ORIGINAL RESEARCH

Testing an objective structured assessment technical skills tool: A pilot study

Samuel E. Dixon *, Shari M. Burns

Nurse Anesthesia Program, Midwestern University, Glendale, AZ, United States

ABSTRACT

Purpose: The purpose of this prospective pilot study was to test the feasibility, reliability and validity of an adapted version of an objective structured assessment technical skills (OSATS) tool using a criterion-reference rating scale while evaluating student registered nurse anesthetists (SRNAs) performing mask ventilation, laryngeal mask airway (LMA) insertion, and endotracheal intubation.
Methods: Prior to testing, three OSATS evaluation tools were developed with nurse anesthesia faculty input. Six faculty-evaluators pilot tested an adapted version of three OSATS tools for feasibility and validity with 23 SRNAs performing mask ventilation, LMA insertion, and endotracheal intubation with an airway management trainer. SRNAs had 15 minutes; five-minute each to perform mask ventilation, insert an LMA, and perform endotracheal intubation. Paired observations from six faculty-evaluators determined inter-rater reliability. Separate survey questionnaires were created for faculty-evaluators and SRNAs; using three four-point Likert-type rating scales to measure feasibility and validity.
Results: Descriptive statistics revealed that faculty-evaluators' found the OSATs tools to be feasible based on six out of seven somewhat favorable to strongly favorable responses. The Kendall's Coefficient of Concordance W showed statistical significance for mask ventilation and endotracheal intubation (*p*-value of .041 and .036). Six faculty-evaluators found the OSATs tools to have face and content validity based on six out of seven somewhat favorable responses. Twenty-three SRNAs found the OSATS tools to have face validity based on somewhat strongly favorable responses.
Conclusions: An adapted version of OSATS tool for evaluating SRNAs performing mask ventilation, LMA insertion, and endotracheal intubation demonstrated feasibility, reliability, and validity. The inter-rater values obtained from Kendall's Coefficient of Concordance W in this project correlated with the Cook and Beckman report.

Key Words: Objective assessment, Criterion-reference, Clinical skill performance, Global skills checklist

1. INTRODUCTION

Practical hands-on skills in anesthesia education are poorly evaluated in contrast to other domains of learning. This is believed to be due to a lack of a comprehensive global assessment tool.[1] An objective structured assessment of technical skills (OSATS) tool is a common instrument used to evaluate the performance of technical skills of medical and surgical residents. In nurse anesthesia programs, the use of an OSATS tool is uncommon. Clinical skills education occurs early in the student registered nurse anesthetist (SRNA) curriculum. Clinical skills evaluations for SRNAs consist of numerous independent assessments by multiple faculty-evaluators. Subjective assessment tools are used in many training surgical resident programs and nurse anesthesia programs to measure skills competence. Strategies for evaluating student clinical performance are frequently

*Correspondence: Samuel E. Dixon; Email: sdixon@midwestern.edu; Address: Nurse Anesthesia Program, Midwestern University, Glendale, AZ, United States.

criticized due to subjectivity.[2] A common rating method for evaluation of clinical skills is the checklist. The scoring is either a pass/fail or yes/no grade. How each checklist is scored can vary. For instance, a checklist may be divided into task specifics and scored as numerical points ranging from zero to four; it can be divided into small components and scored as present, partially present, or absent. The checklist may consist of a list of procedures to be performed and assigned a dichotomous pass/fail assessment or yes/no assessment. Finally, a checklist may consist of a list of steps required to complete each specific procedure and assigned a not done/incorrect or done correctly assessment, as well as a yes/no assessment for adverse events. Nevertheless, the checklist is an invaluable in-training assessment tool for learning. But, the checklist does not provide formative feedback.[3] An assessment tool without a criterion-reference rating scale, the evaluator may exhibit personal bias, present a wide range of opinions, or have personality issues with a student.[4] Subjective evaluations do not always provide detailed feedback or identify proficiencies or deficiencies in a timely manner.[5]

According to Hall, standardize models for evaluating student nurses' skills performance are not in existence and clinical evaluations are usually subjective.[5] The evaluation of healthcare professional's clinical proficiency was insufficient and there was lack of consistency in the evaluation of clinical outcomes amongst instructors in the clinical setting.[6] The authors also shared that a well developed criterion-referenced rating scale was more reliable and valid than a checklist for clinical evaluation of a student's performance. Laeeq *et al.* described the following benefits of using an objective tool for training assessment of surgical residents' skills: 1) it provided documentation of poor performance; 2) it identified specific areas of weak performance; and, 3) it enabled faculty members to map deficiencies, mentor and, offer counseling.[8] The investigators anticipated that an adapted version of an OSATS tools would provide improvements that lead to standardized grading, improvements in consistency and satisfaction in the evaluation of skills performances, and provide immediate feedback. The research questions that guided our pilot study are: To what extent does an adapted version of an OSATS tool measure SRNAs airway skills performances? What effect does an OSATS tool have on faculty-evaluators' grading and feedback practices?

1.1 Aim

The investigators aimed to develop and perform a pilot test that determined the feasibility, reliability and validity of an adapted version of an OSATS tool using a criterion-reference rating scale evaluating SRNAs performing mask ventilation, laryngeal mask airway (LMA) insertion, and endotracheal intubation.

