

# Development and Validation of an Objective Structured Assessment of Technical Skill (OSATS) Tool for Basic Optical Stylet Tracheal Intubation in a Clinical Setting

Zhiwei Tao<sup>1,\*</sup>, Weixing Li<sup>1,\*</sup>, Lili Feng<sup>1,\*</sup>, Yilei Shen<sup>1</sup>, Jiali Jiao<sup>2</sup>, Wenxian Li<sup>1</sup>, Yuan Han<sup>1</sup>, Danyun Fu<sup>1</sup>

<sup>1</sup>Department of Anesthesiology, Eye & ENT Hospital of Fudan University, Shanghai, People's Republic of China; <sup>2</sup>Center for Biomedical Data Science, Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Yuan Han; Danyun Fu, Department of Anesthesiology, Eye & ENT Hospital of Fudan University, Xuhui District, Shanghai, 200031, People's Republic of China, Tel +86 021 643 771 34, Fax +86 021 643 77151, Email yuan.han@fdeent.org; fudanyun0929@163.com

**Purpose:** Optical stylets are increasingly used by anesthesiologists to effectively manage difficult airways in ear, nose and throat (ENT) surgeries; however, no dedicated assessment tool for this technique currently exists. This study aims to develop and test the validity and reliability of a new objective structured assessment of technical skill (OSATS) tool for optical stylet tracheal intubation.

**Participants and Methods:** A new OSATS-based tool for optical stylet intubation was developed at the Airway Management Center of Fudan University, Shanghai, China. A total of 30 anesthesiologists were recruited and divided into three groups based on prior experience with optical stylets: novice (0–5 procedures), intermediate (11–50), and expert ( $\geq 100$ ), with 10 in each group. Each participant performed basic optical stylet intubation on a patient with normal airway in a clinical setting. The tool was assessed using video recordings to evaluate internal consistency, intrarater and interrater reliability, and construct validity. Construct validity was evaluated using a known-groups comparison approach, based on predefined levels of prior optical stylet experience (novice, intermediate, expert). Two raters independently assessed the videos, with one rater performing a repeat assessment to evaluate intrarater reliability.

**Results:** Cronbach's alpha was 0.92, indicating high internal consistency. The intraclass correlations (ICCs) for intrarater and interrater reliability were 0.97 and 0.94, and the mean kappa coefficients were 0.84 and 0.72, respectively. Mean total scores differed significantly across the three groups: novice ( $10.50 \pm 3.24$ ), intermediate ( $18.30 \pm 4.12$ ), and expert ( $25.95 \pm 2.06$ ) ( $P < 0.001$ ), demonstrating construct validity through known-groups comparison.

**Conclusion:** This study demonstrates that the scoring tool has satisfactory reliability and validity for assessing basic optical stylet intubation skills, with the ability to discriminate between different levels of expertise. This scoring tool may serve as a useful formative assessment tool for anesthesiologists in training.

**Keywords:** ENT, optical stylet, tracheal intubation, anesthesia, assessment

## Introduction

In recent years, optical stylets (eg, Bonfils, Shikani) have been increasingly applied among anesthesiologists for tracheal intubation in various clinical settings. Their flexible design also facilitates difficult airways management in patients with limited neck extension,<sup>1</sup> reduced mouth opening,<sup>2</sup> large epiglottic cyst<sup>3</sup> or those undergoing head and neck surgery.<sup>4</sup> Compared with conventional laryngoscopy, optical stylets have been associated with higher first-attempt success rate, improved hemodynamic stability, and fewer complications such as mucosal trauma, soft palate injury and sore throat.<sup>5</sup> While awake fiberoptic intubation remains the gold standard for difficult airways, optical stylets offer the advantages of



shorter intubation time and less training.<sup>6</sup> Given these benefits, anesthesiologists are encouraged to master this technique to enhance airway management. However, the learning process may be impeded by the lack of experienced instructors and standardized assessment tools.

Currently, the assessment of optical stylet intubation skills in most training programs relies on subjective and unstructured clinical observation by senior anesthesiologists, with evaluation criteria limited to basic outcomes such as success rate, intubation time, or complication rates.<sup>4,5</sup> Although these measures reflect overall clinical performance, they do not provide step-specific criteria for evaluating technical proficiency or ensure objective and consistent feedback. Despite the general improvement in performance with experience, standardized methods for objectively assessing skill acquisition are lacking, resulting in largely informal and inconsistent training across institutions. Furthermore, existing intubation assessment instruments have primarily been developed for laryngoscopy<sup>7</sup> or fiberoptic techniques<sup>8</sup> and are not designed for the unique insertion methods of optical stylets. Given these procedural differences, a dedicated and validated assessment tool is needed to support structured training and competency development in clinical practice.

