Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/techniques (UM-OSCAARS): a validation study

Ségolène Chagnon-Monarque, MD
Owen Woods, MD
Apostolos Christopoulos, MD, MSc
Eric Bissada, MD, DMD
Christian Ahmarani, MD
Tareck Ayad, MD

Accepted May 12, 2020

Correspondence to:
T. Ayad
Centre hospitalier de l'Université de Montréal
1001 Sanguinet
Montréal QC H2W 1T8
tareck.ayad@umontreal.ca

DOI: 10.1503/cjs.018418

Background: Use of videos of surgical and medical techniques for educational purposes has grown over the last years. To our knowledge, there is no validated tool to specifically assess the quality of these types of videos. Our goal was to create an evaluation tool and study its intrarater and interrater reliability and its acceptability. We named our tool UM-OSCAARS (Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/techniques).

Methods: UM-OSCAARS is a grid containing 10 criteria, each of which is graded on an ordinal Likert-type scale of 1 to 5 points. We tested the grid with the help of 4 voluntary otolaryngology – head and neck surgery specialists who individually viewed 10 preselected videos. The evaluators graded each criterion for each video. To evaluate intrarater reliability, the evaluation took place in 2 different phases separated by 4 weeks. Interrater reliability was assessed by comparing the 4 top-ranked videos of each evaluator.

Results: There was almost-perfect agreement among the evaluators regarding the 4 videos that received the highest scores from the evaluators, demonstrating that the tool has excellent interrater reliability. There was excellent test–retest correlation, demonstrating the tool's intrarater reliability.

Conclusion: The UM-OSCAARS has proven to be reliable and acceptable to use, but its validity needs to be more thoroughly assessed. We hope this tool will lead to an improvement in the quality of technical videos used for educational purposes.

Contexte : Au fil des ans, l'utilisation de vidéos pour l'enseignement de techniques chirurgicales et médicales s'est répandue. À notre connaissance, il n'existe aucun outil pour évaluer spécifiquement la qualité de ces types de vidéos. Notre objectif était de créer un outil d'évaluation et d'analyser sa fiabilité interévaluateurs et son acceptabilité. Notre outil a pour nom UM-OSCAARS (Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/Techniques).

Méthodes : L'outil UM-OSCAARS est une grille qui contient 10 critères; chacun est noté sur une échelle de type Likert de 1 à 5 points. Nous avons testé la grille avec l'aide de 4 volontaires, spécialistes en otorhinolaryngologie/chirurgie de la tête et du cou, qui ont visionné 10 vidéos présélectionnées. Les évaluateurs ont noté chacun des critères pour chaque vidéo. Afin de vérifier la fiabilité interévaluateurs, l'évaluation s'est déroulée en 2 phases, à 4 semaines d'intervalle. La fiabilité interévaluateurs a été mesurée en comparant les 4 vidéos les mieux cotées par chaque évaluateur.

Résultats : La concordance a été quasi parfaite entre les évaluateurs pour les 4 vidéos qu'ils ont les mieux cotées, ce qui montre que l'outil a une excellente fiabilité interévaluateurs. La corrélation test–retest a été excellente, ce qui démontre la fiabilité interévaluateurs de l'outil.

Conclusion : L'outil UM-OSCAARS et son utilisation se sont révélés fiables et acceptables, mais il faut évaluer davantage sa validité. Nous espérons que cet outil permettra d'améliorer la qualité des vidéos techniques destinées à l'enseignement.

In this age of electronics and communication, emerging technologies are the key to education in modern medicine. Medical education must evolve at the same pace as the digitally oriented world in which we live. One of the pioneers in this field has been the Stanford University School of Medicine, which collaborated with the Khan Academy to develop a flipped classroom model, where students learn from home with a series of short videos and do their homework in the classroom.[1] It is common practice to offer educational alternatives to medical students, and the use of videos seems to meet the needs of the current digital generation of learners.[2] In a pilot study evaluating the impact of otology surgery videos in otolaryngology residency education, residents considered the videos highly useful and perceived them as a high priority for a resident's surgical preparation.[3] A recent study evaluating the impact of a flipped-classroom, video-based surgical curriculum on the surgical skills of dermatology residents showed that the use of videos in that model significantly improved the residents' surgical ability as measured by an objective structured assessment of technical skills (OSATS) instrument on simulation models.[4]

