Adaptation of "Objective Structured Assessment of Technical Skills" for Adenotonsillectomy into Turkish: A Validity and Reliability Study

Cüneyt Orhan Kara[1] , Erdem Mengi[1] , Funda Tümkaya[1] , Fazıl Necdet Ardıç[1] , Hande Şenol[2]

Original Investigation ▶

[1]Department of Otorhinolaryngology, Pamukkale University School of Medicine, Denizli, Turkey
[2]Department of Biostatistics, Pamukkale University School of Medicine, Denizli, Turkey

Abstract ▶

Objective: To adapt an objective surgical assessment tool (OSATS) for adenotonsillectomy/tonsillectomy to Turkish.

Methods: A prospective longitudinal validation study was planned. Adenoidectomy items were added to the tonsillectomy OSATS. Data were collected from 80 real-time observed surgical performances. To examine interrater reliability, 25 performances were simultaneously scored by two evaluators. In total, 105 test forms were filled, and reliability and validity studies were conducted.

Results: Regarding reliability, Cronbach's alpha values were found to be >0.95, indicating very high internal consistency. For interrater reliability, a statistically significant positive correlation was found between the two evaluators' task-specific checklist (TSC) (p=0.0001; r=0.969) and global rating checklist (GRC) (p=0.0001; r=0.957). Between the TSC and GRC, there was a statistically significant positive correlation (p=0.0001; r=0.951). A model was developed using the generalizability theory, and the highest variances belonged to surgical processes. Regarding validity, the trainers' scores were significantly higher than the trainees' scores (p=0.0001). As surgical experience increased, both the TSC and the GRC scores significantly increased (p=0.0001). A trainee's first-ever 34 adenotonsillectomy performances were scored. A scatter plot of the obtained data showed that the trainee had gained more scores as the number of operations increased.

Conclusion: The results indicate that adenotonsillectomy OSATS is a reliable and valid tool for the assessment of residents' surgical skills. This is the first OSATS developed for adenotonsillectomy in the Turkish literature. It is ready to be used in the operating room to provide constructive feedback and monitor the development of trainees' surgical skills in adenotonsillectomy.

Keywords: Surgical skills, tonsillectomy, adenoidectomy, workplace-based assessment, residency training

ORCID IDs of the authors:
C.O.K. 0000-0003-2219-4283;
E.M. 0000-0001-8813-2443;
F.T. 0000-0002-3213-8106;
F.N.A. 0000-0003-4230-3141;
H.Ş. 0000-0001-6395-7924.

Introduction

Residency programs all over the world has shifted away from a traditional time-based approach to a competency-based approach (1, 2). In competency-based surgical training, residents are expected to demonstrate their technical competency, hence document their progress in procedures foreseen in the core curriculum, and to that end, validated and reliable measurement and assessment tools are needed (2-5). Studies aimed at developing such measurement and assessment tools for various surgical procedures in the department of otorhinolaryngology are reported in the literature (5-10).

Objective Structured Assessment of Technical Skills (OSATS), Procedure-Based Assessment (PBA) and checklists are used for measuring and assessing performance in residency training. Among these, OSATS tools are the most commonly used for measuring and assessing surgical performance instances in otorhinolaryngology. As there is no Turkish synonym for OSATS in the literature, with this report we would like to propose "Cerrahi Beceri Cetveli (CBC)" as a first. This study is also the first example of its kind in the Turkish literature. Surgical skills assessment tools consist of two separate checklists; namely, a task-specific checklist and a global checklist.

As defined in the name, the task-specific checklist consists of the steps of a surgical procedure. The global checklist consists of almost the same items in all OSATS tools and observes the overall stages of the surgical performance.

A global checklist has been first developed by Martin et al. (11) and used as part of all OSATS tools with various changes throughout the years. Two OSATS were found in the literature for tonsillectomy (5, 6). While the first OSATS was designed by Roberson et al. (5), in the following years Ahmed et al. (6) designed a two-list tool to be used for tonsillectomy.