1.2 Review of the literature

A review of the literature was initiated to discover the availability of a feasible, reliable, and a valid objective assessment of technical skills tool. A systematic search of the literature was conducted using three databases. The search was confined to objective assessment, objective structured assessment of technical skills, clinical skills assessment, procedural skills, and global skills checklist. The investigators limited the search to relevant English language studies published in PubMed, CINAHL, and the Cochrane Database of Systematic Reviews. The search revealed more than 18 studies between 2008 and 2014. The investigators used the following study designs to guide this pilot study:

- Systematic Reviews (Level I Evidence)
- Case Studies, which included prospective longitudinal studies, therapeutic II, instructional, a cross-sectional validation study, a descriptive study (Level V Evidence)
- Quality Improvement Report (Level V Evidence)

The primary investigator found several medical and surgical studies describing the use of an OSATS tool. Kogan, Holmboe, and Hauer performed a systematic review that identified assessment tools used in direct observation of medical trainees with patients.[9] The authors identified 55 observational tools that were used to assess medical trainees' clinical skills; 21 tools were used to study students 32 tools were used to study residents and fellows and two tools were used to study medical students and residents/fellows across the educational continuum. In addition, there were 32 tools created for formative assessments, 26 tools were designated for rater training, and 11 tools had evidence of validity that were based on internal structure and relationship with other variables. The investigators also found that the trainee or observer attitudes usually measured outcomes; nine tools described self assessed transformations in trainee knowledge, skills, or attitudes; five tools described objective measured transformations in knowledge or skills; and 20 Mini Clinical Evaluation Exercise (Mini-CEX) tools had strong evidence of validity. Kogan *et al.* found that direct observation of medical trainees' clinical skills appeared in disciplines such as Internal medicine (47%), Family medicine (13%), Surgery/surgical specialties (11%), Emergency medicine (7%), Pediatrics (6%), Psychiatry (2%), and Anesthesiology (0%).[9]

The authors also concluded that there were many tools available for direct observations of clinical skills, but there was a

lack of validity evidence and a description of educational outcomes. Noland *et al.* selected 10 senior expert hand surgeons and identified eight hand surgical procedures that should be mastered by graduating orthopedic surgery residents. These eight hand surgical procedures received overwhelming support from 23 of the 155 Accreditation Council for Graduate Medical Education (ACGME) approved orthopedic surgery program directors. However, there was no consensus on which hand surgery procedures should be mastered. The authors' concluded that eight hand surgical procedures could be used as a guideline for developing an OSATS for teaching and documenting technical skills in hand surgery.[10] Iyer *et al*. conducted a pilot study that assessed the validity of an OSATS using a global rating scale evaluating 62 pediatric residents performing neonatal lumbar punctures in a video delayed format over a one year period in a simulation setting. The results revealed that the instrument could provide real time formative and summative feedback, which could lead to improved patient care. The results demonstrated evidence of "reasonable validity".[11] House *et al.* developed and tested a modified version of an OSATS tool that collected performance data on 46 emergency medicine residents in postgraduate years one, three, and four performing pediatric rapid sequence intubation during live and digital video recording sessions. The authors found inter-rater reliability for both live and video evaluations (0.75 and 0.79) respectively. Intra-rater reliability between live evaluations and video evaluations was also very good (0.81) in assessing emergency medicine residents' procedural skills. The researchers concluded that there was some evidence of validity in discriminating experience levels of the residents.

In conclusion, the OSATS was determined to be an effective evaluation tool.[12] Ishman *et al*. developed and evaluated two versions of an OSATS tool for blinded evaluation of residents' and fellows' performing pediatric direct laryngoscopy and rigid bronchoscopy on animals or high fidelity manikins in a simulation setting. The first version used a checklist for direct laryngoscopy and rigid bronchoscopy and a global rating tool for bronchoscopy. The second version used a global rating tool for direct laryngoscopy and rigid bronchoscopy. The evaluators determined that the tools were effective in evaluating residents' and fellows' performance. Inter-rater reliability was "acceptable".[13] Feasibility was found with both the checklist and global rating scale when completion times were between three and five minutes. Inter-rater reliability for both the checklist and global rating tools were excellent. Internal consistency of both the checklist and global rating tools were excellent (Cronbach's alpha rating of 0.97). The authors concluded that the evaluators were able to provide immediate feedback to residents regarding their

skills performance.[13] VanHeest *et al.* developed, tested, and published the first OSATS tool using a procedure specific detailed checklist, a validated global rating scale, and a pass/fail assessment for 27 residents performing surgery on a cadaveric specimen at three stations involving trigger finger release (TFR), open carpal tunnel release (CTR), and distal radius fracture fixation (DTFF). Construct validity was found between the year in training and the following assessments: procedure-specific detailed checklist scores for TFR and open CTR; global rating scores for TFR and DTFF; and, pass/fail assessment for TFR. In addition, criterion validity was demonstrated between the procedure specific detailed checklist scores, global rating scale scores, and pass/fail assessment scores for TFR, open CTR, and DTFF. The residents rated the multiple-station OSATS format as highly educational.[14] Tun *et al.* evaluated the face and content validity of a modified version of an OSATS tool using a task specific checklist, a global rating scale, and a direct observation of procedural skills assessment tool. A total of 20 physicians (10 interns and emergency medicine residents and 10 expert clinicians) were evaluated using a simple hybrid or complex hybrid simulations (*i.e.*, a simulated patient and part-task trainer) to assess competency in the management of traumatic skin lacerations under different levels of clinical challenge in a clinical skills laboratory. Each participant performance was video recorded the rating performance. The findings revealed high face validity for simple hybrid and complex hybrid scenarios and a high content validity (4.4 and 4.2 respectively). In addition, the findings revealed a high content validity for simple hybrid and complex hybrid scenarios (4.5 and 4.3 respectively). In addition, all three rating tools demonstrated a high level of internal consistency for both hybrid scenarios.

