AMERICAN ACADEMY OF
OTOLARYNGOLOGY–
HEAD AND NECK SURGERY
F O U N D A T I O N

Randomized Controlled Pilot Study of Video Self-assessment for Resident Mastoidectomy Training

Ashok R. Jethwa, MD[1], Christopher J. Perdoni, MD[2],
Elizabeth A. Kelly, MD[3], Bevan Yueh, MD[1], Samuel C. Levine, MD[1],
and Meredith E. Adams, MD[1]

Abstract

A prospective randomized controlled pilot study was performed to determine if video self-assessment improves competency in mastoidectomy and to assess interrater agreement between expert and resident evaluations of recorded mastoidectomy. Sixteen otolaryngology residents were recorded while performing cadaveric mastoidectomy and randomized into video self-assessment and control groups. All residents performed a second recorded mastoidectomy. Performance was evaluated by blinded experts with a validated assessment scale. Video self-assessment did not lead to greater skill improvement between the first and second mastoidectomy. Interrater agreement was fair to substantial between the expert evaluators and between resident self-evaluations by recall and video review. Agreement between experts and residents was only slight to fair; residents consistently rated their performance higher than experts ($P < .05$). In conclusion, 1 session of video self-review did not lead to improved competence in mastoidectomy over standard practice. While experts agree on assessments, residents may overestimate their competency in performing cadaveric mastoidectomy.

Mastoidectomy is a critical skill for the surgical management of the ear, temporal bone, and skull base. Historically, cadaveric temporal bone simulation has been the standard method of acquiring skills to become proficient in mastoidectomy.

One novel approach to improving mastoidectomy education is through video self-assessment, similar to a postgame review by athletes. Video self-assessment of simulated and real-time procedures has proven effective in improving technical skills in several surgical specialties, including general surgery, urology, and gynecology.[1,2]

In this pilot work, our goals were to (1) determine if video self-assessment improves resident skill in cadaveric mastoidectomy over standard training and (2) establish the interrater reliability between expert and resident assessments of skill based on recorded mastoidectomy.

Methods

Study Design

This was a prospective randomized pilot study among otolaryngology residents at the University of Minnesota (N = 16). Participation was voluntary. The University of Minnesota Institutional Review Board deemed this study exempt from review (No. 1501E59582).

After reviewing the performance evaluation tools, all residents performed a cadaveric mastoidectomy recorded on a microscope-mounted camera, followed by self-assessment via recall. Participants were block randomized by training

[1]Department of Otolaryngology–Head and Neck Surgery, University of Minnesota, Minneapolis, Minnesota, USA
[2]Department of Otolaryngology–Head and Neck Surgery, Eastern Virginia Medical School, Norfolk, Virginia, USA
[3]Department of Otolaryngology–Head and Neck Surgery, Boys Town National Research Hospital, Omaha, Nebraska, USA

Corresponding Author:
Meredith E. Adams, MD, Department of Otolaryngology–Head and Neck Surgery, University of Minnesota, 420 Delaware St SE, Mayo Mail Code 396, Minneapolis, MN 55455, USA.
Email: meadams@umn.edu

Table 1. Change in Scores on Expert Evaluations and Time to Completion between First and Second Mastoidectomy in the Video Self-assessment and Control Groups.

