Full citation: Kerwin T, Wiet G, Hittle B, Stredney D, Moberly A, De Boeck P, Andersen SA. Standard setting of competency in mastoidectomy for the Cross-Institutional Mastoidectomy Assessment Tool. Ann Otol Rhinol Laryngol. 2020 Apr;129(4):340–346.
DOI: 10.1177/0003489419889376

Title: Standard setting of competency in mastoidectomy for the Cross-Institutional Mastoidectomy Assessment Tool

Authors: Thomas Kerwin, PHD[1], Gregory Wiet, MD[2,3], Brad Hittle, BS[1], Don Stredney, MA[1], Paul De Boeck, PHD[4], Aaron Moberly, MD[2], Steven Arild Wuyts Andersen, MD, PHD[2,3]
1. Office of Research, Ohio State University, Columbus, Ohio, United States.
2. Department of Otolaryngology, Ohio State University, Columbus, Ohio, United States.
3. Department of Otolaryngology, Nationwide Children's Hospital, Columbus, Ohio, United States.
4. Department of Psychology, Ohio State University, Columbus, Ohio, United States.

Corresponding Author: Thomas Kerwin, Ohio State University, 1305 Kinnear Rd, Suite 194, Columbus, Ohio, 43212, United States; Email address: kerwin.6@osu.edu; Phone (614) 360-3357.

Financial disclosure: The authors have no financial interest in the content of this work.

Conflict of interest: The authors have no conflicts of interest for this work.

Funding: This work was supported by The National Institute for Deafness and other Communication Disorders, National Institutes of Health, USA, R01DC011321. Steven Andersen was supported by an international postdoc grant from the Independent Research Fund Denmark, 8026-00003B.

Abstract

Objective: Competency-based surgical training involves progressive autonomy given to the trainee. This requires systematic and evidence-based assessment with well-defined standards of proficiency. The objective of this study is to develop standards for the cross-institutional mastoidectomy assessment tool to inform decisions regarding whether a resident demonstrates sufficient skill to perform a mastoidectomy with or without supervision.

Methods: A panel of fellowship-trained content experts in mastoidectomy was surveyed in relation to the 16 items of the assessment tool to determine the skills needed for supervised and unsupervised surgery. We examined the consensus score to investigate the degree of agreement among respondents for each survey item as well as additional analyses to determine whether the reported skill level required for each survey item was significantly different for the supervised versus unsupervised level.

Results: Ten panelists representing different US training programs responded. There was considerable consensus on cut-off scores for each item and trainee level between panelists, with moderate (0.62) to very high (0.95) consensus scores depending on assessment item. Further analyses demonstrated that the difference between supervised and unsupervised skill levels was significantly meaningful for all items. Finally, minimum passing scores for each item was established.

Conclusion: We defined performance standards for the cross-institutional mastoidectomy assessment tool using the Angoff method. These cut-off scores that can be used to determine when trainees can progress from performance under supervision to performance without supervision. This can be used to guide training in a competency-based training curriculum.

Keywords: Mastoidectomy, assessment, standard setting, curriculum development, evidence-based medical education.

Introduction

The surgical educational paradigm has over the last couple of decades shifted towards competency-based education rather than the logged number of surgical cases performed. During the years of surgical training, performance at a certain level of technical skill in the operating room (OR) is required before further responsibilities are entrusted to the trainee. The competency-based paradigm requires not only the integration of competency evaluation into the surgical training programs but also evidence that the evaluations are valid and reliable. Consequently, precise definition of competency levels and evidence-based assessment are key components of the competency-based surgical training paradigm.