There was no difference in the performance rating of the experts compared to the novice participants in the simple hybrid scenario for the modified OSATS using all three rating assessment tools. Novices' performance ratings in the complex hybrid scenario were significantly lower than experts using the modified OSATS with all three rating skills tools. The performance rating was insignificant between the experts when both simple and complex hybrid scenarios were assessed. The performance rating was significant between the novices when both simple and complex hybrid scenarios were used with the OSATS task specific checklist and a direct observation of procedural skills tools. The performance rating was insignificant between the novices when both simple and complex hybrid scenarios were used with the OSATS global rating scale tool.[15] Griffin, Hoesli, and Thorne evaluated validity and efficacy of an OSATS instrument. The instrument was designed to evaluate both

cognitive and technical performances of 12 junior and senior otolaryngology residents performing pediatric airway foreign body management on a high fidelity simulation mannequin (Laerdal SimBaby). Inter-rater reliability of the instrument was good (0.71). The authors found the pediatric airway foreign body course had good face validity. The OSATS instrument possessed construct validity, as evidenced by the cognitive, technical, and interpersonal task performances noted amongst junior and senior residents performing pediatric airway foreign body management.[16] Jabbour et al. developed and used a 15-point OSATS checklist to evaluate 17 otolaryngology residents in different years of training performing pediatric airway bronchoscopic of foreign body retrieval from multiple models and different infant airway mannequins in a simulation laboratory. The authors found otolaryngology residents' showed confidence in assembling the foreign body instruments. The residents were successful with identification and extraction of the airway foreign body from multiple models, which was statistically significant ($p < .001$). In addition, the average pre-course score was seven and the average post-course score was 11.3 when a 15-point OSATS grading system was used. Furthermore, the authors asserted that a simulation-based training could serve as a crucial component for developing resident's confidence and improving their ability to perform bronchoscopic of foreign body retrieval.[17] Francis et al. evaluated the feasibility and validity of a modified version of a two part OSATS tool using a task based checklist and a global rating scale to evaluate 15 otolaryngology residents in postgraduate years two through six performing mastoidectomy procedures over a three year period in the operating room. The OSATS instrument was found to be feasible, as evidenced by faculty feedback and the time used to complete the evaluation. There was a strong correlation between the task based checklist and the global rating scale ($r = 0.93$; $p < .001$). Construct validity and internal consistency were high for the task based checklist and the global rating scale. The reliability coefficient for the task based checklist was 0.98 and 0.95 for the global rating scale.[18] Ishman et al. developed and pilot tested an OSATS tool to conduct non-blind evaluations of 19 residents' in different postgraduate levels performing pediatric direct laryngoscopy and rigid bronchoscopy on live animals in a laboratory and in children in the operating room over a three year period. The authors developed two OSATS instruments using a checklist and a global rating scale. The OSATS tool was to be feasible. The instrument took seven minutes to complete. Faculty-evaluators found the tool easy to understand, comprehensive, and practical. Residents thought that the instrument was helpful in providing immediate feedback of their performance. Internal consistency and internal relia-

bility was high for both the task specific checklist and global rating checklist. Construct validity was found with both the task specific and the global rating scale ($p < .001$).[4] Laeeq et al. developed and evaluated the reliability and validity of an training objective assessment tool using a global rating scale and a checklist for endoscopic sinus surgery skills in the operating room. The authors found both the global rating scale and checklist instruments possessed a high internal consistency and construct validity.[3] Laeeq et al. (2009) developed and assessed the feasibility, validity, and reliability of a two part OSATS tool using a global rating scale and task-based checklist for 23 residents in all postgraduate years performing a cortical mastoidectomy on a cadaveric temporal bone over a three year period. Construct validity scores increased each year of clinical training. Faculty found the assessment tool easy to use, it had a short completion time, and overall, it was feasible. Inter-rater reliability scores for the global rating scale was 74.4% and the task-based checklist was 78.5%. Construct validity was high for both the global rating scale and the task based checklist tools.[8] Lin et al. designed and tested the feasibility, reliability and validity of an OSATS tool using a global rating scale and a checklist for 28 residents performing endoscopic sinus surgery on cadavers over a three year period. The authors found the checklist and the global assessment tool feasible; seven minutes to complete, easy, and practical. Internal consistency and internal reliability and construct validity were high for both instruments. In addition, the surgical residents thought the tool was helpful in providing immediate informative feedback of their performance.[19] Chipman and Schmitz created and utilized an OSATS instrument using a task specific checklist and a global rating scale for 38 residents in postgraduate years one, two, and three performing tasks from the first five modules of an instructional unit covering asepsis and skin preparation, gowning and gloving, knot-tying, suturing, and excision of a benign skin lesion. The authors found the internal consistency reliability coefficient to be 0.896 for scores produced by an OSATS rating tool. There was evidence of construct validity, which was determined by lower scores from novice residents and higher scores from experienced residents. In addition, the task specific checklist and the global rating scale were found useful. The checklist was also useful in documenting whether or not discrete steps or behaviors occurred. The global rating scale was useful in communicating how well behaviors were executed.[20]

1.3 Theoretical framework

The theoretical framework selected was the Criterion-referenced Conceptual and Measurement Framework was used to guide the development of an OSATS tool and how the evaluate the process in determining feasibility, reliability,

and validity of an OSATS tool using a criterion-referenced rating scale.[6, 7] A predetermined criteria provides a format for faculty-evaluators to evaluate SRNAs performing mask ventilation, laryngeal mask airway, and endotracheal intubation. This framework would guide faculty-evaluators with providing direct observation of the SRNAs airway management skills performance. More importantly, this theoretical framework will assist in identifying standardized criteria and essential steps, critical elements or task to be assessed and for designing an OSATS tool that would minimize extraneous and faculty-evaluator influences and biases. The incorporation of standardized criterion-reference rating scale is intended to improve SRNA satisfaction, correct grading consistency and standardize grading practices amongst faculty-evaluators. The performance evaluation is in align with the International Nursing Association for Clinical Simulation and Learning (INACSL) published standards on the quality of simulation in nursing programs. Simulation is regarded as an acceptable method for evaluating learning in three major domains of learning (cognitive, affective, and psychomotor), but "the achievement of expected outcomes of a simulation experience should be based on valid and reliable instrumentation, tools, and methodologies".[21] The American Association of Colleges of Nursing (AACN) supports the use of a criterion-referenced rating scales for measuring levels of competencies and expected performances of students in their clinical course and at the end of their nursing program.[6] The American Board of Anesthesiology (ABA) has mandated simulated-based training as a part of the practice performance improvement and assessment component for the maintenance of certification in anesthesiology (MOCA).[22] The Accreditation Council for Graduate Medical Education (ACGME) proposed that specific outcomes should assess competency and improve education and performance feedback.[4]

