

Is YouTube a Reliable Source for 3D Face Scanning Application in Orthodontics? A Cross- Sectional Assessment of the Quality, Reliability, and Educational Value of Videos

İdris Onur YALÇIN

idris.yalcin@usak.edu.tr

Uşak University

İlayda ATASOY DEMİRTAŞ

Uşak University

Research Article

Keywords: Three-dimensional facial scanning, Orthodontics, YouTube, Health information quality, Patient education

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10848330/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Three-dimensional facial scanning has become increasingly accessible in orthodontics through smartphone-based applications, enabling scans to be performed by patients or staff outside direct clinical supervision. As the usability of such scans depends on specific technical requirements, the quality of instructional content available online is clinically consequential. This study assessed the quality, reliability and educational value of YouTube videos on three-dimensional facial scanning in orthodontics.

Methods

Search terms were identified using Google Trends. Of 143 screened videos, 91 met the inclusion criteria. Two independent raters assessed each video using the Global Quality Scale (GQS), the Video Information and Quality Index (VIQI), the modified DISCERN instrument, and a purpose-designed ten-item Total Content Score (TCS) capturing the procedure-specific technical steps of facial scanning. Video metrics, source category and engagement parameters were recorded and analysed using non-parametric methods.

Results

Commercial companies were the most frequent source, followed by laypersons and specialists. Whereas 93.4% of videos were rated high quality on VIQI, only 8.8% reached the high category on TCS. All eight high-TCS videos were specialist-authored, and no video from any other source reached this category. Data privacy was addressed in 11.0% of videos, measurement accuracy in 20.9% and lighting conditions in 34.1%, whereas scanning purpose and device specification were addressed in all videos. None of the 28 engagement–quality correlations reached significance.

Conclusions

YouTube content on three-dimensional facial scanning is technically well produced but clinically incomplete. Specialist authorship was the only factor associated with high content quality, yet such videos were scarce and among the least viewed, and no engagement metric indicated quality.

Introduction

Aesthetic considerations have become central to contemporary orthodontic practice, shaping both patients' expectations and clinicians' treatment goals. Accordingly, a considerable proportion of the individuals presenting to the clinic frequently express aesthetic concerns¹. In addition, although the specific complaints differ across the various types of malocclusions, there are patient groups who harbor concerns about their facial appearance². Consequently, orthodontists have come to focus at the end of treatment not only on the occlusion but also on the changes that may occur in facial form. To enhance facial aesthetics, both orthodontic and combined orthodontic–surgical treatments are widely employed today. For this reason, many clinicians now incorporate soft-tissue records into their orthodontic treatment records. These records are not used solely for soft-tissue analysis³ or to demonstrate to the patient the differences between the beginning and the end of treatment⁴. They may also be used for orthodontic and orthognathic treatment planning and simulation^{4–6}, for radiation-free patient follow-up⁷, for the monitoring of asymmetry and growth⁸, and for integration with other imaging modalities.

Keeping pace with evolving technology is a concern for every clinician. As technology has been adopted into clinical practice, clinicians who wish to offer their patients more predictable treatment outcomes have turned their attention to three-dimensional imaging and scanning methods⁹. Whereas three-dimensional facial records were traditionally obtained using devices that entailed considerable cost, the rapid proliferation of high-resolution smartphone cameras and the introduction of consumer-oriented three-dimensional scanning applications have ushered in an era in which patients are now able, even at home or under minimal clinical conditions, to perform their own facial scans. These devices, which also save time, have been the subject of a great deal of research^{8,10,11}.

Every device used within this developing technology requires a degree of training. Although facial scanning devices have become readily accessible and may be used for orthodontic records, adherence to the manufacturer's or clinician's instructions remains important. Such guidance may be provided face to face or, increasingly, through instructional videos recorded by manufacturers or by clinicians themselves

and uploaded to the internet for easier access by clinicians, auxiliary personnel, and patients alike. Because users differ in their needs and only comprehensive demonstrations yield reliable results, the quality, content, and completeness of these videos are critical.

YouTube, which ranks second in global internet traffic and on which users spend an average of 18 minutes and 55 seconds per visit¹², has become an important platform to which people turn when seeking health-related information. Research has shown that approximately 80% of internet users regards online resources as their principal source of health information¹³. Within the orthodontic patient population—which largely comprises adolescents and young adults with high levels of digital literacy and heavy smartphone use—YouTube is a frequently consulted resource for procedural information, treatment guidance, and general oral health education. Although its visual format and accessibility make YouTube an attractive platform for health communication, the platform lacks any mechanism of editorial oversight or peer review. This makes it difficult for users to appraise the scientific validity and clinical accuracy of the content they encounter. Previous research conducted across numerous health disciplines, such as medicine and dentistry has consistently demonstrated considerable variability in the quality of health-related content on YouTube, showing that a substantial proportion of videos contain inaccurate information, omit critical clinical details, or present information in a misleading manner^{14–16}.

Previous studies evaluating YouTube video content in the field of orthodontics have reported differing findings. In a 2023 study by Tamošiūnaitė et al., content-quality analyses of videos relating to orthodontic pain concluded that YouTube provided inadequate information on this subject¹⁷. By contrast, in a 2022 study by Yavan and Gökçe investigating video content on adult orthodontic treatment, YouTube videos were found to constitute a reliable source of information, even though the content was generally of moderate quality¹⁸. These divergent findings indicate that the reliability of YouTube as a source of health information is highly topic-dependent and cannot be generalized across subjects, which underscores the need for content-specific evaluation. Although several studies have evaluated the quality of YouTube videos concerning orthodontic treatment procedures¹⁹, retainer use²⁰, oral hygiene care²¹, and the use of orthodontic appliances²², no study has to date systematically evaluated the quality, reliability, and educational value of YouTube videos demonstrating three-dimensional facial scanning for orthodontic purposes—whether the scan is performed with a professional scanner or with a consumer, smartphone-based application, and irrespective of whether the intended viewer is a clinician, an auxiliary staff member, or a patient. The aim of the present study was therefore to evaluate the quality, reliability, and educational value of YouTube videos on three-dimensional facial scanning in orthodontics. To this end, four complementary instruments were employed: the Global Quality Scale (GQS) to assess overall video quality, the Video Information and Quality Index (VIQI) to evaluate the audiovisual presentation, and the modified DISCERN instrument to examine the reliability of the health information. As these generic instruments do not capture the procedure-specific technical adequacy of a facial scan, a customized Total Content Score (TCS) was additionally developed to assess the completeness of the key technical steps—namely the use of the natural head position, device positioning and distance, lighting and background conditions, multi-angle image capture, and awareness of motion artefacts and scanning errors.

Methods

Ethical approval was not required for the present study, as it evaluated only publicly available data and did not involve any human participants or animals. The study was designed as a cross-sectional examination of a video-based social media source.

The Google Trends application was used to identify the most commonly used search terms relating to three-dimensional (3D) facial scanning. The terms “3D face scanner” and “face scanner” were screened using the parameters “worldwide,” “past 5 years,” and “Google Web Search.” On this basis, “face scanner app” was identified as the most relevant search term. Because individuals tend to use different queries, and in order to increase video diversity and broaden the range of retrieved results, the second most relevant term, “3D face scanner,” was also included (Figure 1). Both terms were searched on YouTube (www.youtube.com) on 1 June 2026.

To avoid the influence of previous searches, a new account was created, the computer's cookies and history were cleared, and location services were disabled to prevent geographically biased search results. “YouTube Shorts” videos were excluded using the platform's filtering feature. All retrieved videos were thoroughly screened and categorized by one independent researcher (IOY) with specialist expertise in orthodontics. For the purposes of this study, sarcastic or satirical videos were defined as those whose primary intent was comedic rather than informational. The exclusion criteria were as follows: (1) videos in a language other than English; (2) a duration of less than 1 minute or greater than 15 minutes; (3) videos without audio; (4) unacceptably poor audio or visual quality; (5) content irrelevant to the study topic; and (6) sarcastic or satirical videos. A total of 143 videos were screened. Of these, 22 were excluded for falling outside the 1–15-minute duration range, 14 for the absence of audio, 9 for poor video quality, and 7 for content unrelated to the topic. Accordingly, 91 videos met the inclusion criteria and were included in the final analysis (Figure 2). After screening, a total of 91 videos were included in the analysis: 46 retrieved by “face scanner app” and 45 by “3D face scanner”.