Since the objective and structured assessment of surgical skills (OSATS) method of creating assessment tools was first introduced in 1997,[1] specific tools have been developed for many surgical procedures. In temporal bone surgery, multiple tools for the structured assessment of mastoidectomy performance have been developed. Different approaches have been used such as task-based checklists, global rating scales, and final-product analysis[2] for the different training settings including intra-operative assessment of performance and for feedback[3], as well as in a cadaveric lab training setting and in virtual reality simulation.[4–7] In a recent systematic review[8], the current validity evidence of different mastoidectomy assessment tools have been scrutinized in relation to Messick's contemporary framework of validity.[9] There is currently good evidence in relation to content (i.e. that assessment content reflects the intended construct), internal structure (how individual assessment items relate to the overall construct) and relations to other variables (correlation with other assessments of the same construct). In contrast, there is almost no validity evidence on response process (alignment between assessment construct and rater-subject thought processes) and consequences of assessment (for example standard-setting for pass/fail decisions).[8]

Structured assessment of temporal bone surgical skills have been used to define milestones towards competency in mastoidectomy.[10] This is an example of how assessment can be integrated into the clinical training curriculum for classification of competency level with the potential identification of trainees needing remediation. Systematic evaluation of competency should be used to determine when the trainee can progress from performing the procedure with supervision from an attending to surgery without direct supervision, i.e. the ability to safely perform mastoidectomy in independent clinical practice. However, such standards need to be well-defined to be used as a guide for determining surgical responsibility and specific feedback. In order to be generalizable across residency programs, these standards must be developed and evaluated across institutions and represent a consensus from multiple experts with various otologic training backgrounds. Since the standards would have wide implications for the and on the overall training curriculum across institutions, broad consensus is a vital part of standards development.

In a previous study, a large panel consisting of the members of American Neurotology and Otological Societies was surveyed to identify the most important items for mastoidectomy performance assessment.[11] This cross-institutional assessment tool represents the key elements that can be used to define competency. In this study, we therefore invited a smaller panel of fellowship trained surgical otologists to define skill thresholds for resident advancement with the purpose of defining surgical standards for trainees learning to perform mastoidectomy safely and effectively. The ultimate goal is to use these standards to determine when a trainee is sufficiently experienced to progress to performance of the procedure without direct supervision.

Methods

Study design and participants

This study was designed as a survey of experts in temporal bone surgery with the purpose of defining standards of performance for competency assessment of ORL trainees for a cross-institutional assessment tool. The survey was conducted in May 2017.

As content experts for our panel, we invited 17 fellowship-trained, attending otologic surgeons associated with ORL residency training programs at different institutions across the United States. The panelists were selected based on their previous involvement in studies of a computerized temporal bone surgery simulation system. The panelists were invited by e-mail.

The Cross-Institutional Mastoidectomy Assessment Tool

Items for a cross-institutional assessment scale for mastoidectomy performance was compiled through a survey of members of the American Neurotology and Otological Societies in a previous study.[11] This resulted in 24 items being ranked as "Important" or "Very Important" by more than 70% of the 88 responding panelists. These select items were later reviewed and validated using a Delphi process[12], resulting in the merger of overlapping items and for informing potential descriptive anchors to guide rating of the individual items.

In this study, we aggregated the information from these previous studies to operationalize the Cross-Institutional Mastoidectomy Assessment Tool (CIMAT) (Table 1 and Supplementary Digital Content 1). This tool consists of a total of 16 items each graded on a 5-point Likert scale according to the trainee demonstrating no skill (0 points), slight skill (1 point), moderate skills (2 points), high skills (3 points), and expert skill (4 points). Descriptive anchors for the extreme and middle values are provided

as a reference for the rater. The items have been ordered according to the steps of the procedure to ease rating with more global items ordered last.

Standard-setting

A number of different approaches can be used for standard-setting of performance.[13] In this study, we chose to use a criterion-based, item-based approach based on the appraisal by content experts (Angoff method). We further chose to examine items individually using an absolute scale instead of using an overall instrument score, since specific elements across items are not interchangeable because the assessment tool had already been reduced.

Our panelists were provided a link to an online survey (Supplementary Digital Content 2) for establishing the standards for performance on the CIMAT. First, the panelists were asked to provide some background information on their experience in temporal bone surgery to verify their content expertise. Next, the panelists were presented for each of the 16 items and the descriptive anchors associated with the specific item. The panelists were then asked to choose the 5-point score they felt should indicate the minimum level of skill for a trainee that is able to perform mastoidectomy (1) with supervision, and (2) unsupervised. This would allow to set a standard for when the trainee is proficient to commence supervised surgery on patients, and also when the trainee is competent for independent surgery. The panelists were also provided the opportunity to comment on each item and the suggested anchors.