To address this gap, we designed a novel objective structured assessment of technical skill (OSATS) tool for basic optical stylet tracheal intubation. The OSATS framework has been widely applied and validated in procedural and surgical education, combining task-specific checklists with global rating scales to provide structured and reproducible evaluation of technical skills.<sup>9</sup> Compared with global rating scales or simple outcome-based measures, OSATS enables comprehensive evaluation of technical performance and facilitates targeted formative feedback.<sup>9</sup> Such an instrument can be used to periodically measure clinicians' performance, assess skill retention, and provide feedback on strengths and weaknesses.<sup>10</sup> It may also be integrated into training programs to support formative assessment, monitor skill progression, and identify trainees requiring additional supervision or remediation. By supporting structured training and competency-based education, this tool aims to improve anesthesiologists' optical stylet intubation skills, with the expectation that improved technical proficiency will ultimately contribute to better patient outcomes.

This study aimed to develop and evaluate the reliability and construct validity of an OSATS-based tool for optical stylet intubation in a clinical setting. We hypothesized that the tool would demonstrate adequate intrarater and interrater reliability, generating reproducible scores when used by independent examiners, and that it would discriminate among novices, intermediates, and experts based on prior optical stylet experience, thereby establishing construct validity.

## Materials and Methods

### Study Design and Setting

This prospective observational study was conducted at the Eye & ENT Hospital of Fudan University in Shanghai, China, in accordance with the STROBE guidelines. The study was conducted from February to December 2021 and included participant recruitment, data collection, and evaluation of the reliability and validity of the assessment tool.

Ethical approval (No. 2021008) was obtained from the Ethics Committee of Shanghai Eye & ENT Hospital of Fudan University on January 15, 2021. The Ethics Committee granted a waiver of patient informed consent because the study posed minimal risk, did not affect routine clinical management, and did not involve the collection of identifiable patient information. All trainees provided written informed consent.

### Population

A total of 30 anesthesiologist volunteers were divided into three groups: novice operators (performed 0–5 times optical stylets intubations,  $n = 10$ ), intermediate operators (11–50,  $n = 10$ ) and expert ( $\geq 100$ ,  $n = 10$ ). Eligible volunteers met the following inclusion criteria: were on a voluntary basis for training, had at least one year practice experience in the department of anesthesiology, had performed at least 100 video laryngoscopic tracheal intubations. In the absence of validated competency standards for optical stylet intubation, we adopted the  $\geq 100$ -procedure benchmark for rigid bronchoscopy reported by Mahmood et al<sup>11</sup> as both are visually guided endoscopic airway techniques requiring fine motor control and spatial orientation. The demographic information of the participants was collected.

## Development of the Assessment Tool

The initial candidate items were generated through a review of published checklists for laryngoscopic endotracheal intubation<sup>7,12</sup> and global rating scale for bronchoscopy performance in clinic,<sup>13</sup> combined with the authors' clinical and teaching experience at the Airway Management Training Center of Fudan University (website: <https://www.linaatp.com>).

To establish content validity, a two-round modified Delphi process was conducted by 10 anesthesiology experts from various affiliated hospitals of Fudan University. In the first round, experts rated each item's relevance on a 9-point Likert scale (1 = not important, 9 = extremely important) and provided qualitative feedback. During this same process, experts also systematically evaluated each item for clarity and appropriateness of phrasing to ensure face validity. Disagreements were addressed using a predefined consensus criterion: items that failed to achieve at least 80% agreement (ratings  $\geq 7$ ) in the first round were revised based on expert feedback and reassessed in the second round; items that still failed to achieve 80% agreement after the second round were excluded, while items meeting the 80% threshold in either round were retained. For quantitative assessment, item-level content validity indices (I-CVI) were calculated as the proportion of experts rating an item  $\geq 7$ . The scale-level content validity index (S-CVI/Ave) was then computed as the average of all I-CVIs across retained items.

At the end of the process, the final tool consisted of 25 items with a maximum score of 28 points (see [Supplementary Table 1](#)). The tool was constructed to be as specific as possible, with items classified into two categories: optimization maneuvers and intubation procedures (including the ability to identify anatomical landmarks and hand-eye coordination). Although the items were organized into these content categories for instructional clarity, the instrument was conceptually designed as a unidimensional scale assessing overall technical competence in optical stylet intubation. Most items were scored dichotomously: a score of "1" was given on the basis of a "yes" response, and a score of "0" was given on the basis of a "no" response. Three items were scored trichotomously: 0 points (not done), 1 point (partially done), and 2 points (correctly performed).