The latter, being the only example in the literature for tonsillectomy, was used for Turkish adaptation. Since tonsillectomy is an intervention mostly performed in children and often together with adenoidectomy, steps for adenoidectomy were added to the tonsil-OSATS that was used as base for the Turkish adaptation and the validity and the reliability of the adapted version was assessed.

Methods

The adaptation plan proposed by Şeker and Gençdoğan (12) in their book entitled 'Development of assessment tools in psychology and education' was used as a guide in the adaptation process of the tonsil-OSATS which is originally written in English. As a first step, permission was obtained from the author of the original English tonsil-OSATS for the Turkish adaptation. As this is a text with technical content, rather than mere scale adaptation a different approach was adopted in the language adaptation phase. As a first step, the researchers on our team independently translated the original text to Turkish. Then, by reviewing and comparing the translations, the team authored a single Turkish text. In general tonsil-OSATS is a technical text that describes the procedure in sequence. Because the text is of technical content rather than a psychopathology or behavior assessment scale, the lingual adaptation process was terminated once the translation was in place. The three items on adenoidectomy were determined with the modified Delphi Method and included in the task-specific checklist (13). In this way the test has been designed to be used for adenotonsillectomy or only for tonsillectomy as needed. The Turkish adaptation has been developed to be comparable with the original tool (Table 1, 2). As the next step a meeting was held to present the Turkish Adenotonsillectomy CBC (AT-CBC) to the trainers and trainees who volunteered to participate in the study. A copy of the test was provided to each participant and the phases of the study were explained. Since the study involved surgical skills and performance, written consent was obtained from every surgeon who volunteered and was observed in the study. The data were collected by observing the routine treatment procedures performed in the operating room. In the assessment process, trainees, as is the routine, operated under the supervision of a senior specialist in charge of the case. Routine surgical skills training processes were maintained in the operating room. The

Table 1. Adenotonsillectomy — OSATS "Task-Specific Checklist"

Adenotonsillectomy OSATS / Task-Specific Checklist						
Name and Surname of Resident:	Number of adenotonsillectomy procedures performed:					
Evaluator:	Date:				Total Score:	
Surgical Procedure: Tonsillectomy/Adenotonsillectomy	Name and Surname of Patient:					
	Unable to perform		Performs w/ minimal prompting		Performs Easily/ Good flow	N/A
Patient positioning and draping	1	2	3	4	5	N/A
Atraumatic mouth gag placement	1	2	3	4	5	N/A
Gag suspension	1	2	3	4	5	N/A
Adequate exposure for adenoids/tonsils	1	2	3	4	5	N/A
Digital palpation of nasopharynx and adenoid tissue	1	2	3	4	5	N/A
Curettage of adenoid tissue	1	2	3	4	5	N/A
Achieving adenoidectomy hemostasis	1	2	3	4	5	N/A
Grasps tonsils appropriately	1	2	3	4	5	N/A
Finds the plane	1	2	3	4	5	N/A
Dissects in the correct plane	1	2	3	4	5	N/A
Dissects the tonsils by preserving the surrounding tissues	1	2	3	4	5	N/A
Lower pole resection using tonsil snare	1	2	3	4	5	N/A
Obtains hemostasis	1	2	3	4	5	N/A
Checks Intraoral gauzes	1	2	3	4	5	N/A
Irrigation and aspiration of oral cavity and oropharynx	1	2	3	4	5	N/A
Removs gag safely while respecting endotracheal tube	1	2	3	4	5	N/A

Table 2. Adenotonsillectomy — OSATS "Global Checklist"