| | Study Group, Mean (SD) | | | | |
| | Video Review (Intervention) | | No Video Review (Control) | | |
Component	First Mastoidectomy	Change[a]	First Mastoidectomy	Change[a]	P Value[b]
Task-Based Checklist					
1a. Placement of superior cut	2.78 (0.93)	0.04 (1.01)	3.00 (0.27)	0.29 (0.40)	.5499
1b. Placement of canal cut	2.81 (0.90)	−0.07 (1.08)	3.05 (0.68)	0.19 (0.69)	.5819
2a. Identification and definition of tegmen	2.33 (0.87)	−0.19 (0.53)	2.43 (1.07)	0.48 (1.51)	.2397
2b. Sharpen posterior EAC cortex	2.48 (0.85)	0.19 (0.67)	2.52 (1.02)	0.24 (0.9)	.8941
2c. Define sigmoid sinus and sinodural angle	2.48 (1.02)	0.07 (1.10)	2.81 (0.72)	−0.05 (1.08)	.8283
3a. Deepen dissection at sinodural angle	2.48 (0.77)	0.07 (0.64)	2.43 (1.01)	−0.24 (1.08)	.4827
3b. Open antrum from posterior to anterior	2.37 (0.73)	−0.22 (0.60)	2.29 (0.97)	−0.14 (1.03)	.8496
3c. Atraumatic exposure of short process of incus	2.48 (0.80)	0.22 (0.87)	2.48 (0.72)	0.62 (1.03)	.4153
4a. View posterior EAC en face	2.22 (0.83)	−0.07 (1.02)	2.14 (0.79)	0.00 (0.77)	.8759
4b. Use side/front of appropriate bur	2.67 (0.76)	0.19 (0.67)	2.29 (0.95)	−0.24 (1.15)	.3701
4c. Saucerization	2.37 (0.84)	0.04 (0.65)	2.10 (0.69)	−0.29 (1.25)	.5149
Global Rating Scale					
1. Use of otologic drills	2.59 (0.74)	0.07 (1.01)	2.33 (1.15)	−0.33(1.33)	.4973
2. Use of irrigation	2.85 (0.5)	0.22 (0.76)	3.0 (0.51)	0.10 (0.37)	.6935
3. Use of microscope	2.70 (0.65)	0.00 (0.91)	2.76 (1.10)	−0.05 (1.35)	.9341
4. Respect for surgical limits	2.41 (1.23)	−0.04 (1.12)	2.57 (1.15)	0.10 (1.63)	.8501
5. Time and motion	2.63 (0.98)	0.15 (0.63)	2.52 (1.14)	−0.24 (1.33)	.4520
6. Flow of operation	2.74 (1.01)	0.22 (0.65)	2.71 (1.11)	−0.05(1.42)	.6180
7. Overall surgical performance	2.44 (1.03)	0.0 (0.69)	2.48 (1.14)	0.05 (1.24)	.9232
Time from cortex to structure, min					
Tegmen	12.7 (10.5)	−1.0 (3.5)	11.5 (6.4)	−2.2 (6.8)	.7074
Sigmoid	14.2 (8.4)	−4.4 (10.6)	16.5 (10.0)	−7.1 (11.6)	.7154
Incus	19.2 (8.2)	−0.3 (4.9)	21.4 (9.0)	−3.8 (5.5)	.3107
Total	26.6 (13.5)	−3.1 (7.3)	37.7 (22.5)	−16.5 (22.9)	.1792

Abbreviation: EAC, external auditory canal.
[a]Change between first and second mastoid. A negative value indicates that the score in the second mastoidectomy was lower (worse).
[b]P values are for the paired *t* tests comparing the change from the first to second mastoidectomy between the study groups.

year to the intervention group (which received the recording) or to the control group (which did not). The intervention group reviewed the recording within 7 days and performed self-assessment. Both groups performed a second mastoidectomy 7 to 10 days after the first on a temporal bone of the same-side ear as the first session, followed by self-assessment via recall. All participants were given the video of their second mastoidectomy to review and self-assess within 7 days.

Two attending neurotologists and 1 neurotology fellow served as expert assessors of each recording. They were blinded to study group, resident year, and order of mastoidectomy. Experts met prior to evaluating the videos to establish consistent evaluation techniques.

Assessment Instruments

The Task-Based Checklist (TBC) and the Global Rating Scale (GRS)—developed by Francis et al for mastoidectomy and modified to enable video review—were used for all

self- and expert assessment (see Supplemental Table S1, available at www.otojournal.org/supplemental).[3] The following data were recorded: (1) time from initiation of cortical drilling to completion of mastoidectomy stages and (2) number of injuries to relevant structures. Demographic and satisfaction surveys were administered.

Statistical Analysis

We compared the change in outcome measures between the first and second mastoidectomy between the study groups with 2-tailed paired *t* tests. The mean of the scores assigned by the 3 expert evaluators was used for analysis. Injury counts were compared with chi-square tests. Significance was set at $P < .05$.

Interobserver agreement was determined with weighted kappa statistics, per the criteria of Landis and Koch for interpretation of levels of agreement.[4] STATA 14.2 (StataCorp, College Station, Texas) was used for all analyses.

Table 2. Weighted Kappa Statistics for Interrater Agreement.