Prior to formal implementation, a pilot study was conducted with a small sample of recordings to evaluate tool clarity and feasibility. Based on pilot results, ambiguous items were revised or removed. Subsequently, both raters—experienced airway management instructors—underwent calibration on the finalized tool, which included a 2-hour training session, independent scoring of pilot videos, and consensus meetings to resolve discrepancies.

## Assessment Procedure

Each participant was asked to perform intubation on one patient in the clinic after receiving a 10-minute training on basic techniques of optical stylet intubation. To minimize variability, standardized protocols were implemented across various aspects, including patient inclusion criteria, anesthesia protocols, assistant responsibilities, and the management of failed intubation, as detailed below. Patients were included if they were ASA grade I or II with normal airway anatomy. Patients were excluded if they were with a known history of difficult intubation, a small mouth opening ( $< 3$  cm), a Mallampati score of  $\geq 3$ , a body mass index  $> 30$ , and/or an increased risk of gastric aspiration. The consultants supervised all intubations performed by novice participants. If intubation was not completed within 3 attempts or 3.5 minutes from the start of the insertion into the mouth, the operation was deemed failure. An attempt was stopped if desaturation ( $\text{SpO}_2 < 90\%$ ) occurred before completing intubation. An attempt was defined as the time from the insertion of an optical stylet into the oral cavity and ended when the stylet was drawn from the oral cavity. If the operation failed, the participants adopted video laryngoscopic intubation as an alternative, or the consultants intervened to complete the intubation to prevent harm to the patient. To minimize variability and ensure the OSATS tool measured only participant skill, we standardized the assistant's role. A single experienced anesthetic nurse served as the assistant for all participants. The assistant remained passive, performing maneuvers (eg, chin lift) only when explicitly requested, and was prohibited from providing proactive assistance, verbal cues, or technical suggestions. Standardization training for the assistant ensured consistent response time and physical maneuvers across all cases.

To further address bias, all procedures were video-recorded and independently evaluated by two blinded experts. Appropriate measures were implemented to ensure patient confidentiality in all video recordings, which were used solely for research purposes and stored securely with restricted access. Two video angles (external view and monitor display)

were merged into a picture-in-picture composite for assessment. Two raters independently scored the procedures to assess interrater reliability, with one rater repeating the scoring after four weeks to evaluate intrarater reliability. They reviewed anonymized intubation recordings and filled out the 25 items of the assessment form.

## Outcomes

The outcomes included: (1) Internal consistency, assessed using Cronbach's alpha, to evaluate the overall reliability of the tool; (2) Intrarater and interrater reliability, analyzed using intraclass correlation coefficients (ICC), to determine consistency within and between raters; (3) Item-level agreement, evaluated using weighted Cohen's kappa coefficients, to assess inter-rater agreement for individual items; (4) Construct validity, examined by comparing the mean total scores, intubation time, and intubation quality scores (total score/second) among operators in the novice, intermediate, and expert groups. Construct validity was assessed using a known-groups comparison approach, based on the predefined hypothesis that more experienced clinicians using optical stylets would perform significantly better than those with less experience (ie, have higher scores). All outcomes were calculated based on the scores from videotaped operations.

## Criteria

Cronbach's alpha  $> 0.70$  was considered acceptable for internal consistency. ICC values  $> 0.75$  were considered indicative of good reliability. Weighted Cohen's kappa coefficients  $> 0.60$  were considered acceptable for item-level agreement.

## Sample

To achieve 80% power to detect a true kappa value of 0.90 in a test of  $H_0: \text{kappa} = 0.50$  versus  $H_1: \text{kappa} \neq 0.50$  with a 2-sided alpha level of 5%, a sample size of 30 participants was sufficient in a test for consistency between raters via the kappa statistic. PASS version 11 (NCSS LLC, Kaysville, Utah) was used to calculate the sample size.

## Statistical Analysis

Continuous variables are presented as mean  $\pm$  standard deviation (SD), and categorical variables as numbers and percentages. Comparisons of continuous variables among the three groups were performed using one-way analysis of variance (ANOVA). When a significant overall difference was observed, post hoc pairwise comparisons were conducted using Tukey's honestly significant difference (HSD) test to adjust for multiple comparisons. Categorical variables were compared using the chi-square test. Two-way ANOVA was additionally used to explore the effects of intubation experience level, clinical role, and working experience on OSATS scores. Statistical analyses were performed using SPSS version 27.0 (IBM Corp., Chicago, IL, USA). A P value of  $< 0.05$  was considered to indicate statistical significance.

## Results

The study sample included 30 volunteers in anesthesiology departments from different hospitals. A total of 30 participants were recruited at the novice, intermediate, and expert levels based on performed optical stylet intubation number with 10 participants in each group. All participants completed the study and 30 (100%) videotaped intubation procedures were received and analyzed. [Table 1](#) shows the demographic data of the participants (age, sex, clinical role and working experience in anesthesiology). No statistically significant differences were observed in baseline demographic characteristics among the three groups.