2. METHODS

2.1 Design, setting, and sample

The investigators prospectively pilot tested three adapted versions of an OSATS tools using a criterion-reference rating scales to evaluate SRNAs performing mask ventilation, laryngeal mask airway insertion, and endotracheal intubation for feasibility, reliability and validity. The Midwestern University Institutional Review Board (IRB) exempted the pilot study. The pilot study was conducted in the simulation laboratory of the nurse anesthesia program, College of Health Sciences, Midwestern University Glendale. Twenty-three first year SRNAs volunteered to participate. The investigator informed the SRNAs that they would have an opportunity to evaluate the feasibility, reliability, and validity of an OSATS

instrument using a criterion-rating scale performing mask ventilation, laryngeal mask airway insertion, and endotracheal tube intubation skills on an airway management trainer. They were also informed that their participation and performance would not affect their grade and feedback would be provided immediately. All SRNAs provided written consent to participate. In addition, six full time nurse anesthesia faculty members were recruited via e-mail and consented to participate as faculty-evaluators.

2.2 Procedures

There were two phases of the pilot study which begun during the winter quarter of 2014 and ending spring quarter of 2015. In phase one of the initial development of the three OSATS tools, six faculty-evaluators' identified the procedural steps for each tool. After subsequent serial refinement of the procedural steps and critical elements, they were evaluated for relevancy, clarity, and consistency based on the six faculty-evaluators clinical experiences and published information from anesthesia references such as Nurse Anesthesia,[23] Morgan & Mikhail's Clinical Anesthesiology,[24] and Miller's Anesthesia Volume 2.[25] A total of 37 procedural sets were selected by the faculty-evaluators: eight procedural steps for mask ventilation, 11 procedural steps for LMA insertion, and 18 procedural steps for endotracheal tube (ETT) intubation (see Tables 1-3). In addition, faculty-evaluators identified a criterion-reference rating scale and selected numerical scores (1 through 5) to assess a level of competences (1 = dependent/unsatisfactory, 2 = marginal/needs improvement, 3 = assisted/satisfactory, 4= supervised/commendable, and 5 = independent/outstanding) and a grading score (0 to 100%), which was based on execution of the essential sequence, completion time, and the need for directions or cues (see Tables 1-3). Each OSATS form included a section to write the total score, comments and a not applicable. Faculty-evaluators identified and approved seven questions in a survey to assess each OSATS tool for mask ventilation, LMA insertion, and ETT intubation. In addition, Faculty-evaluators identified and approved five questions in a survey for SRNAs assess each OSATS tool for mask ventilation, LMA insertion, and ETT intubation. The primary investigator discussed the contents and Likert scale. Oral one on one and group instructions as well as written instructions were provided. In phase two of testing, all SRNAs either received one-on-one or small group instructions on the contents and Likert scale and how to complete the student airway management –learning questionnaire, and each of the three survey questionnaires for mask ventilation, LMA insertion, and ETT intubation. SRNA was given a brown envelope with a number (1 to 23) and a sticker with differentiating characteristics one week prior to testing. Each envelope contained a

request for the SRNA to mark the number of times they had performed, practiced, or studied airway management (see Table 4), as well as the airway skills performance objectives and three scripted scenarios (see Table 5). The primary investigator asked faculty-evaluators and SRNAs to select one of three days for testing. On each of the three days, there were 10 SRNAs, six SRNAs and seven SRNAs observed by two faculty-evaluators simultaneously. The primary investigator selected and used the same airway management trainer and the following airway equipment for three SRNA groups: (a) adult Ambu Bag, (b) adult mask, (c) 5.0 Unique LMA, (d) 7.0-mm or 7.5-mm endotracheal tube, and a (e) 90-mm oral airway. Additional equipment included an anesthesia machine, stethoscope, tongue depressor, adhesive tape, small, medium, and large latex-free gloves, and airway lubricant for the mannequin. Each SRNA assembled and inspected the airway equipment prior to testing. Each SRNA was presented with a scenario for mask ventilation, LMA insertion, and ETT intubation prior to testing (see Table 5). Testing began when the SRNA verbalized they were prepared and ready. Each SRNA had 15 minutes to complete the skills performance; 5-minutes each to perform mask ventilation, insertion of a laryngeal airway mask, and to perform endotracheal intubation.

2.3 Data collection

Paired faculty-evaluators observed each SRNA airway management skills performance in succession: mask ventilation,

insertion of a laryngeal airway mask, and endotracheal intubation using an OSATS tool. Testing was completed when the student verbalized they were finished or 15 minutes had elapsed. Upon completion of the SRNA airway management skills performance, the faculty-evaluator provided immediate feedback in the form of a numerical score (1 through 5) indicating their level of competence, as well as how many steps were completed, how long the skill took, and how much directing or cueing was needed. Three separate survey questionnaires were created for six faculty-evaluators and 23 SRNAs to use a four-point Likert-type rating scale with scores [1 (strongly disagree) to 4 (strongly agree)] to rate the OSATS tools' ability to evaluate airway management skills performances (see Tables 6 and 7). All survey questions were reviewed for comprehension and analyzed by the faculty-evaluators for content validity. All OSATS evaluations were completed in the simulation laboratory and submitted to the primary investigator. The primary investigator entered the data onto an Excel spreadsheet and forwarded the compiled data to the Midwestern University statistician.

2.4 Data analysis

Descriptive statistics [mean, median, and standard deviation (SD)] determined the distribution of faculty-evaluators' and SRNAs survey responses. The data was analyzed using the Statistical Package for the Social Sciences (SPSS) software version 22.

