International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-026-03605-3

ORIGINAL ARTICLE

Analysis and objective assessment of transoral robotic surgery

Dervishan Sezer[1] · **Doga Demirel**[2] · **Madeleine Leake**[3] · **Mouhsin M. Shafi**[3] · **Suvranu De**[4] · **Floyd Christopher Holsinger**[5] · **Andrés M. Bur**[6] · **Scharukh M. Jalisi**[7,8] · **Ernest Gomez**[7,8]

Received: 22 October 2025 / Accepted: 13 March 2026

Abstract

Purpose Transoral robotic surgery (TORS) has become a well-established surgical technique for the treatment of oropharyngeal cancer, but the significant learning curve and lack of standardized credentialing have resulted in wide variability in surgical outcomes. This study aims to define procedure-specific competence standards for TORS and test whether a hierarchical task analysis (HTA)-derived procedure-based assessment (PBA) distinguishes experience levels. We examined the ability of PBA and Global Evaluative Assessment of Robotic Skills (GEARS) scores to discriminate between novice and experienced surgeons and to assess their association with operative efficiency and margin quality.

Methods We built an HTA by deconstructing the TORS lateral oropharyngectomy into tasks and subtasks. Then, PBA metrics for mucosal incision and deep dissection were developed. Two independent raters scored 40 porcine tongue TORS videos (20 novice, 20 experienced) using PBA and GEARS and recorded global and phase times.

Results Experienced surgeons scored higher on total PBA (39.98 vs 35.35, $p = 0.0055$) and GEARS (22.60 vs 19.63, $p = 0.0009$) and showed less score variability. The largest gaps were lateral tasks: lateral mucosal incision 4.65 vs 3.60 ($p = 0.0015$) and lateral deep dissection 4.58 vs 3.85 ($p = 0.0115$). Margin scores were higher in experienced surgeons (4.38 vs 3.80, $p = 0.0149$). Procedures were faster overall (298.47 s vs 466.43 s, $p = 0.0003$) with shorter mucosal incision and deep dissection times.

Conclusions An HTA-derived PBA reliably differentiates TORS expertise, aligns with speed and margin quality, and identifies lateral tasks as high-yield training targets. These metrics support standardized training, assessment, and integration into VR simulation for competency-based credentialing.

Keywords Transoral robotic surgery · Training · Assessment · Task analysis · Skill evaluation

Introduction

The increasing incidence of oropharyngeal squamous cell carcinoma (OPSCC) has been strongly associated with high-risk human papillomavirus (HPV) infection, particularly HPV-16 [1–3]. This epidemiologic shift has led to a paradigm change in clinical management, emphasizing function-preserving treatment strategies for younger and generally healthier patients. Within this context, transoral robotic surgery (TORS) has emerged as a transformative surgical modality, enabling minimally invasive en bloc resection of oropharyngeal tumors [4, 5]. Compared to open

---

✉ Doga Demirel
doga@ou.edu

[1] Data Science and Analytics Institute, University of Oklahoma, Norman, OK, USA

[2] School of Computer Science, University of Oklahoma, Devon Energy Hall, 110 W. Boyd St, Norman, OK, USA

[3] Beth Israel Deaconess Medical Center, Berenson-Allen Center for Noninvasive Brain Stimulation, Department of Neurology, Harvard Medical School, Boston, MA, USA

[4] College of Engineering, Florida A&M University-Florida State University, Tallahassee, FL, USA

[5] Department of Otolaryngology-Head and Neck Surgery, Stanford School of Medicine, Stanford, CA, USA

[6] Department of Otolaryngology-Head and Neck Surgery, University of Kansas School of Medicine, Kansas City, KS, USA

[7] Division of Otolaryngology–Head and Neck Surgery, Beth Israel Deaconess Medical Center, Boston, MA, USA

[8] Department of Otolaryngology–Head and Neck Surgery, Harvard Medical School, Boston, MA, USA

approaches such as lip-split mandibulotomy, TORS is associated with decreased postoperative morbidity, improved swallowing function, and reduced rates of tracheostomy and feeding tube dependence [6, 7].

Early surgical approaches to the oropharynx included extensive soft tissue dissection and mandibular disruption, which can contribute to prolonged tracheostomy and gastrostomy tube dependence in addition to functional deficits such as dysphagia and speech impairment [7, 8]. TORS provides a less invasive approach to oropharyngeal tumors with enhanced visualization and dexterity through stereoscopic magnification and wrist-articulating miniature instruments. TORS may be associated with decreased operative times, lower complication rates, improved swallowing outcomes, and shorter hospital stays compared to open approaches [6, 9]. The use of TORS for the surgical treatment of HPV-related OPSCC has facilitated treatment de-intensification strategies such as the reduction in the dose and volume of adjuvant radiation therapy. Emerging data suggest that TORS, when combined with risk-adapted adjuvant therapy, can achieve excellent oncologic outcomes while reducing radiation field sizes and total doses compared to primary chemoradiation protocols, thereby minimizing long-term toxicity in this favorable-risk population [10].

As with any surgical technique, TORS has an associated learning curve. Published studies estimate that approximately 20–30 cases are required for a surgeon to achieve procedural proficiency and consistent oncologic outcomes [9, 11]. Complication rates during an early experience may be higher, but overall TORS demonstrates favorable safety [12], with systematic reviews reporting major complication rates—including hemorrhage, fistula, and prolonged gastrostomy or tracheostomy dependence—ranging from 5 to 10% [13–15]. However, there is considerable variability in achieving negative surgical margins across institutions. A systematic review and meta-analysis of published TORS series reported an overall positive margin rate of 7.8% [16]. However, an analysis of the National Cancer Database (NCDB) reported that 16.9% of patients treated with TORS for OPSCC have close or positive surgical margins, with positive margin rates being generally higher in lower-volume centers [17]. Surgical quality discrepancies are mirrored in clinical trials. In ECOG 3311, a randomized phase II de-intensification trial that included a rigorous surgeon credentialing process, the reported major surgical complication rate was 5.3% [18]. Conversely, the ORATOR2 trial reported a 12% rate of serious surgical morbidity—including two treatment-related deaths—in the TORS arm, leading to early closure of the surgical arm [19]. These contrasting findings in high-profile clinical trials highlight the variability in surgical outcomes across institutions and emphasize the importance of surgeon training, credentialing, and skill maintenance.

Hierarchical task analysis (HTA) is a well-established methodology within human factors and ergonomics for systematically decomposing complex tasks into goals, subgoals, and sequential operations [20]. Since its introduction, HTA has been widely adopted for analyzing safety–critical activities, supporting training design, workflow evaluation, and human–system interaction studies [21–24]. In healthcare, HTA has become a common framework for modeling clinical workflows and complex procedures because it provides a transparent representation of task structure and decision pathways [25]. Within surgery, HTA has been used to formally characterize operative procedures, identify critical steps and errors, and derive procedure-specific training and assessment metrics [26–29]. Prior work has shown that HTA-derived task structures can be used to construct objective assessment tools, guide simulator development, and support structured surgical training curricula [30–32]. In simulation-based surgical education, HTA provides the foundational framework for defining procedural steps, mapping performance metrics to each step, and ensuring that simulator tasks reflect real operative workflows [28, 33]. For these reasons, HTA has been widely adopted as a methodological basis for developing procedure-specific performance assessments and simulation training systems in surgery. Building on this established approach, we applied HTA to deconstruct TORS into structured tasks and subtasks, enabling the development of objective procedure-based assessment (PBA) metrics and supporting standardized training and evaluation of surgical performance.

