 1-5	Skills Lesson videos 1-2, On-line assessment (checklist, Global)
High (Naturalization)	Combining skills, effectively and efficiently resolves laparoscopic ectopic pregnancy in simulated setting	Minimum summed score of 25 on checklist, and score 15/25 on Global	Skills Lesson videos 1-2, On-line assessment (checklist, Global)

^b Dave R.H. (1975). Developing and Writing Behavioral Objectives. (R J Armstrong, ed.) Educational Innovators Press.

Affect – Attitude ^c			
Domain (KSA) Miller's Level (relative)	Learning Objective	Outcome	Content Source
Mid (Responding)	Report improved confidence toward their own ability to laparoscopically manage an ectopic pregnancy presented at their own institution	Improved self-reported confidence rating (change in pre- post-intervention ratings)	On-line survey
High (Valuing)	Attaches value of peer feedback by <i>providing their own feedback</i> on other learners' performance of salpingostomy in simulated setting	Assessment ratings of other learners' procedural skills and global performance	On-line assessment (page or URL?)
High (Valuing)	Attaches value of peer feedback by <i>accepting others' feedback</i> on their own performance of salpingostomy in simulated setting demonstrated by added practice and performance submissions	Submission of at least 2 video performances	On-line assessment (page or URL?)

^c Krathwohl, D.R., Bloom, B.S., and Masia, B.B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain*. NY: David McKay Co.