Production of videos showing technical procedures or surgical techniques is gaining in popularity, as witnessed by the increase in the number of articles accompanied by videos being published in peer-reviewed journals and the number of video sessions being held at international conferences. However, surgical skills and expertise do not always carry over into skilful video production. Some articles have been published that provide guidance to clinicians on how to optimize the quality of educational videos.[5,6] However, to the best of our knowledge, there are no validated tools to assess the quality of surgical and technical videos, even though their use for educational purposes has been democratized with free online resources. Our objectives were to develop a tool to assess the quality of videos focusing on surgical procedures or medical techniques and to study its intrarater reliability, interrater reliability and acceptability.

Methods

Creation of the tool

We created an evaluation tool for videos about surgical procedures and medical techniques in the form of a checklist we named the Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/techniques (UM-OSCAARS). The checklist was developed by 4 expert surgeons and 1 audiovisual professional. The 4 expert surgeons were otolaryngology – head and neck surgeons who had produced videos depicting surgeries and techniques for publication or teaching purposes. The criteria were chosen through the use of a modified Delphi method with a series of 3 rounds. The checklist contains 10 criteria focusing on clinical relevance and audiovisual quality, with 5 criteria in each category (Box 1). Each of the criteria are graded on an ordinal Likert-type scale of 1 to 5 points with descriptors for scores 1, 3 and 5 (Table 1). The descriptors were also part of the validation process, which was done using a modified Delphi method. A more thorough description of each criterion is also provided as a guide to enable users to fully understand each criterion (Table 2).

Choice of assessors

Four otolaryngology – head and neck surgery specialists who were not involved in developing the tool volunteered to participate in the project as assessors. These 4 staff physicians had academic practices in different subspecialties and a wide range of years of experience (Table 3).

Choice of videos

The senior author (T.A.) chose 10 videos from various sources, including videos from YouTube and ones published in peer-reviewed journals (*New England Journal of Medecine*, *Plastic and Reconstructive Surgery*, *Head & Neck*). The use of videos from a video-sharing website (YouTube) and from peer-reviewed journals created an opportunity to present the assessors with videos of possibly different levels of quality. The senior author (T.A.) selected videos with a range content related to several subspecialties of otolaryngology – head and neck surgery or relevant to a general practice. The characteristics of the selected videos are provided in Table 4. The order in which the videos were shown to the assessors was randomly selected; the order was the same for each assessor. The total viewing time of the 10 videos was 59 minutes and 57 seconds.

Data collection

To assess intrarater reliability, we organized a 2-phase evaluation plan: the evaluation took place in 2 different

Box 1. Criteria of the UM-OSCAARS

Clinical criteria

- Relevance of topic
- Clinical setting or indications
- Quality of technique or operative flow
- Quality of comments
- Cleanliness of technical or operative field

Audiovisual criteria

- Structured presentation of the procedure
- Choice of image capture technique
- Quality of audio technique
- Quality of filming technique
- Spatial orientation

UM-OSCAARS = Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/techniques.

Table 1. Université de Montréal Objective and Structured Checklist for Assessment of Audiovisual Recordings of Surgeries/techniques (UM-OSCAARS)

Criterion	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Score awarded (between 1 and 5)
Relevance of topic	Limited audience and limited clinical impact		Limited audience but with substantial clinical impact		Large audience and substantial clinical impact	
Clinical setting or indications	No clinical setting or indications provided		Approximate clinical setting or indications		Clinical setting or indications well demonstrated	
Quality of technique or operative flow	Many imprecise and unnecessary moves		Adequate, but some imprecise and unnecessary moves		Excellent operative technique: precise and well executed	
Quality of comments	Poor content and timing of comments		Comments help viewers to understand the procedure but they are sometimes mistimed or inaccurate or not optimal for complete understanding of the procedure		Comments well-timed, accurate and enable viewers to completely understand the images depicted	
Cleanliness of technical or operative field	Focus on the procedure hard to maintain because of distracting elements (e.g.,gauzes, bleeding)		Adequate for most of the procedure		Complete absence of distracting elements Clean technical field	
Structured presentation of the procedure	Procedure not demonstrated in a structured manner		Step-by-step approach seems to have been used but not clearly highlighted		Use of a step-by-step approach that is clearly highlighted	
Choice of image capture technique	Inappropriate for the technique depicted		Appropriate but suboptimal Could have used an additional point of view or filming technique		Optimal	
Quality of audio technique	Low (e.g., unintelligible words, interference, sound volume too low or too high, artefacts, noises)		Adequate		Optimal	
Quality of filming technique	Low (e.g., blurry image, low-quality image, excessive shakes, excessive zoom in and out)		Some technical mishaps but adequate overall		Optimal No technical errors	
Spatial orientation	Recurrent loss of spatial orientation No anatomical landmarks provided		Spatial orientation adequate for much of the procedure but inconsistent Some anatomical landmarks provided		Spatial orientation consistently maintained throughout the video All relevant anatomical landmarks provided	
Total score						