## Reliability

The tool demonstrated high internal consistency, with a Cronbach's alpha of 0.92. The ICCs were 0.97 (95% CI 0.94 to 0.99) for intrarater reliability and 0.94 (95% CI 0.86 to 0.97) for interrater reliability (both almost perfect). [Supplementary Table 2](#) shows weighted Cohen's kappa coefficients for the intrarater and interrater reliability of 25 individual items. Mean weighted Cohen's kappa coefficients for intrarater and interrater reliability were 0.84 (95% CI 0.77 to 0.90) and 0.72 (95% CI 0.63 to

**Table 1** Demographics of the Participants

Category	Novice (n = 10)	Intermediate (n = 10)	Expert (n = 10)	P value
Sex (male/female)	4/6	4/6	3/7	1.000
Age (21–30/31–40/41–50, years)	2/7/1	1/8/1	1/7/2	1.000
Clinical role (resident/attending/professor)	2/7/1	2/8/0	2/7/1	1.000
Working experience (1–3/3–5/5–10/> 10, years)	1/3/2/4	1/3/4/2	2/3/3/2	0.959

**Notes:** Data are presented as n. P-values were calculated using Fisher's exact test.

0.80), respectively. Individual items showed variable inter-rater reliability, ranging from moderate to almost perfect ( $\kappa = 0.41$ – $1.00$ ).

## Content Validity

I-CVI ranged from 0.80 to 1.00. The S-CVI/Ave was 0.94, indicating excellent content validity. These results suggest that the items were deemed relevant and representative by the expert panel, confirming that the tool adequately covers the key domains of optical stylet intubation.

## Construct Validity

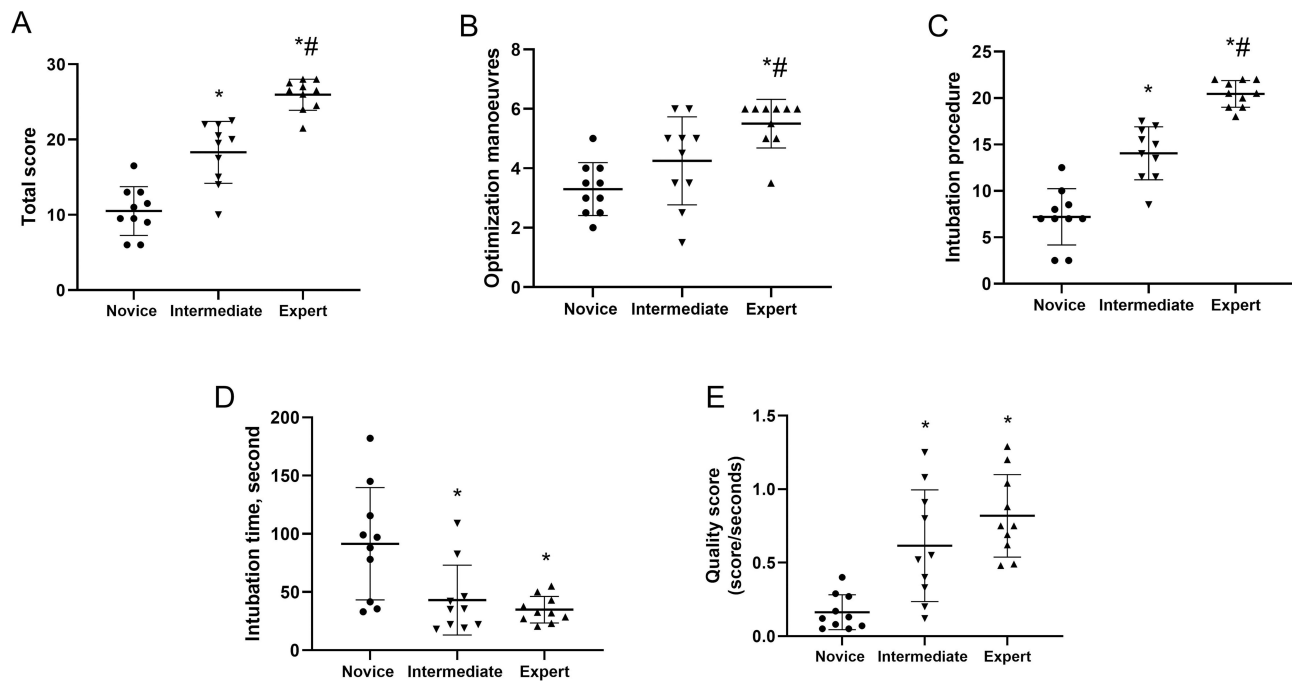
The mean total scores of the operators in the novice, intermediate, and expert groups were  $10.50 \pm 3.24$ ,  $18.30 \pm 4.12$ , and  $25.95 \pm 2.06$  points, respectively, as shown in Table 2 and Figure 1A. The total scores were significantly different among the three groups, and post hoc tests revealed that the intermediate operators performed significantly better than novice operators ( $18.30 \pm 4.12$  versus  $10.50 \pm 3.24$ ,  $P < 0.001$ ), and the expert operators also performed significantly better than the intermediate operators did ( $25.95 \pm 2.06$  versus  $18.30 \pm 4.12$ ,  $P < 0.001$ ). As shown in Table 2 and Figure 1B, the expert operators scored significantly higher than the intermediate operators for the optimization manoeuvres. However, no difference was found between the novice and intermediate operators. Furthermore, the data revealed significant differences among the three groups (novice, intermediate and expert) for the intubation procedure ( $P < 0.001$ , Table 2 and Figure 1C).