Table 1. OSATS rating scale evaluation form for mask ventilation

	1 Dependent/ unsatisfactory • Executed 0 to 24% of the essential sequence • Complete the performance sequence in nine (9) minutes • Required directions or cues 75% of the time	2 Marginal/needs improvement • Executed 25% to 49% of the essential sequence with guidance • Completed the performance sequence in eight (8) minutes • Required directions or cues 50% of the time	3 Assisted/satisfactory • Executed 50% to 74% of the essential sequence • Completed the performance sequence performance in seven (7) minutes • Required directions or cues 25% of the time	4 Supervised/commendable • Executed 75% to 90% of the essential sequence • Completed the performance sequence in six (6) minutes • Required directions or cues 10% of the time	5 Independent/ outstanding • Executed all the essential sequence, • Completed the performance exercise in the allotted time, • Required no direction or cues to complete the performance sequence	N/A
Selected airway equipment and prepared emergency drugs						
Donned gloves						
Selected adult face mask						
Created a mask-to-face seal						
Ventilated patient with enough volume to provide visible chest rise						
Inserted an adult oral airway for difficult airway						
Adjusted oxygen liter flow						
Provided two-handed mask ventilation with an assistant for difficult airway						
Total score						
Comments						

Note. Adapted from Holaday SD, Buckley KM. (2008). Chapter 7, a standardized clinical evaluated tool-kit: Improving nursing education and practice. In M. H. Oermann (Ed.), *Annual Review of Nursing Education*. 6: 123-149; Bondy KM. (1993) Criterion-referenced definitions for rating scales in clinical evaluation. *Journal of Nursing Education*. 22: 376-382.

Table 2. OSATS rating scale evaluation form for insertion of laryngeal mask airway (LMA)

	1 **Dependent/ unsatisfactory** • Executed 0 to 24% of the essential sequence • Complete the performance sequence in nine (9) minutes • Required directions or cues 75% of the time	**2** **Marginal/needs improvement** • Executed 25% to 49% of the essential sequence with guidance • Completed the performance sequence in eight (8) minutes • Required directions or cues 50% of the time	**3** **Assisted/ satisfactory** • Executed 50% to 74% of the essential sequence • Completed the performance sequence performance in seven (7) minutes • Required directions or cues 25% of the time	**4** **Supervised/ commendable** • Executed 75% to 90% of the essential sequence • Completed the performance sequence in six (6) minutes • Required directions or cues 10% of the time	**5** **Independent/ outstanding** • Executed all the essential sequence, • Completed the performance exercise in the allotted time, • Required no direction or cues to complete the performance sequence	**N/A**
Selected airway equipment and prepared emergency drugs for airway management						
Selected LMA based on patient weight in kilograms						
Deflated and inflated LMA cuff and inspected valve						
Lubricated LMA posterior surface						
Donned gloves						
Placed patient head in sniffing position						
Used the index finger with cuff/tube interface						
Inserted LMA against hard palate of the posterior pharyngeal cavity						
Exerted cephalad pressure with other hand LMA prior to removing index finger						
Inflated up to 30 ml in the cuff and secured LMA in place						
Performed Bag and Mask ventilation						
Total score						
Comments						

Note. Adapted from Holaday SD, Buckley KM. (2008). Chapter 7, a standardized clinical evaluated tool-kit: Improving nursing education and practice. In M. H. Oermann (Ed.), *Annual Review of Nursing Education*. 6: 123-149; Bondy, K. M. (1993) Criterion-referenced definitions for rating scales in clinical evaluation. *Journal of Nursing Education.* 22: 376-382.

3. RESULTS

A total of 138 assessments with immediate feedback carried out 23 SRNAs performing mask ventilation, LMA insertion, and endotracheal intubation using an airway management trainer in the simulation laboratory. One hundred and thirty-eight OSATS tools were completed by six faculty-evaluators, which resulted in a completion rate of 100%. Face and content validity were found when five of the six faculty-evaluators responded favorable to the OSATS tools comprehensibility, ability to differentiate level of performance, assess other clinical training activities, and complete the tool and provide immediate feedback.[4] Kendall's Coefficient of Concordance W (inter-rater reliability) yielded a statistical significance for mask ventilation and ETT intubation (w = 0.80, χ^2 = 35.3, *p*-value = .036) (see Table 8). The Kendall's Coefficient of Concordance W was not significant for LMA insertion (w = 0.64, χ^2 = 28.3, *p*-value was .166)

(see Table 8). Overall, the feasibility responses were favorable, except for two faculty-evaluators unfavorable response that the learning outcomes were not stated in terms of observable behavior for mask ventilation, LMA insertion and endotracheal intubation and the form was not user friendly for mask ventilation and endotracheal intubation (see Table 6). Twenty-three SRNAs completed 69 OSATS tool assessment surveys. There was a 100% favorable response to the OSATS tools evaluating mask ventilation, LMA insertion, and endotracheal intubation. Fifty percent of the SRNAs agreed that the OSATS tools provided objective assessments and 50% agreed that the OSATS tools provided subjective assessment of their skills performing mask ventilation, LMA insertion, and endotracheal intubation (see Table 7). The face-to-face administration of the survey questionnaires to faculty-evaluators and SRNAs resulted in a 100% response rate.