Statistical methods

Statistical analyses were conducted with the R statistical software (R Foundation for Statistical Computing, Vienna, Austria) using the '*effsize*' and '*agrmt*' packages. To explore the degree of

consensus among participants for each item, we used Tastle and Wierman's[14] consensus measure *c*. The *c* score ranges from 0 (maximal non-consensus) to 1 (perfect agreement) and is related to the narrowness of the histogram of the respondents' answers. In contrast to older measures such as agreement percentage , this is well suited for multiple raters. We interpret *c* scores between 0.60–0.75 as a moderate degree of consensus, 0.75–0.85 as a high degree of consensus and 0.85–1.0 as a very high degree of consensus. The Wilcoxon test was used to assess whether there were differences between the distributions of scores for the two skill levels (supervised/unsupervised), using the Benjamini–Hochberg correction to reduce false positives. To determine whether the reported skill level differed for the two skills levels for each item, a Wilcoxon signed-rank test was performed. Cliff's δ was computed to measure the strength of the difference of results between the two progress categories[15]: $\delta<0.3$ is considered a small effect size, $0.3<\delta<0.6$ is a medium effect size, and $\delta\geq0.6$ is considered a large effect size. Finally, we determined the cut-off score for each item for the two levels (supervised and unsupervised surgery).

Ethics

The study was approved by our institutional review board (#2017E0328).

Results

Ten out of the 17 invited panelists provided complete responses, resulting in a 59 % response rate. The number of years in practice for the respondents ranged from 4 to 30 (median 22 years). The panel reported substantial experience in temporal bone surgery, verifying their content expertise: six respondents performed over 100 mastoidectomies per year, three indicated 81–100 per year, one indicated 61–80 per year, and none reported performing fewer than 60 mastoidectomies per year.

For standard setting of performance level, each panelist was asked to specify, which score they determined should be required for a trainee that is at the level (1) able to perform the surgery with supervision and (2) able to perform the surgery unsupervised/without attending present. The results are presented as histograms for each item (Figure 1).

There was considerable consensus on cut-off scores for each item and trainee level between panelists, with moderate to very high consensus scores (Figure 2). The item *creates appropriate depth of cavity* had the lowest consensus (*c*) score of 0.62, interpreted as a moderate degree consensus, whereas the majority of items had *c* scores in the range 0.75–0.85, interpreted as a high degree of consensus. Several items demonstrated a very high degree of consensus with *c* scores of ≥0.92 such as for example *avoids violation of the sigmoid sinus, drills in best direction, correct identification of chorda tympani nerve,* and *posterior external auditory canal wall is thinned appropriately.*

The distributions of the scores for the two skill levels (supervised/unsupervised) were distinct for each of the items (Wilcoxon tests, $p<0.01$ for each item). The mean difference between the anchor values assigned as cut-off for supervised and unsupervised performance was 1.2, with a minimum of 0.9 and maximum of 1.6 points. For all items, Cliff's δ was found to be ≥0.6, thus demonstrating a large effect size of the difference in skill level for each of the items: in other words, there was a substantial difference in what was determined to be the cut-off for supervised and unsupervised performances.

Finally, the minimum score for passing for each item (cut-off) was calculated as the median of the panelists responses rounded up to ensure sufficient performance. These are marked in Figure 1.

All items have 2 as the minimum score for supervised performance except for Item 7, *avoids drill contact with ossicles*, which has a 3 minimum score. Most items have 3 as the minimum score for independent performance, except for Items 8, 10, 6, 7, 12 and 9, which have 4 as the minimum score. A more detailed summary of the findings is provided in *Supplementary Digital Content 3*.

Discussion

In this study, we explored standard-setting of the cross-institutional mastoidectomy assessment tool using the Angoff method and a content expert panel of fellowship-trained otologists from across the US. This resulted in cut-off scores that can be used to determine when trainees can progress from performance under supervision to performance without supervision for the use in a competency-based training curriculum. Consensus scores among respondents were overall high. As expected, the content experts indicated overall higher minimum skill scores required for each item to classify ready for unsupervised surgery.