The purpose of this study is to identify the benchmarks for competence required to perform TORS in the treatment of oropharyngeal cancer, with a focus on training and skill development. To address the challenges associated with TORS, a comprehensive HTA was conducted, leading to the development of objective, PBA metrics. Despite the complexity of TORS procedures, no prior published work has presented a hierarchical task analysis specific to TORS.

Materials and methods

Hierarchical task analysis

In this study, an HTA was used to identify and describe every task and procedural step involved in transoral lateral oropharyngectomy [34]. While the procedure had been previously described and broken down into component steps [35], here we used a structured approach to objectively parse TORS into a hierarchy of tasks and subtasks while systematically detailing their relationships and the sequential steps needed to achieve the surgical goals. Task steps are shown as rectangles with forward arrows indicating linear surgical progression. Optional branches are depicted as diamonds, marking points where surgeons may choose alternative paths based on preference or patient needs. The complete HTA can be found in the supplementary material.

The HTA was developed by reviewing real TORS surgical videos, narrative walkthroughs, and textbook materials [36, 37]. Tasks were organized into a hierarchical task tree reflecting their chronological order. A head and neck surgical oncologist with over a decade of TORS experience provided the initial narrative walkthrough. Two additional senior oncologists, each with more than twenty years of experience and involvement since TORS's early development, reviewed and refined the structure. The final HTA provides a structured guide to the procedure and the foundation for objective PBA and detailed time-performance analysis.

The HTA organizes the surgery into three main phases (see Fig. 1): Setup and Exposure, Tumor Resection, and Post-Resection Procedures. Setup and Exposure cover room and equipment preparation, airway, eye and tooth protection, tongue retraction, mouth gag placement, and robot setup. Tumor Resection includes mucosal incisions, deep dissection, and tumor removal. Post-Resection Procedures finalize the case through hemostasis, reconstruction, enteral access, stomach decompression, specimen orientation, and airway management. Hemostasis is shown with a bidirectional arrow because the surgeon may return to it at any point before completing the operation.

Procedure

The subtasks of the full procedure including Setup and Exposure and Post-Resection Procedures are shown in the supplementary material.

Tumor resection (see Fig. 2) starts by marking mucosal incision points on the oropharynx, then completing the planned superior, anterior, medial, and lateral incisions. The superior incision may include the tonsillar pillars or go directly over the palatine tonsil. Monopolar energy in cut mode is used for these superficial cuts. At the tongue base, the mucosa is incised at least 1 cm anterior to the visible tumor. Medial cuts open the posterior pharyngeal wall; lateral cuts release the pterygomandibular raphe and the posterior floor of the mouth.

After superficial margins are defined, deep dissection begins. If the soft palate is involved, the palatoglossus and palatopharyngeus are divided until the pharyngeal constrictors are reached. For cases limited to standard palatine tonsillectomy, dissection follows the subcapsular plane using coagulation. Anteriorly, tongue-base muscles are resected with attention to depth and margins. Medially, constrictor fibers are divided, stopping either at the prevertebral fascia for full-thickness resections or leaving a thin constrictor layer if full thickness is not required.

Lateral dissection depends on whether a radical tonsillectomy is needed. If so, palatoglossus fibers are divided to expose the medial pterygoid and parapharyngeal fat. Deeper muscles including inferior palatoglossus, styloglossus, and stylopharyngeus are divided. Sacrifice of the glossopharyngeal and lingual nerves is considered if necessary for adequate margins.

Once fully mobilized, the tumor specimen is released and removed. It is inspected, and if margins are uncertain, additional resection is performed.

Metrics

Currently, there are no widely accepted procedure-specific performance metrics for measuring the quality or tasks of TORS. A Procedure-Based Assessment (PBA) for the TORS procedure was developed with the help of experienced attending surgeons in the field and the data gathered from the HTA. Each hierarchical task and its subtasks were examined separately. The PBA metric scoring system incorporates performance components that account for optimal and suboptimal actions, as well as the time required to complete the procedure and its tasks.

PBA uses Likert-style scores (e.g., 1 to 5) for each task and subtask (see Fig. 3). In any task, the optimal action is given the highest score. For example, in the "Mucosal Incision" metric, the highest score (e.g., 5) reflects "Ideal incision location along the entire length," showing precise and safe execution. 3 indicates a suboptimal but acceptable outcome, such as an "Adequate incision with minor deviations." Critical errors, including causing damage or incomplete incision, receive the lowest score (e.g., 1). The "Deep Dissection" and the "Tumor Removal" metrics were scored on the same 5-point Likert scale. The PBA metrics were reviewed and validated by experienced attending head and neck surgical oncologists (EG, AB, CH, SJ), to ensure alignment with clinical standards and procedural expectations.

In addition to the developed PBA metrics, the Global Evaluative Assessment of Robotic Skills (GEARS) is incorporated as a normalized measure for evaluating robotic surgical performance [38]. The "Autonomy" metric was excluded as it was not applicable to the recorded videos. Scores derived from GEARS were validated using the dataset of a prior study to ensure alignment with the PBA metrics and experienced attending surgeon evaluations [39].

Analysis of TORS videos

Two observers graded and timed 40 full-length porcine tongue TORS videos performed on the da Vinci robotic surgical system. These videos were collected during a prior study that developed a porcine tongue simulator for TORS [39].