3-week phases separated by 4 weeks (test–retest model). After each evaluation round, namely after the completion of all the evaluation checklists, the assessors had to choose the 3 best videos according to their personal impressions, without looking at the scores they had given to the videos. After the first evaluation round, the assessors were also asked to complete a short acceptability survey regarding the use of the UM-OSCAARS. We evaluated the acceptability of the tool with questions regarding the time required to complete the survey, the relevance of the criteria assessed, the quality of the scoring system and directions and the relevance of the tool in the context of evaluating submissions to a video contest or as a tool for peer review.

To assess interrater reliability, we compared the scores given to each video by the 4 assessors for each criterion. We calculated a correlation coefficient for each of the 10 criteria for the first phase of evaluation and then calculated an overall correlation coefficient that would represent the global interrater agreement for the 10 criteria.

Because we did not have a gold standard tool against which to compare the UM-OSCAARS, we could not thoroughly assess its external validity. We chose to compare the 4 videos that received the highest scores from the assessors with the 4 videos that assessors most frequently ranked among their top 3 videos according to their general impression.

Statistical analysis

We calculated intraclass correlation coefficients (ICCs) with 95% confidence intervals to evaluate intra- and interrater agreement. The formula used to calculate the ICC was a 1-way random effects–absolute agreement–single

Table 2. Description of the criteria	
Criterion	Description
Relevance of topic	The video should depict a technique that could be beneficial to a large audience or to a large number of patients. Alternatively, the video should depict a technique that reaches a more limited audience but that offers great clinical benefits to the few patients to which it could be applied.
Clinical setting or indications	The audience will be more receptive and captivated if a clinical setting is given. If the clinical setting does not seem relevant or practical in this video, the authors could give the main indications for this technique as an alternative.
Quality of technique or operative flow	The technique depicted should be completed in a timely manner with minimal or no distracting motions or steps. Unusual anatomy or unexpected findings should be explicitly stated.
Quality of comments	There should be minimal or no gap in time between the comments (subtitles or verbal comments) and the corresponding images. Important comments, such as comments related to anatomic landmarks or important steps of the technique, should be accompanied by text superimposed over the images (illustration or video).
Cleanliness of technical or operative field	There should be no distracting elements in the technical or surgical fields such as dirty gauzes, drapes or unnecessary instruments.
Structured presentation of the procedure	The technique depicted will be best understood if a step-by-step approach is demonstrated and clearly stated verbally, with or without the support of text superimposed on the images.
Choice of image capture technique	The image capture technique should reproduce the surgeon's or technician's point of view (e.g., endoscope for endoscopic approaches, external camcorder for open approaches). A combination of image capture techniques could be used if it would enhance the audience's general comprehension of the technique. For example, a regular camcorder could be used to show the setting of an endoscopic, microscopic or robotic approach before adopting the surgeon's point of view. Another example would be the use of an endoscopic or microscopic view to enhance the audience's understanding of technical or anatomic details of an open surgery.
Quality of audio technique	The audio technique should help rather than impede the audience's understanding of the procedure. Unintelligible words, interference, sound that is too low or too high, artefacts and noises are examples of poor audio technique.
Quality of filming technique	The filming technique should help rather than impede the audience's understanding of the procedure. Blurry images, low-quality definition, instability of the camera and excessive use of the zoom effect are examples of poor filming technique.
Spatial orientation	The authors should make every effort to keep the audience spatially oriented, keeping in mind that the audience might not be familiar with the procedure. To do this, the authors could use illustrations, add visual landmarks during the procedure and verbally point out landmarks several times during the video, for example.

