



ALL-SAFE PRELIMINARY VALIDITY EVIDENCE

Laparoscopic Appendectomy Module 2

Analysis by DM Rooney (dmrooney@med.umich.edu) 10/7-30/22

Part A. Internal Evaluation of Measures of ALL-SAFE Laparoscopic Appendectomy (Lap Appy) Cognitive Testing Tool

METHODS

Study. 24 participants from 4 sites completed the web-based module. Participants included 15 novice, 6 intermediate, and 3 expert participants. All participating sites were represented (Mbingo, n=9; MRS, n=1; Soddo, n=6; 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, with a maximum score of 10. Pre- and post-module summed scores were compared using paired student-test, while differences between novice, intermediate, and expert participants was tested using one-way ANOVA, both with SPSS Statistics for Windows v.25 (IBM, Armonk, NY) Item-level analyses were performed 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

Test of Score Change Following Training.

For all. Paired Student T-Test Comparison of pre- and post-intervention Quiz (Appendix A) summed scores from all 24 participants indicated that there was not a statistically significant improvement in mean summed scores from Pre (M=7.13, SD=1.6) to Post (M=7.15, SD = 1.9), $p=.55$

Rasch analysis at item-level was consistent with this finding, indicating no statistical difference from pre- (M=0.7) and post (M=.8) training, $p=.99$. Deeper analysis indicated score improvement for both novice and intermediate participants.

*****For novice participants.*** Paired Student T-Test Comparison of pre- and post-intervention Quiz summed scores from all 15 novices indicated that there was a statistically significant improvement in mean summed scores from Pre (M=5.87, SD=2.03) to Post (M=7.47, SD = 1.58), $p<.001$.

*****For intermediate participants.*** Paired Student T-Test Comparison of pre- and post-intervention Quiz summed scores from all 6 intermediate participants indicated that there was a statistically significant improvement in mean summed scores from Pre (M=7.83, SD=.98) to Post (M=9.50, SD = .84), $p=.032$.

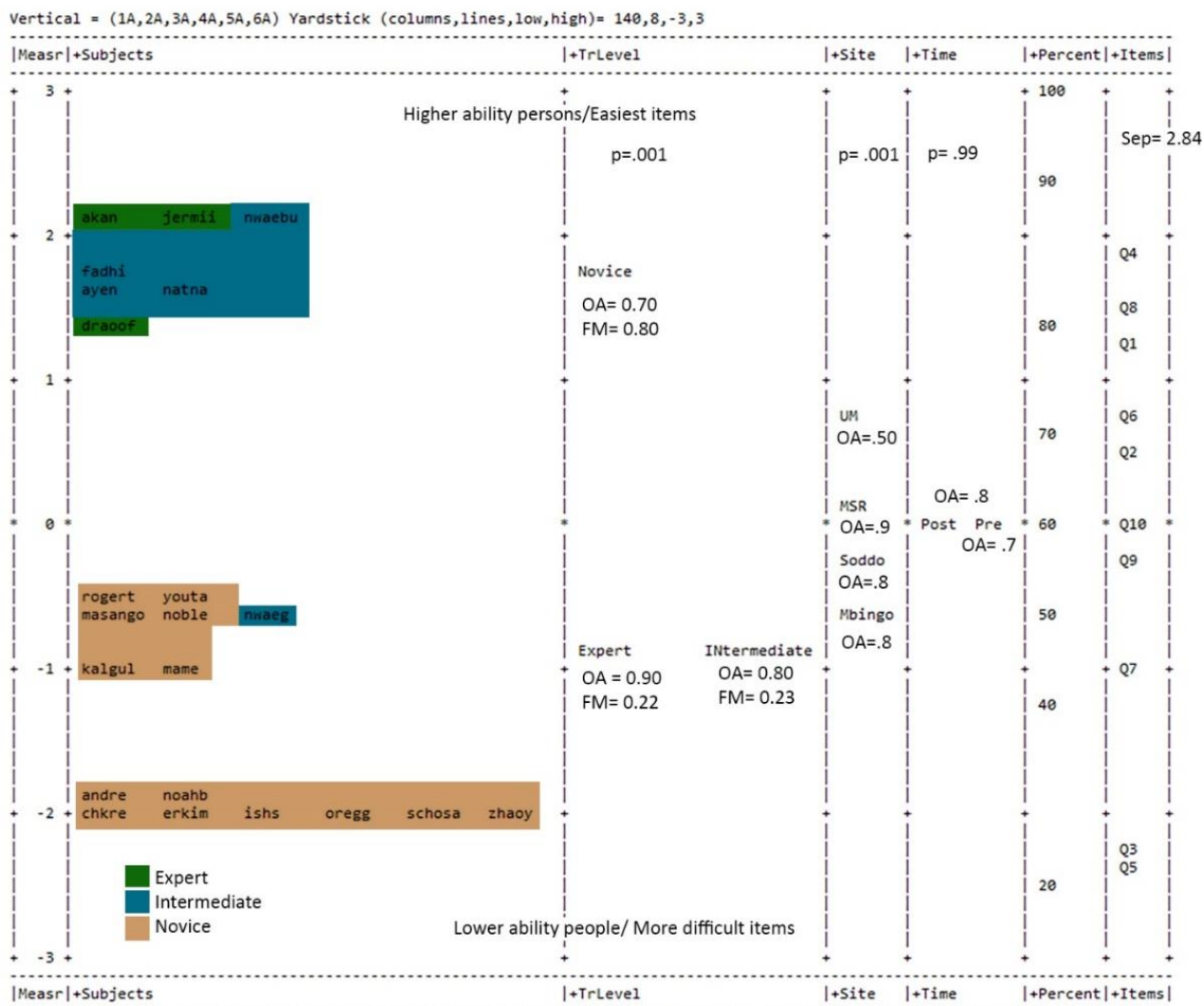
*****Discrimination across novice, intermediate, and expert participants.*** One-way ANOVA test indicated summed scores were able to discriminate between novice (M=6.67, SD=1.97), intermediate (M=8.67, SD=1.23), and expert (M=8.83, SD=.40) participants, $p<.001$. Rasch analysis supported this finding, $X^2(2,X)=70.8$, $p=.001$