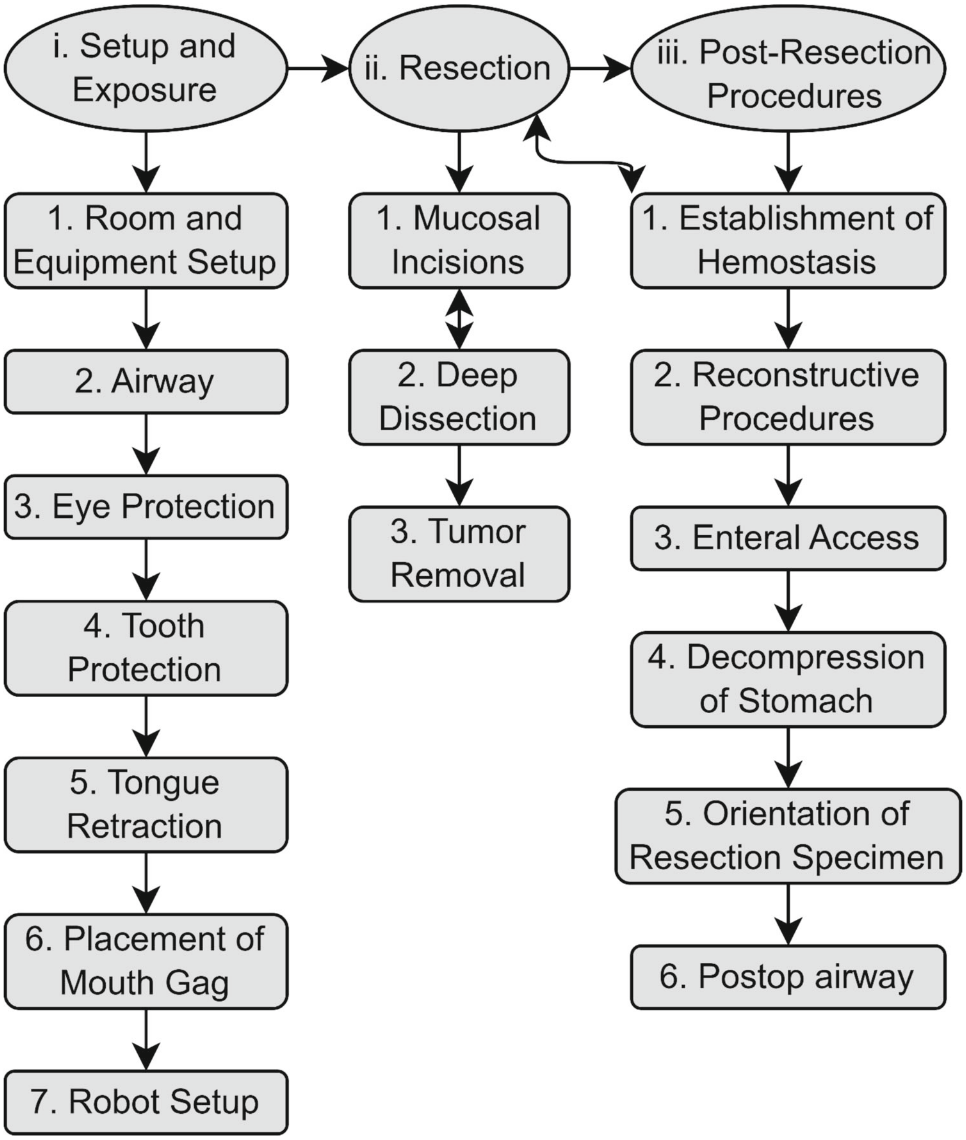
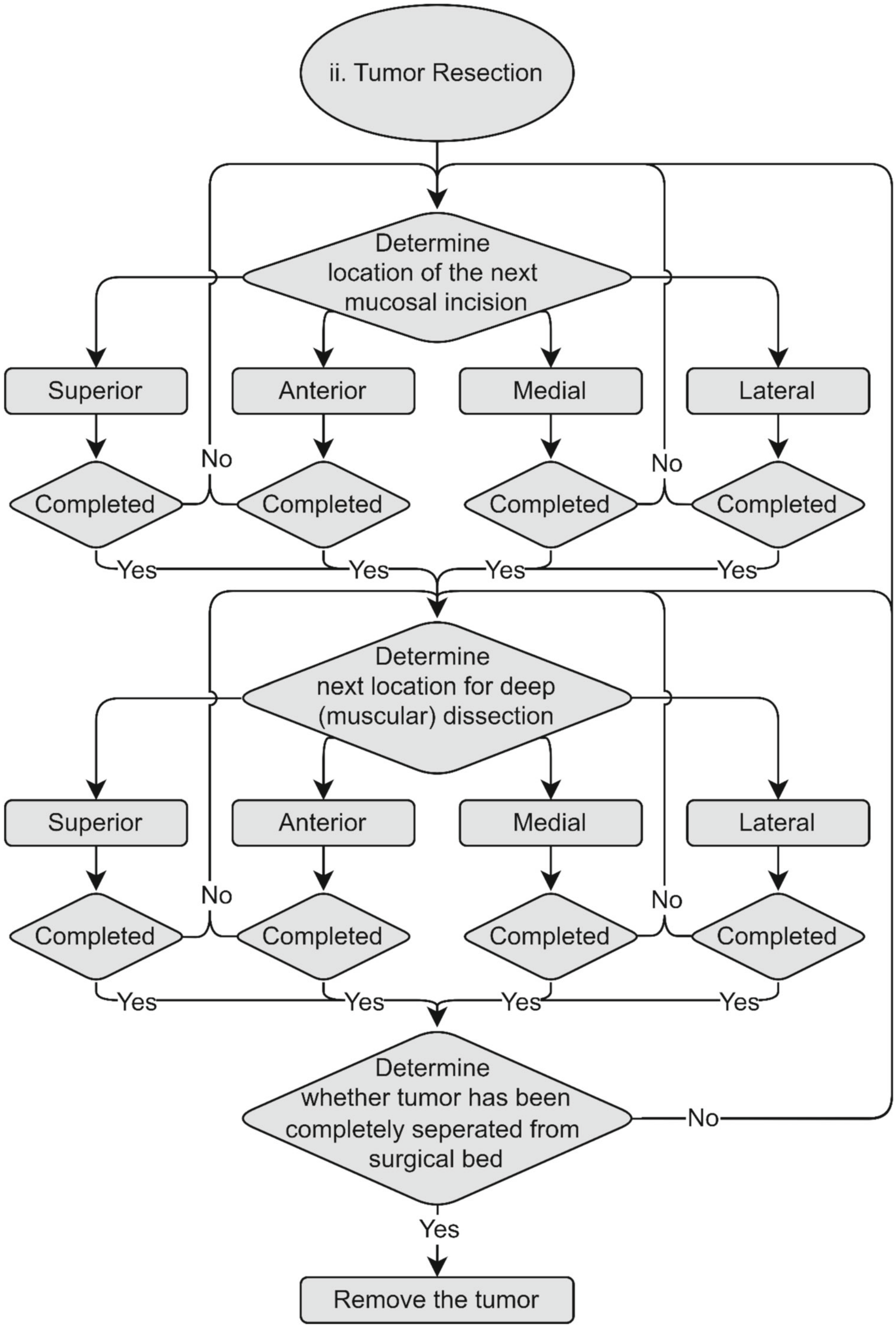


Fig. 1 Overview of the transoral robotic surgery (TORS) procedure

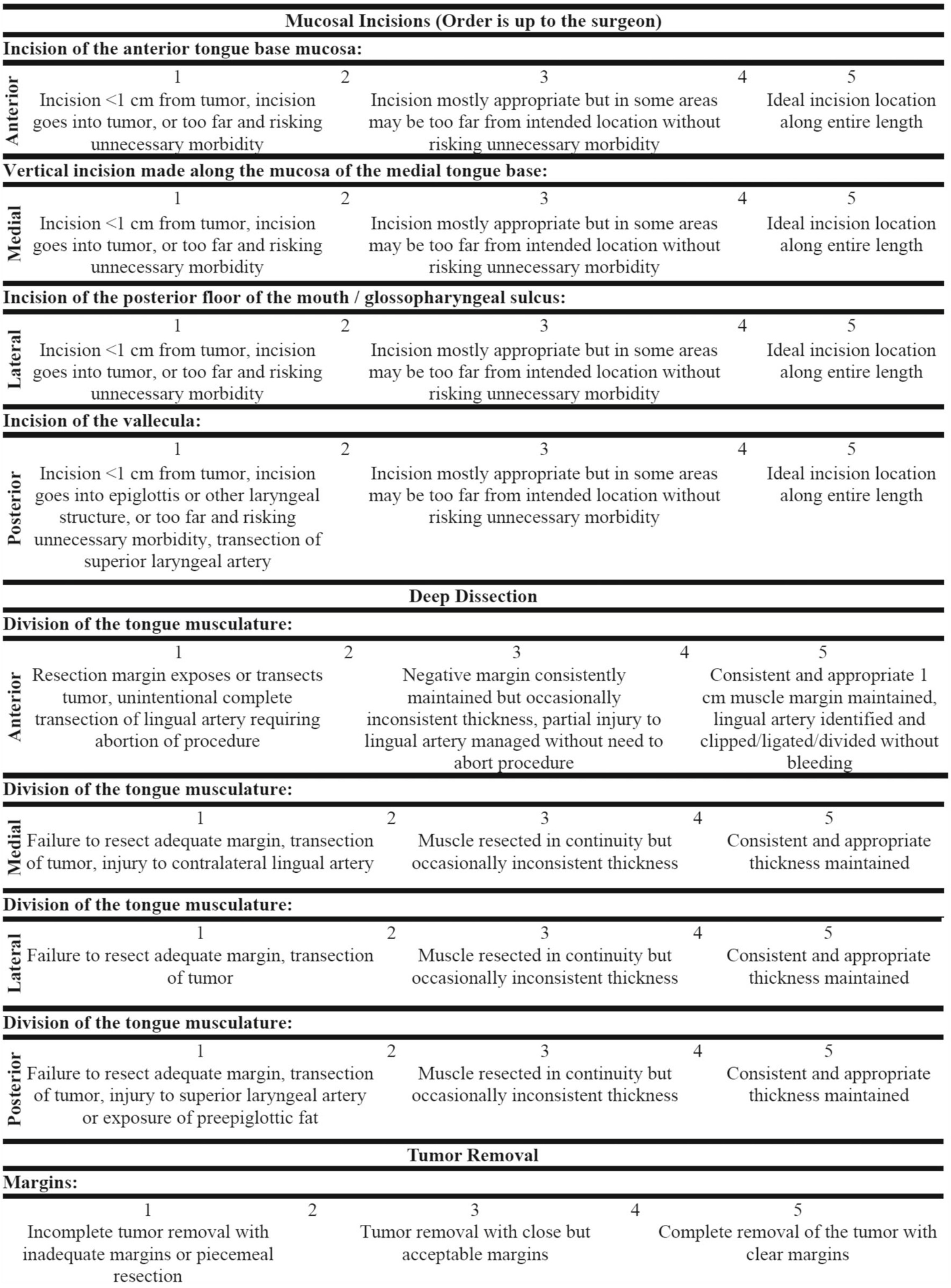
Both observers were trained and calibrated by an experienced attending otolaryngologist, who explained and graded two full-length videos. Two observers then independently graded three additional videos, after which the expert provided feedback and calibrated their assessments to maximize inter-rater reliability.