Table 3. OSATS rating scale evaluation form for endotracheal tube (ETT) intubation

	1 Dependent/ unsatisfactory • Executed 0 to 24% of the essential sequence • Complete the performance sequence in nine (9) minutes • Required directions or cues 75% of the time	2 Marginal/needs improvement • Executed 25% to 49% of the essential sequence with guidance • Completed the performance sequence in eight (8) minutes • Required directions or cues 50% of the time	3 Assisted/ satisfactory • Executed 50% to 74% of the essential sequence • Completed the performance sequence performance in seven (7) minutes • Required directions or cues 25% of the time	4 Supervised/ commendable • Executed 75% to 90% of the essential sequence • Completed the performance sequence in six (6) minutes • Required directions or cues 10% of the time	5 Independent/ outstanding • Executed all the essential sequence, • Completed the performance exercise in the allotted time. • Required no direction or cues to complete the performance sequence	N/A
Selected airway equipment and prepared emergency drugs for endotracheal intubation						
Selected endotracheal tube for an adult male or female						
Inspected endotracheal tube cuff						
Donned gloves						
Placed the patient head in a sniffing position						
Pre-oxygenated with 100% oxygen for 3 to 5 minutes or 8 deep breaths						
Pretreated to protect against increased sympathetic activity & anxiety						
Performed a Rapid Sequence Induction						
Administered induction agents						
Applied cricoid pressure prior to the administration of induction agents and held until tube placement confirmed						
Checked for paralysis with a peripheral nerve stimulator						
Inserted the laryngoscope blade into the mouth, displaced the tongue laterally						
Introduced the endotracheal tube to depth of 21 cm (female) and 22 cm (male)						
Inflated the cuff (5-10 ml) of air						
Disconnected the syringe from ETT cuff inlet port						
Ventilated with enough volume to produce visible chest rise						
Confirmed ETT placement; auscultated lungs bilaterally & epigastrium, checked for EtCO$_2$, & tube condensation						
Secured the endotracheal tube						
Total score						
Comments						

Note. Adapted from Holaday SD, Buckley KM. (2008). Chapter 7, a standardized clinical evaluated tool-kit: Improving nursing education and practice. In M. H. Oermann (Ed.), *Annual Review of Nursing Education*. 2008; 6: 123-149; Bondy K.M. (1993) Criterion-referenced definitions for rating scales in clinical evaluation. *Journal of Nursing Education*. 22: 376-382.

Table 4. Student airway management-learning questionnaire

	Mask	LMA	ETT	≤ 2 weeks	≤ 1 month	1.5 months	2 months	3 months	4 months
Simulation manikin									
Live patient									
Video exposure									
Reading of literature									
Advanced Cardiac Life Support Recertification									

Table 5. Airway skills performance objectives and case scenarios for student registered nurse anesthetists (SRNAs)

Airway skills performance objectives:

(1) Perform mask ventilation within 5 minutes

(2) Perform laryngeal mask airway insertion within 5 minutes

(3) Perform an endotracheal tube intubation within 5 minutes

(4) Describe your experience by answering a survey questionnaire on each of the three skills performed upon completion of the scenarios

Scenario for Mask Ventilation (MV)

A 34-year-old male construction worker scheduled for a wound irrigation and avulsion of the right first phalange nail plate. He has an American Society of Anesthesiologists (ASA) physical classification status of 2 and weighs 70 kg. He smokes a one of a pack of cigarettes per day for 10 years, no known drug allergies, and no history of surgery. He has requested general anesthesia.

Scenario for Laryngeal Mask Airway (LMA)

A 34-year-old female professional singer scheduled for a right buninectomy. She has an American Society of Anesthesiologists (ASA) physical classification status of 2 and weighs 56 kg. She smokes a half of a pack of cigarettes per day for 10 years, no known drug allergies, and previous had a left buninectomy without anesthesia complications. She has requested a laryngeal mask airway (LMA) under general anesthesia.

Scenario for Endotracheal Tube (ETT) Intubation

A 34-year-old male athlete scheduled for a lateral release modified mcbrides procedure of the right first metatarsal. He has an American Society of Anesthesiologists (ASA) physical classification status of 2, and weighs 80 kg. He consumes one can of smokeless tobacco per day for 10 years, no known drug allergies, and no history of surgery. He has requested
general anesthesia.

Table 6. Faculty-evaluators assessment of the OSATS tools (n-6)

Description	Mask Ventilation			Insertion of LMA			Endotracheal Intubation		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
The rating scale learning outcomes were stated in terms of observable behavior.	3.33	4.00	1.21	3.33	4.00	1.21	3.33	4.00	1.21
The numbers assigned to each level of scoring were comprehensible.	3.17	3.00	0.75	3.33	3.50	0.82	3.33	3.50	0.82
I was able to use the assessment of the rating scale ability to differentiate among levels of performance.	3.17	3.50	0.98	3.17	3.50	0.98	3.17	3.50	0.98
The rating scale evaluation form was user-friendly.	2.83	3.00	1.17	3.00	3.50	1.26	2.83	3.00	1.17
I would use the rating scale evaluation tool to assess other clinical training activities.	3.50	3.50	0.55	3.50	3.50	0.55	3.50	3.50	0.55
I was able to complete the tool during the assessment of the student.	3.67	4.00	0.52	3.50	3.50	0.55	3.50	3.50	0.55
I was able to provide the student with immediate feedback/debriefing.	3.83	4.00	0.41	3.83	4.00	0.41	3.67	4.00	0.52

Note. Likert scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = strongly agree.

Table 7. SRNAs assessment of the OSATS tools (n-23)

Description	Mask Ventilation			Insertion of LMA			Endotracheal Intubation		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
The rating scale provided me with an accurate assessment of my skills.	3.43	4.00	0.73	3.60	4.00	0.50	3.45	4.00	0.76
The rating scale provided me with an objective assessment of my skills.	3.52	4.00	0.67	3.48	4.00	0.60	3.43	4.00	0.75
The rating scale provided me with a subjective assessment of my skills.	3.48	4.00	0.67	3.48	4.00	0.60	3.43	4.00	0.68
The rating scale provided me with valuable feedback during the debriefing.	3.70	4.00	0.56	3.71	4.00	0.56	3.86	4.00	0.36
Having knowledge of the learning outcomes enhance my ability to meet the course objective.	3.73	4.00	0.54	3.76	4.00	0.44	3.67	4.00	0.58

Note. Likert scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = strongly agree.

Table 8. Paired inter-rater reliability

Kendall's Coefficient of Concordance W	Mask Ventilation			Laryngeal Mask Airway Insertion			Endotracheal Intubation		
	w	χ^2	*p*-value	w	χ^2	*p*-value	w	χ^2	*p*-value
	0.79	34.8	.041***	0.64	28.3	.166 N/S	0.80	35.3	.036***

Note. *** $p < .5$; N/S = Not Significant.