Furthermore, the data demonstrated that the mean intubation time of operators in the novice, intermediate, and expert groups were  $91.45 \pm 48.27$ ,  $43.10 \pm 30.05$  and  $34.90 \pm 11.41$  seconds, respectively (Table 2 and Figure 1D). Intermediate and expert operators spent less time completing intubation than novice operators did ( $P = 0.009$  and  $P = 0.002$ , respectively); however, no difference was found between intermediate and expert operators. The trend of quality score (score/second) data was consistent with the intubation time (Table 2 and Figure 1E).

**Table 2** Comparisons of Total Score, Optimization Manoeuvres and Intubation Procedure, Intubation Time, and Quality Score

Group	Total Score (28 Points)	Optimization Manoeuvres (6 Points)	Intubation Procedure (22 Points)	Intubation Time (s)	Quality Score (Score/s)
Novice (n=10)	$10.50 \pm 3.24$	$3.30 \pm 0.89$	$7.20 \pm 3.03$	$91.45 \pm 48.27$	$0.16 \pm 0.12$
Intermediate (n=10)	$18.30 \pm 4.12$	$4.3 \pm 1.48$	$14.05 \pm 2.86$	$43.10 \pm 30.05$	$0.62 \pm 0.38$
Expert (n=10)	$25.95 \pm 2.06$	$5.50 \pm 0.82$	$20.45 \pm 1.44$	$34.90 \pm 11.41$	$0.82 \pm 0.28$
Overall P value	$< 0.001$	$< 0.001$	$< 0.001$	0.002	$< 0.001$
Novice vs Intermediate	$< 0.001$	0.150	$< 0.001$	0.009	0.003
Novice vs Expert	$< 0.001$	$< 0.001$	$< 0.001$	0.002	$< 0.001$
Intermediate vs Expert	$< 0.001$	0.044	$< 0.001$	0.849	0.254

**Notes:** Values are presented as mean  $\pm$  SD. Overall P values were calculated using one-way ANOVA. Post-hoc pairwise comparisons were performed using Tukey's honestly significant difference (HSD) test. The time to intubation was recorded within 3 attempts or 3.5 minutes (an attempt was defined as the time from the insertion of an optical stylet into the oral cavity and ended when the stylet was drawn from the oral cavity).



**Figure 1** Performance of participants in tracheal intubation using optical stylests. (A) Total score; (B) Optimization manoeuvres; (C) Intubation procedure; (D) Intubation time (s); (E) Quality score. A higher score indicates better performance.

**Notes:** Comparisons among the individual groups concerning total score, optimization manoeuvres and intubation procedure, intubation time and quality score. Higher score depicts better performance. \* $P < 0.05$  vs novice. # $P < 0.05$  vs intermediate. The scores are shown as mean  $\pm$  SD.

## Exploratory Analysis Adjusted for Clinical Role and Working Experience

To further explore the influence of demographic variables, two separate two-way ANOVA models were constructed. In Model 1, intubation experience level (novice, intermediate, expert) and clinical role (resident, attending, professor) were included as fixed factors. In Model 2, intubation experience level and working experience (1–3, 3–5, 5–10, >10 years) were entered as fixed factors. OSATS total score was used as the dependent variable in both models. As shown in Table 3, in both models, the main effect of intubation experience remained highly significant (Model 1:  $F(2, 22) = 31.97$ ,  $P < 0.001$ , partial  $\eta^2 = 0.744$ ; Model 2:  $F(2, 18) = 57.35$ ,  $P < 0.001$ , partial  $\eta^2 = 0.864$ ), confirming that procedure-specific experience was the primary determinant of technical performance. The main effect of clinical role was not significant ( $F(2, 22) = 0.01$ ,  $P = 0.986$ , partial  $\eta^2 = 0.001$ ). Similarly, working experience did not reach statistical significance ( $F(3, 18) = 1.00$ ,  $P = 0.414$ , partial  $\eta^2 = 0.143$ ), although the observed effect size suggests that a modest influence cannot be entirely excluded given the limited sample size. No significant interaction effects were observed in either model (Model 1:  $P = 0.556$ ; Model 2:  $P = 0.189$ ), indicating that the effect of intubation experience on OSATS performance was consistent across different clinical roles and levels of working experience. In both models, the estimated marginal means demonstrated a consistent stepwise increase in OSATS scores with increasing levels of intubation experience, further supporting the ability of the OSATS tool to distinguish between different levels of technical skill (Table 4).