For each video, a pre-questionnaire was facilitated to receive demographic data. Only the post-graduate years (PGY) data were used in the analysis. Surgeons ranged from PGY-2 to PGY-6. Following ACGME criteria [40], PGY-2–3 surgeons were classified as novice and PGY-4 or higher as experienced. On a global basis, there is significant program-to-program variability for the expectations for robotic surgical skill for otolaryngology trainees. It is not uncommon for an otolaryngology resident to graduate from their training program with no more than a few hours of console operating time and zero procedures performed beginning-to-end as the primary surgeon. Resident surgeons from three otolaryngology training programs participated in the study for which the video and demographic data were initially collected. At the highest surgical volume institution whose residents participated in the study, PGY-2–3 residents were expected to have little to no experience as the primary console surgeon and more often had experience serving as bedside assistants for approximately 20 to 30 cases. PGY-4–5 residents at the highest volume institution were expected to have several hours of operative time as the primary console surgeon with zero procedures performed beginning-to-end. Fellows are expected to have several cases of robotic oropharyngeal resection performed beginning-to-end, and attending surgeons are expected to have dozens to hundreds of cases performed. Furthermore,

Fig. 2 Decision-making and procedural steps for tumor resection



experience varies within PGY grouping based on their individual rotation schedules. A PGY-2 resident who happens to have their head and neck cancer surgery rotation early during their academic year will have significantly more experience than a PGY-2 resident who has that rotation at the end of the academic year. There are no standards established by the American College of Graduate Medical Education (ACGME) for graduation from otolaryngology residency in terms of robotic surgical case volume [41]. Thus, grouping by PGY-2–3 and PGY-4–5 was the best division that could be created for the purposes of this study because robotic console operation at the primary institution starts at the PGY-4 year. Of the 40 videos, 20 were performed by novice surgeons and 20 by experienced surgeons. All procedures were recorded using the da Vinci system's camera. Figure 4 shows a repre-



Mucosal Incisions (Order is up to the surgeon)

Incision of the anterior tongue base mucosa:

	1	2	3	4	5
Anterior	Incision <1 cm from tumor, incision goes into tumor, or too far and risking unnecessary morbidity		Incision mostly appropriate but in some areas may be too far from intended location without risking unnecessary morbidity		Ideal incision location along entire length

Vertical incision made along the mucosa of the medial tongue base:

	1	2	3	4	5
Medial	Incision <1 cm from tumor, incision goes into tumor, or too far and risking unnecessary morbidity		Incision mostly appropriate but in some areas may be too far from intended location without risking unnecessary morbidity		Ideal incision location along entire length

Incision of the posterior floor of the mouth / glossopharyngeal sulcus:

	1	2	3	4	5
Lateral	Incision <1 cm from tumor, incision goes into tumor, or too far and risking unnecessary morbidity		Incision mostly appropriate but in some areas may be too far from intended location without risking unnecessary morbidity		Ideal incision location along entire length

Incision of the vallecula:

	1	2	3	4	5
Posterior	Incision <1 cm from tumor, incision goes into epiglottis or other laryngeal structure, or too far and risking unnecessary morbidity, transection of superior laryngeal artery		Incision mostly appropriate but in some areas may be too far from intended location without risking unnecessary morbidity		Ideal incision location along entire length

Deep Dissection

Division of the tongue musculature:

	1	2	3	4	5
Anterior	Resection margin exposes or transects tumor, unintentional complete transection of lingual artery requiring abortion of procedure		Negative margin consistently maintained but occasionally inconsistent thickness, partial injury to lingual artery managed without need to abort procedure		Consistent and appropriate 1 cm muscle margin maintained, lingual artery identified and clipped/ligated/divided without bleeding

Division of the tongue musculature:

	1	2	3	4	5
Medial	Failure to resect adequate margin, transection of tumor, injury to contralateral lingual artery		Muscle resected in continuity but occasionally inconsistent thickness		Consistent and appropriate thickness maintained

Division of the tongue musculature:

	1	2	3	4	5
Lateral	Failure to resect adequate margin, transection of tumor		Muscle resected in continuity but occasionally inconsistent thickness		Consistent and appropriate thickness maintained

Division of the tongue musculature:

	1	2	3	4	5
Posterior	Failure to resect adequate margin, transection of tumor, injury to superior laryngeal artery or exposure of preepiglottic fat		Muscle resected in continuity but occasionally inconsistent thickness		Consistent and appropriate thickness maintained

Tumor Removal

Margins:

1	2	3	4	5
Incomplete tumor removal with inadequate margins or piecemeal resection		Tumor removal with close but acceptable margins		Complete removal of the tumor with clear margins

Fig. 3 Performance-based assessment metrics

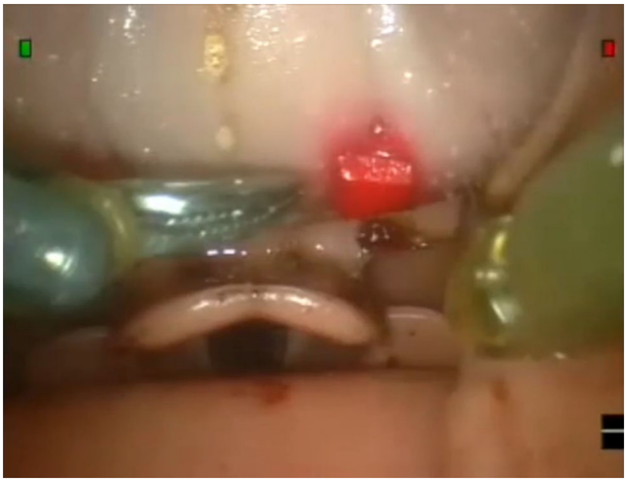


Fig. 4 Intraoperative image of TORS

sentative intraoperative TORS view with robotic instruments performing soft tissue dissection during tumor resection.