Table 3. Description of the evaluators		
Assessor	Subspecialty	No. of years in practice
1	Otology and pediatrics	2
2	Head and neck surgery and microvascular reconstruction	8
3	Facial plastics	24
4	Head and neck surgery	8

rater/measurement according to the McGraw and Wong (1996) convention.[7]

To interpret the ICCs, we used the Landis and Koch interpretation[8] of the κ statistic. Values under 0 indicate poor agreement, values from 0.0 to 0.2 indicate slight agreement, values from 0.21 to 0.40 indicate fair agreement, values from 0.41 to 0.60 indicate moderate agreement, values from 0.61 to 0.80 indicate substantial agreement and values 0.81 to 1.0 indicate almost perfect to perfect agreement.

We calculated a global ICC for each criterion for the first phase of evaluation with the scores of the 4 assessors. We also calculated a global ICC encompassing the 10 criteria. In addition, we calculated the Cronbach α to evaluate the internal consistency of the items in the 2 phases of evaluation. Statistical analysis was performed with SPSS version 24.

Results

All of the assessors' evaluations were included in the analysis. The scores assigned by the assessors to the videos ranged from 11 to 50 (the maximum score was 50) (Table 5).

Intrarater reliability

Table 6 shows the global intrarater correlation of each assessor. Every ICC was greater than or equal to 0.888, indicating almost perfect agreement. The excellent test–retest correlation confirmed the intrarater reliability of the UM-OSCAARS.

Interrater reliability

The global ICC of each criterion for the first phase of evaluation varied between 0.352 (lowest value) and 0.770 (highest value). The global interrater agreement of the 10 criteria of the first phase was 0.754, indicating

Table 4. Description of the videos

Video title	Source	Duration	Subspeciality	Type of video
Retromolar flexible fibreoptic orotracheal intubation	*Head & Neck*	3 min 12 s	General OTL-HNS	Technique
Supracricoid partial laryngectomy	*Head & Neck*	8 min 49 s	Head and neck	Surgery
Hand hygiene	*New England Journal of Medicine*	14 min	General medicine, OTL-HNS	Technique
Laryngeal replacement with an artificial larynx after total laryngectomy	*Head & Neck*	3 min 21 s	Laryngology	Surgery
Transoral robot-assisted carbon dioxide laser surgery for hypopharyngeal cancer	*Head & Neck*	2 min	Head and neck	Surgery
Robotic facelift thyroidectomy	*Head & Neck*	8 min 20 s	Head and neck	Surgery
Submental flap	*Plastic and Reconstructive Surgery*	5 min 32 s	Reconstructive surgery	Surgery
Trapezius flap	YouTube	6 min 52 s	Reconstructive surgery	Surgery
Endoscopic sinus surgery	YouTube	3 min 10 s	Rhinology	Surgery
Stapedotomy	YouTube	4 min 41 s	Otology	Surgery

OTL-HNS = otolaryngology – head and neck surgery.

Table 5. Scores assigned to the 10 videos by each assessor for each phase of evaluation

	Score; assessor; phase of evaluation							
	Assessor 1		Assessor 2		Assessor 3		Assessor 4	
Video title	First phase	Second phase	First phase	Second phase	First phase	Second phase	First phase	Second phase
Retromolar flexible fibreoptic orotracheal intubation	38	43	36	38	39	39	39	43
Supracricoid partial laryngectomy	27	31	31	32	43	42	21	21
Hand hygiene	46	45	48	48	44	46	49	50
Laryngeal replacement with an artificial larynx after total laryngectomy	13	18	19	19	21	20	12	11
Transoral robot-assisted carbon dioxide laser surgery for hypopharyngeal cancer	26	36	34	32	38	37	22	22
Robotic facelift thyroidectomy	34	42	25	23	27	26	27	30
Submental flap	44	44	40	38	45	47	41	40
Trapezius flap	34	35	28	28	36	35	23	33
Endoscopic sinus surgery	33	36	27	27	44	43	30	36
Stapedotomy	33	29	44	45	41	43	39	43

Table 6. Global intrarater correlation of each assessor

Assessor	ICC (95% CI)	Interpretation of the agreement
1	0.888 (0.523–0.962)	Almost perfect
2	0.988 (0.954–0.997)	Almost perfect
3	0.986 (0.944–0.996)	Almost perfect
4	0.955 (0.829–0.989)	Almost perfect

CI = confidence interval; ICC = intraclass correlation coefficient.

substantial agreement, which confirms the good interrater reliability of the checklist (Table 7).