Some items for which we expected very high consensus scores this was not found: an example is item 12, *avoids violation of the facial nerve,* which intuitively should be easy to determine but only demonstrated moderate consensus (*c*-score 0.72). This suggests that the item can potentially be improved in relation to clarity of the descriptive anchors so that the ambiguity of the middle anchor ("*facial nerve partially exposed*"), representing moderate skill, more clearly distinguishes between the undesirable partial exposure of the nerve sheath during dissection or the more positive skill progress relating to partial successful identification of the nerve in the mastoid segment. Altogether, this highlights the inherent nature of assessment: for some items, some degree of interpretation is difficult to prevent and in addition, variation in what experts consider acceptable skill exists. We further found a

clear and significant difference in what otologists believe are acceptable skill levels for performing supervised and unsupervised surgery (Wilcoxon tests and a high Cliff's δ). This supports that the items represent essential skills that are useful in determining the progression of trainees across a wide range of expert opinion.

Few of the other available mastoidectomy performance assessment tools have defined standards of performance based on empirical data. For final-product assessment of performances in a virtual reality simulator, a cut-off of 19.5/26 points was determined based on the expert performance approach.[16] This has two major limitations: first of all, final-product analysis reflects only the end product and does not consider process. Secondly, the use of overall score does not reveal major shortcomings on some item conditions that are absolutely necessary for acceptability: for example, major violation of the facial nerve is inacceptable in the OR but represents for the specific final-product tool only a 1-point deduction from the 26-point total. Weighted scores could lessen this problem but will not entirely eliminate it.

A strength of our approach is therefore the absolute criteria with clear minimum standards are better at ensuring a safe and adequate performance and that this standard was defined by a panel representing different training programs and traditions across the US. A limitation is the number of panelists, however, qualitative approaches such as our modified Angoff typically reaches saturation at 10–15 participants[17] and our data indicate that we achieved sufficient saturation of responses. However, even though increasing the number of panelists could be speculated to yield equivalent results, a larger sample would provide stronger evidence for our results. Another limitation relates to the standard-setting of the level of unsupervised surgery: many items showed a split between 3 and 4, which makes it difficult to firmly conclude which is the appropriate anchor number for unsupervised performance of

that item. Finally, although this study provides a first step towards developing performance standards in mastoidectomy for otolaryngology training programs, it cannot inform us as to how to implement those standards in the training curriculum, real life application for assessment and feedback in the operating room, nor the most appropriate teaching methods.

Competency-based surgical training has been introduced to ensure that all surgeons can provide safe service to the patients, which is the objective of any surgical training program.[18] In such a curriculum, it is essential to define milestones of progression[10] and implement systematic assessment as a measurement of competency as well as specific and measurable criteria that can be used to inform curriculum design and support deliberate practice.[19] Furthermore, assessment that can be used for monitoring throughout residency should be developed and has potential for quality evaluation of different training programs.[20] To ensure consistency, such universal standards should be defined based on broad expert input and adopted across all training programs.

Conclusion

We have investigated standard-setting of performance for the cross-institutional mastoidectomy assessment tool that represent the key items identified by a large panel as the most important skills in mastoidectomy. Using a sub-panel representing different training programs across the US, we defined skill thresholds that determine when a trainee is sufficiently experienced to progress to surgery without direct supervision. These cut-offs demonstrated moderate-to-very high consensus and substantial difference in the performance level of supervised and unsupervised surgery. Such standard-setting with clear criteria is important for competency-based surgical training and has potential implications for training curricula and design of training programs.

Acknowledgements

This work was supported by The National Institute for Deafness and other Communication Disorders, National Institutes of Health, USA, R01DC011321.