4. DISCUSSION

The investigators conducted a unique simulation-based pilot study with an adapted version of three OSATS tools using a criterion-reference rating scale to assess SRNAs airway skills management performance. The results substantiate the investigators hypothesis that an OSATS tool would provide improvement in standardize grading, consistency and satisfaction in evaluation of skills performances as well as, provide immediate feedback. The investigators adapted version of an OSATS module using a criterion-reference rating scale findings were similar to previous airway skills performance assessments.[13, 16, 17] The inter-rater reliability results show that each of the three OSATS tools have face validity.[1] Each OSATS tool enabled faculty-evaluators to focus on task completion and provide immediate feedback.[10, 11, 20] The favorable responses by SRNAs indicate that each OSATS tool enhanced faculty-evaluators abilities to provide formal, real-time and immediate feedback.[4, 19]

The investigators were able to standardize practical skills for evaluating airway management skills performances for simulation laboratory testing.[2] The 50/50 survey responses by the SRNAs to questions asking whether their skills were assessed objectively or subjectively is unknown but, may be due to a perception that the tools were biased, unrealistic, and did not reflect their previous graded skills performance. The OSATS tool went through several transformations. An OSATS tool is not confined to one procedural skill; it can be adapted to assess different practice skills.[1] The potential

use of an OSATS tool at this institution is very high considering it minimizes subjectivity, allows for immediate feedback or debriefing, and improves grading consistency amongst faculty evaluators. The OSATS tool can be modified to meet a specific topic, learning objectives, or learning outcomes while standardizing assessments, consistency in grading and feedback practices. Because this was a pilot study, further investigations and a larger sample size is needed to determine the feasibility, reliability and validity of the OSATS tool for formative and/or summative assessments. We observed some SRNAs struggling with sequential steps during their performance of either mask ventilation, laryngeal mask airway insertion, and/or endotracheal intubation, which is in contrast to Bould et al. findings that procedural skills are sequential and predictable.[1]

Limitations

The study sample was limited to one small cohort of SRNAs at one nurse anesthesia program. Each SRNA knew that their level of performance would not impact their student status or prior grades and may have impacted their survey responses. There were no previous feasibility, reliability, and validity studies of an OSATS tool assessing anesthesia students' airway management skills performance. The investigators were not able to conduct pilot testing on days without interfering with SRNAs study time and preparation for examinations. There were insufficient time and number of faculty-evaluators to evaluate 23 SRNAs in eight hours to eliminate or minimize student collaboration. The inves-

tigators discovered that it was difficult determining which steps were clinically essential versus those steps that were nonessential for each of the OSATS tool. This was due to the practice differences amongst the faculty-evaluators and the step variance in anesthesia textbooks. It was also difficult to design an OSATS tool that combined the skills for mask ventilation, LMA insertion, and endotracheal intubation into one page. Also, it was difficult to design an OSATS tool with fewer numbers of steps, more readable, user-friendly, and supported criteria that enable the faculty-evaluator with providing immediate feedback. The investigators recognized that SRNA-to-SRNA collaboration could not be prevented during the assessment phase of each OSATS tools. Student registered nurse anesthetist communicated similar concerns and asked similar questions about the airway management trainer malfunction during testing. The airway management trainer malfunctioned (failed to provide enough air to inflate the lungs) during SRNAs performance of mask and LMA ventilations. The investigators suggest testing the airway trainer for air leaks and use an airway management trainer that is designed for advanced students and mask ventilation, endotracheal intubation, LMA insertion, and assessment of adequate cuff inflation, such as the "Airway Larry" an airway management trainer torso from Life/form® developed by Nasco. The investigators choose an airway management trainer because of their professional experience and knowledge that a low fidelity partial task trainer is just as effective as high fidelity mannequin.[1]

To determine predictive validity, Boulder *et al.* suggested using a Human Patient Simulator for tracheal intubation with a high-fidelity mannequin such as a METIman.[1] The student airway management-learning questionnaire results were not used, because the primary investigator saw inconsistencies in their documentations and possible inaccuracies in the number of times they had practiced one or more of the airway skills techniques. For example, many of the SRNAs responded with a yes, single or multiple check marks, wrote a percent-age, used a greater than symbol or less than symbol with a numerical number ranging from 0 to 750. In addition, there was criticism by many of the SRNAs that the questionnaire was confusing and was designed poorly. Going forward, this pilot study is the foundation for a follow-up study examining the validity and efficacy using an OSATS tool for standardized assessment of airway management skills performance of SRNAs in nurse anesthesia programs. The results of this pilot study provided nurse anesthesia educators with a blueprint for evaluating SRNAs level of technical skills proficiency. The OSATS tool may provide documentation of performance progression from a formative test to summative test, thereby eliminating reliance on memory and recall, improper and inaccurate assessment of essential details, which would benefit both the SRNA and the nurse anesthesia program.

5. CONCLUSION

The researchers conclude that feasibility, reliability, and validity is achievable with a adapted version of an OSATS tool using a criterion-reference rating scale when evaluating SRNAs performing mask ventilation, LMA insertion, and endotracheal intubation. The OSATS tool can be adapted to evaluate a specific practical skill, learning objective, or learning outcome while standardizing scoring based on skills performance, and feedback practices.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest.