The results of the Cronbach α calculations are presented in Table 8. The α values were all above 0.9 except for 1 that was above 0.8, demonstrating good to excellent internal consistency between the items and the 2 phases, according to George and Mallery's rule of thumb (> 0.9 is excellent, > 0.8 is good, > 0.7 is acceptable, > 0.6 is questionable, > 0.5 is poor and < 0.5 is unacceptable).[9]

Table 7. Global intraclass correlation coefficient of each criterion in the first phase of evaluation

Criterion	ICC (95% CI)
Relevance of topic	0.612 (0.308–0.863)
Clinical setting or indications	0.770 (0.527–0.927)
Quality of technique or operative flow	0.616 (0.312–0.865)
Quality of comments	0.740 (0.480–0.915)
Cleanliness of technical or operative field	0.352 (0.049–0.722)
Structured presentation of the procedure	0.593 (0.285–0.854)
Choice of image capture technique	0.504 (0.188–0.811)
Quality of audio technique	0.610 (0.306–0.862)
Quality of filming technique	0.403 (0.092–0.754)
Spatial orientation	0.645 (0.349–0.877)
Total	0.754 (0.502–0.921)

CI = confidence interval; ICC = intraclass correlation coefficient.

Table 8. Cronbach α of the items for the 2 phases of evaluation

	Cronbach α; phase	
Assessor	First phase	Second phase
1	0.939	0.915
2	0.947	0.942
3	0.893	0.922
4	0.974	0.955
Mean for the 4 assessors	0.967	0.958

Validity

Table 9 shows the 3 top videos ranked by each of the 4 assessors on the basis of their general impression, for each of the 2 phases of evaluation. The videos most often ranked among the top 3 were videos 1, 3, 7 and 10. These 4 videos were also the ones that received the highest mean scores (Table 10). Thus, the evaluation tool correlated well with the general impression of the assessors, which could indicate good external validity.

Acceptability

The assessors all completed the acceptability survey on the use of the UM-OSCAARS less than a week after the end of the first phase of evaluation. All 4 assessors found that they were able to complete the checklist rapidly, that the directions regarding the use of the checklist were adequate and that the criteria were well defined. One assessor would not recommend the use of the checklist for the evaluation of videos for contests or journal publication, even though he found the checklist easy to use and well conceived. This assessor did not provide any information about the reason for this opinion. All of the assessors found the tool easy to use.

Table 9. Top 3 videos chosen by each assessor in the 2 phases of evaluation

Assessor	First phase	Second phase
1	Video 7	Video 7
	Video 3	Video 3
	Video 1	Video 1
2	Video 3	Video 3
	Video 10	Video 1
	Video 1	Video 10
3	Video 7	Video 7
	Video 3	Video 3
	Video 1	Video 1
4	Video 3	Video 3
	Video 7	Video 10
	Video 10	Video 7

Discussion

The Internet has become the largest, most up-to-date source for medical information, with freely accessible audiovisual material that can be used for medical education. Videos provide the opportunity for students to have more control over their learning, enabling them to engage in ubiquitous learning; in other words, videos give learners the opportunity to learn anywhere at any time. Video-based learning offers a cost-effective, location-independent method of flexible study, allowing students to learn at their own pace and view the material as often as they wish.

Studies have evaluated the quality and accuracy of open-access video content designed for health care provider use. Even though many high-quality videos are available, the quality of video clips is inconsistent and can be poor.[10,11] In a recent pilot study evaluating the impact of otology surgery videos on otolaryngology resident education, residents reported that they found that videos were highly useful and promoted self-efficacy and that they should be a high priority for a resident's surgical preparation.[3]

In light of the increased use of videos in medical and surgical teaching, more high-quality medical learning videos must be made available. However, the vast majority of videos shown to medical students or residents or as part of scientific sessions in international conferences are not peer reviewed with an objective tool. We designed the UM-OSCAARS to standardize the assessment of the quality of videos on surgical procedures or medical techniques. Our study results demonstrate that it is a reliable and acceptable tool.