References

1. 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.
2. Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective Assessment of Temporal Bone Drilling Skills. *Ann Otol Rhinol Laryngol.* 2007;116(11):793-798. doi:10.1177/000348940711601101.
3. Francis HW, Masood H, Chaudhry KN, et al. Objective Assessment of Mastoidectomy Skills in the Operating Room. *Otol Neurotol*. 2010;31(5):759-765. doi:10.1097/MAO.0b013e3181e3d385.
4. Butler NN, Wiet GJ. Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance. *Laryngoscope*. 2007;117(10):1803-1808. doi:10.1097/MLG.0b013e31811edd7a.
5. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope*. 2009;119(12):2402-2410. doi:10.1002/lary.20678.
6. Zhao YC, Kennedy G, Yukawa K, Pyman B, O'Leary S. Improving temporal bone dissection using self-directed virtual reality simulation: results of a randomized blinded control trial. *Otolaryngol Head Neck Surg*. 2011;144(3):357-364. doi:10.1177/0194599810391624.
7. Andersen SAW, Caye-Thomasen P, Sorensen MS. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. *Laryngoscope*. 2015;125(2):431-435. doi:10.1002/lary.24838.
8. Sethia R, Kerwin TF, Wiet GJ. Performance Assessment for Mastoidectomy. *Otolaryngol Head Neck Surg*. 2017;156(1):61-69. doi:10.1177/0194599816670886.
9. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *Am J Med*. 2006;119(2):166.e7-166.e16. doi:10.1016/j.amjmed.2005.10.036.

10. Francis HW, Masood H, Laeeq K, Bhatti NI. Defining milestones toward competency in mastoidectomy using a skills assessment paradigm. *Laryngoscope*. 2010;120(7):1417-1421. doi:10.1002/lary.20953.
11. Wan D, Wiet GJ, Welling DB, Kerwin T, Stredney D. Creating a cross-institutional grading scale for temporal bone dissection. *Laryngoscope*. 2010;120(7):1422-1427. doi:10.1002/lary.20957.
12. Kerwin T, Hittle B, Stredney D, De Boeck P, Wiet G. Multi-Institutional Development of a Mastoidectomy Performance Evaluation Instrument. *J Surg Educ*. 2017;74(6):1081-1087. doi:10.1016/j.jsurg.2017.05.006.
13. Thinggaard E, Bjerrum F, Strandbygaard J, *et al*. Ensuring Competency of Novice Laparoscopic Surgeons—Exploring Standard Setting Methods and their Consequences. *J Surg Educ*. 2016;73(6):986-991. doi:10.1016/j.jsurg.2016.05.008.
14. Tastle WJ, Wierman MJ. Consensus and dissention: A measure of ordinal dispersion. *Int J Approx Reason*. 2007;45(3):531-545. doi:10.1016/j.ijar.2006.06.024.
15. Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol Bull*. 1993;114(3):494-509. doi:10.1037/0033-2909.114.3.494.
16. Andersen SAW, Mikkelsen PT, Sørensen MS. Expert sampling of VR simulator metrics for automated assessment of mastoidectomy performance. *Laryngoscope*. 2019 Jan 9 [Epub ahead of print]. doi:10.1002/lary.27798.
17. Hurtz GM, Hertz NR. How Many Raters Should be Used for Establishing Cutoff Scores with the Angoff Method? A Generalizability Theory Study. *Educ Psychol Meas*. 1999;59(6):885-897. doi:10.1177/00131649921970233.
18. Bhatti NI, Cummings CW. Viewpoint: Competency in Surgical Residency Training: Defining and Raising the Bar. *Acad Med*. 2007;82(6):569-573. doi:10.1097/ACM.0b013e3180555bfb.

19. Bhatti NI, Ahmed A. Improving skills development in residency using a deliberate-practice and learner-centered model. *Laryngoscope*. 2015;125 Suppl 8:S1-S14. doi:10.1002/lary.25434.

20. Awad Z, Hayden L, Muthuswamy K, Ziprin P, Darzi A, Tolley NS. Does direct observation of procedural skills reflect trainee's progress in otolaryngology? *Clin Otolaryngol*. 2014;39(3):169-173. doi:10.1111/coa.12251.