All included videos were assessed independently by the second researcher (IAD) after the initial assessment (made by IOY) in order to establish inter-rater reliability, and 25 randomly selected videos were re-evaluated 15 days later to assess intra-observer reliability. For every included video, the following parameters were recorded: video duration, source, number of views, number of likes, number of comments, and time since upload (in months). Because YouTube ceased to publicly display dislike counts in December 2021, a modified interaction index was calculated as $(\text{number of likes} / \text{total number of views}) \times 100$. In addition, the viewing rate was calculated as $(\text{number of views} / \text{number of days since upload}) \times 100$, with the number of days derived from the number of months since upload $\times 30.44$.

The video source was classified into the following categories: (1) specialist/dentist/orthodontist; (2) academic institution; (3) commercial company; (4) layperson; (5) patient experience; (6) hospital/university hospital; and (7) government agency.

The quality, reliability, and content of the videos were assessed using four scoring systems: the Video Information and Quality Index (VIQI), the Global Quality Scale (GQS), the modified DISCERN instrument, and a customized Total Content Score (TCS).

The overall audiovisual quality of the videos was assessed using the Video Information and Quality Index (VIQI). This index employs a 5-point Likert scale, ranging from 1 (poor quality) to 5 (high quality), to evaluate four video features: flow of information, information accuracy, audiovisual quality, and overall quality. The total VIQI score ranges from 4 to 20. Scores of 1–6 were classified as low, 7–13 as moderate, and 14–20 as high quality.

The educational quality of the videos was examined using the Global Quality Scale (GQS), a 5-point scale that assesses whether the content provides useful, complete, and reliable information for the patient. Scores of 1–2 were classified as low quality, 3 as moderate quality, and 4–5 as high quality.

The reliability of the videos was evaluated using the modified DISCERN instrument, a standardized questionnaire originally developed by Charnock et al.²³ and Singh et al.²⁴ (https://www.discern.org.uk/discern_instrument.php). This five-item tool assesses the quality and reliability of health information. The criteria were as follows: clarity of purpose; reliability of the information sources; balance and objectivity of the content; recommendation of professional consultation or additional sources; and acknowledgement of the uncertainties and limitations of home-based scanning. These five dimensions were selected because they represent the core reliability and quality indicators applicable to short-form video health content, in which the detailed referencing and risk discussion central to the original instrument are largely absent. Each item was scored as 1 (positive answer) or 0 (negative answer), yielding a total score ranging from 0 to 5. Scores of 1–2 were classified as low reliability, 3 as moderate, and 4–5 as high reliability.

The completeness of the technical content was assessed using a customized Total Content Score (TCS). Because the generic instruments described above do not capture the procedure-specific technical adequacy of a facial scan, the ten TCS items were developed by the authors to reflect the essential steps required for a reliable facial scan and were finalized by consensus among the researchers (IOY and IAD) prior to video evaluation. The items assessed whether the video: (a) explained the purpose of the 3D scan (e.g., orthodontic assessment, monitoring, treatment follow-up); (b) specified the device or application used; (c) described camera/phone positioning (distance, angle, and device position); (d) addressed lighting and background conditions; (e) explained head-position standardization (natural head position or a fixed position); (f) recommended image capture from multiple angles (frontal, profile, or other views); (g) mentioned the risk of motion artefact or scanning error; (h) stated the measurement accuracy or limitations; (i) addressed data privacy or security; and (j) described a clinical use scenario (e.g., follow-up, progress assessment, or tele-orthodontics). Each item was scored as met (1) or unmet (0), giving a total possible score of 10. Scores of 1–4 were classified as low content, 5–7 as moderate content, and 8–10 as high content.

Statistical Analysis

Statistical analyses were performed using SPSS (version 29.0, SPSS, Chicago, Ill). The Shapiro–Wilk test indicated that all continuous variables departed significantly from normality (all $p < 0.001$), and non-parametric methods were therefore used throughout. Descriptive statistics are reported as minimum, maximum, mean, standard deviation, median and interquartile range (Q1–Q3). Because videos were retrieved through two search queries, the comparability of the two subsamples was assessed first, using the Mann–Whitney U test for continuous and ordinal variables and the chi-square test for categorical variables; the subsamples were pooled only after comparability had been established.

Differences across TCS and GQS categories were examined with the Kruskal–Wallis H test, followed by Bonferroni-corrected post-hoc Mann–Whitney U tests where the omnibus test was significant. Associations between categorical variables were assessed with the chi-square test; because expected cell frequencies below five were present in every cross-tabulation, each test was accompanied by a Monte Carlo permutation test (100,000 replicates, row and column totals fixed) and the principal associations were additionally verified with

Fisher's exact test on collapsed 2×2 tables. Correlations were quantified with Kendall's tau-b, chosen because all four quality instruments are ordinal with a very high density of tied ranks (25.6–35.3% of all possible pairs), for which tau-b applies a direct correction whereas Spearman's rho provides only an approximation. Finally, a sensitivity analysis repeated every primary analysis separately within each search-query subsample, supplemented by partial Kendall's tau coefficients controlling for query group, together with stratum-weighted (common) tau coefficients tested by within-stratum permutation, and by a Cochran–Mantel–Haenszel test stratified by query group. Statistical significance was set at $p < 0.05$; no global correction for multiple comparisons was applied, and findings with p-values close to the threshold should therefore be interpreted with caution.

Results

IAD and IOY are both specialists in orthodontics with more than 5 year of clinical experience. The first observer (IOY) conducted the video search and evaluation on June 1, 2026, and recorded the URLs of the eligible videos. Ten days later, the second observer (IAD) independently reviewed the same videos and evaluated them using the same methodology. The intraclass correlation coefficient (ICC) values ranged from 0.891 to 0.938 for intra-rater reliability and from 0.872 to 0.926 for inter-rater reliability.

Comparability of the two search-query subsamples

The two subsamples did not differ significantly in any of the four quality instruments, in video duration or time since upload, or in any engagement metric, nor in the categorical distribution of the four quality classifications (all $p > 0.05$). The only significant difference concerned source composition ($\chi^2 = 9.644$, $p = 0.022$): commercial content predominated in the “3D face scanner” subsample, whereas layperson and academic content was concentrated in the “face scanner app” subsample, a pattern consistent with the complementary device- versus application-oriented focus of the two queries. The subsamples were therefore analyzed as a single pooled sample (Table 1).

Table 1
Comparability of the Two Search-Query Subsamples

	"face scanner app" n = 46	"3D face scanner" n = 45	U / χ^2	p	r
	mean $\pm$ SD	mean $\pm$ SD			
Quality instruments (Mann–Whitney U)					
TCS	6.13 $\pm$ 1.65	6.60 $\pm$ 1.39	886.0	0.215	—
GQS	3.41 $\pm$ 0.86	3.49 $\pm$ 0.69	978.5	0.630	—
VIQI	15.33 $\pm$ 2.34	15.96 $\pm$ 1.57	888.5	0.227	—
DISCERN	3.11 $\pm$ 0.99	2.96 $\pm$ 0.74	1126.0	0.445	—
Video characteristics and engagement (Mann–Whitney U)					
Duration (min)	5.65 $\pm$ 3.96	5.85 $\pm$ 3.66	970.5	0.611	—
Number of views	49,699 $\pm$ 127,148	88,425 $\pm$ 218,631	1051.5	0.899	—
Number of likes	698 $\pm$ 1,738	2,291 $\pm$ 7,799	1054.5	0.880	—
Number of comments	38.2 $\pm$ 70.2	395.6 $\pm$ 1,309.2	1043.0	0.951	—
Months since upload	59.70 $\pm$ 50.84	71.31 $\pm$ 53.48	879.0	0.215	—
Interaction Index (%)	1.25 $\pm$ 1.40	2.32 $\pm$ 3.14	964.0	0.576	—
Viewing Rate	4,375 $\pm$ 10,000	4,278 $\pm$ 8,082	1100.0	0.609	—
View Index	1.44 $\pm$ 1.37	1.90 $\pm$ 1.23	789.5	0.052	—
Categorical distributions (chi-square) — n					
Source type ☒	9 / 15 / 19 / 3	5 / 28 / 12 / 0	9.644	0.022*	—
TCS category ☒	10 / 32 / 4	3 / 38 / 4	4.273	0.118	—
GQS category ☒	6 / 20 / 20	2 / 22 / 21	2.109	0.348	—
DISCERN category ☒	12 / 19 / 15	13 / 21 / 11	0.744	0.689	—
VIQI category ☒	0 / 5 / 41	0 / 1 / 44	1.536	0.215	—
Mann–Whitney U test for continuous and ordinal variables; chi-square test for categorical variables; *p < 0.05, **p < 0.01, ***p < 0.001. r: rank-biserial correlation, reported as a measure of effect size for significant comparisons. ☒ Source-type counts are given as specialist / commercial / layperson / academic. ☒ Categorical distributions are given as low / moderate / high counts. ☒ Fisher's exact test for the 2 $\times$ 2 VIQI comparison yielded p = 0.203. The only significant difference between the subsamples concerned source composition; the two groups were equivalent with respect to all four quality instruments, all engagement metrics and all four categorical quality classifications.					