Before analyzing, we hypothesized that procedure times would correlate with scores. Prior research supports this, showing that surgeon qualifications and experience affect both total task durations [28, 33, 42].

Statistical design

To assess performance differences between novice and experienced surgeons, Mann–Whitney U tests were conducted across all key outcome variables, including Total PBA Score, Total GEARS Score, Total Time, and individual task scores for mucosal incisions, deep dissections, and GEARS subdomains. Levene's tests were also applied to examine differences in performance variability between groups. These analyses made it possible to identify the significant differences in central tendency and to evaluate the consistency of performance within each group across multiple surgical domains.

To further investigate the relationship between procedural time and performance scores, Spearman's rank correlation coefficient was used to examine the associations between Total Time, Total PBA Metric Score, and Total GEARS Score. Spearman's correlation was chosen due to its ability to capture monotonic relationships without assuming linearity.

A significance threshold of $p < 0.05$ was applied to determine whether the observed differences and correlations were statistically significant.

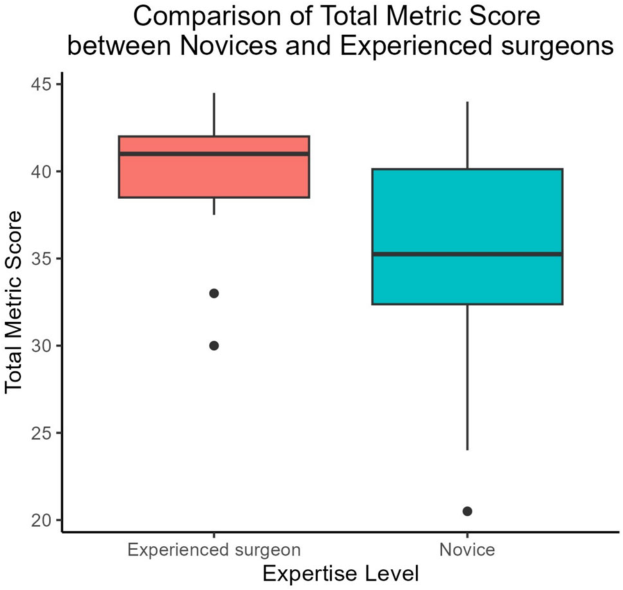


Fig. 5 Comparison of average total PBA metric scores between novice and experienced surgeons

Results

The HTA and PBA metrics provided a structured breakdown of the TORS procedure. Experienced surgeons consistently outperformed novices across all evaluated metrics (see Table 1). For the total PBA score, experienced surgeons ($\bar{x}$ = 39.98 and median = 41) achieved a higher score compared to novices ($\bar{x}$ = 35.35 median = 35.25) with a p value of 0.0055 (see Fig. 5). A Levene's test indicated a significant difference in score variance between the two groups (p = 0.0493). The detailed subtask findings can be found in the supplementary material.

Experienced surgeons achieved higher scores across most mucosal incision subregions. The highest difference was in the lateral subregion (4.65 vs 3.60, $p = 0.0015$), with a lower variance (Levene's test $p = 0.0077$). Differences in the anterior (4.48 vs 3.85, $p = 0.0971$) and medial (4.18 vs 3.83, $p = 0.1575$) subregions were not significant.

For the deep dissection, the largest performance gap appeared in the lateral subregion (4.58 vs 3.85, $p = 0.0115$). Anterior and medial subregions showed non-significant results. All dissection and incision results can be seen in Table 2.

In GEARS, experienced surgeons received a higher mean score (22.60 vs 19.63, $p = 0.0009$) (see Fig. 6) and showed greater consistency (Levene's test $p = 0.0317$). Depth perception ($p = 0.0039$) and efficiency ($p = 0.0118$) significantly differentiated groups, while bimanual dexterity and robotic control did not (see Table 3).

Table 1 Experienced and novice of average scores and total time

	Margin	Total metric score	Total time	Total GEARS score
Experienced	4.38	39.98	298.47	22.60
Novice	3.80	35.35	466.43	19.63
Mann–Whitney U *p* value	0.0149	0.0055	0.0003	0.0009
Levene’s *p* value	0.1461	0.0493	0.2668	0.0317

Table 2 Experienced and novice comparison of mucosal incision and deep dissection tasks

	Mucosal incision				Deep dissection			
	Anterior	Medial	Lateral	Posterior	Anterior	Medial	Lateral	Posterior
Experienced	4.48	4.18	4.65	4.15	4.75	4.48	4.58	4.33
Novice	3.85	3.83	3.6	4.1	4.13	3.98	3.85	3.95
Mann–Whitney U *p* value	0.0971	0.1575	0.0015	0.7567	0.0618	0.0761	0.0115	0.2063
Levene’s *p* value	0.0424	0.1303	0.0077	0.2555	0.0553	0.1303	0.2516	0.077

Table 3 Average GEARS scores of experienced and novices and their Mann–Whitney U and Levene’s *p* values

	Depth perception	Bimanual dexterity	Efficiency	Force sensitivity	Robot control
Experienced	4.45	4.7	4.55	4.45	4.45
Novice	3.85	3.825	3.95	3.975	4.025
Mann–Whitney U *p* value	0.0039	0.0578	0.0118	0.3316	0.1066
Levene’s *p* value	0.6534	0.0004	0.2710	0.0134	0.0534

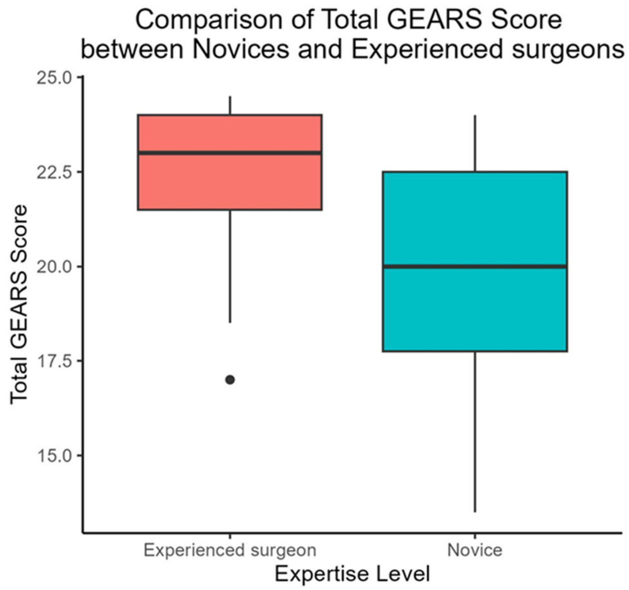


Fig. 6 Comparison of average total gears scores between novice and experienced surgeons

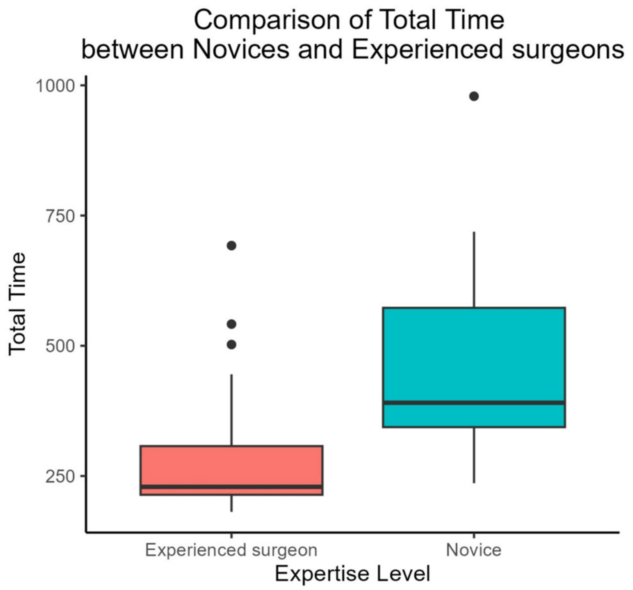


Fig. 7 Comparison of average total times between novice and experienced surgeons

Time efficiency was aligned with the findings from the PBA metrics. Experienced surgeons completed the procedure faster (298.47 s vs 466.43 s, $p = 0.0003$) (Fig. 7), and task-level analysis revealed shorter times in both mucosal incisions (120.95 s vs 166.54 s, $p = 0.0020$) and deep dissection (173.28 s vs 304.06 s, $p = 0.0002$) phases.