Component	Weighted Kappa for Rater Agreement		
	Residents[a]	Experts[b]	Experts vs Residents[c]
Task-Based Checklist			
1a. Placement of superior cut	0.64	0.31	0.12 (0.05 to 0.18)
1b. Placement of canal cut	0.42	0.44	0.17 (0.15 to 0.18)
2a. Identification and definition of tegmen	0.51	0.51	0.21 (0.17 to 0.27)
2b. Sharpen posterior EAC cortex	0.56	0.36	0.13 (0.09 to 0.15)
2c. Define sigmoid sinus and sinodural angle	0.44	0.46	0.09 (0.02 to 0.18)
3a. Deepen dissection at sinodural angle	0.57	0.45	0.18 (0.10 to 0.29)
3b. Open antrum from posterior to anterior	0.40	0.26	0.12 (0.04 to 0.17)
3c. Atraumatic exposure of short process of incus	0.77	0.48	0.13 (0.09 to 0.18)
4a. View posterior EAC en face	0.78	0.40	0.12 (0.06 to 0.17)
4b. Use side/front of appropriate bur	0.57	0.38	0.08 (0.07 to 0.09)
4c. Saucerization	0.63	0.39	0.03 (0.02 to 0.04)
Global Rating Scale			
1. Use of otologic drills	0.52	0.50	0.03 (0.01 to 0.06)
2. Use of irrigation	0.56	0.23	0.04 (−0.02 to 0.15)
3. Use of microscope	0.73	0.45	0.07 (0.02 to 0.1)
4. Respect for surgical limits	0.51	0.61	0.22 (0.21 to 0.24)
5. Time and motion	0.45	0.42	0.18 (0.12 to 0.27)
6. Flow of operation	0.64	0.46	0.25 (0.16 to 0.32)
7. Overall surgical performance	0.68	0.62	0.19 (0.13 to 0.26)

Abbreviation: EAC, external auditory canal.
[a]Self-recall vs self-video assessment.
[b]Agreement among 3 expert assessments.
[c]Weighted kappa was calculated for agreement between each expert (n = 3) and resident self-video assessments. The mean (range) of the 3 kappas is presented here.

Results

Study groups were balanced on resident experience based on training year and mastoids previously drilled in the laboratory (mean $\pm$ SD, 6 $\pm$ 4 vs 5 $\pm$ 4) and operating room (30 $\pm$ 24 vs 31 $\pm$ 22). There were no significant differences between the first and second mastoidectomy in TBC or GRS scores, completion time of mastoidectomy stages, or injury counts between study arms or within either study arm (**Table 1**). Injury counts were low for all structures except the tegmen (11 injuries).

Interrater agreement (**Table 2**) was fair to substantial among the expert evaluators (κ = 0.23-0.62) with highest agreement on overall surgical performance. Interrater agreement was fair to substantial among resident self-evaluations by recall and video review (κ = 0.40-0.78). There was only slight to fair agreement between expert ratings and resident ratings (κ = 0.03-0.25). For all TBC and GRS items, resident ratings were 0.41 to 1.51 points higher than the mean expert ratings (all $P <$ .007; **Figure 1**).

Residents rated satisfaction with video assessment highly (4 $\pm$ 1.4 out of 5), and 78% said that they would repeat the study.

Discussion

In this study, 1 episode of video self-review did not produce improved competence in mastoidectomy over standard training. Expert and resident assessments conflicted, as residents consistently rated themselves higher than experts did.

Our findings should be considered in the context of other literature. Malik et al demonstrated a negative association between time spent in the temporal bone laboratory and resident competence in mastoidectomy.[5] In their study and ours, residents did not receive expert feedback or coaching. Expert feedback is critical for complex task learning. Hu et al demonstrated that, in contrast to standard practices, video-based coaching sessions of surgical procedures generate more questions and detailed discussions between attendings and residents about intraoperative decision making.[6] Feedback is also important for development of accurate self-estimations of skill.[7-9] Thus, these findings suggest that mastoidectomy simulation without feedback may not promote new skill development or it may lead to bad habits, as residents may not recognize or correct their own technical errors and inefficiencies.