REFERENCES

[1] Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. Br J Anaesth. 2009; 103(4): 472-483. PMid:19720612 http://dx.doi.org/10.1093/bja/aep241

[2] Hayden J, Keegan M, Kardong-Edgren S, *et al.* Reliability and validity testing of the Creighton competency evaluation instrument for use in the national council of state boards of nursing national simulation study. Nurs Edu Perspect. 2014; 35(4): 244-252. http://dx.doi.org/10.5480/13-1130.1

[3] Laeeq K, Waseem R, Weatherly RA, *et al.* In-training assessment and predictors of competency in endoscopic sinus surgery. Laryngoscope. 2010; 120: 2540-2445. PMid:21082748 http://dx.doi.org/10.1002/lary.21134

[4] Ishman SL, Brown DJ, Boss EF, *et al.* Development and pilot testing of an operative competency assessment tool for pediatric direct laryngoscopy and rigid bronchoscopy. Laryngoscope. 2010; 120: 2294-2300. PMid:20939072 http://dx.doi.org/10.1002/lary.21067

[5] Hall MA. An expanded look at evaluating clinical performance: Faculty use of anecdotal notes in the U.S. and Canada. Nurse Edu in Practic. 2013; 13: 271-276. http://dx.doi.org/10.1016/j.nepr.2013.02.001

[6] Holaday SD, Buckley KM. Chapter 7, A standardized clinical evaluated tool-kit: Improving nursing education and practice. In M. H. Oermann (Ed.), Annual Review of Nursing Education. 2008; 6: 123-149.

[7] Waltz C, Strickland O. Measurement in Nursing and Health Research. New York: Springer; 2005.

[8] Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. Laryngoscope. 2009; 119: 2402-2410. PMid:19885831 http://dx.doi.org/10.1002/lary.20678

[9] Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. JAMA. 2009; 302(12): 1316-1326. PMid:19773567 http://dx.doi.org/10.1001/jama.2009.1365

[10] Noland SS, Fisher LH, Lee GK, et al. Essential hand surgery procedures for mastery by graduating orthopedic surgery residents: A survey of program directors. J Hand Surg. 2013; 38: 760-765. http://dx.doi.org/10.1016/j.jhsa.2012.12.035

[11] Iyer MS, Santen SA, Nypaver M, et al. Assessing the validity evidence of an objective structured assessment tool of technical skills for neonatal lumbar punctures. Acad Emerg Med. 2013; 20: 321-324. http://doi:10.1111/acem.12093

[12] House JB, Dooley-Hash S, Kowalenko T, et al. Prospective comparison of live evaluation and video review in the evaluation of operator performance in a pediatric emergency airway simulation. J Grad Med Educ. 2012; 4(3): 312-316. http://dx.doi.org/10.4300/JGME-D-11-00123.1

[13] Ishman SL, Benke JR, Johnson KE, et al. Blinded evaluation of interrater reliability of an operative competency assessment tool for direct laryngoscopy and rigid bronchoscopy. Arch Otolaryngol Head Neck Surg. 2012; 138(10): 916-922. PMid:22986639 http://dx.doi.org/10.1001/2013.jamaoto.115

[14] VanHeest A, Kuzel B, Agel J, et al. Objective structured assessment of technical skill in upper extremity surgery. J Hand Surg. 2012; 37A: 332-337. PMid:22281169 http://dx.doi.org/101016/j.jhsa.2011.10.050

[15] Tun JK, Granados A, Mavroveli S, et al. Simulating various levels of clinical challenge in the assessment of clinical procedure competence. Ann Emerg Med. 2012; 60(1): 112-120. PMid:22542307 http://dx.doi.org/10.1016/j.annemergmed.2012.01.036

[16] Griffin GR, Hoesli R, Thorne MC. Validity and efficacy of a pediatric airway foreign body training course in resident education. Ann Otol Rhinol Laryngol. 2011; 120(10): 635-640. PMid:22097148

[17] Jabbour N, Reihsen T, Sweet RM, et al. Psychomotor skills training in pediatric airway endoscopy simulation. J Otolaryngol Head Neck Surg. 2011; 145(1): 43-50. PMid:21493270 http://dx.doi.org/10.1177/0194599811403379

[18] Francis HW, Masood H, Chaudhry KN, et al. Objective assessment of mastoidectomy skills in the operating room. Otol Neurotol. 2010; 31(5): 759-765. PMid:20517169 http://dx.doi.org/10.1097/MAO.0b013e3181e3d385

[19] Lin SY, Laeeq K, Ishii M, et al. Development and piloting of a feasible, reliable, and valid operative competency assessment tool for endoscopic sinus surgery. Am J Rhinol Allergy. 2009; 23: 354-359. http://doi.org/10.2500/ajra.2009.23.3275

[20] Chipman JG, Schmitz CC. Using objective structured assessment of technical skills to evaluate a basic skills simulation curriculum for first-year surgical residents. J Am Coll Surg. 2009; 209: 364-370. PMid:19717041 http://dx.doi.org/10.1016/j.jamcollsurg.2009.05.005

[21] The INACSL Board of Directors. Standard VIII: Evaluation of expected outcomes. Clinical Simulation in Nursing. 2011; 7(4S): s18-s19.

[22] Mudumbai SC, Gaba DM, Boulet JR, et al. External validation of simulation-based assessments with other performance measures of third-year anesthesiology residents. Simul Healthc. 2012; 7: 73-80. PMid:22374230 http://dx.doi.org/10.1097/SIH.0b013e31823d018a

[23] Heiner JS, Gabot MH. Airway management. In: Nagelhout JJ, Karen P, eds. Nurse Anesthesia. 5th ed. St. Louis, MO: Elsevier Saunders; 2014: 423-469.

[24] Butterworth JF, Mackey DC, Wasnick JD. Airway management. In: Butterworth JF, Mackey DC, Wasnick JD, eds. Morgan and Mikhail's Clinical Anesthesiology. 5th ed. United States: McGraw Hill Companies, Inc.; 2013: 309-341. PMid:22729487

[25] Henderson J. Airway management in the adult. In: Miller, RD, Erikson LI, Fleisher, LA, Wiener-Kronish JP, Young WL, eds. Miller's Anesthesia. Vol. 2. 7th ed. Philadelphia, PA: Churchill Livingstone Elsevier; 2010. 1573-1610 p. http://dx.doi.org/10.1016/B978-0-443-06959-8.00050-9