Bias analysis

One-way ANOVA indicated statistical differences in mean summed scores across sites, $M_{UM}=5.38$, $M_{MSR}= 8.50$, $M_{Soddo}=8.08$, $M_{Mbingo}= 8.50$, $p < .001$. Rasch analysis supported this finding, $M_{UM}=0.5$, $M_{MSR}=0.9$, $M_{Soddo}=0.8$, $M_{Mbingo}=0.8$, $p=.001$.

Although these findings could suggest item or test bias, because the majority of medical students (the lowest performers) in the UM cohort, differences can be rationalized.

Figure 1. Rasch Variable Map Cognitive Test, questions 1-10



Cognitive Test Item discrimination

Review of item discrimination showed reasonable distribution of item difficulty for items, with items 3 and 5 as the most difficult item (item discrimination=.15 and .67, respectively), and Qs 4 and 8 as the easiest (item discrimination= .81 and .86, respectively) (Table 1)

Table 1. Item discrimination values for ALL_SAFE cognitive test items, ordered highest to lowest.

Item No.	Item Difficulty	Estimated Discrimination*	Discrimination Power	Notes	Suggested Action
Q7	Moderate	1.66	High		—
Q6**	Easier	1.40	High	All Intermediate=1.0	—
Q1	Easiest	1.17	High	All Intermediate=1.0	—
Q2	Easier	1.16	High		—
Q9	Moderate	0.97	Good		—
Q10	Moderate	0.95	Good	Intermediate score declined after training (MPre=1.0, MPost=.83)	Review question for clarity/alignment with content. Focus groups with Residents to ID problem
Q8	Easiest	0.86	Good	All Intermediate=1.0	—
Q4	Easiest	0.81	Good	All Intermediate=1.0	—
Q5	Most Difficult	0.67	Low	All Intermediate=1.0 Novices; Remained difficult (Mpre=.27, Mpost=.53)	High Rasch MnSq Infit (1.67) suggests guessing from lower ability participants, so review question/content to ensure they align; review question to ensure clear
Q3	Most Difficult	0.15	Low	Novices; Remained difficult (Mpre=.13, Mpost=.47) Intermediate; No change in pre-post score.	Simply a difficult question that seems to be too hard for this targeted group of participants

*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.

** Question 6 is the only question all novices answered correctly following training.

Considerations include:

- a) Review/modification of Q5 to avoid ambiguity. Ensure question target is indeed covered within content.
- b) Review Q3 to ensure it's clearly written and targeted content is covered
- c) ** Q6 is the only question all novices answered correctly following training.
- d) Likely, Intermediate participants came in with set knowledge (Pre-test means for Qs 1,4, 5, 6 = 1.0, SD=.00), which is expected
- e) Given that mean post-test scores are still low (M=7.47, SD = 1.58) for novices, it might be expected that they review the content until they achieve mastery (100%) or some expected target, after ensuring content indeed aligns with QUESTIONS 3 and 5.