Descriptive characteristics

Of the 91 videos analyzed, 46 (50.5%) were retrieved by "face scanner app" and 45 (49.5%) by "3D face scanner". The most frequent source was commercial companies (n = 43, 47.3%), followed by laypersons (n = 31, 34.1%), specialists (n = 14, 15.4%) and academic institutions (n = 3, 3.3%); no video originated from the remaining source categories. The median number of views was 6,400 (Q1–Q3: 553.50–38,000), mean duration was 5.75 $\pm$ 3.79 minutes and median time since upload was 48 months (Q1–Q3: 24–96).

Mean scores were 3.45 for GQS, 15.64 for VIQI, 3.03 for modified DISCERN and 6.36 for TCS. Applying the pre-defined categories, 13 videos (14.3%) showed low, 70 (76.9%) moderate and 8 (8.8%) high TCS content quality. For GQS, 8 (8.8%) were low, 42 (46.2%) moderate and 41 (45.1%) high quality; for modified DISCERN, 25 (27.5%) low, 40 (44.0%) moderate and 26 (28.6%) high reliability; and for VIQI, no video fell into the low category, 6 (6.6%) were moderate and 85 (93.4%) high quality (Table 2).

Table 2

Distribution of the YouTube™ Videos on 3D Facial Scanning According to Their Characteristics and Quality-Instrument Scores (n = 91)

Number of views	Min	Max	Mean	SD	Median	Q1	Q3
	3	1,100,000	68,849.23	178,412.20	6,400	553.50	38,000
Duration in minutes	1.12	14.42	5.75	3.79	4.23	2.61	8.27
Months since upload	1	216	65.44	52.20	48	24	96
Number of comments	0	6,533	214.96	934.20	3	0	44.50
Number of likes	0	51,000	1,486.02	5,646.62	23	3.50	652.50
Viewing Rate	0.68	50,235.44	4,327.04	9,051.87	328.52	44.82	3,513.29
Interaction Index (%)	0.00	13.42	1.78	2.47	1.00	0.39	1.89
View Index	0.02	7.00	1.67	1.31	1.22	0.61	2.48
TCS	3	10	6.36	1.54	7	6	7
VIQI	9	20	15.64	2.01	16	14	17
DISCERN	1	5	3.03	0.87	3	2	4
GQS	2	5	3.45	0.78	3	3	4
					n	%	
Search query							
“face scanner app”					46	50.5%	
“3D face scanner”					45	49.5%	
Source type							
Specialist / dentist / orthodontist					14	15.4%	
Commercial company					43	47.3%	
Layperson					31	34.1%	
Academic institution					3	3.3%	
TCS category (Low 1–4 / Moderate 5–7 / High 8–10)							
Low					13	14.3%	
Moderate					70	76.9%	
High					8	8.8%	
GQS category (Low 1–2 / Moderate 3 / High 4–5)							
Low					8	8.8%	
Moderate					42	46.2%	
High					41	45.1%	
DISCERN category (Low 1–2 / Moderate 3 / High 4–5)							
Low reliability					25	27.5%	
Moderate reliability					40	44.0%	
High reliability					26	28.6%	
Q1: first quartile (25th percentile); Q3: third quartile (75th percentile). Together these define the interquartile range (IQR = Q3 – Q1), which contains the middle 50% of observations and is reported alongside the median as the appropriate measure of dispersion for non-normally distributed data. Viewing Rate = (views / days since upload) × 100, where days = months × 30.44. Interaction Index = (likes / views) × 100; likes only, as YouTube™ removed public dislike counts in December 2021. A GQS score of 1 and the low VIQI category (1–6) were not observed.							

Number of views	Min	Max	Mean	SD	Median	Q1	Q3
	3	1,100,000	68,849.23	178,412.20	6,400	553.50	38,000
VIQI category (Low 1–6 / Moderate 7–13 / High 14–20)							
Low					0	0.0%	
Moderate					6	6.6%	
High					85	93.4%	
Q1: first quartile (25th percentile); Q3: third quartile (75th percentile). Together these define the interquartile range (IQR = Q3 – Q1), which contains the middle 50% of observations and is reported alongside the median as the appropriate measure of dispersion for non-normally distributed data. Viewing Rate = (views / days since upload) × 100, where days = months × 30.44. Interaction Index = (likes / views) × 100; likes only, as YouTube™ removed public dislike counts in December 2021. A GQS score of 1 and the low VIQI category (1–6) were not observed.							

Sub-item analysis revealed a consistent gradient across the three multi-item instruments. Within the TCS, explanation of the purpose of scanning and specification of the device or application were addressed in every video (both 100.0%), followed by camera positioning (87.9%) and description of a clinical use scenario (83.5%); by contrast, lighting and background conditions were addressed in only 34.1% of videos, measurement accuracy or limitations in 20.9% and data privacy or security in 11.0%. A parallel pattern was observed among the modified DISCERN items, where clarity of aims was most frequently satisfied (82.4%) and acknowledgement of areas of uncertainty least frequently (29.7%). Among the VIQI sub-items, audiovisual quality achieved the highest mean score (4.66 ± 0.70) and precision the lowest (3.09 ± 0.49) (Table 3).

Table 3
Distribution of VIQI Sub-Item Scores and Proportion of Videos Meeting Each Modified DISCERN and TCS Criterion (n = 91)

VIQI sub-item	1		2		3		4		5	
	n	%	n	%	n	%	n	%	n	%
1. Flow of information	2	2.2	2	2.2	58	63.7	3	3.3	26	28.6
2. Information accuracy	0	0.0	0	0.0	27	29.7	5	5.5	59	64.8
3. Audiovisual quality	0	0.0	0	0.0	12	13.2	7	7.7	72	79.1
4. Precision	1	1.1	1	1.1	81	89.0	5	5.5	3	3.3

DISCERN sub-item (criterion met)	n	%
a. Are the aims clear and achieved?	75	82.4%
b. Are reliable sources of information used?	69	75.8%
c. Is the information balanced and unbiased?	62	68.1%
d. Are additional sources of support provided?	43	47.3%
e. Are areas of uncertainty mentioned?	27	29.7%
TCS sub-item (criterion met)	n	%
a. Purpose of 3D facial scanning explained	91	100.0%
b. Device or application specified	91	100.0%
c. Camera / phone positioning described	80	87.9%
d. Lighting and background conditions addressed	31	34.1%
e. Head position standardisation explained	64	70.3%
f. Multiple angles recommended	67	73.6%
g. Motion artefact or error risk mentioned	50	54.9%
h. Measurement accuracy or limitations stated	19	20.9%
i. Data privacy / security addressed	10	11.0%
j. Clinical use scenario described	76	83.5%
Modified DISCERN and TCS sub-items are scored 0 (criterion not met) or 1 (criterion met); n and % denote videos meeting the criterion.		

Video characteristics according to TCS content-quality groups

All three remaining quality instruments differed significantly across TCS groups: modified DISCERN ($H = 26.233$, $p < 0.001$), GQS ($H = 20.235$, $p < 0.001$) and VIQI ($H = 13.649$, $p = 0.001$), with mean scores rising progressively from the low to the high group. Modified DISCERN was the only instrument for which all three pairwise contrasts were significant (all $p \leq 0.005$), increasing from 2.23 ± 0.73 to 4.38 ± 0.52 . For GQS and VIQI, the high TCS group differed from both the low and the moderate group (all $p \leq 0.004$), whereas the low-versus-moderate contrast was non-significant for both.