Guidelines in peer-reviewed journals have been proposed to optimize video quality in the setting of medical and surgical teaching.[5,6,12] Iorio-Morin and colleagues have identified 4 workflow interventions to improve the effectiveness of video content in the context of medical education on the basis of Mayer and Moreno's cognitive theory of multimedia learning: (1) choosing appropriate content, (2) optimizing the voiceover, (3) optimizing the supporting visuals and (4) planning the shooting schedule in advance.[12,13] The authors also recommend that content creators should aim to improve their work by applying evidence-based principles.

Table 10. Comparison of the videos with the best mean scores and the number of assessors who ranked them among their top 3 choices

Video	Total mean score	No. of assessors who ranked the video among their top 3 choices*
Video 3	46.75	4
Video 7	42.50	3
Video 10	39.25	2
Video 1	38.00	3

The UM-OSCAARS could be used to facilitate video selection for events such as video sessions at conferences, for example. We aimed to create a checklist that would be easy to use and understand. If the validity of the UM-OSCAARS is confirmed in future studies, faculties, medical departments or scientific groups could use it to objectively evaluate the quality of videos submitted for a selection process and rate the videos. Having a validated objective tool could help evaluators to discriminate between videos in the case of a tie instead of engaging in a deliberation.

The UM-OSCAARS could also help video creators to improve their work. Having a checklist with objective and detailed criteria allows video creators to focus on different aspect of their videos to improve the quality of their production. This is especially relevant because most videos on surgical procedures or medical techniques are conceived by medical or surgical experts with little or no training in videography. The UM-OSCAARS could help to partially fill this gap by serving as a guide for these video creators in the making of high-quality medical and surgical technique videos.

In addition, our tool could be used for the evaluation of articles accompanied by videos submitted to peer-reviewed journals that have already adopted this format (such as the *New England Journal of Medecine* and *Head & Neck*) after more thorough assessment of if its validity. Videos submitted for online publication in scientific journals should go through the same degree of rigorous peer review and evaluation as manuscripts do. To the best of our knowledge, videos are currently being assessed for possible publication in peer-reviewed journals by reviewers who have not been provided with a tool or specific training. Our results showed that the UM-OSCAARS scores given to the study videos by the assessors were consistent with their general impression of the videos. However, the fact that the tool enables evaluators to detail and break down their assessment should allow a more thorough review and give better guidance to authors on how to improve their video material.

Knowledge of the specific evaluative criteria of UM-OSCAARS might have also helped the assessors to come to a more meaningful general impression. Our assessors had different areas of expertise in the field of otolaryngology – head and neck surgery, but the good interrater reliability in our study leads us to think that video reviewers need only to be familiar with the procedures depicted to be able to judge not only the overall quality of videos but also the quality of technique and operative flow more specifically.

We chose not to provide a formal training course for the assessors on how to use the checklist; we expected the instructions provided with the UM-OSCAARS to be self-sufficient. We wanted to assess if the descriptions of the criteria were clear enough for every assessor to understand and to be able to score the videos adequately. The global interrater agreement of the 10 criteria in the first phase of evaluation was 0.754, which confirms the good interrater reliability of the UM-OSCAARS. However, the ICC of each criterion for the first phase was very heterogeneous and might be improved with better instructions.

The UM-OSCAARS needs to be validated with a wider range of asessors from different medical specialties and with videos encompassing a broader set of medical and surgical procedures. We plan to pursue the validation of this tool by forming 2 groups of assessors and using the Delphi method: 1 group of experts would discuss the quality of each video until they reached a consensus on a quality score from 0 to 10, and the other group would use the checklist. We would compare the results of the 2 approaches to thoroughly assess the validity of our instrument. Given that in the present study 1 of the 4 assessors indicated he would not recommend the use of this checklist for peer review purposes or for use in video contests, we also plan on assessing again the acceptability of the tool to improve this parameter.