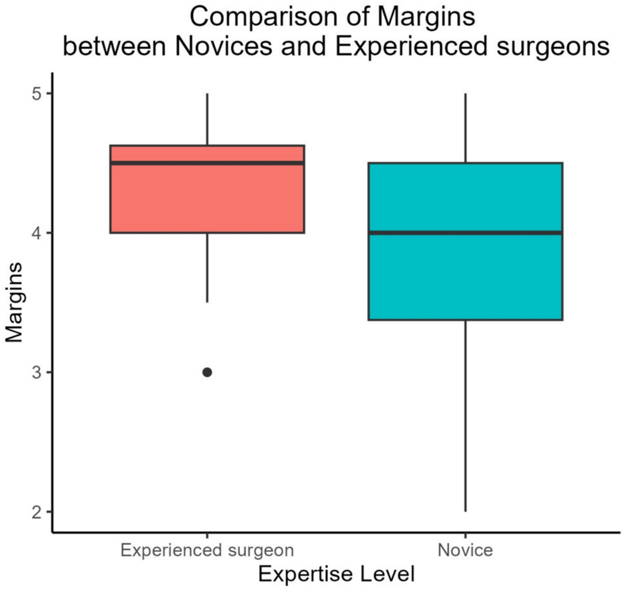


Fig. 8 Comparison of average margin scores between novice and experienced surgeons

Margin scores highlighted the performance gap. Experienced surgeons scored higher on average margins (4.38 vs 3.80, $p = 0.0149$) (Fig. 8). No significant difference in variance was observed (Levene's $p = 0.1461$).

The reliability of the observers was evaluated using: Cronbach's alpha for internal consistency and intraclass correlation coefficient (ICC) for inter-rater agreement. Full reliability matrices and extended interpretation are provided in the supplementary material.

Discussion

The results demonstrate the importance of structured training and assessment in TORS. By developing a TORS-specific HTA, this study decomposes a robotic lateral oropharyngectomy into discrete tasks, enabling detailed comparisons between novice and experienced surgeons. Coupling task-based PBA metrics with GEARS provides a dual framework for evaluating technical proficiency and procedural quality.

To our knowledge, this is the first HTA developed for a head and neck oncology procedure. Previous HTAs in otolaryngology have focused on otologic surgery, thyroidectomy [43], sentinel lymph node dissection for endometrial cancer [44], and laparoscopic adrenalectomy [45]. A unique quality of the HTA presented in this work is its acknowledgment of variability between both tumor characteristics and surgeon approach. Factors such as tumor size, growth pattern, mouth opening, and anatomic variation can influence the operative path, and the HTA was designed to capture this complexity.

Construct validity was demonstrated by significantly higher total PBA scores (mean = 39.98 vs 35.35, $p = 0.00551$) and GEARS scores (mean = 22.60 vs 19.63, $p = 0.00095$) among experienced surgeons, consistent with prior findings by Bur et al. [39].

Lateral regions emerged as good indicators of expertise. Experienced surgeons outperformed novices in lateral mucosal incision (mean = 4.65 vs 3.60, $p = 0.0015$) and lateral deep dissection (mean = 4.58 vs 3.85, $p = 0.0115$) and showed the largest performance disparities and greater consistency. They also achieved higher margin scores (4.38 vs 3.80, $p = 0.0149$). These findings suggest that lateral regions are particularly sensitive to surgeon expertise due to limited visualization and difficult access.

Some subregion scores did not reach significance (e.g., anterior mucosal 4.48 vs 3.85, $p = 0.0971$; medial mucosal 4.18 vs 3.83, $p = 0.1575$), but the consistent trends favoring experienced surgeons imply that a broader sample may reveal additional significant differences.

In GEARS, depth perception ($p = 0.0039$) and efficiency ($p = 0.0118$) also distinguished expertise, highlighting the value of fine motor control and spatial awareness in robotic surgery.

The time metrics reinforced these conclusions. Experienced surgeons completed the procedure significantly faster (298.47 s vs 466.43 s, $p = 0.00033$), with shorter times in mucosal incision (120.95 s vs 166.54 s, $p = 0.0020$) and deep dissection phases (173.28 s vs 304.06 s, $p = 0.0002$), without compromising technical performance. Efficient task execution is clinically relevant, as shorter operative times are associated with fewer complications and better patient outcomes [12].

The normalized PBA metrics are a notable contribution of this study. The PBA metrics provide an objective way to assess technical skills and efficiency, making it easier to identify areas for improvement. The observed negative correlation between task time and performance scores shows that time management is a meaningful indicator of surgical competence.

Credentialing for TORS remains institution-dependent and typically includes case log review, online didactic modules, cadaver laboratories, video review, and proctored cases [46, 47]. Although structured and simulation-based curricula can reduce the learning curve, adoption is inconsistent [39, 48].

Access to simulation-based training for the da Vinci Single-port (Sp) Surgical System (Intuitive Surgical, Sunnyvale, California, USA) has been limited to virtual reality (VR) modules that require a clinical console. Furthermore, while robotic systems are FDA cleared for oropharynx tumors, use in adjacent regions such as the parapharyngeal space, nasopharynx, and larynx remains off-label [49–51].

These applications underscore the necessity of comprehensive perioperative planning and surgeon familiarity with region-specific anatomy to reduce the risk of complications.

TORS offers significant benefits for patients with oropharyngeal cancer [51]. However, to reach its full potential, addressing current challenges in surgeon training and perioperative care is essential [39, 52]. To achieve this objective, the ultimate goal is to develop the Virtual Transoral Robotic Surgical (VTORS) simulator. This simulator will be run through commercial VR headsets that do not require a robotic console. VTORS will focus on enhancing surgical outcomes by offering a risk-free training alternative using PBA metrics developed in this study and providing a risk-free, high-fidelity environment to practice the TORS procedure [48, 53–55].

This study has several limitations worth noting. The reliance on video analysis may introduce subjectivity in the interpretation of surgical actions, despite efforts to calibrate scoring between raters. Additionally, the sample size could be expanded to include a more diverse range of surgeons and surgical scenarios to further validate the findings. Another limitation is, despite HTA addressing differences in the extent of resection (e.g., resection of tonsillar pillars), the assessment tools cannot determine whether the correct cuts were made for complete oncologic resection.

Conclusion

This study provides the first full HTA and PBA framework for TORS and shows that surgical experience strongly affects performance. Experienced surgeons achieved clearer margins, demonstrated more consistent anatomy handling, and completed tasks with better control. Novices showed uneven dissection and longer task times. Higher performance scores correlated with shorter procedure times, especially during tumor resection where precision and coordination mattered most. The results highlight the need for structured, standardized training to improve TORS outcomes.