Among the engagement metrics, only duration ($H = 9.033$, $p = 0.011$) and number of likes ($H = 8.177$, $p = 0.017$) reached significance in the omnibus test. For duration, only the low-versus-moderate contrast survived correction ($p = 0.036$), moderate-TCS videos being longer than low-TCS videos (6.36 ± 3.90 versus 3.63 ± 2.24 minutes). For likes, no pairwise contrast survived correction and the distribution was non-monotonic, moderate-TCS videos attracting the highest engagement and high-TCS videos the lowest. The remaining engagement metrics did not differ across TCS groups (all $p > 0.05$) (Table 4).

Table 4
Comparison of Video Characteristics and Quality-Instrument Scores According to TCS Content-Quality Groups

Number of views	TCS group	Min	Max	Mean	SD	Median	H	p	Pairwise	p (adj.)
	Low (1–4)	71	71,000	13,318.00	23,296.05	1,500	5.101	0.078	L – M	0.352
	Moderate (5–7)	3	1,100,000	86,313.57	200,166.44	8,650			L – H	1.000
	High (8–10)	24	33,000	6,274.50	11,407.92	973.50			M – H	0.222
Duration (min)	Low (1–4)	1.40	8.17	3.63	2.24	2.68	9.033	0.011*	L – M	0.036*
	Moderate (5–7)	1.40	14.42	6.36	3.90	5.63			L – H	1.000
	High (8–10)	1.12	10.67	3.82	3.20	2.72			M – H	0.159
Months since upload	Low (1–4)	8	204	57.85	54.81	36	1.178	0.555	L – M	1.000
	Moderate (5–7)	1	216	68.84	53.45	60			L – H	1.000
	High (8–10)	12	120	48.00	33.94	36			M – H	1.000
Number of comments	Low (1–4)	0	39	6.77	11.41	1	4.980	0.083	L – M	0.627
	Moderate (5–7)	0	6,533	277.76	1,058.76	5.5			L – H	0.866
	High (8–10)	0	23	3.75	8.05	0			M – H	0.157
Number of likes	Low (1–4)	0	855	102.69	237.30	4	8.177	0.017*	L – M	0.173
	Moderate (5–7)	0	51,000	1,910.43	6,386.38	38			L – H	1.000
	High (8–10)	0	129	20.38	44.13	5.5			M – H	0.055
Interaction Index (%)	Low (1–4)	0.00	2.63	0.85	0.77	0.57	2.262	0.323	L – M	0.584
	Moderate (5–7)	0.00	13.42	1.97	2.63	1.06			L – H	1.000
	High (8–10)	0.00	8.00	1.59	2.67	0.75			M – H	1.000
Viewing Rate	Low (1–4)	2.78	5,201.49	1,219.24	1,885.64	86.24	4.001	0.135	L – M	0.449
	Moderate (5–7)	0.78	50,235.44	5,313.46	10,090.43	399.24			L – H	1.000
	High (8–10)	0.68	3,011.39	746.06	1,194.53	77.57			M – H	0.371
GQS	Low (1–4)	2	4	3.08	0.95	3	20.235	< 0.001***	L – M	1.000

Kruskal–Wallis H test with Bonferroni-corrected post-hoc Mann–Whitney U tests, shown in the two right-hand columns; *p < 0.05, **p < 0.01, ***p < 0.001. Group sizes: low n = 13, moderate n = 70, high n = 8. L: low; M: moderate; H: high. p (adj.): p-value multiplied by three to correct for the three pairwise comparisons per variable, capped at 1.000. Modified DISCERN was the only variable for which all three pairwise contrasts reached significance. For number of likes the omnibus test was significant but no pairwise contrast survived correction.

Number of views	TCS group	Min	Max	Mean	SD	Median	H	p	Pairwise	p (adj.)
	Low (1–4)	71	71,000	13,318.00	23,296.05	1,500	5.101	0.078	L – M	0.352
	Moderate (5–7)	2	5	3.37	0.62	3			L – H	0.002**
	High (8–10)	4	5	4.75	0.46	5			M – H	< 0.001***
VIQI	Low (1–4)	9	16	14.15	2.67	16	13.649	0.001**	L – M	0.436
	Moderate (5–7)	12	18	15.64	1.65	16			L – H	0.002**
	High (8–10)	16	20	18.00	1.51	18			M – H	0.004**
DISCERN	Low (1–4)	1	3	2.23	0.73	2	26.233	< 0.001***	L – M	0.005**
	Moderate (5–7)	2	5	3.03	0.74	3			L – H	< 0.001***
	High (8–10)	4	5	4.38	0.52	4			M – H	< 0.001***

Kruskal–Wallis H test with Bonferroni-corrected post-hoc Mann–Whitney U tests, shown in the two right-hand columns; *p < 0.05, **p < 0.01, ***p < 0.001. Group sizes: low n = 13, moderate n = 70, high n = 8. L: low; M: moderate; H: high. p (adj.): p-value multiplied by three to correct for the three pairwise comparisons per variable, capped at 1.000. Modified DISCERN was the only variable for which all three pairwise contrasts reached significance. For number of likes the omnibus test was significant but no pairwise contrast survived correction.

Association between TCS groups, source type and quality categories

Source type was strongly associated with TCS content quality ($\chi^2 = 49.309$, $p < 0.001$; Monte Carlo $p < 0.001$). All eight high-TCS videos were specialist-authored, corresponding to 57.1% of all specialist videos, and no video uploaded by a commercial company, layperson or academic institution reached the high-TCS category; conversely, commercial and layperson upload together accounted for 92.4% of the low-TCS group. Collapsing the table to specialist versus non-specialist by high versus non-high TCS and applying Fisher's exact test confirmed the association ($p < 0.001$).

Modified DISCERN category was equally strongly associated with TCS group ($\chi^2 = 30.336$, $p < 0.001$; Monte Carlo $p < 0.001$): no high-reliability video fell into the low-TCS group and all eight high-TCS videos showed high reliability. GQS category showed a comparable association ($\chi^2 = 29.545$, $p < 0.001$; Monte Carlo $p < 0.001$), with all high-TCS videos classified as high GQS quality. The association with VIQI category reached significance in the asymptotic test ($\chi^2 = 6.905$, $p = 0.032$) but not in the Monte Carlo permutation test ($p = 0.053$) and should therefore be regarded as unstable (Table 5).

Table 5
Cross-Tabulation of TCS Content-Quality Groups With Source Type and With Modified DISCERN, GQS and VIQI Categories

Cross-tabulation of TCS Content-Quality Groups with Source Type and With Modified DISCERN, GQS and VIQI Categories													
		Low TCS n = 13			Moderate TCS n = 70			High TCS n = 8			χ²	p	MC p
		n	row%	col%	n	row%	col%	n	row%	col%			
Source ☒	Specialist / dentist	1	7.1%	7.7%	5	35.7%	7.1%	8	57.1%	100.0%	49.309	< 0.001***	< 0.001***
	Commercial company	6	14.0%	46.2%	37	86.0%	52.9%	0	0.0%	0.0%			
	Layperson	6	19.4%	46.2%	25	80.6%	35.7%	0	0.0%	0.0%			
	Academic institution	0	0.0%	0.0%	3	100.0%	4.3%	0	0.0%	0.0%			
DISCERN ☒	Low (1–2)	8	32.0%	61.5%	17	68.0%	24.3%	0	0.0%	0.0%	30.336	< 0.001***	< 0.001***
	Moderate (3)	5	12.5%	38.5%	35	87.5%	50.0%	0	0.0%	0.0%			
	High (4–5)	0	0.0%	0.0%	18	69.2%	25.7%	8	30.8%	100.0%			
GQS ☒	Low (1–2)	5	62.5%	38.5%	3	37.5%	4.3%	0	0.0%	0.0%	29.545	< 0.001***	< 0.001***
	Moderate (3)	2	4.8%	15.4%	40	95.2%	57.1%	0	0.0%	0.0%			
	High (4–5)	6	14.6%	46.2%	27	65.9%	38.6%	8	19.5%	100.0%			
VIQI ☒	Moderate (7–13)	3	50.0%	23.1%	3	50.0%	4.3%	0	0.0%	0.0%	6.905	0.032*	0.053
	High (14–20)	10	11.8%	76.9%	67	78.8%	95.7%	8	9.4%	100.0%			
*p < 0.05, **p < 0.01, ***p < 0.001. row%: percentage within the source or quality category (row total). col%: percentage within the TCS group (column total). MC p: Monte Carlo permutation p-value based on 100,000 replicates with row and column totals held fixed, reported because expected cell frequencies below five were present in every table; the asymptotic chi-square p-value is therefore an approximation. ☒ Source × TCS: 8/12 cells with expected frequency < 5, minimum expected 0.264; collapsing to specialist versus non-specialist by high versus non-high TCS and applying Fisher's exact test yielded p < 0.001. ☒ DISCERN × TCS: 5/9 cells, minimum expected 2.198 (collapsed 2 × 2: p = 0.006); GQS × TCS: 4/9 cells, minimum expected 0.703 (collapsed 2 × 2: p = 0.001). ☒ VIQI × TCS: 3/6 cells, minimum expected 0.527; the low VIQI category (1–6) was not observed and is therefore omitted. This association reached significance in the asymptotic test but not in the Monte Carlo test (p = 0.053) and should be regarded as unstable.													