Part B. Internal Evaluation of Validity Evidence for use of ALL-SAFE Laparoscopic Appendectomy VOP Checklist (13 items) and Global (5 items)

Internal Structure: Comparison of Novice v. Intermediate v. Expert Performance Ratings

Description: Comparison of performance scores across 20 performances, including 11 Novice and 8 Intermediate, and 1 Expert. Each performance was judged by 4 randomly selected participants, including the operator as “self”. Kruskal-Wallis was used to test performance differences across the 3 groups. Checklist items 1-3,5-7,9-11,12 rescored max score=2. Items 4, 8, 12 max score=3. Global items all max score=5. Max combined sum=29+25=54. (Appendix B)

Summary for Checklist. Analyses indicated mixed findings.

- First, comparison of novice /intermediate/ expert performance ratings at the item-level was not helpful to discriminate performance levels, consistent with findings from previous module
- In spite of this, summed total of the checklist (SUMMED) *did* discriminate novice and expert performances ($p=.005$), which was consistent with previous findings for Module 1.

Table 2. Comparison of novice (n=11), intermediate (n=8), and expert (n=1) mean performance ratings

item	Checklist item	Novice Mean (SD) N=11	Intermediate Mean (SD) N=8	Expert Mean (SD) N=1	P-value
1	Identifies anatomy of appendix, cecum and ileum by pointing to <i>each</i> with an instrument	1.60 (.81)	1.73 (.70)	2.00 (.00)	.53
2	Carefully grasps and elevates appendix	1.60 (.81)	1.82 (.59)	2.00 (.00)	.35
3	Mobilizes appendix by sharply taking down sidewall attachments	1.72 (.70)	1.91 (.43)	2.00 (.00)	.39
4	Avoids injury to appendix by excessive grasping or traction	2.10 (1.39)	2.32 (1.29)	3.00 (.00)	.39
5	Creates window in mesoappendix bluntly by spreading with laparoscopic Maryland dissector	1.40 (.97)	1.45 (.92)	2.00 (.00)	.45
6	Ligates appendiceal artery by placing figure of eight suture laparoscopically	1.29 (.99)	1.36 (.95)	2.00 (.00)	.26
7	Performs intracorporeal knot with a surgeon’s knot followed by two additional throws	1.12 (1.03)	1.73 (.70)	2.00 (.00)	.01
8	Avoids tearing the mesoappendix while placing ligating suture	1.80 (1.49)	2.45 (1.18)	3.00 (.00)	.07
9	Cuts remainder of mesoappendix off of appendix using laparoscopic scissors.	1.68 (.74)	1.91 (.45)	2.00 (.00)	.30
10	Places two suture loops/endoloops at base of appendix.	1.64 (.78)	2.00 (.00)	2.00 (.00)	.07
11	Transects appendix sharply	1.68 (.74)	1.73 (.70)	2.00 (.00)	.69
12	Avoids leaving residual appendix on cecum (<3mm)	1.80 (1.49)	2.32 (1.29)	2.25 (1.43)	.35
13	Removes appendix from abdomen	1.88 (.48)	2.00 (.00)	2.00 (.00)	.45
–	SUMMED	21.02 (6.30)	23.64 (3.42)	28.25 (1.50)	.005

Supplemental Analyses for Checklist:

Many-facet Rasch model which examined ratings differences using a 9-facet Rasch model (ID x Operator Expertise x Operator Continent X Judge Expertise x Judge Continent x Judge/Evaluator X Final Rating x Item).