	Item	*Demonstrates no skill*	*Demonstrates moderate skill*	*Demonstrates expert skill*
1	**Creates appropriate depth of cavity**	Antrum not entered or horizontal canal not visualized	Antrum opened without damage to horizontal canal or tegmen	Antrum widely opened with adequate thinning of tegmen and posterior superior canal wall
2	**Avoids violation of the sigmoid sinus**	Penetrates sigmoid, unaware of its location	Exposes sigmoid enough to identify sufficiently to avoid violation but may leave overlying air cells	Sigmoid well defined for procedure, may expose sigmoid for retraction and better exposure
3	**Avoids holes in tegmen**	Tegmen and dura violated	Dura exposed without violation	Tegmen thinned appropriately for surgical approach, possibly removing tegmen to retract dura for better exposure
4	**Avoids violation of dura**	Dura violated (opened)	May overthin tegmen and expose dura without violation	Tegmen completely dissected to sinodural angle, dura may be exposed intentionally
5	**Maintains a complete saucerization**	Penetrates sigmoid, residual air cells, facial ridge not identified, antrum not opened appropriately, tegmen and sigmoid not defined at their locations	Adequate air cell removal to avoid damage to critical structures, antrum opened sufficient for visualization of horizontal canal, fossa incudis, etc.	All necessary air cells removed and critical structures well defined
6	**Avoids violation of the horizontal (lateral) semi-circular canal**	Horizontal canal violated	Horizontal canal accidentally bluelined	Horizonal canal easily identified, may blueline as needed for exposure
7	**Avoid drill contact with ossicles**	Contacts ossicles with cutting burr	Contacts ossicles with diamond burr	Drills close to ossicles with appropriate burr
8	**External auditory canal remains up**	Significant posterior canal wall lowering deep to lateral extent.	Minor lowering of posterior canal wall	Maintains posterior canal wall completely intact
9	**Posterior external auditory canal wall is thinned appropriately**	Posterior canal wall thick, poor definition of facial ridge if any	Partial definition of facial ridge	Sufficient thinning to fully expose facial nerve and chorda tympani for recess approach
10	**Avoids holes in external auditory canal**	Multiple or large holes in posterior canal wall	Minor, clinically insignificant holes in posterior canal wall	No holes in posterior canal wall

11	**Correct identification of chorda tympani nerve**	Does not identify or expose chorda	Opens facial recess but does not identify choda completely	Exposes and widely opens facial recess without violations of chorda or facial nerves
12	**Avoids violation of the facial nerve**	Facial nerve violated	Facial nerve partially exposed	Facial nerve exposed completely throughout its course in the mastoid
13	**Maintains visibility of burr while removing bone**	Burr view dangerously obstructed for example beneath bony ledge when thinning tegmen	Burr visibility safely unobstructed throughout procedure	Burr visibility obstructed but at times when it is safe to do so, during decortication of mastoid
14	**Selects appropriate burr**	Uses too small or too large of a burr for task at hand, uses diamond burr for decortication or when not necessary, or uses cutting burr in close proximity to critical structure	May use too small or too large a burr for occasion or using diamond burr excessively	Uses cutting and diamond burrs appropriately, such as using cutting burr to enhance speed of dissection but in a safe manner
15	**Drills in best direction**	Drills perpendicular to critical strucutes or without regard to trajectory of critical structure	Maintains drill direction parallel to critical structures	Alternates drill direction rapidly for efficient removal of bone without jeopardizing critical structures
16	**Avoids excessive force near critical structures**	Does not alter force when approaching critical structures	Alternates between minimal force and moderate force throughout case	Uses high force when appropriate, such as decortication of mastoid

Table 1. The Cross-Institutional Mastoidectomy Assessment Tool (CIMAT). Each item is rated on a 0-4 scale. The "no skill", "moderate skill", and "expert skill" anchor descriptions are for the 0, 2, and 4 ratings on the scale.