**Table 3** Effects of Intubation Experience, Clinical Role, and Working Experience on OSATS Scores

Model	Effect	F (df1, df2)	P value	Partial $\eta^2$
Model 1 (Intubation experience $\times$ Clinical role)	Intubation experience	$F(2, 22) = 31.97$	<0.001	0.744
	Clinical role	$F(2, 22) = 0.01$	0.986	0.001
	Intubation experience $\times$ Clinical role	$F(3, 22) = 0.71$	0.556	0.088
Model 2 (Intubation experience $\times$ Working experience)	Intubation experience	$F(2, 18) = 57.35$	<0.001	0.864
	Working experience	$F(3, 18) = 1.00$	0.414	0.143
	Intubation experience $\times$ Working experience	$F(6, 18) = 1.66$	0.189	0.356

**Notes:** Model 1  $R^2 = 0.824$ . Model 2  $R^2 = 0.890$ . Partial  $\eta^2$  represents effect size.  $\eta^2 \geq 0.01$  indicates a small effect,  $\geq 0.06$  a medium effect, and  $\geq 0.14$  a large effect.

**Table 4** Estimated Marginal Means of OSATS Scores by Intubation Experience Level

Model	Intubation Experience	Mean (SE)	95% Confidence Interval
Model 1 (Adjusted for clinical role)	Novice	9.60 (1.47)	6.55–12.64
	Intermediate	19.03 (1.36)	16.21–21.85
	Expert	26.14 (1.47)	23.10–29.19
Model 2 (Adjusted for working experience)	Novice	10.41 (1.10)	8.11–12.71
	Intermediate	17.96 (1.10)	15.66–20.26
	Expert	26.08 (0.98)	24.03–28.14

**Notes:** Estimated marginal means represent predicted OSATS scores after controlling for demographic variables. Model 1 controls for clinical role (resident, attending, professor). Model 2 controls for working experience (1–3, 3–5, 5–10, >10 years).

## Discussion

### Main Findings

This is the first study to develop and validate an OSATS-based assessment tool for optical stylet intubation in clinical practice. Our results demonstrated that the tool has good reliability and validity when used by two trained raters, with high intrarater and interrater reliability, high internal consistency ( $\alpha = 0.92$ ), and clear construct validity in discriminating operators across three experience levels.

### Reliability and Validity

The tool showed robust reliability, with an overall weighted kappa of 0.72 and high ICC values. Several individual items showed moderate inter-rater agreement ( $\kappa = 0.4$ – $0.6$ ). These items involved subtle and dynamic technical maneuvers that require interpretation of fine motor control and spatial relationships. Such procedural micro-skills are inherently difficult to assess consistently, particularly in video-based evaluations. Despite the moderate agreement observed for some technically complex items, the overall scale exhibited clear discriminative validity across experience levels, confirming its ability to distinguish between novice, intermediate, and expert performers. Items with perfect agreement ( $\kappa = 1.00$ ) were safety-critical steps that reflect universal compliance with mandatory protocols rather than item triviality, and are essential for ensuring the tool captures both routine safety behaviors and more variable technical skills.

### Construct Validity

The Delphi method was used to develop a standardized checklist covering key technical steps of optical stylet intubation, establishing face and content validity. To assess construct validity, we hypothesized that greater intubation experience would be associated with higher performance scores.<sup>11</sup> Our results confirmed a significant positive relationship between optical stylet intubation experience and OSATS scores, independent of job title or working years. Two-way ANOVA further demonstrated that intubation experience had a large, significant effect on performance, whereas clinical role and working experience showed no significant main or interaction effects. Although working experience exhibited a moderate effect size (partial  $\eta^2 = 0.143$ ), it did not reach statistical significance, likely due to limited sample size. The stepwise increase in scores across novice, intermediate, and expert groups reinforces the instrument's discriminative validity. Overall, our findings suggest that skill assessment and training should focus more on procedure-specific experience rather than relying solely on clinical role or working experience as indicators of competence.