Adenotonsillectomy OSATS / Global Checklist					
Understanding of indications for surgery / objectives of surgery					
1	2	3	4	5	n/a
Deficient understanding of indication for surgery/ objectives of surgery		Understands most indications/objectives of surgery		Fully understands indications/objectives of surgery	
Communication with the anesthesiologist					
1	2	3	4	5	n/a
Unable to communicate essential points of care and concern		Able to communicate the majority of essential points of care and concern		Able to communicate all the essential points of care and concern	
Instrument Handling					
1	2	3	4	5	n/a
Repeatedly makes tentative or awkward movements		Competent use of instruments but occasionally stiff or awkward		Fluid moves with instruments and no awkwardness	
Respect for Tissue					
1	2	3	4	5	n/a
Uses unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionalinadvertent damage to tissue		Consistently handled tissues appropriately with minimal damage	
Time and Motion					
1	2	3	4	5	n/a
Many unnecessary moves		Efficient time/ motion but some unnecessary moves		Clear economy of movement and maximum efficiency	
Amount of Tension					
1	2	3	4	5	n/a
Frequently uses excessive or inadequate tension		Occasionally uses excessive or inadequate tension		Always uses adequate tension	
Direction of Tension					
1	2	3	4	5	n/a
Rarely uses correct angle of tension		Occasionally has tension at incorrect angle		Always has tension at correct angle	
Cautery Technique					
1	2	3	4	5	n/a
Cauterizes other tissues causing trauma		Minimal cautery of other tissue		Cauterizes only necessary tissue preserved the surrounding tissues	
Knowledge of specific procedure					
1	2	3	4	5	n/a
Deficient knowledge and needs instruction in most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation	
Flow of Operation					
1	2	3	4	5	n/a
Frequently stopped and unsure of next move		Some forward planning with reasonable progression		Obviously planned course of operation with effortless flow	
Overall Surgical Performance					
1	2	3	4	5	n/a
Poor		Performs majority of surgery acceptable		Outstanding	
n/a; could not be assessed					
Did you provide formative feedback to the resident?				Yes	No
Do you think the resident is compotent to perform the surgery independently				Yes	No

next step after data collection was to analyze the psychometric properties (validity and reliability) of the AT-CBC. In this process, the validity and reliability studies on similar OSATS were used as reference (5-10). Permission was obtained from the Ethics Committee of Noninvasive Clinical Studies of the Pamukkale University (nr. 60116787-020/37940).

Data were analyzed using the Statistical Package for the Social Sciences version 24 (IBM Corp., NY, USA) software. A p-value of less than 0.05 was accepted as significant in all statistical analyses which was performed by a biostatistician.

Results

Fifty-four procedures performed by five trainees and 26 procedures performed by four trainers from the Otorhinolaryngology Department of the Pamukkale University were observed in real-time by Kara and Mengi, and their performances were scored. No results are available for 2nd and 4th-year trainees because the participants were in their 1st, 3rd or 5th-year. To assess interrater reliability, 25 procedures were simultaneously observed and scored by two evaluators. To test whether AT-CBC contributed to achievement, the first 34 adenotonsillectomy procedures were observed in one trainee. In total 80 procedures were observed, and 105 performances were scored. Of the 69 patients that underwent an adenotonsillectomy 22 were female and 47 were male with a mean age of 8.03±7.2. Of the 36 patients that underwent a tonsillectomy 13 were female and 23 were male with a mean age of 32.13±11.81. While in six patients a ventilation tube was placed in the same session, the placement procedure was performed during the initial stage of the procedure and did not have an influence on the assessment process. Surgical indications of patients included adenotonsillar-tonsillar hypertrophy and/or recurrent infection.

For reliability: Cronbach's alpha was calculated to assess internal consistency for both the global checklist and the task-specific checklist to demonstrate the reliability of the test. Cronbach's alpha value for the 11 items on the task-specific checklist was 0.983, and Cronbach's alpha value for the 16 items on the global checklist was 0.989. Cronbach's alpha values were found higher than 0.95 and indicated very high internal consistency.

Regarding interrater reliability, a statistically significant and strong positive correlation was found between the two evaluators for the task-specific checklist (p=0.0001; r=0.989). For the global checklist, a statistically significant and strong positive correlation was found between the two evaluators (p=0.0001; r=0.950). This demonstrated that the measurement results of AT-CBC did not vary between evaluators.

Generalizability (G theory) was also examined with respect to reliability. The G theory is an approach to simultaneously assess all potential sources of error in the measurement including facets such as raters, time, test forms, items, and tasks. The G theory is a statistical theory that allows to assess behavioral reliability, to design, examine and conceptualize reliable observations, and is based on variance analysis (ANOVA) (14). In this study, generalizability was separately analyzed for the task-specific checklists and the global checklists for adenotonsillectomy and the tonsillectomy results received from two evaluators (Table 3, 4). A model has been designed to focus on procedure, items, evaluators, procedure x items, procedure x evaluator, item x evaluator, and oddment. The highest variance among the four assessments was found in surgical procedures, where the G values were greater than 0.9, indicating high reliability of the test. Regarding the variances for adenotonsillectomy and tonsillectomy separately, the highest variances were found to be related to surgical procedures (Table 3, 4).