Objective PBA metrics and HTA-based evaluations create a foundation for training programs, especially when paired with simulators like VTORS. VTORS aims to deliver an immersive, high-fidelity environment that builds skill while avoiding patient risk. Normalized, objective PBAs support reliable assessment and improvement of TORS performance.

As robotic surgery grows, future work should validate these metrics in wider clinical use, refine simulation-based training, and support competency-based credentialing. Integrating validated PBA metrics into VR systems offers a scalable training pathway that helps narrow the gap between novice and expert performance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-026-03605-3.

Acknowledgements This project was supported by grants from the National Institutes of Health (NIH)/NIBIB R01EB032820, R01EB005807, R01EB033674, and NIGMS P20GM130423. The authors thank Dr. Gregory Weinstein for his help and guidance during the preparation of this manuscript.

Data availability Data are available upon reasonable request to the corresponding author.

Declarations

Conflict of interests Authors have no conflicts of interest or financial ties to disclose.

Ethics approval All procedures performed in this study were in accordance with the ethical standards as of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. IRB approval was obtained from the University of Pennsylvania Human Subjects Electronic Research Application system.

References

1. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF et al (2010) Human Papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 363:24–35. https://doi.org/10.1056/NEJMoa0912217
2. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E et al (2011) Human Papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 29:4294–4301. https://doi.org/10.1200/JCO.2011.36.4596
3. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C (2015) Epidemiology of Human Papillomavirus–positive head and neck squamous cell carcinoma. J Clin Oncol 33:3235–3242. https://doi.org/10.1200/JCO.2015.61.6995
4. Blanco RGF, Fakhry C, Ha PK, Ryniak K, Messing B, Califano JA et al (2013) Transoral robotic surgery experience in 44 cases. J Laparoendosc Adv Surg Tech 23:900–907. https://doi.org/10.1089/lap.2013.0261
5. O'Malley BW, Weinstein GS, Snyder W, Hockstein NG (2006) Transoral robotic surgery (TORS) for base of tongue neoplasms. Laryngoscope 116:1465–1472. https://doi.org/10.1097/01.mlg.0000227184.90514.1a
6. Weinstein GS, O'Malley BW, Snyder W, Sherman E, Quon H (2007) Transoral robotic surgery: radical tonsillectomy. Arch Otolaryngol Head Neck Surg 133:1220. https://doi.org/10.1001/archotol.133.12.1220
7. Moore EJ, Olsen KD, Kasperbauer JL (2009) Transoral robotic surgery for oropharyngeal squamous cell carcinoma: A prospective study of feasibility and functional outcomes. Laryngoscope 119:2156–2164. https://doi.org/10.1002/lary.20647
8. Urken ML, Buchbinder D, Costantino PD, Sinha U, Okay D, Lawson W et al (1998) Oromandibular reconstruction using microvascular composite flaps: report of 210 Cases. Arch Otolaryngol Head Neck Surg 124:46. https://doi.org/10.1001/archotol.124.1.46
9. De Almeida JR, Li R, Magnuson JS, Smith RV, Moore E, Lawson G et al (2015) Oncologic outcomes after transoral robotic surgery: a multi-institutional study. JAMA Otolaryngol Head Neck Surg 141:1043. https://doi.org/10.1001/jamaoto.2015.1508

10. An Y, Holsinger FC, Husain ZA (2016) De-intensification of adjuvant therapy in Human Papillomavirus-associated oropharyngeal cancer. Cancers Head Neck 1:18. https://doi.org/10.1186/s41199-016-0016-7
11. Albergotti WG, Gooding WE, Kubik MW, Geltzeiler M, Kim S, Duvvuri U et al (2017) Assessment of surgical learning curves in transoral robotic surgery for squamous cell carcinoma of the oropharynx. JAMA Otolaryngol Head Neck Surg 143:542. https://doi.org/10.1001/jamaoto.2016.4132
12. White HN (2013) Learning curve for transoral robotic surgeryA 4-year analysisLearning curve for transoral robotic surgery. JAMA Otolaryngol Head Neck Surg. https://doi.org/10.1001/jamaoto.2013.3007
13. Hutcheson KA, Holsinger FC, Kupferman ME, Lewin JS (2015) Functional outcomes after TORS for oropharyngeal cancer: a systematic review. Eur Arch Otorhinolaryngol 272:463–471. https://doi.org/10.1007/s00405-014-2985-7
14. Sethi RKV, Chen MM, Malloy KM (2020) Complications of transoral robotic surgery. Otolaryngol Clin North Am 53:1109–1115. https://doi.org/10.1016/j.otc.2020.07.017
15. Stokes W, Ramadan J, Lawson G, Ferris FRL, Holsinger FC, Turner MT (2021) Bleeding complications after transoral robotic surgery: a meta-analysis and systematic review. Laryngoscope 131:95–105. https://doi.org/10.1002/lary.28580
16. Gorphe P, Simon C (2019) A systematic review and meta-analysis of margins in transoral surgery for oropharyngeal carcinoma. Oral Oncol 98:69–77. https://doi.org/10.1016/j.oraloncology.2019.09.017
17. Hanna J, Morse E, Brauer PR, Judson B, Mehra S (2019) Positive margin rates and predictors in transoral robotic surgery after federal approval: a national quality study. Head Neck 41:3064–3072. https://doi.org/10.1002/hed.25792
18. Ferris RL, Flamand Y, Weinstein GS, Li S, Quon H, Mehra R et al (2022) Phase II randomized trial of transoral surgery and low-dose intensity modulated radiation therapy in resectable p16+ locally advanced oropharynx cancer: an ECOG-ACRIN cancer research group trial (E3311). JCO 40:138–149. https://doi.org/10.1200/JCO.21.01752
19. Nichols AC, Theurer J, Prisman E, Read N, Berthelet E, Tran E et al (2024) Radiotherapy versus transoral robotic surgery for oropharyngeal squamous cell carcinoma: final results of the ORATOR randomized trial. JCO 42:4023–4028. https://doi.org/10.1200/JCO.24.00119
20. Stanton NA (2006) Hierarchical task analysis: developments, applications, and extensions. Appl Ergon 37:55–79
21. Shepherd A (1998) HTA as a framework for task analysis. Ergonomics 41:1537–1552. https://doi.org/10.1080/001401398186063
22. Salmon P, Jenkins D, Stanton N, Walker G (2010) Hierarchical task analysis vs cognitive work analysis: comparison of theory, methodology and contribution to system design. Theor Issues Ergon Sci 11:504–531. https://doi.org/10.1080/14639220903165169
23. Phipps DL, Meakin GH, Beatty PC (2011) Extending hierarchical task analysis to identify cognitive demands and information design requirements. Appl Ergon 42:741–748
24. Lane R, Stanton NA, Harrison D (2006) Applying hierarchical task analysis to medication administration errors. Appl Ergon 37:669–679
25. Phipps D, Meakin GH, Beatty PC, Nsoedo C, Parker D (2008) Human factors in anaesthetic practice: insights from a task analysis. Br J Anaesth 100:333–343
26. Sarker SK, Chang A, Albrani T, Vincent C (2008) Constructing hierarchical task analysis in surgery. Surg Endosc 22:107–111. https://doi.org/10.1007/s00464-007-9380-z
27. Demirel D, Yu A, Cooper-Baer S, Dendukuri A, Halic T, Kockara S et al (2017) A hierarchical task analysis of shoulder arthroscopy for a virtual arthroscopic tear diagnosis and evaluation platform (VATDEP). Int J Med Robot 13:e1799. https://doi.org/10.1002/rcs.1799
28. Demirel D, Butler KL, Halic T, Sankaranarayanan G, Spindler D, Cao C et al (2016) A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer simulator. Am J Surg 212:475–484. https://doi.org/10.1016/j.amjsurg.2015.08.029
29. Cetinsaya B, Gromski MA, Lee S, Xia Z, Demirel D, Halic T et al (2019) A task and performance analysis of endoscopic submucosal dissection (ESD) surgery. Surg Endosc 33:592–606. https://doi.org/10.1007/s00464-018-6379-6
30. Knight S, Aggarwal R, Agostini A, Loundou A, Berdah S, Crochet P (2018) Development of an objective assessment tool for total laparoscopic hysterectomy: a Delphi method among experts and evaluation on a virtual reality simulator. PLoS ONE 13:e0190580
31. Corbett M, O'Connor P, Byrne D, Thornton M, Keogh I (2019) Identifying and reducing risks in functional endoscopic sinus surgery through a hierarchical task analysis. Laryngoscope Investig Otolaryngol 4:5–12. https://doi.org/10.1002/lio2.220
32. Ballini L, Minozzi S, Negro A, Pirini G, Grilli R (2010) A method for addressing research gaps in HTA, developed whilst evaluating robotic-assisted surgery: a proposal. Health Res Policy Syst 8:27. https://doi.org/10.1186/1478-4505-8-27
33. Dials J, Demirel D, Halic T, De S, Ryason A, Kundumadam S et al (2022) Hierarchical task analysis of endoscopic sleeve gastroplasty. Surg Endosc 36:5167–5182. https://doi.org/10.1007/s00464-021-08893-1
34. Holsinger FC, McWhorter AJ, Ménard M, Garcia D, Laccourreye O (2005) Transoral lateral oropharyngectomy for squamous cell carcinoma of the tonsillar region: I. Technique, complications, and functional results. Arch Otolaryngol Head Neck Surg 131:583–591. https://doi.org/10.1001/archotol.131.7.583
35. Holsinger FC (2016) A flexible, single-arm robotic surgical system for transoral resection of the tonsil and lateral pharyngeal wall: next-generation robotic head and neck surgery. Laryngoscope 126:864–869. https://doi.org/10.1002/lary.25724
36. Weinstein GS, O'Malley BW (2012) TransOral robotic surgery (TORS). Plural Pub, San Diego
37. Grillone GA, Jalisi S (eds) (2015) Robotic surgery of the head and neck: a comprehensive guide [Internet]. Springer New York, New York https://doi.org/10.1007/978-1-4939-1547-7
38. Sánchez R, Rodríguez O, Rosciano J, Vegas L, Bond V, Rojas A et al (2016) Robotic surgery training: construct validity of global evaluative assessment of robotic skills (GEARS). J Robot Surg 10:227–231. https://doi.org/10.1007/s11701-016-0572-1
39. Bur AM, Gomez ED, Newman JG, Weinstein GS, O'Malley BW, Rassekh CH et al (2017) Evaluation of high-fidelity simulation as a training tool in transoral robotic surgery. Laryngoscope 127:2790–2795. https://doi.org/10.1002/lary.26733
40. Accreditation Council for Graduate Medical Education (ACGME) (2016). Resident/fellow levels of training as defined by specialty [Internet]. [cited 2025 Jan 9] https://www.acgme.org/globalassets/pdfs/datacollectionsystems/leveloftrainingbyspecialty.pdf. Accessed 9 Jan 2025
41. Review Committee for Otolaryngology-Head and Neck Surgery (2018) Required minimum number of key indicator procedures for graduating residents [Internet]. ACGME https://www.acgme.org/globalassets/PFAssets/ProgramResources/280_Required_Minimum_Number_of_Key_Indicator_Procedures.pdf
42. Demirel D, Yu A, Cooper-Baer S, Dendukuri A, Halic T, Kockara S et al (2017) A hierarchical task analysis of shoulder arthroscopy for a virtual arthroscopic tear diagnosis and evaluation platform (VATDEP). Robot Comput Surg 13:e1799. https://doi.org/10.1002/rcs.1799