### Internal Consistency

A high Cronbach's alpha can sometimes raise concerns about item redundancy. In our study, the internal consistency was high ( $\alpha = 0.92$ ). We interpret this finding as primarily reflecting the sequential nature of the intubation procedure rather than item redundancy. Although procedurally related, each item is designed to capture a distinct skill. For example, "Railroads ETT into the trachea smoothly" assesses avoidance of vocal cord injury, while

“Stabilizes the ETT in situ with one hand” evaluates prevention of accidental dislodgement. The Delphi process ensured content validity and minimized redundancy to the extent possible. More importantly, as established in the foundational OSATS literature (Martin et al<sup>9</sup>), the main sources of error in procedural assessments are rater stringency and case specificity, therefore, inter-rater reliability is considered the more relevant outcome than internal consistency. On this basis, Cronbach’s alpha is reported primarily as supportive evidence of internal consistency, while inter-rater agreement—which was excellent in this study (ICC = 0.94)—remains the primary reliability metric.

## Mixed Scoring Structure

An important methodological feature is the instrument’s mixed scoring structure. All 25 items are equally weighted, contributing equally to the total score without differential coefficients. The scoring method for each item was selected based on the type of performance it assesses: dichotomous (0/1) for binary clinical endpoints without meaningful intermediate states (eg, successful intubation), and trichotomous (0/1/2) for technical quality items that involve gradations of proficiency. The three-level scale distinguishes inadequate, partially adequate, and optimal performance—essential for formative feedback. Such mixed scoring formats have been used in prior airway skill assessment instruments to improve discrimination while maintaining scoring simplicity.<sup>7,12</sup>

## Intubation Time and Quality Score

While intubation time and quality scores effectively differentiate novices from intermediate and expert operators, they did not distinguish between intermediate and expert trainees. Experts, although taking longer, focus on deliberate anatomical identification and avoiding injury. Intubation time should not be prioritized over careful technique, especially in high-risk cases.<sup>14</sup>

## Generalizability and Practical Applicability

The assessment was conducted in a real-world clinical setting—an approach consistent with the development of other validated tools, such as clinical bronchoscopy.<sup>13</sup> By deconstructing optical stylet intubation into its constituent steps, it promotes consistent instruction and enables detailed, constructive feedback. Its straightforward design minimizes grader subjectivity, enhancing acceptability in busy training programs. An earlier version of the tool has already been used to measure training effects and improve competency in a distance training program,<sup>15</sup> supporting its potential for broader educational application.

## Limitations

This study has several limitations. First, by excluding anticipated difficult airways, we assessed only basic competency; our tool establishes a foundational milestone, but advanced training remains necessary for difficult airway scenarios. Second, we did not assess participants’ prior training quality or recency of practice, relying instead on cumulative procedural numbers to define experience levels, which may introduce heterogeneity within groups. Third, the single-center design and relatively small sample size may limit generalizability; relatedly, the sample size was insufficient to perform factor analysis to empirically confirm the unidimensionality of the instrument—although conceptually designed as a unidimensional scale. Fourth, each participant performed only one intubation, providing a single-point assessment that may not reflect consistent competency. Fifth, the 10-minute training before assessment may have introduced short-term recall bias.

## Future Directions

Multicenter studies with larger samples are needed to validate these findings and to perform factor analysis confirming the instrument’s dimensionality. Future research should incorporate repeated measures per participant, include washout periods to distinguish memory from skill retention, and evaluate the tool’s utility in difficult airway scenarios and its impact on training outcomes. Additionally, incorporating detailed assessments of training quality and learning curve analyses would help validate experience-based competency thresholds.

## Conclusion

In summary, this tool demonstrates high intrarater and interrater reliability and can validly discriminate between operators with different levels of intubation experience in patients with normal airways. This scoring tool may serve as a useful formative assessment tool for anesthesiologists in training. Future research is needed to validate the tool in multicenter settings and among patients with difficult airway.

## Abbreviations

ENT, ear, nose and throat; OSATS, objective structured assessment of technical skill; ICCs, intraclass correlations; I-CVI, item-level content validity indices; S-CVI, scale-level content validity index; SD, standard deviation; ANOVA, one-way analysis of variance; HSD, honestly significant difference.

## Data Sharing Statement

The data are available from the corresponding author upon reasonable request.

## Ethics Approval and Consent to Participate

This study was conducted as a prospective observational study at the Eye & ENT Hospital of Fudan University in Shanghai, China. The study period spanned from February 2021 to December 2021. Ethical approval (No. 2021008) was obtained from the Ethical Committee of Shanghai Eye & ENT Hospital of Fudan University on January 15, 2021. The Ethics Committee granted a waiver of patient informed consent because the study posed minimal risk, did not affect routine clinical management, and did not involve the collection of identifiable patient information. Written informed consent was obtained from all trainee participants prior to their participation in this study.