For validity: Comparison of the mean scores received by the trainers and the trainees in achievement tests, the scores of the trainers were seen to be statistically significantly higher than those of the trainees (Table 5). The AT-CBC allows to discriminate between the apprentice and the master, between the qualified and the unqualified.

Furthermore, both the task-specific scores and the global scores received by trainees were seen to enjoy a significant and strongly positive increase in line with their experience (p=0.0001; r=0.898), (p=0.0001; r=0.875). AT-CBC demonstrates the performance enhancement in the process of mastery (Figure 1).

Table 3. Variance and percentage of factors in generalizability calculations for adenotonsillectomy

Factor	Task-Specific Checklist		Global Checklist	
	Variance	Percentage	Variance	Percentage
Adenotonsillectomy	0.452*	60.9	0.532*	69.3
Item	0.011	1.5	0.023	2.9
Evaluators	0.019	2.6	0.006	0.8
Adenotonsillectomy x Item	0.034	4.5	0.037	4.8
Adenotonsillectomy x Evaluator	0.024	3.3	0.041	5.3
Item x Evaluator	0.000	0	0.0005	0.1
Oddment	0.202	27.2	0.129	16.8

Generalizability model G values for adenotonsillectomy were G=0.956 (for Task-specific) and G=0.947 (for Global)
*Highest variances come from the surgical procedure of adenotonsillectomy

Table 4. Variance and percentage of factors in generalizability calculations for tonsillectomy

Factor	Task-Specific Checklist		Global Checklist	
	Variance	Percentage	Variance	Percentage
Tonsillectomy	0.716*	65.4	0.836*	69.6
Item	0.034	3.1	0.028	2.3
Evaluators	0.01	0.9	0.015	1.2
Tonsillectomy x item	0.069	6.3	0.106	8.8
Tonsillectomy x Evaluator	0.026	2.3	0.063	5.2
Item x Evaluator	0	0	0.006	0.5
Oddment	0.241	22	0.147	12.2

Generalizability model G values for tonsillectomy were G=0.967 (for Task-Specific) and G=0.946 (for Global)
*Highest variances are related to the surgical procedure of tonsillectomy

Table 5. Average scores of trainees and trainers from task-specific and global checklists that constitute the two sections under the adenotonsillectomy surgical skills chart

	Students		Professors		
	AM±SD	Med (min - max)	AM±SD	Med (min - max)	p
Task-Specific Checklist	3.93±0.94	4.19 (2.06-5)	4.88±0.29	4.94 (3.38-5)	0.0001*
Global Checklist	3.66±1.02	3.64 (1.45-5)	4.84±0.33	4.95 (3.27-5)	0.0001*

AM: arithmetic mean of scores; SD: standard deviation; med: median; min: minimum; max: maximum *p<0.05 statistically significant difference
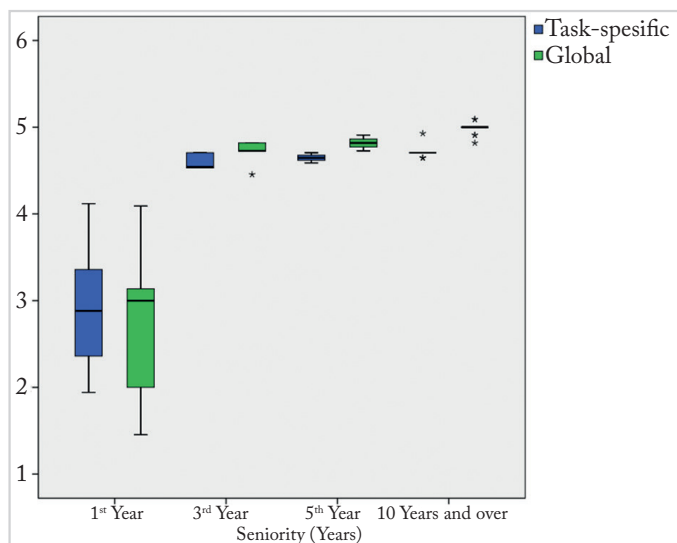Mann-Whitney U test