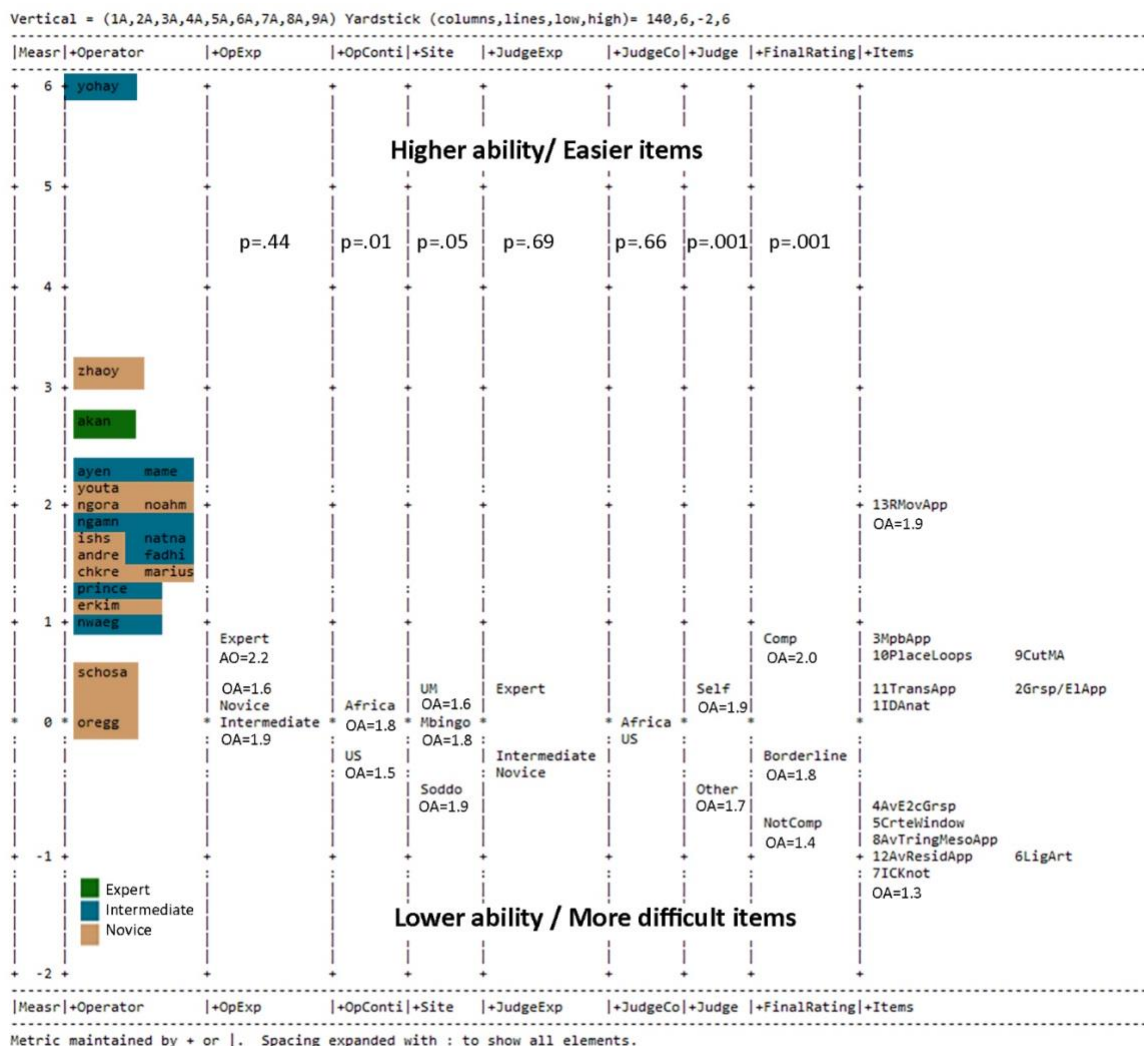
Consistent with classical analysis, Rasch analysis indicated that the Checklist items *could not discriminate* across Novice (M=1.6) v Intermediate (M=1.9) v expert (M=2.2) performances, $p=.44$, although increasing scores with experience was positive (Figure 1)

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=2.0) → Borderline (M=1.8) → Not Competent (M=1.4)

$\chi^2(85) = 32.3, p=.001$, suggesting that these three response options could adequately discriminate subjects.

Figure 2. Variable Map, Rasch Analyses Laparoscopic Appendectomy VOP Checklist



Summary for Global (OSATS) ratings

Summary for Global ratings. First, according to initial Kruskal-Wallis test, comparison of novice /intermediate/expert global ratings indicated the 5 domains were able discriminate across the 3 performance levels, $p = [.001, .002]$. The Global Summed, Total Sum, and Final Rating were all able to discriminate across these 3 levels of performance, $p < .001$ for all.