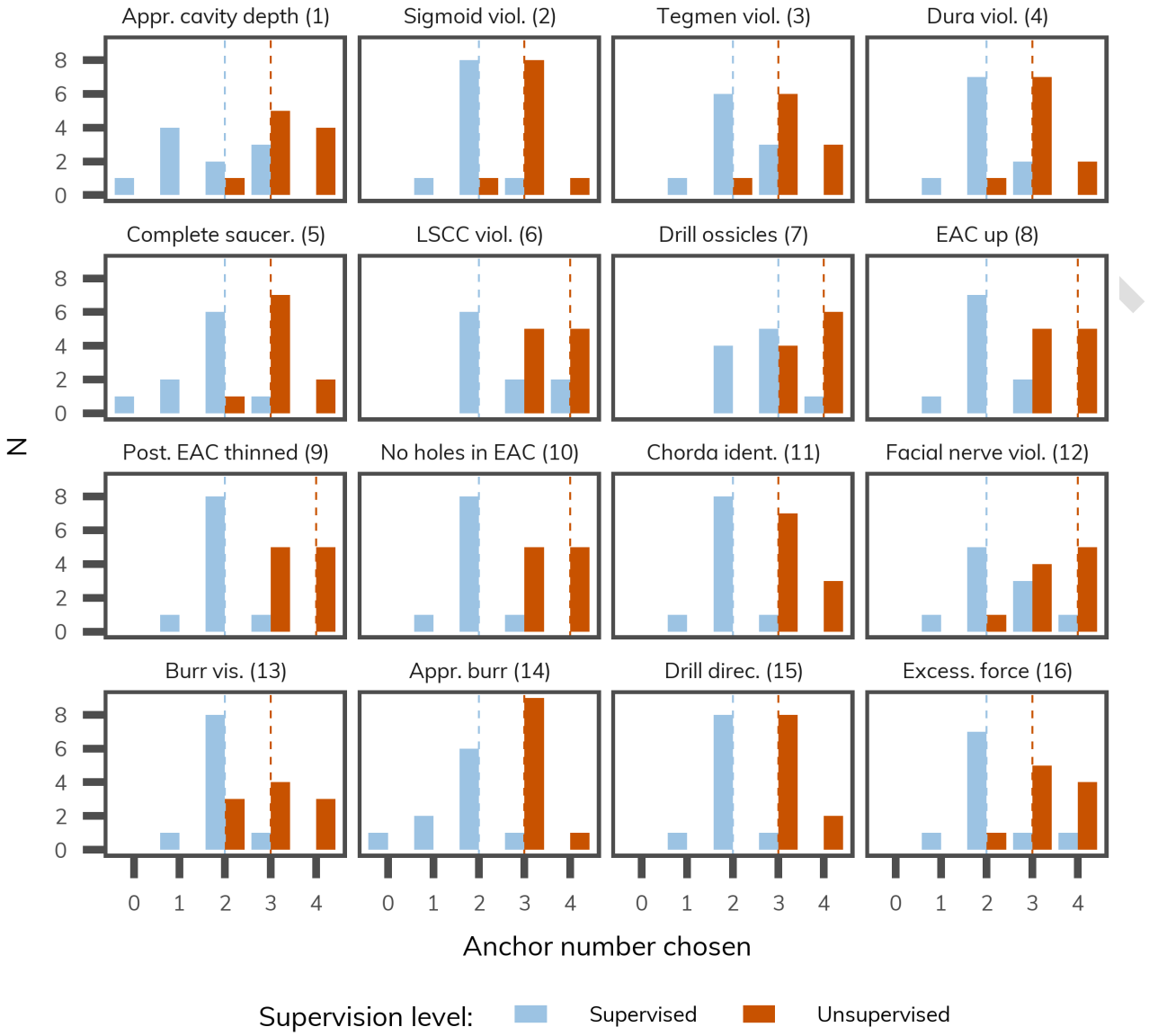

Figure 1. Histogram of the panelists' selected cut-off scores for performance in relation to each item according to the trainee level (level of supervision). The height of the bars indicate the number of responses. Dashed vertical lines indicate the determined cut-off scores.