43. Hamour A, Mendez A, Harris JR, Biron VL, Seikaly H, Côté DWJ (2017) Thyroidectomy teaching module. VideoEndocrinology 4:ve.2017.0092. https://doi.org/10.1089/ve.2017.0092
44. Moloney K, Janda M, Frumovitz M, Leitao M, Abu-Rustum NR, Rossi E et al (2021) Development of a surgical competency assessment tool for sentinel lymph node dissection by minimally invasive surgery for endometrial cancer. Int J Gynecol Cancer 31:647–655. https://doi.org/10.1136/ijgc-2020-002315
45. Madani A, Grover K, Kuo JH, Mitmaker EJ, Shen W, Beninato T et al (2020) Defining the competencies for laparoscopic transabdominal adrenalectomy: an investigation of intraoperative behaviors and decisions of experts. Surgery 167:241–249. https://doi.org/10.1016/j.surg.2019.03.035
46. Gross ND, Holsinger FC, Magnuson JS, Duvvuri U, Genden EM, Ghanem TAh, et al (2016) Robotics in otolaryngology and head and neck surgery: recommendations for training and credentialing: a report of the 2015 AHNS education committee, AAO-HNS robotic task force and AAO-HNS sleep disorders committee. Head Neck [Internet] 38 https://doi.org/10.1002/hed.24207
47. Ferris RL, Flamand Y, Holsinger FC, Weinstein GS, Quon H, Mehra R et al (2020) A novel surgeon credentialing and quality assurance process using transoral surgery for oropharyngeal cancer in ECOG-ACRIN cancer research group trial E3311. Oral Oncol 110:104797. https://doi.org/10.1016/j.oraloncology.2020.104797
48. Zhang N, Sumer BD (2013) Transoral robotic surgery: simulation-based standardized training. JAMA Otolaryngol Head Neck Surg 139:1111. https://doi.org/10.1001/jamaoto.2013.4720
49. O'Malley BW Jr., Quon H, Leonhardt FD, Chalian AA, Weinstein GS (2010) Transoral robotic surgery for parapharyngeal space tumors. ORL J Otorhinolaryngol Relat Spec 72:332–336. https://doi.org/10.1159/000320596
50. King WWK, Ku PKM, Mok C, Teo PML (2000) Nasopharyngectomy in the treatment of recurrent nasopharyngeal carcinoma: a twelve-year experience. Head Neck 22:215–222
51. Park YM, Lee WJ, Lee JG, Lee WS, Choi EC, Chung SM et al (2009) Transoral robotic surgery (TORS) in laryngeal and hypopharyngeal cancer. J Laparoendosc Adv Surg Tech 19:361–368. https://doi.org/10.1089/lap.2008.0320
52. Murr AT, Lumley CJ, Feins RH, Hackman TG (2022) Evaluation of a 3D- PRINTED transoral robotic surgery simulator utilizing artificial tissue. Laryngoscope 132:1588–1593. https://doi.org/10.1002/lary.29981
53. Kwong FNK, Puvanendran M, Paleri V (2015) Transoral robotic surgery in head neck cancer management. B-ENT Suppl 24:7–13
54. Cammaroto G, Stringa LM, Zhang H, Capaccio P, Galletti F, Galletti B et al (2020) Alternative applications of trans-oral robotic surgery (TORS): a systematic review. JCM 9:201. https://doi.org/10.3390/jcm9010201
55. Lawson G, Matar N, Remacle M, Jamart J, Bachy V (2011) Transoral Robotic Surgery for the management of head and neck tumors: learning curve. Eur Arch Otorhinolaryngol 268:1795–1801. https://doi.org/10.1007/s00405-011-1537-7