## Consent for Publication

Written informed consent for publication of their individual details and any accompanying images (eg, in photographs or videos) was obtained from all participants in this study.

## Acknowledgments

Zhiwei Tao, Weixing Li, and Lili Feng are co-first authors study. The authors acknowledge the clinical experts who participated in drafting and revising the assessment tool. The authors also acknowledge the participants from different hospitals in this study for their assistance with the study.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This work was supported by the Continuing Education Program of Shanghai Medical College Fudan University [grant number FDYXYBJ-20222004], the Perioperative Airway Refinement Management Program of the Health Human Resources Development Center, the National Health Commission [grant number RCLX2315001] and the Chinese Alliance of Airway Management Program, Shanghai Medical College Fudan University [grant number IDH6282016/049] to Wenxian Li.

## Disclosure

The authors declare that they have no competing interests in this work.

## References

1. Shih TL, Koay KP, Hu CY, et al. The use of the Shikani video-assisted intubating stylet technique in patients with restricted neck mobility. *Healthcare*. 2022;10(9). doi:10.3390/healthcare10091688
2. Pius J, Ioanidis K, Noppens RR. Use of the novel C-MAC video stylet in a case of predicted difficult intubation: a case report. *A a Pract*. 2019;13(3):88–90. doi:10.1213/XAA.0000000000000998
3. Seo H, Lee G, Ha SI, Song JG. An awake double lumen endotracheal tube intubation using the Clarus video system in a patient with an epiglottic cyst: a case report. *Korean J Anesthesiol*. 2014;66(2):157–159. doi:10.4097/kjae.2014.66.2.157
4. Cheng T, Wang LK, Wu HY, et al. Shikani optical stylet for awake nasal intubation in patients undergoing head and neck surgery. *Laryngoscope*. 2021;131(2):319–325. doi:10.1002/lary.28763
5. Liu L, Yue H, Li J. Comparison of three tracheal intubation techniques in thyroid tumor patients with a difficult airway: a randomized controlled trial. *Med Principles Pract*. 2014;23(5):448–452. doi:10.1159/000364875
6. Desai N, Ratnayake G, Onwochei DN, El-Boghdady K, Ahmad I. Airway devices for awake tracheal intubation in adults: a systematic review and network meta-analysis. *Br J Anaesth*. 2021;127(4):636–647. doi:10.1016/j.bja.2021.05.025
7. Kuijpers L, Binkhorst M, Yamada NK, et al. Validation of an instrument for real-time assessment of neonatal intubation skills: a randomized controlled simulation study. *Am J Perinatol*. 2022;39(2):195–203. doi:10.1055/s-0040-1715530
8. Bube S, Dagnaes-Hansen J, Mahmood O, et al. Simulation-based training for flexible cystoscopy - A randomized trial comparing two approaches. *Heliyon*. 2020;6(1):e03086. doi:10.1016/j.heliyon.2019.e03086
9. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273–278. doi:10.1046/j.1365-2168.1997.02502.x
10. Asif H, McInnis C, Dang F, et al. Objective structured assessment of technical skill (OSATS) in the surgical skills and technology elective program (SSTEP): comparison of peer and expert raters. *Am J Surg*. 2022;223(2):276–279. doi:10.1016/j.amjsurg.2021.03.064
11. Mahmood K, Wahidi MM, Osann KE, et al. Development of a tool to assess basic competency in the performance of rigid bronchoscopy. *Ann Am Thorac Soc*. 2016;13(4):502–511. doi:10.1513/AnnalsATS.201509-593OC
12. Hart D, Clinton J, Anders S, et al. Validation of an assessment tool for field endotracheal intubation. *Mil Med*. 2016;181(11):e1484–e1490. doi:10.7205/MILMED-D-15-00570
13. Konge L, Larsen KR, Clementsen P, et al. Reliable and valid assessment of clinical bronchoscopy performance. *Respiration*. 2012;83(1):53–60. doi:10.1159/000330061
14. Hung KC, Tan PH, Lin VC, Wang HK, Chen HS. A comparison of the Trachway intubating stylet and the Macintosh laryngoscope in tracheal intubation: a manikin study. *J Anesth*. 2013;27(2):205–210. doi:10.1007/s00540-012-1491-6
15. Fu D, Li W, Li W, Han Y. Improved skill for tracheal intubation using optical stylets through remote training model: a before and after interventional study. *BMC Med Educ*. 2022;22(1):668. doi:10.1186/s12909-022-03715-x

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

**Dovepress**  
Taylor & Francis Group