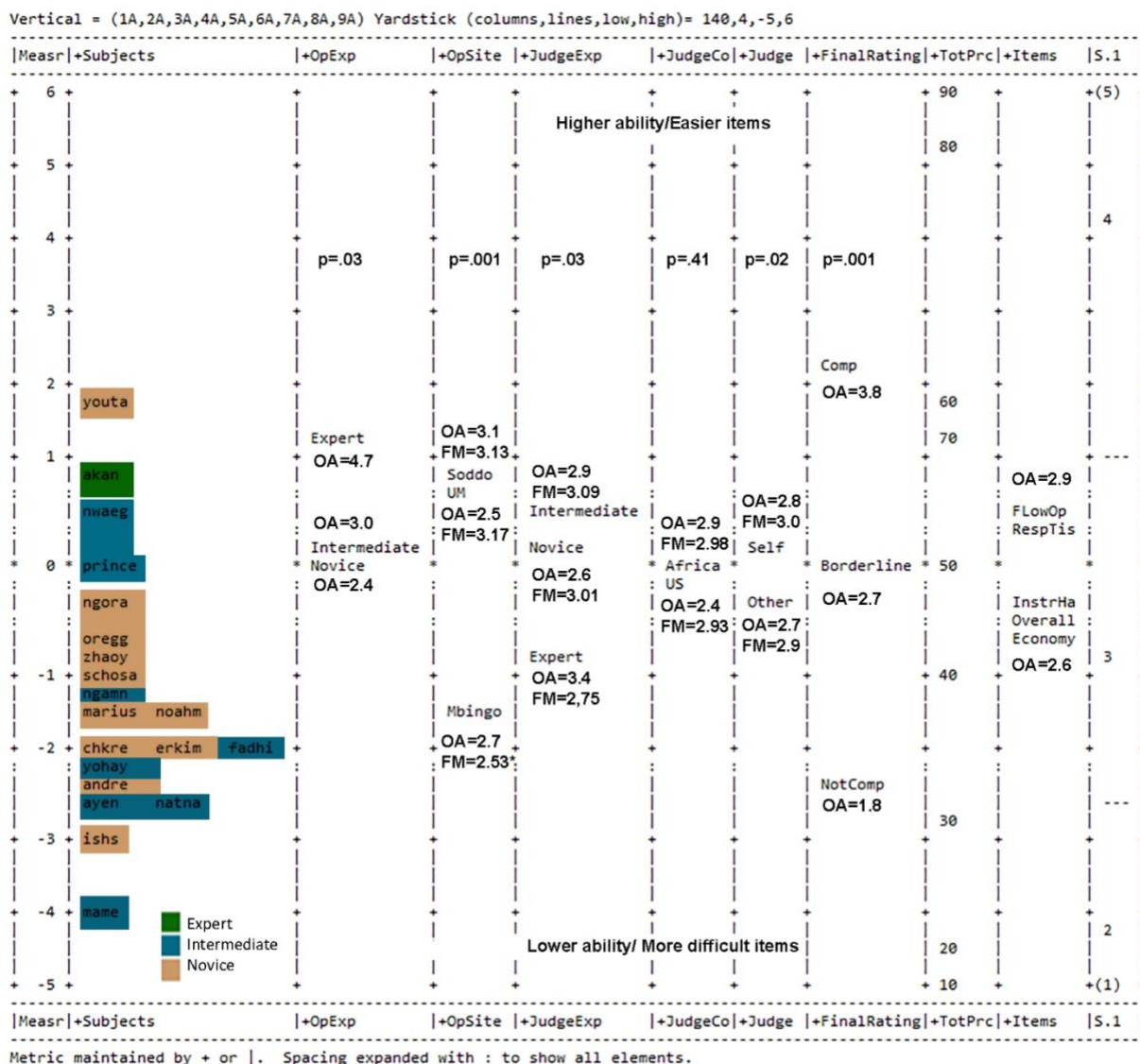
Table 3. Global ratings Novice v Intermediate v Expert

item	Global Domain	Novice Mean (SD) n=11	Intermediate Mean (SD) n=8	Expert Mean (SD) n=1	P-value
1	Respect for Tissue	2.73 (.98)	3.18 (.66)	4.25 (.50)	.002
2	Economy of Time and Motion	2.30 (1.04)	2.86 (.77)	4.75 (.50)	<.001
3	Instrument Handling	2.36 (1.05)	2.95 (.99)	5.00 (.00)	<.001
4	Flow of Operation	2.66 (1.0)	3.07 (1.20)	5.00 (.00)	<.001
5	Overall Performance	2.34 (1.06)	3.05 (.79)	4.25 (.50)	<.001
–	GLOBAL SUMMED	12.28 (4.82)	15.05 (3.76)	23.25 (.50)	<.001
–	TOTAL SUMMED	33.30 (10.07)	39.68 (5.58)	51.50 (1.73)	<.001
	Final Rating	1.76 (.77)	2.32 (.57)	3.00 (.00)	<.001

Supplemental Rasch Analyses for OSATS

Deeper, item-level analyses of OSATS with Many-facet Rasch model examined rating differences using same 9-facet Rasch model (Operator/Subject x Operator Expertise x Judge Expertise x Judge Continent x Final Rating x Total Percent x Item) were consistent with classical test findings and indicated some minor issues associated with sampling, but no significant issues or biases determined (Figure 3, following page).