Figure 1. This graph includes all scores obtained in the study and shows the score averages from the task-specific and global checklists in the adenotonsillectomy surgical skills chart versus seniority. Scores corresponding to 1, 3, 5 years show the scores of trainees, and scores corresponding to +10 years show the scores of educators. A strongly positive correlation is seen between trainees experience and task-specific scores (r=0.834; p=0.0001*) A strongly positive correlation is also seen between trainees experience and global scores (r=0.861; p=0.0001*)

One trainee was observed in his first 34 adenotonsillectomy or tonsillectomy procedures. A scatter plot of the data collected in these 34 procedures showed that his score increased in line with the number of performed procedures (Figure 2). This finding

demonstrates that AT-CBC can objectively show whether performance improves as more procedures are performed.

Discussion

Competency-based training requires trainees to demonstrate their competency in the core areas of their specialization. Valid and reliable tests are needed to assess competency in the surgical branches foreseen in the core specialization program.

Tonsillectomy is mostly performed in children and together with adenoidectomy. In the adaptation steps to assess adenoidectomy were added to the surgical assessment tool developed by Ahmed et al. (6) for tonsillectomy. In its final form, the AT-CBC tool developed in this study can be used for adenotonsillectomy or only for tonsillectomy as needed. Then, validity and reliability-two of the most important properties needed in an achievement test were demonstrated with various statistical methods (5-10).

Validity is the level of accuracy to which the feature to be assessed. In this AT-CBC our goal is to discriminate between those who have and those who not competent, and to demonstrate the improvement in their surgical performance. In our study, we compared the scores of the trainees with the experienced trainers and found that the scores of experienced trainers were higher than those of the trainees (Table 5). Moreover, comparison of the scores received by the 1st, 3rd or 5th-year trainees showed that more senior and experienced trainees received higher scores (Figure 1). A valid test should, among others,
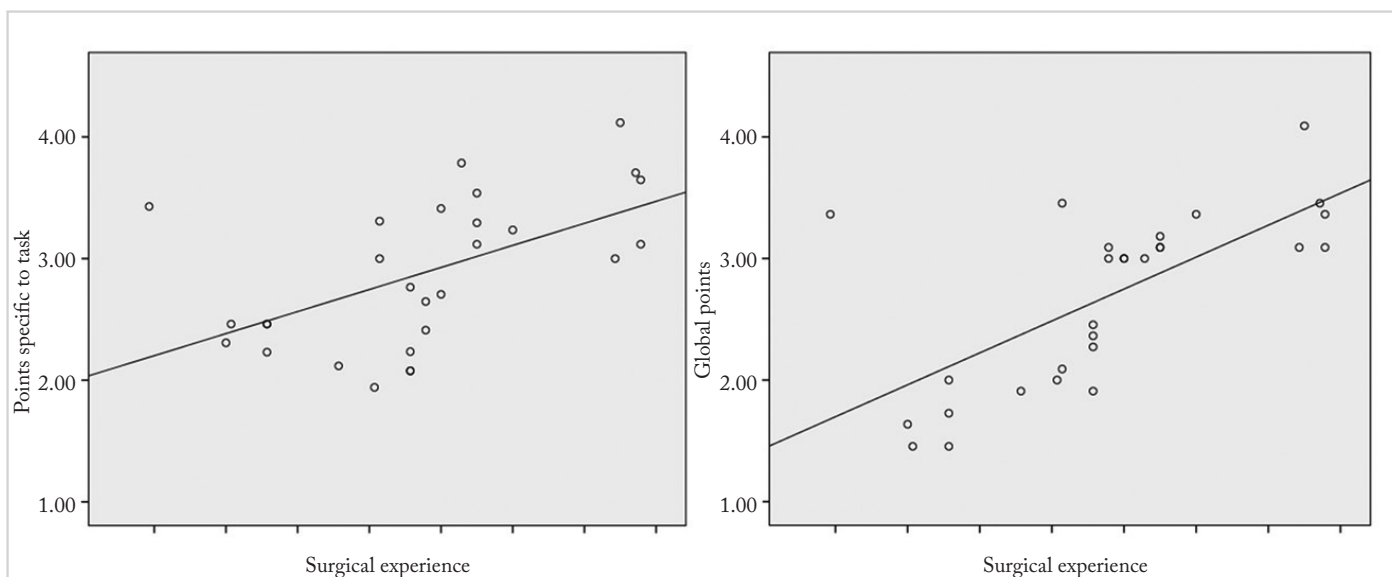
Figure 2. Thirty four first-ever adenotonsillectomy or tonsillectomy procedures performed by one trainee: scatter plot of scores from task-specific (left) and global (right) checklists for adenotonsillectomy surgical skills. Both plots show that scores have increased in both checklists in line with surgical experience