Video review also allows experts to reflect on educational techniques, identifying areas that require focus. For example, we increased our emphasis on early definition of
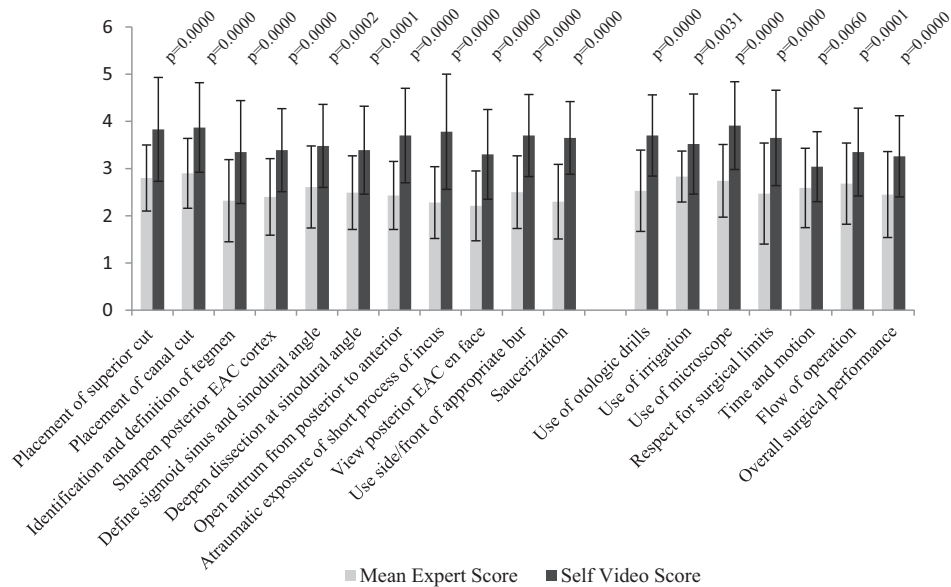
Figure 1. Expert vs resident self-assessment of recorded mastoidectomy performance with the Task-Based Checklist and Global Rating Scale. EAC, external auditory canal. Mean values are presented, with error bars indicating SD.

tegmen contours to address the high rates of injury observed. Furthermore, we learned that recordings provide an accessible and efficient means of observing resident drilling and may augment evaluation and feedback when attendings cannot be present during each laboratory dissection.

The primary limitation of our study is its small sample size, drawn from a single institution. Enrollment in multiple programs would increase study power and external validity. Furthermore, as self-video review was limited to 1 session, additional video review sessions with multiple cadaveric temporal bone dissections may lead to improvements that could not be detected in this study. Future work will determine if joint video review with an expert may afford greater educational value.

Author Contributions

Ashok R. Jethwa, substantial contributions to the conception and design of the work, drafting the work and revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Christopher J. Perdoni**, substantial contributions to the acquisition of the data for the work; revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Elizabeth A. Kelly**, substantial contributions to the acquisition and interpretation of the data for the work; revising it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Bevan Yueh**, substantial contributions to the design of the work; revising the work critically for important intellectual content; revising it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Samuel C. Levine**, substantial contributions to the acquisition of the data for the work; revising it critically for important intellectual content; final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Meredith E. Adams**, substantial contributions to the conception and design of the work, drafting the work and revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Disclosures

Supplemental Material

Additional supporting information is available in the online version of the article.

References

1. Jamshidi R, LaMasters T, Eisenberg D, et al. Video self-assessment augments development of videoscopic suturing skill. *J Am Coll Surg.* 2009;209:622-625.

2. Carter SC, Chiang A, Shah G, et al. Video-based peer feedback through social networking for robotic surgery simulation: a multicenter randomized controlled trial. *Ann Surg.* 2015;261:870-875.

3. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope.* 2009;119:2402-2410.

4. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics.* 1977;33:363-374.

5. Malik MU, Varela DA, Park E, et al. Determinants of resident competence in mastoidectomy: role of interest and deliberate practice. *Laryngoscope.* 2013;123:3162-3167.

6. Hu YY, Mazer LM, Yule SJ, et al. Complementing operating room teaching with video-based coaching. *JAMA Surg.* 2017; 152:318-325.

7. Zevin B. Self versus external assessment for technical tasks in surgery: a narrative review. *J Grad Med Ed.* 2012;4:417-424.

8. Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof.* 2008;28:14-19.

9. Evans AW, Leeson RM, Newton John TR, et al. The influence of self-deception and impression management upon self-assessment in oral surgery. *Br Dent J.* 2005;198:765-769.