Figure 3. Variable Map, Rasch analyses Laparoscopic Appendectomy VOP Global Ratings



The same Many-facet Rasch Model was used to examine Global 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=3.8) → Borderline (M=2.7) → Not Competent (M=1.8)

$\chi^2(85) = 243.3, p=.001$, suggesting that these three response options could adequately discriminate subjects.

Positive Findings:

- 1) No rating differences across judge expertise, $p=.69$ (no bias-good)
- 2) No rating differences across judge continent, $p=.66$ (no bias-good)



Potential Concern: Results indicate differences in ratings across Site facet

- Soddo: OA=3.1, FM=3.13 (good alignment)
- Mbingo: OA=2.7, FM=2.53 (slight shift, indicating ratings for Mbingo subjects are slightly less severe than expected)
- UM: OA=2.5, FM=3.17 (big shift, indicating ratings for UM subjects were more severe than expected)

Although relatively high Fair-M Average for UM subjects might indicate that that raters were more severe than expected for US subjects, deeper biases analysis indicated no statistical bias across Site (Rasch Bias measures = [.13, -.01|*]) suggesting differences could be explained by the fact that all US subjects were novices. Given small sample size this is not a concern at this time.

*Values over 1.0 indicate bias

Internal Structure: Rater agreement across novice and experienced judges

Table 4. Estimated ICC values for VOP Components

item	Domain	ICC
Checklist		
-	Checklist Summed	
Global		
1	Respect for Tissue	.70
2	Economy of Time and Motion	.45
3	Instrument Handling	.62
4	Flow of Operation	.45
5	Overall Performance	.67
-	GLOBAL SUMMED	.60
-	TOTAL SUMMED	.83
-	Percent	.75
	Final Rating	.65

Description: Review of inter-rater agreement of Global ratings of a subsample that included 10 performances (Novice, n =9, Experienced, n=1) that had been judged by both novice and experienced raters.* Raters were recategorized as novice versus experienced. Agreement was measured by averaged two-way mixed Intraclass correlation.

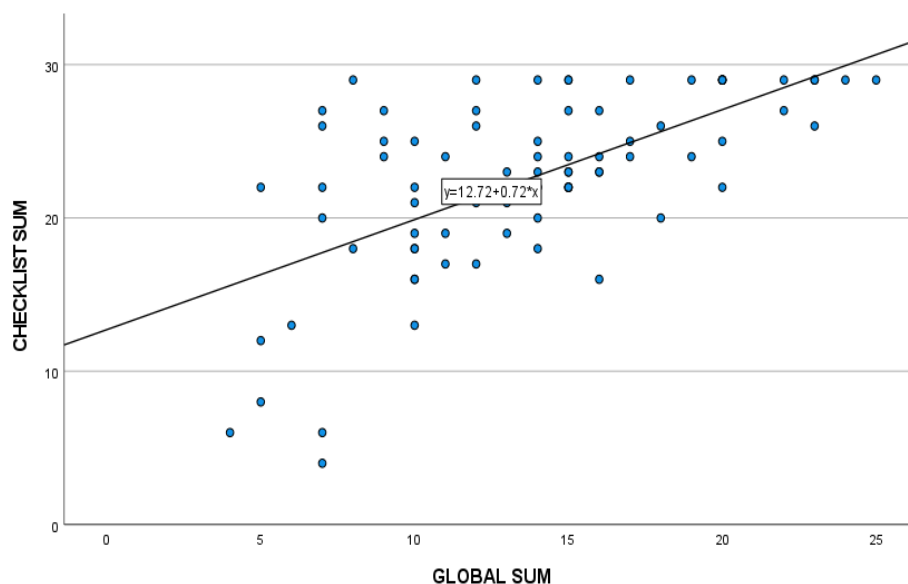
Summary: There was mixed rater agreement across novice and experienced judges, ranging from poor to moderate for the OSATS domains. Poorest inter-rater agreement was estimated for 2 domains, Economy of Time and Motion and Flow of Operation, while higher (moderate agreement) for Respect for Tissue, Total Summed scores.

*Note: Many participants’ performances had been reviewed by only novices OR intermediate participants, limited sample for this analysis

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

Description: Correlation of all participants' summed checklist scores with OSATS scores estimated by Pearson's r .

Figure 4. Correlation between Checklist Sum and Global Sum Scores



Summary. Findings suggest a positive correlation between summed ALL-SAFE Checklist score (CHECKLIST SUM) and established OSATS summed score (GLOBAL SUM), $r(83) = .63$, $p < .001$, supporting use of ALL-SAFE Lap App checklist summed score to measure performance skill.