demonstrate the improvement of a trainee. To that end, starting with his first-ever procedure, 34 adenotonsillectomy or tonsillectomy procedures performed by one trainee were scored, and this trainee was seen to improve his score with experience. This implementation demonstrated the structural validity of the test (Figure 2).

Reliability, on the other hand, shows the stability and the consistency of test results. The most commonly used method, Cronbach's alpha was calculated for these purposes. The >0.95 Cronbach's alpha value has demonstrated high internal consistency for the CBC tool.

Another reliability indicator is the interrater reliability. To test interrater reliability two evaluators have simultaneously assessed 25 procedures. Review of the correlations between the results received from the two evaluators showed r=0.969 for the task-specific checklist (p=0.0001), and r=0.957 for the global checklist (p=0.0001). These findings demonstrate very strong interrater reliability for both checklists with correlation values very close to 1.

In this study generalizability has also been observed for reliability and found (Task-specific, G=0.956; Global, G=0.947) for adenotonsillectomy and (Task-specific, G=0.967, Global, G=0.946) for tonsillectomy. The G values, found to be very close to 1, are yet another strong evidence demonstrating the generalizability and the reliability of the checklists.

Adenotonsillectomy and tonsillectomy are relatively less-complex procedures. Trainees can quickly achieve high competency scores in such easy-to-learn and easy-to-perform surgical procedures. Figure 1 shows that the participating trainees have already achieved a sufficient level in adenotonsillectomy after their 3rd year of training. Trainees are likely to achieve this level

of competency in their 2nd year of training; however, this could not be verified in our study because no 2nd year trainees participated in our study.

These results demonstrate that AT-CBC is a surgical competency test with high reliability and validity. OSATS are instruments developed for workplace-based assessment. Like other workplace-based assessment tools, OSATS are used for two different purposes. Using for formative approach is suggested for surgical procedures aimed at workplace-based assessment. This allows OSATS tools to be used for observing the surgical performance of the trainee trainee, identifying their strengths and weaknesses, and finally providing feedback. This constitutes a formative assessment approach which uses the assessment results for shaping the training process. Utilization of measurement and assessment tools designed for performance are especially recommended in competency-based training (15-17). Utilization of such tests will objectively demonstrate the development of trainees and whether they have achieved sufficient competency. This also will allow to document the competency level students achieve in the course of their specialty training. These documents should be filed in the student's portfolio throughout their training.

Conclusion

In scope of this study, we have adapted the OSATS tool and developed a Turkish version including items for adenoidectomy, analyzed its psychometric properties, and finally demonstrated the validity and reliability of the assessment tool that we developed. This assessment tool is the first Turkish surgical competency test adapted for the purposes of evaluating surgical skills. We recommend using the AT-CBC test in the operating room to monitor the improvements in the surgical skills of trainees in adenotonsillectomy procedures, and to provide constructive feedback.

Ethics Committee Approval: Ethics committee approval was received for this study from the Ethics Committee of Pamukkale University Non-Interventional Clinical Trials (60116787-020/37940).