Video characteristics according to GQS quality groups

Significant differences across GQS groups were observed for VIQI (H = 29.028, p < 0.001) and TCS (H = 11.600, p = 0.003). High-GQS videos differed from both low- and moderate-GQS videos in VIQI (both p < 0.001), whereas the low-versus-moderate contrast was non-significant. For TCS, low-GQS videos differed from both moderate- and high-GQS videos (p = 0.003 and p = 0.008), whereas moderate versus high was non-significant. Modified DISCERN did not reach significance (H = 5.973, p = 0.051), nor did duration or any engagement metric (all p > 0.05) (Table 6).

Table 6
Comparison of Video Characteristics and Quality-Instrument Scores According to GQS Quality Groups

Number of views	GQS group	Min	Max	Mean	SD	Median	H	p	Pairwise	p (adj.)
	Low (n = 8)	21	71,000	16,823.62	25,826.40	4,205	1.547	0.462	L – M	0.701
	Moderate (n = 42)	3	1,100,000	107,793.26	243,475.12	8,150			L – H	1.000
	High (n = 41)	15	375,000	39,106.68	87,750.54	6,200			M – H	1.000
Duration (min)	Low	1.40	8.17	3.28	2.14	2.96	3.997	0.136	L – M	0.186
	Moderate	1.40	14.42	6.09	3.99	4.46			L – H	0.164
	High	1.12	14.42	5.88	3.73	5.40			M – H	1.000
Months since upload	Low	8	180	56.50	55.94	42	1.111	0.574	L – M	1.000
	Moderate	6	204	61.74	48.00	42			L – H	1.000
	High	1	216	70.98	56.22	48			M – H	1.000
Number of comments	Low	0	39	8.12	14.10	0.5	2.656	0.265	L – M	0.513
	Moderate	0	6,533	426.43	1,350.43	5.5			L – H	1.000
	High	0	392	38.68	83.46	3			M – H	0.742
Number of likes	Low	0	3,500	558.38	1,224.32	24.5	1.618	0.445	L – M	0.942
	Moderate	0	51,000	2,473.10	8,098.64	31			L – H	1.000
	High	0	7,100	655.88	1,532.37	18			M – H	0.995
Interaction Index (%)	Low	0.00	9.52	2.76	3.96	0.86	0.121	0.941	L – M	1.000
	Moderate	0.00	13.42	1.93	2.77	0.93			L – H	1.000
	High	0.00	8.00	1.44	1.66	1.00			M – H	1.000
Viewing Rate	Low	2.78	4,859.29	1,068.53	1,716.20	180.00	1.613	0.446	L – M	0.749
	Moderate	0.78	50,235.44	6,311.44	11,631.76	629.65			L – H	1.000
	High	0.68	25,665.24	2,930.05	6,074.73	325.09			M – H	1.000
VIQI	Low	9	16	13.25	2.43	14	29.028	< 0.001***	L – M	0.237
	Moderate	10	18	14.98	1.70	14			L – H	< 0.001***
	High	14	20	16.78	1.46	16			M – H	< 0.001***
DISCERN	Low	2	4	2.50	0.76	2	5.973	0.051	L – M	0.378
	Moderate	1	4	2.93	0.78	3			L – H	0.095
	High	1	5	3.24	0.94	3			M – H	0.392
Kruskal–Wallis H test with Bonferroni-corrected post-hoc Mann–Whitney U tests, shown in the two right-hand columns; *p < 0.05, **p < 0.01, ***p < 0.001. Group sizes: low n = 8, moderate n = 42, high n = 41. L: low; M: moderate; H: high. p (adj.): p-value multiplied by three to correct for the three pairwise comparisons per variable, capped at 1.000. No engagement metric differed significantly across GQS groups in either the omnibus or the pairwise tests.										

Number of views	GQS group	Min	Max	Mean	SD	Median	H	p	Pairwise	p (adj.)
	Low (n = 8)	21	71,000	16,823.62	25,826.40	4,205	1.547	0.462	L – M	0.701
TCS	Low	3	7	4.50	1.41	4	11.600	0.003**	L – M	0.003**
	Moderate	4	7	6.36	0.82	7			L – H	0.008**
	High	4	10	6.73	1.86	7			M – H	1.000
Kruskal–Wallis H test with Bonferroni-corrected post-hoc Mann–Whitney U tests, shown in the two right-hand columns; *p < 0.05, **p < 0.01, ***p < 0.001. Group sizes: low n = 8, moderate n = 42, high n = 41. L: low; M: moderate; H: high. p (adj.): p-value multiplied by three to correct for the three pairwise comparisons per variable, capped at 1.000. No engagement metric differed significantly across GQS groups in either the omnibus or the pairwise tests.										

Correlations between video characteristics and quality instruments

None of the 28 engagement–quality correlations reached significance (all $p > 0.05$); the association closest to significance was the negative correlation between the Interaction Index and modified DISCERN ($\tau\text{-}b = -0.156$, $p = 0.053$).

Among the quality instruments, five of the six intercorrelations were significant. The strongest was GQS–VIQI ($\tau\text{-}b = 0.528$, $p < 0.001$), followed by TCS–modified DISCERN ($\tau\text{-}b = 0.413$, $p < 0.001$), GQS–modified DISCERN ($\tau\text{-}b = 0.296$, $p = 0.001$), GQS–TCS ($\tau\text{-}b = 0.280$, $p = 0.002$) and TCS–VIQI ($\tau\text{-}b = 0.177$, $p = 0.044$). Only VIQI–modified DISCERN did not reach significance ($\tau\text{-}b = 0.158$, $p = 0.076$). These findings indicate two partially distinct dimensions of video quality: a presentation dimension, in which overall quality is closely aligned with audiovisual and informational presentation, and a content dimension, in which content completeness is most strongly aligned with information reliability (Table 7).

Table 7
Correlations Between Video Characteristics and Quality-Instrument Scores (Kendall's tau-b)

	GQS		TCS		VIQI		DISCERN	
	$\tau\text{-}b$	p	$\tau\text{-}b$	p	$\tau\text{-}b$	p	$\tau\text{-}b$	p
Number of views	-0.046	0.577	0.011	0.894	-0.003	0.966	-0.053	0.511
Duration in minutes	0.062	0.446	0.056	0.481	0.117	0.140	0.006	0.936
Months since upload	0.046	0.588	-0.046	0.578	0.081	0.320	0.144	0.085
Number of comments	-0.071	0.411	0.028	0.739	-0.059	0.477	-0.079	0.350
Number of likes	-0.059	0.475	0.015	0.848	-0.011	0.895	-0.123	0.129
Interaction Index	-0.031	0.707	0.069	0.388	0.041	0.606	-0.156	0.053
Viewing Rate	-0.037	0.653	0.042	0.593	-0.025	0.750	-0.099	0.217
GQS	—	—	0.280**	0.002	0.528***	< 0.001	0.296**	0.001
TCS	0.280**	0.002	—	—	0.177*	0.044	0.413***	< 0.001
VIQI	0.528***	< 0.001	0.177*	0.044	—	—	0.158	0.076
DISCERN	0.296**	0.001	0.413***	< 0.001	0.158	0.076	—	—

*p < 0.05, **p < 0.01, ***p < 0.001. $\tau\text{-}b$: Kendall's tau-b rank correlation coefficient, selected because all quality instruments are ordinal with a very high density of tied ranks (25.6–35.3% of all possible pairs), for which tau-b provides a direct correction whereas Spearman's rho applies only an approximation. — denotes self-correlation. No engagement metric correlated significantly with any quality instrument; the association closest to significance was that between the Interaction Index and modified DISCERN ($\tau\text{-}b = -0.156$, $p = 0.053$). Five of the six quality-instrument intercorrelations were significant; only VIQI–DISCERN ($p = 0.076$) was not.