Similarly, these summed scores (CHECKLIST SUM) correlated with the combined summed score of the Checklist and OSATS (TOTAL SUM), $r(83) = .92$, $p < .001$, as well as the Final Rating, scored on 3-point scale (1 = Does not demonstrate competence, 2 = Borderline, 3 = Does demonstrate competence), $r(83) = .58$, $p < .001$.

FINAL SUMMARY

Case Scenario/ Associated Cognitive Test:

- Cognitive test effectively discriminated between novice, intermediate, and expert participants, and demonstrated benefit to novice and intermediate participants with statistically significant score improvements for novice and intermediate groups, $p \leq .032$.
- Item discrimination analysis suggests review/potential modification of 2 questions (Qs 3/5) to ensure questions target is indeed covered within content, and language is clear.
- Evidence suggests Intermediate participants came in with set knowledge (Pre-test means for Qs 1,4, 5, and 6 = 1.0, $SD=.00$), which is expected
- Given that mean post-test scores are still low ($M=7.47$, $SD = 1.58$) for novices, it might be expected that they review the content until they achieve mastery (100%) or some expected target, after ensuring content indeed aligns with QUESTIONS 3 and 5.

Psychomotor Performance Checklist VOP Assessment:

- Comparison of novice /intermediate/ expert performance ratings at the VOP Checklist item-level was not helpful to discriminate performance levels, consistent with findings from previous module
- Summed total of the checklist (SUMMED) *did* discriminate novice and expert performances ($p=.005$), which was consistent with previous findings for Module 1.
- Overall, checklist scores were able to discriminate across 3 levels of ability; Competent ($M=2.0$) → Borderline ($M=1.8$) → Not Competent ($M=1.4$), $\chi^2(85) = 32.3$, $p=.001$, suggesting that these three response options could adequately discriminate subjects.

Psychomotor Performance Global VOP Assessment:

- All 5 Global VOP domains were able to discriminate across the 3 (novice, intermediate, and expert) performance levels, $p = [.001, .002]$. The Global Summed, Total Sum, and Final Rating were all able to discriminate across these 3 levels of performance, $p < .001$ for all.
- Many-facet Rasch Model analysis indicated statistical scoring differences across Final Rating (e.g. Competent, Borderline, and Not Competent); Competent ($M=3.8$) → Borderline ($M=2.7$) → Not Competent ($M=1.8$)
 $\chi^2(85) = 243.3$, $p=.001$, suggesting that Global scoring could adequately discriminate subjects across these three response options.
- Potential concerns included ratings differences across site (Soddo, Mbingo, and UM). Although Rasch analysis indicated UM subjects were scored more severely than expected, deeper bias analyses indicated no statistically significant bias. Primary inference: scoring differences might easily be explained by the fact that all US subjects were novices (lower performances, hence lower scores).
- Inter-rater reliability estimated by averaged two-way mixed Intraclass correlation, indicated mixed reliability of Global ratings, ranging from .45 (poor) to .83 (moderate-high). Variability in reliability estimates could be caused by small sample size combined with need for supportive materials.



Suggested Next Steps

- 1) evaluate 1) low, 2) borderline, and 3) high performer, with automated feedback based on AI to test alignment with scoring/competency decisions
- 2) Improve evaluation matrix to maximize distribution, minimize nesting which could introduce unexpected score patterns/biases.
- 3) To minimize future bias from experienced “novice” participants, recruit from new/virgin “novice” groups if possible, or include true experts as “gold standard”
- 4) Also, to avoid potential issues from nesting, consider recruitment of residents at UM/SUI, and if possible, novices (med students?) from all participating sites (Soddo/ Mbingo/ Kijabi) and ensure that each operator that submits a video is evaluated by a) judges from another site, b) these judges are ideally, balanced
- 5) To test judging quality of novice (medical students), add attendings to allow comparison to ‘gold standards.’



Part C. Validity Evidence Relevant to Test Content

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

Case scenario and associated questions:

1. Originally drafted by a GS resident (Ngoin Ngam, Mbingo) to ensure relevance to targeted learners
2. Reviewed by 2 Co-Is (Jeffcoach, GS; Snell, GS; Mbingo) for content and relevance
3. Reviewed by 2 M4 students-research assistants (Anidi, Bidwell 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, UM) for stem clarity, response option bias, for content clarity, relevance, alignment of questions with scenario, and flow
6. Disseminated to entire research team for review
7. Final review and approval by PI/Co-I (Kim, UM; Barnard, SIU)

Assessment Tool: Pre/post Multiple Choice Quiz:

1. Originally drafted by MS3 (Reynolds, UM), based on the case scenario
2. Reviewed and edited for relevance/content by PI (Kim)
3. Reviewed by 2 M4 students-research assistants (Anidi, Bidwell) for clarity
4. Reviewed by psychometrician (Rooney) 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)

Assessment Tool: 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
4. Disseminated to entire research team for review
5. 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"
6. Review by psychometrician (Rooney) for clarity, relevance, alignment of questions with skills



Appendix A. Pre/Post-Intervention Quiz

1. **Which of the following is the most common cause of appendicitis?**
 - a. Abdominal adhesions
 - b. Colitis
 - c. Obstruction (fecalith)**
 - d. Traumatic injury

2. **Which of the following patient criteria would most strongly cause you to choose open surgical management over laparoscopic management of acute appendicitis?**
 - a. Patient has a leukocyte count greater than 20,000
 - b. Patient has a morbid obesity with BMI>40
 - c. Patient presents with a fever >40 C
 - d. Patient has multiple adhesions from prior lower abdominal surgeries**

3. **How urgently must laparoscopic appendectomy be performed following diagnosis of acute appendicitis?**
 - a. Within 1 hour
 - b. Within 12 hours
 - c. Within 24 hours**
 - d. Within 1 week

4. **Which of the following is an advantage of laparoscopic over open management of acute appendicitis?**
 - a. No requirement for perioperative antibiotics
 - b. Use of spinal anesthesia
 - c. Lower equipment cost
 - d. Decreased risk of wound infection**

5. **Which of the following is true regarding the use of antibiotic treatment versus surgical management of appendicitis?**
 - a. Overall complication rates are significantly lower in patients treated with antibiotics
 - b. Antibiotic therapy for appendicitis decreases length of hospital stay
 - c. Among patients who receive antibiotics for appendicitis, 100% will eventually undergo appendectomy
 - d. Antibiotics are a reasonable management option for both complicated and uncomplicated appendicitis**

6. **What is the standard pressure of pneumoperitoneum for a standard transabdominal laparoscopic procedure in adults?**
 - a. 5 mmHg
 - b. 10 mmHg
 - c. 15 mmHg**
 - d. 20 mmHg



7. **What is the proper positioning of a patient undergoing laparoscopic appendectomy?**
 - a. Supine position, Trendelenburg, left arm tucked with surgeon and camera on right side
 - b. Supine position, Trendelenburg, left arm tucked with surgeon and camera on left side**
 - c. Prone position, reverse Trendelenburg, both arms tucked with surgeon and camera on right side
 - d. Supine position, reverse Trendelenburg, right arm tucked with surgeon and camera on right side

8. **Which of the following trocars should be placed first during a laparoscopic appendectomy?**
 - a. Camera port at RUQ
 - b. Camera port at umbilicus**
 - c. Instrument port at LLQ
 - d. Instrument port at suprapubic region

9. **Which of the following demonstrates the basic steps of a laparoscopic appendectomy?**
 - a. Trocar placement, locate appendix, bluntly dissect cecum base, divide the mesoappendix, divide appendix base, remove appendix**
 - b. Locate appendix, trocar placement, bluntly dissect cecum base, divide appendix base, remove appendix
 - c. Trocar placement, locate appendix, divide the mesoappendix, divide appendix base, bluntly dissect cecum base, remove appendix
 - d. Trocar placement, locate appendix, divide appendix base, divide the mesoappendix, bluntly dissect cecum base, remove appendix

10. **While performing a laparoscopic appendectomy, you quickly dissect the mesoappendix and notice mild bleeding. Which of the following is the next best step?**
 - a. Convert to open appendectomy
 - b. Apply direct pressure with the instrument to control the bleeding**
 - c. Control the bleeding by ligating the ileocolic artery
 - d. Turn down pneumoperitoneum and observe



Tasks		Y	N	Comments	
Identifies anatomy of appendix, cecum and ileum by pointing to <i>each</i> with an instrument					
Carefully grasps and elevates appendix					
Mobilizes appendix by sharply taking down sidewall attachments					
Avoids injury to appendix by excessive grasping or traction					
Creates window in mesoappendix bluntly by spreading with laparoscopic Maryland dissector					
Ligates appendiceal artery by placing figure of eight suture laparoscopically					
Performs intracorporeal knot with a surgeon's knot followed by two additional throws					
Avoids tearing the mesoappendix while placing ligating suture					
Cuts remainder of mesoappendix off of appendix using laparoscopic scissors.					
Places two suture loops/endoloops at base of appendix.					
Transects appendix sharply					
Avoids leaving residual appendix on cecum (<3mm)					
Removes appendix from abdomen					
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>)		Other Summative Comments:			
Yes Borderline No					