Informed Consent: Written informed consent was obtained from trainers and trainees who participated in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept - C.O.K., E.M., F.T., F.N.A, H.Ş.; Design - C.O.K., E.M., F.T., F.N.A., H.Ş.; Supervision - C.O.K., E.M., F.T., F.N.A.; Resource - C.O.K., E.M., F.T., F.N.A.; Materials - C.O.K., E.M., F.T., F.N.A.; Data Collection and/or Processing - C.O.K., E.M., H.Ş.; Analysis and/or Interpretation - C.O.K., E.M., H.Ş.; Literature Search - C.O.K., E.M., H.Ş.; Writing - C.O.K., E.M.; Critical Reviews - C.O.K., E.M., F.N.A., H.Ş.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

References

1. Kulak burun boğaz hastalıkları müfredat v.2.1. Avaible from: URL: http://www.tuk.saglik.gov.tr/muf2.1/kulak_burun_bogaz_hastaliklari/kulak_burun_bogaz_hastaliklari_mufredat_v.2.1.pdf.

2. Wagner N, Fahim C, Dunn K, Reid D, Sonnadara RR. Otolaryngology residency education: a scoping review on the shift towards competency-based medical education. Clin Otolaryngol 2017; 42: 564-72. [CrossRef]

3. Kara CO. Kulak ameliyatları için ölçme değerlendirme araçları. KBB-Forum 2015; 14: 65-70.

4. Kara CO, Özdemir S. Assessment of technical skills in postgraduate surgical education. Türkiye Klinikleri J Med Educ-Special Topics 2016; 1: 41-6

5. Roberson DW, Kentala E, Forbes P. Development and validation of an objective instrument to measure surgical performance at tonsillectomy. Laryngoscope 2005; 115: 2127-37. [CrossRef]

6. Ahmed A, Ishman SL, Laeeq K, Bhatti NI. Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. Laryngoscope 2013; 123: 1639-44. [CrossRef]

7. Lin SY, Laeeq K, Ishii M, Kim J, Lane AP, Reh D, et al. Development and pilot-testing of a feasible, reliable, and valid operative competency assessment tool for endoscopic sinus surgery. Am J Rhinol Allergy 2009; 23: 354-9. [CrossRef]

8. Francis HW, Masood H, Chaudhry KN, Laeeq K, Carey JP, Della Santina CC, et al. Objective assessment of mastoidectomy skills in the operating room. Otol Neurotol 2010; 31: 759-65. [CrossRef]

9. Obeid AA, Al-Qahtani KH, Ashraf M, Alghamdi FR, Marglani O, Alherabi A. Development and testing for an operative competency assessment tool for nasal septoplasty surgery. Am J Rhinol Allergy 2014; 23: e163-7. [CrossRef]

10. Awad Z, Hayden L, Robson AK, Muthuswamy K, Tolley NS. Reliability and valaidity of procedure-based assessments in otolaryngology training. Laryngoscope 2015; 125: 1328-35. [CrossRef]

11. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997; 84: 273-8. [CrossRef]

12. Şeker H, Gençdoğan B. Psikoloijide ve eğitimde ölçme aracı geliştirme. 2. Basım Ankara: Nobel Kitabevi; 2014.

13. Aydın HC. Eğitim iletişimi alanında Delfi tekniğinin uygulanışı. Kurgu Dergisi 1999; 16: 225-41.

14. Atılgan H. Genellenebilirlik kuramı ve puanlayıcılar arası güvenirlik için örnek bir uygulama. Eğitim Bilimleri ve Uygulama 2005; 4: 95-108.

15. Setna Z, Jha V, Boursicot KA, Roberts TE. Evaluating the utility of workplace-based assessment tools for speciality training. Best Pract Res Clin Obstet Gynaecol 2010; 24: 767-82. [CrossRef]

16. Ali JM. Getting lost in translation? Workplace based assessments in surgical training. Surgeon 2013; 11: 286-9. [CrossRef]

17. Kara CO, Mengi E, Tümkaya F, Topuz B, Ardıç FN. Direct observation of procedural skills in otorhinolaryngology training. Turk Arch Otorhinolaryngol 2018; 56: 7-14. [CrossRef]