Sensitivity analysis

All primary analyses were repeated within each search-query subsample (Tables 8 and 9). No quality-instrument intercorrelation changed direction in either stratum, and partial correlations controlling for query group differed from the zero-order coefficients by no more than 6.7%, well below the conventional 10% threshold for confounding. The principal finding was reproduced identically in both strata, with all high-TCS videos within each subsample being specialist-authored (Fisher's exact $p < 0.001$ in both; Cochran–Mantel–Haenszel statistic = 48.002, $p < 0.001$), as were the differences in modified DISCERN and GQS across TCS groups, the difference in VIQI across GQS groups, and the GQS–VIQI, TCS–modified DISCERN and GQS–modified DISCERN correlations.

Table 8
Sensitivity Analysis: Principal Findings Reproduced Separately Within Each Search-Query Subsample

	Test	Pooled (n = 91)		“face scanner app” (n = 46)		“3D face scanner” (n = 45)		Consistent
		stat	p	stat	p	stat	p	
Panel A – Across TCS groups								
DISCERN	KW H	26.233	< 0.001***	19.612	< 0.001***	8.749	0.013*	Yes
GQS	KW H	20.235	< 0.001***	10.848	0.004**	11.215	0.004**	Yes
VIQI	KW H	13.649	0.001**	8.320	0.016*	5.878	0.053	Partial
Duration	KW H	9.033	0.011*	6.715	0.035*	1.880	0.391	Partial
Number of likes	KW H	8.177	0.017*	4.453	0.108	4.430	0.109	No
Panel B – Across GQS groups								
VIQI	KW H	29.028	< 0.001***	17.800	< 0.001***	10.035	0.007**	Yes
TCS	KW H	11.600	0.003**	11.288	0.004**	1.131	0.568	Partial
DISCERN	KW H	5.973	0.051	2.503	0.286	5.141	0.077	Yes (null)
Duration	KW H	3.997	0.136	3.399	0.183	0.948	0.623	Yes (null)
Panel C – Cross-tabulations against TCS groups								
Source type	χ²	49.309	< 0.001***	19.970	0.003**	37.473	< 0.001***	Yes
DISCERN category	χ²	30.336	< 0.001***	21.188	< 0.001***	14.124	0.007**	Yes
GQS category	χ²	29.545	< 0.001***	22.813	< 0.001***	5.905	0.206	Partial
VIQI category	χ²	6.905	0.032*	4.970	0.083	0.188	0.910	No
Specialist × High TCS	Fisher	100%	< 0.001***	100%	< 0.001***	%100	< 0.001***	Yes
Panel D – Instrument intercorrelations (Kendall's τ)								
GQS – VIQI	τ	0.528	< 0.001***	0.577	< 0.001***	0.446	0.001**	Yes
TCS – DISCERN	τ	0.413	< 0.001***	0.518	< 0.001***	0.294	0.026*	Yes
GQS – DISCERN	τ	0.296	0.001**	0.291	0.023*	0.324	0.018*	Yes
GQS – TCS	τ	0.280	0.002**	0.332	0.008**	0.219	0.101	Partial
TCS – VIQI	τ	0.177	0.044*	0.234	0.055	0.085	0.511	No
VIQI – DISCERN	τ	0.158	0.076	0.204	0.102	0.135	0.307	Yes (null)
*p < 0.05, **p < 0.01, ***p < 0.001. KW: Kruskal–Wallis H test; χ²: chi-square test. “Consistent” denotes agreement between the pooled result and both stratum-specific results: Yes = significant in the pooled analysis and in both strata; Yes (null) = non-significant in the pooled analysis and in both strata; Partial = same direction in both strata but significance reached in one stratum only; No = significant in the pooled analysis but not reproduced in either stratum. For the specialist × high-TCS association the figure shown is the percentage of high-TCS videos authored by specialists (pooled 8/8; “face scanner app” 4/4; “3D face scanner” 4/4), tested with Fisher's exact test. Findings reproduced in only one stratum were in every case significant in the “face scanner app” subsample and non-significant, though directionally consistent, in the “3D face scanner” subsample.								

Table 9
Partial and Stratum-Weighted Kendall's tau Coefficients Between Quality Instruments, Controlling for Search-Query Group

Instrument pair	Zero-order τ -b	Partial τ	Change	Stratified τ	p
GQS – VIQI	0.528	0.526	0.3%	0.513	true
GQS – TCS	0.280	0.277	1.2%	0.277	undefined
GQS – DISCERN	0.296	0.301	1.6%	0.307	undefined
TCS – DISCERN	0.413	0.426	3.2%	0.409	undefined
TCS – VIQI	0.177	0.165	6.6%	0.161	undefined
VIQI – DISCERN	0.158	0.169	6.6%	0.170	false

*p < 0.05, **p < 0.01, ***p < 0.001. Zero-order τ -b: Kendall's tau-b in the pooled sample. Partial τ : first-order partial Kendall's tau adjusted for search-query group. Change: absolute percentage change from the zero-order to the partial coefficient; a change exceeding 10% is conventionally regarded as indicative of confounding. Stratified τ : stratum-weighted (common) Kendall's tau, computed within each search-query subsample and weighted by the number of pairs, with p-values obtained by within-stratum permutation (20,000 replicates). Kendall's tau was used throughout for consistency with the primary correlation analysis (Table 7). All six associations changed by less than 7% after adjustment, indicating that search query did not act as a confounder for any of them. In addition, the association between specialist authorship and high TCS content quality remained highly significant after stratification by search-query group (Cochran–Mantel–Haenszel statistic = 48.002, p < 0.001).

Four findings were reproduced in the “face scanner app” subsample only, in each case with the same direction in both strata and differing only in magnitude and statistical power. Three findings were not reproduced in either stratum and should be regarded as unstable: the difference in number of likes across TCS groups, the VIQI category $\times$ TCS cross-tabulation and the TCS–VIQI correlation. For the VIQI cross-tabulation this is consistent with the Monte Carlo permutation test, which also failed to reach significance.

Discussion

In this study, the quality of 91 YouTube™ videos on 3D facial scanning was assessed using four complementary instruments: GQS, VIQI, modified DISCERN and a purpose-designed content score (TCS). Every video achieving high content quality had been produced by a specialist, and none uploaded by a commercial company, layperson or academic institution reached this category. Three further patterns emerged — a dissociation between production and content quality, the systematic omission of clinically critical information, and the independence of quality from audience engagement — together describing content that is technically accomplished yet clinically incomplete.

Health professionals increasingly turn to online platforms alongside traditional sources, since procedural knowledge is learned more readily through visual demonstration than written text^{25–27}. Three-dimensional facial scanning nevertheless differs from much of the online health content examined to date: the procedure is often performed by users themselves, outside direct clinical supervision, and an incorrectly performed scan may generate a superficially plausible but inaccurate measurement requiring quality control before clinical use^{28–30}. Acquisition is moreover frequently delegated to auxiliary staff while responsibility for recognising errors remains with the clinician, so such content addresses patients, staff and clinicians alike; as the procedure also generates biometric data, its informational scope extends beyond technical execution. Evaluation therefore requires not only general quality instruments but a measure interrogating the clinical requirements of the procedure.