Limitations

Our study has several limitations. First, a limited number of videos were included. The total viewing time of the 10 videos was 59 minutes and 57 seconds, excluding the time required to fill in the checklist. To extrapolate and conduct an analytic study, we could have increased the number of videos, but this would also have increased the evaluation time and might have weakened the rigor shown by the evaluators in this study. Another potential limitation may have been caused by asking the assessors to choose their top 3 videos after having viewed and scored the 10 videos. The evaluators could have remembered the scores they gave to the videos, which might have tainted their overall impression of the best videos. We could not have proceeded otherwise as it was essential that the evaluators fill in the checklist immediately after they viewed each video. Another possible limitation is that we tested the UM-OSCAARS with a group of surgeons from the same specialty, which could limit the generalizability of our results. However, we chose surgeons with different areas of interest in the same specialty, who graded videos encompassing a wide variety of medical and surgical techniques, sometimes not specific to their specialty (e.g., hand hygiene, intubation).

Conclusion

To our knowledge, UM-OSCAARS is the first checklist developed to evaluate videos depicting medical and surgical techniques. Our evaluation tool has proven to be reliable and acceptable to use, but its validity needs to be more thoroughly assessed. Use of UM-OSCAARS does not require specific training other than reading the instructions provided with it. We hope this tool will lead to an improvement in the quality of technical videos used for educational purposes in medicine.

Affiliations: From the Faculty of Medecine, Université de Montréal, Montréal, Que. (Chagnon-Monarque); the Department of Otolaryngology – Head and Neck Surgery, Hôpital Maisonneuve–Rosemont, Montréal, Que. (Woods, Bissada, Ahmarani); the Department of Otolaryngology – Head and Neck Surgery, Hôpital Sainte-Justine, Montréal, Que. (Woods, Ayad); the Department of Otolaryngology – Head and Neck Surgery, Centre hospitalier de l'Université de Montréal (CHUM), Montréal, Que. (Christopoulos, Bissada, Ahmarani, Ayad); and the Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM), Montréal, Que. (Christopoulos, Ayad).

Competing interests: None declared.

Contributors: S. Chagnon-Monarque, A. Christopoulos and T. Ayad conceived the study. S. Chagnon-Monarque, A. Christopoulos, E. Bissada and T. Ayad acquired the data, which S. Chagnon-Monarque, O. Woods, A. Christopoulos and C. Ahmarani analyzed. S. Chagnon-Monarque and T. Ayad wrote the article, which O. Woods, A. Christopoulos, E. Bissada, C. Ahmarani and T. Ayad critically revised. All authors agreed to be accountable for all aspects of the work.

References

1. Dong C, Goh PS. Twelve tips for the effective use of videos in medical education. *Med Teach* 2015;37:140-5.
2. Prober CG, Khan S. Medical education reimagined: a call to action. *Acad Med* 2013;88:1407-10.
3. Poon C, Stevens SM, Golub JS, et al. Pilot study evaluating the impact of otology surgery videos on otolaryngology resident education. *Otol Neurotol* 2017;38:423-8.
4. Liu KJ, Tkachenko E, Waldman A, et al. A video-based, flipped-classroom, simulation curriculum for dermatologic surgery: a prospective, multi-institution study. *J Am Acad Dermatol* 2019;81:1271-6.
5. Gordon SL, Porto DA, Ozog DM, et al. Creating and editing video to accompany manuscripts. *Dermatol Surg* 2016;42:249-50.
6. Rehim SA, Chung KC. Educational video recording and editing for the hand surgeon. *J Hand Surg Am* 2015;40:1048-54.
7. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30-46.
8. Landis JR. Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
9. George D, Mallery P. *SPSS for Windows step by step: a simple guide and reference*. 4th ed. (11.0 update.) Boston: Allyn & Bacon; 2003.
10. Rössler B, Lahner D, Schebesta K, et al. Medical information on the Internet: quality assessment of lumbar puncture and neuroaxial block techniques on YouTube. *Clin Neurol Neurosurg* 2012;114:655-8.
11. Urch E, Taylor SA, Cody E, et al. The quality of open-access video-based orthopaedic instructional content for the shoulder physical exam is inconsistent. *HSS J* 2016;12:209-15.
12. Iorio-Morin C, Brisebois S, Becotte A, et al. Improving the pedagogical effectiveness of medical videos. *J Vis Commun Med* 2017;40:96-100.
13. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. *Educ Psychol* 2003;38:43-52.