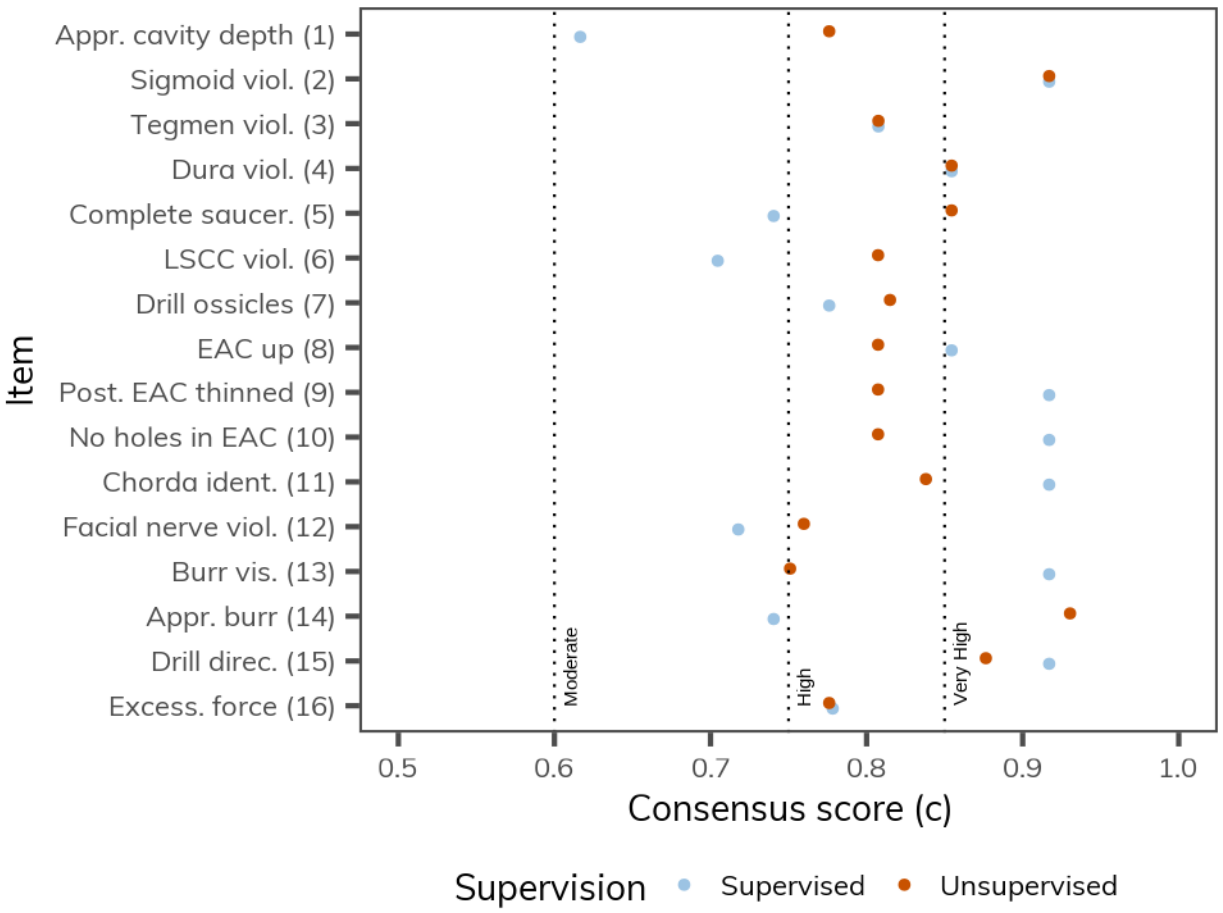

Figure 2. The consensus measure (*c*) for all items for trainees at a supervised and unsupervised levels. Possible *c* values range from 0 to 1, with 1 being complete consensus.

Supplementary Digital Content

The raw data for this manuscript is located on FigShare at

https://doi.org/10.6084/m9.figshare.6328292.v2

There are three supplemental data files mentioned in the text which refer to documents that are on FigShare at the link above:

Supplementary Digital Content 1. The Cross-Institutional Assessment Tool in a two-page format. (CIMAT.pdf)

Supplementary Digital Content 2. The text of the survey given to the panelists. The survey was provided online. (surveytext.pdf)

Supplementary Digital Content 3. Detailed notes and comments on the responses for each item. (SDC3 - Notes on responses.docx**)**