The strongest determinant of content quality was not any quality score but the source of the video. An absolute boundary rather than a graded tendency was observed, so the association behaves as a threshold; it survived permutation testing, exact testing of the collapsed table, and was reproduced identically in each search-query subsample. Only a proportion of specialist videos themselves reached the highest category, indicating that professional authorship is necessary but not sufficient for comprehensive coverage. This accords with reports linking professional authorship to accuracy in other medical topics^{31–33}, although the sharpness of the separation here differs from previously described graded differences, possibly because criteria governing correct execution can be conveyed only through domain experience. Identifying which source produces adequate content is, however, only part of the picture: specialist videos formed only a small proportion of the sample, while commercial and layperson content made up the great majority. Compounding this, those of highest content quality were also among the least viewed, and are therefore less likely to attract user attention^{33,34}. Scarcity and low visibility together constitute a problem not only of content quality but of accessibility, since reliable material does not necessarily reach its intended audience — with direct implications for how professional bodies approach digital patient and staff education.

The ceiling effect in the presentation index likely reflects the subject matter itself: facial scanning is grounded in a contemporary imaging technology, and those producing such content typically already possess advanced camera and editing capabilities. Visual quality here is not merely a presentational choice but a demonstration of the technology shown, a tendency reinforced by the predominance of commercial content in which production quality serves a marketing function. High presentation quality should therefore be regarded less as a distinguishing merit than as a baseline standard imposed by the technological context.

The cross-tabulations reveal a hierarchy consistent with the group comparisons. The reliability instrument ranked second after source type and produced a boundary in both directions: no high-reliability video fell into the lowest content-quality group, and none of low or moderate reliability reached the highest. The general quality rating, though comparable in strength, was clean at the upper end but broke down at the lower, where almost half of the lowest content-quality videos were nonetheless rated high – the most concrete illustration of the dissociation described above. The presentation index reached significance only in the asymptotic test and was not reproduced in either subsample. What emerges is an ordering of explanatory value – authorship, then reliability, then general quality, with presentation contributing little – reproduced by three methodologically distinct approaches, suggesting a structural property of the content rather than an analytical artefact.

Sub-item analysis revealed that the gaps in content coverage were not randomly distributed: the videos were rich in procedural information yet markedly poor in critical information. The purpose of the scan, the device used and camera positioning were addressed in almost every video, whereas lighting conditions, confirmation of measurement accuracy and the limitations of the procedure appeared in far fewer. This distinction is clinically consequential, since adequate illumination is among the principal determinants of accuracy in smartphone-based facial capture³⁵; a substantial proportion of the videos that instruct viewers on how to position the camera make no mention of the fact that lighting may compromise the measurement. The most striking of these gaps concerns data privacy: although 3D facial scanning generates data of a biometric nature and can be performed with a simple, universally available application, the fate of the captured data was addressed in only a very small number of videos. Since individuals have the right to know what becomes of a scan they have voluntarily performed and to keep that record confidential, users should be directed towards applications providing the necessary safeguards, regardless of whether the scan is carried out at home or in a clinical setting ³⁴.

As in previous studies, no association was identified between the descriptive parameters of the videos and the quality instruments ^{36,37}. This result supports the view that metrics such as likes or view counts are poor indicators of a video's educational value; popularity is not equivalent to quality. The popularity of YouTube videos is largely determined by the platform algorithm, with ranking based on watch time, click-through rate and engagement, while attractive titles and thumbnails draw viewers and confirmation bias and social proof reinforce this effect. Although quality content may gain popularity, high popularity does not indicate high quality, particularly in fields such as medicine and science, where misleading content can overshadow reliable sources. While some studies have reported an association between duration and quality ³⁸, no consistent relationship between video duration and the quality instruments was found in the present study. Taken together, these findings support the view that video parameters cannot serve as reliable indicators of content quality, since they are directly influenced by viewer behavior and algorithmic ranking.

Platforms such as YouTube have become indispensable for the rapid dissemination of technical information concerning both interventional and non-interventional dental procedures. This study nevertheless demonstrated that only a very small proportion of videos fully met the clinical content criteria, illustrating the difficulty of accessing reliable educational content. As the reach of YouTube continues to grow, it is important that videos which are both reliable and instructive be prioritized when the platform is used as a supplementary resource. In this context, the establishment of an integrated peer-review system could enhance the educational value of such content. Contributions from clinicians, specialists and academic institutions to these platforms are of importance for the dissemination of evidence-based practice, facilitating access to instructive content and raising awareness.

Limitations:

This study has several limitations. It provides a cross-sectional snapshot of a platform whose content and ranking algorithms change continuously, and the results are specific to the search terms used; although two complementary queries were employed and shown to yield comparable subsamples, other formulations might retrieve different material. Only English-language content was assessed. The highest content-quality group comprised few videos and expected cell frequencies in the cross-tabulations were low; although Monte Carlo permutation and Fisher's exact tests were applied, estimates involving this group remain imprecise. Finally, the TCS was designed for this study and, although its items derive from established requirements for facial capture, it has not been independently validated. Although not formally validated, its discriminative performance in this sample is described in the Discussion.

Conclusions

YouTube™ content on 3D facial scanning is technically well produced but clinically incomplete. Specialist authorship was the only factor associated with high content quality, yet such videos constituted a small minority of the available material and were among the least viewed. Information on measurement accuracy and data privacy was largely absent, and no engagement metric provided any indication of quality. Patients, auxiliary staff and clinicians seeking guidance are therefore likely to encounter well-produced content that teaches the procedure without conveying its limitations, and no platform-level means exists by which they might identify the small proportion of material that does.

Abbreviations

GQS: Global Quality Scale

VIQI: Video Information and Quality Index

TCS: Total Content Score

DISCERN: Standardized health information quality instrument

ICC: Intraclass Correlation Coefficient

IQR: Interquartile Range

SPSS: Statistical Package for the Social Sciences

Declarations

Ethics approval and consent to participate: This study did not involve human participants, identifiable personal data, or experimental interventions. Therefore, ethical approval was not required and was not sought.

Consent for publication: Not applicable.

Availability of data and materials: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions: İAD: Conceptualization, Formal Analysis, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation. İOY: Data Curation, Investigation, Methodology, Formal Analysis, Software, Writing – Review & Editing. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

1. Samsonyanová L, Broukal Z. A systematic review of individual motivational factors in orthodontic treatment: Facial attractiveness as the main motivational factor in orthodontic treatment. *Int J Dentistry Hindawi Publishing Corporation*. 2014. 10.1155/2014/938274 . PubMed PMID: 24963296.
2. Ishizuka-Nambu H, Aida J, Kiuchi S, Kato C, Ishizuka K, Ono T. Differences in chief complaints based on the molar relationship: A quantitative text analysis of open-ended responses from orthodontic patients. *J World Fed Orthod*. 2026. 10.1016/j.ejwf.2026.04.004.
3. Baik HS, Jeon JM, Lee HJ. Facial soft-tissue analysis of Korean adults with normal occlusion using a 3-dimensional laser scanner. *Am J Orthod Dentofac Orthop*. 2007;131:759–66. 10.1016/j.ajodo.2005.08.038 . PubMed PMID: 17561054.
4. Uranbey Ö, Aktaş S, Cennetoğlu MN, Şeker H, Gürsoytrak B, Nelke K. Three-dimensional facial scanning in oral and maxillofacial surgery: a scoping review of clinical applications, accuracy, and outcomes. *Oral and maxillofacial surgery*. 2026. 10.1007/s10006-026-01530-y PubMed PMID: 41826787.
5. Ayoub AF, Xiao Y, Khambay B, Siebert JP, Hadley D. Towards building a photo-realistic virtual human face for craniomaxillofacial diagnosis and treatment planning. *Int J Oral Maxillofac Surg*. 2007;36:423–8. 10.1016/j.ijom.2007.02.003 . PubMed PMID: 17428638.

6. Tomášik J, Zsoldos M, Oravcová L, Lifková M, Pavleová G, Strunga M, et al. AI and Face-Driven Orthodontics: A Scoping Review of Digital Advances in Diagnosis and Treatment Planning. AI (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI); 2024. pp. 158–76. 10.3390/ai5010009.
7. Ritschl LM, Wolff KD, Erben P, Grill FD. Simultaneous, radiation-free registration of the dentoalveolar position and the face by combining 3D photography with a portable scanner and impression-taking. *Head Face Med*. 2019;15. 10.1186/s13005-019-0212-x . PubMed PMID: 31767030.
8. Pellitteri F, Albertini P, Brucculeri L, Cremonini F, Guiducci D, Falconi V, et al. Soft tissue changes during orthopedic therapy: An in vivo 3-dimensional facial scan study. *Am J Orthod Dentofac Orthop*. 2025;167:154–65. .016 PubMed PMID: 39412468.
9. Jheon AH, Oberoi S, Solem RC, Kapila S. Moving towards precision orthodontics: An evolving paradigm shift in the planning and delivery of customized orthodontic therapy. *Orthod Craniofac Res*. 2017;20:106–13. <https://doi.org/10.1111/ocr.12171>.
10. Nehru A, Nagesh S. Smartphone-based 3D facial scanning applications in orthodontic diagnosis: A cross-sectional study. *J Dent Res Dent Clin Dent Prospects*. 2025;19:97–103. 10.34172/joddd.025.42109.
11. Bor S, Oğuz F, Özdemir D. Evaluation of trueness and precision of 3 face-scanning devices. *Am J Orthod Dentofac Orthop*. 2025. <https://doi.org/10.1016/j.ajodo.2025.04.009>.
12. Similar Web. YouTube Internet Traffic [Internet]. [cited 2026 Jul 15]. Available from: https://pro.similarweb.com/?action=website_performance&domain=youtube#/digitalsuite/websiteanalysis/overview/website-performance/*/999/1m?webSource=Total&key=youtube.com
13. Sezici YL, Gediz M, Dindaroğlu F. Is YouTube an adequate patient resource about orthodontic retention? A cross-sectional analysis of content and quality. *Am J Orthod Dentofac Orthop*. 2022;161:e72–9. 10.1016/j.ajodo.2020.12.023 . PubMed PMID: 34509330.
14. Guzman AK, Wang RH, Nazarian RS, Barbieri JS. Evaluation of YouTube as an educational resource for treatment options of common dermatologic conditions. *Int J Dermatol*. 2019;59:e65. 10.1111/ijd.14693.
15. Yildirim G, Kocaelli HA. Assessment of the content and quality of YouTube videos related zygomatic implants: A content–quality analysis. *Clin Implant Dent Relat Res*. 2023;25:605–12. 10.1111/cid.13194.
16. Firth F, Bennani F, Stacknik S, Farella M. Orthodontic patient co-operation: A review of the clinician’s role in predicting and improving patient compliance. *Australasian Orthodontic J*. 2019;35:5–12.
17. Tamosiūnaite I, Vasiliauskas A, Dindaroglu F. Does YouTube provide adequate information about orthodontic pain? *Angle Orthod*. 2023;93:403–8. 10.2319/072822-527.1 . PubMed PMID: 36820821.
18. Yavan MA, Gökçe G. YouTube as a source of information on adult orthodontics: a video analysis study. *J World Fed Orthod*. 2022;11:41–6. 10.1016/j.ejwf.2021.09. .001 PubMed PMID: 34756709.
19. Ozturk T, Gumus H. YouTube as an information and education source for early orthodontic treatment. *Am J Orthod Dentofac Orthop*. 2022;162:e123–32. 10.1016/j.ajodo.2022. 06.013 PubMed PMID: 35835702.
20. Meade MJ, Sooriakumaran P, Dreyer CW. Orthodontic retention and retainers: Quality of information provided by dental professionals on YouTube. *Am J Orthod Dentofac Orthop*. 2020;158:229–36. 08.013 PubMed PMID: 32616322.
21. Topsakal KG, Duran GS, Görgülü S, Eser Misir S. Is YouTubeTM an adequate source of oral hygiene education for orthodontic patients? *Int J Dent Hyg*. 2022;20:504–11. 10.1111/idh.12557 . PubMed PMID: 34687593.
22. Selvaraj M, Sennimalai K, Parija PP, Siddiqui HP, Singh A, Mohaideen K. Content quality and reliability of YouTube videos on oral appliance therapy for obstructive sleep apnea: A systematic analysis. *Special Care in Dentistry*. John Wiley and Sons Inc; 2024. pp. 1307–16. 10.1111/scd.12995 . PubMed PMID: 38520058.
23. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* (1978). 1999;53:105. 10.1136/jech.53.2.105
24. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis—a wakeup call? *J Rheumatol*. 2012;39:899–903. 10.3899/jrheum.111114.
25. Kritsanaviparkporn C, Kitcharoen N, Yimcharoen N, Thaweesuk S, Maneerit S, Treesirichod A. A Comparative Study of Online Medical Procedure Learning via Video Versus Text Modalities During the COVID-19 Pandemic. *South-East Asian J Med Educ*. 2021;15:69–74. 10.4038/seajme.v15i2.369.
26. Ghasroddashti A, Sorefan-Mangou F, Fernandes R, Del, Williams E, Choi K, Zevin B. Learning Outcomes and Educational Effectiveness of Social Media as a Continuing Professional Development Intervention for Practicing Surgeons: A Systematic Review and Narrative Synthesis. *Annals Surg Open*. 2024;5:e470. 10.1097/as9.000000000000470.
27. Buch SV, Treschow FP, Svendsen JB, Worm BS. Video- or text-based e-learning when teaching clinical procedures? A randomized controlled trial. *Adv Med Educ Pract*. 2014;5:257–62. 10.2147/AMEP.S62473.

28. Papadopoulou AK, Di Santo F, Antonarakis GS, Ghislanzoni LH. The Validity of a Smartphone-Based Method for Acquiring 3D Images of the Face. *J Clin Med*. 2024;13. 10.3390/jcm13216362.
29. Luo Y, Zhao M, Lu J. Accuracy of Smartphone-Based Three-Dimensional Facial Scanning System: A Systematic Review. *Aesthetic Plastic Surgery*. Springer; 2024. pp. 4500–12. 10.1007/s00266-024-04121-y PubMed PMID: 38831068.
30. Chong Y, Liu X, Shi M, Huang J, Yu N, Long X. Three-dimensional facial scanner in the hands of patients: validation of a novel application on iPad/iPhone for three-dimensional imaging. *Ann Transl Med*. 2021;9:1115–1115. 10.21037/atm-21-1620.
31. Ng MK, Emara AK, Molloy RM, Krebs VE, Mont M, Piuze NS. YouTube as a source of patient information for total knee/hip arthroplasty: quantitative analysis of video reliability, quality, and content. *JAAOS-Journal Am Acad Orthop Surg*. 2021;29:e1034–44. 10.5435/JAAOS-D-20-00910.
32. Chen AY, Azizi B, Borba AJ, Armstrong AW. Rosacea videos on social media: A comparison of accuracy, quality, and viewer engagement. *Dermatol Online J*. 2021;27. 10.5070/d3272052381 . PubMed PMID: 33818977.
33. Desai T, Shariff A, Dhingra V, Minhas D, Eure M, Kats M. Is content really king? An objective analysis of the public's response to medical videos on YouTube. *PLoS ONE*. 2013;8. 10.1371/journal.pone.0082469 . PubMed PMID: 24367517.
34. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice. *IEEE Access*. 2018;6:9390–403. 10.1109/ACCESS.2018.2799522.
35. Ntovas P, Ladia O, Barmak AB, Fehmer V, Sailer I, Revilla-León M. Influence of ambient light color temperature and illuminance on the accuracy of 3-dimensional patient representation using different facial scanning technologies: A clinical study. *J Prosthet Dent*. 2026. 10.1016/j.prosdent.2026.06.004ExternalLink.
36. Barhate UH, Sharan J, Jena AK. Dynamics of Social Media Popularity of Do It Yourself Orthodontic Aligners: A YouTube Case Study. *J Indian Orthodontic Soc*. 2023;57:171–7. 10.1177/03015742221125969.
37. Gorgy A, Hawary H, El, Galli R, MacDonald M, Barone N, Thibaudeau S. Evaluating the educational quality of surgical YouTube® videos: A systematic review. *Health Sci Rev*. 2022;5:100067. <https://doi.org/10.1016/j.hsr.2022.100067>.
38. Topbaş C, Paksoy T, İslamoğlu AG, Çağlar K, Kul AK. Is it safe to learn about vital pulp capping from YouTube™ videos? A content and quality analysis. *Int J Med Inf*. 2024;185. 10.1016/j.ijmedinf.2024.105409 . PubMed PMID: 38471408.

Figures

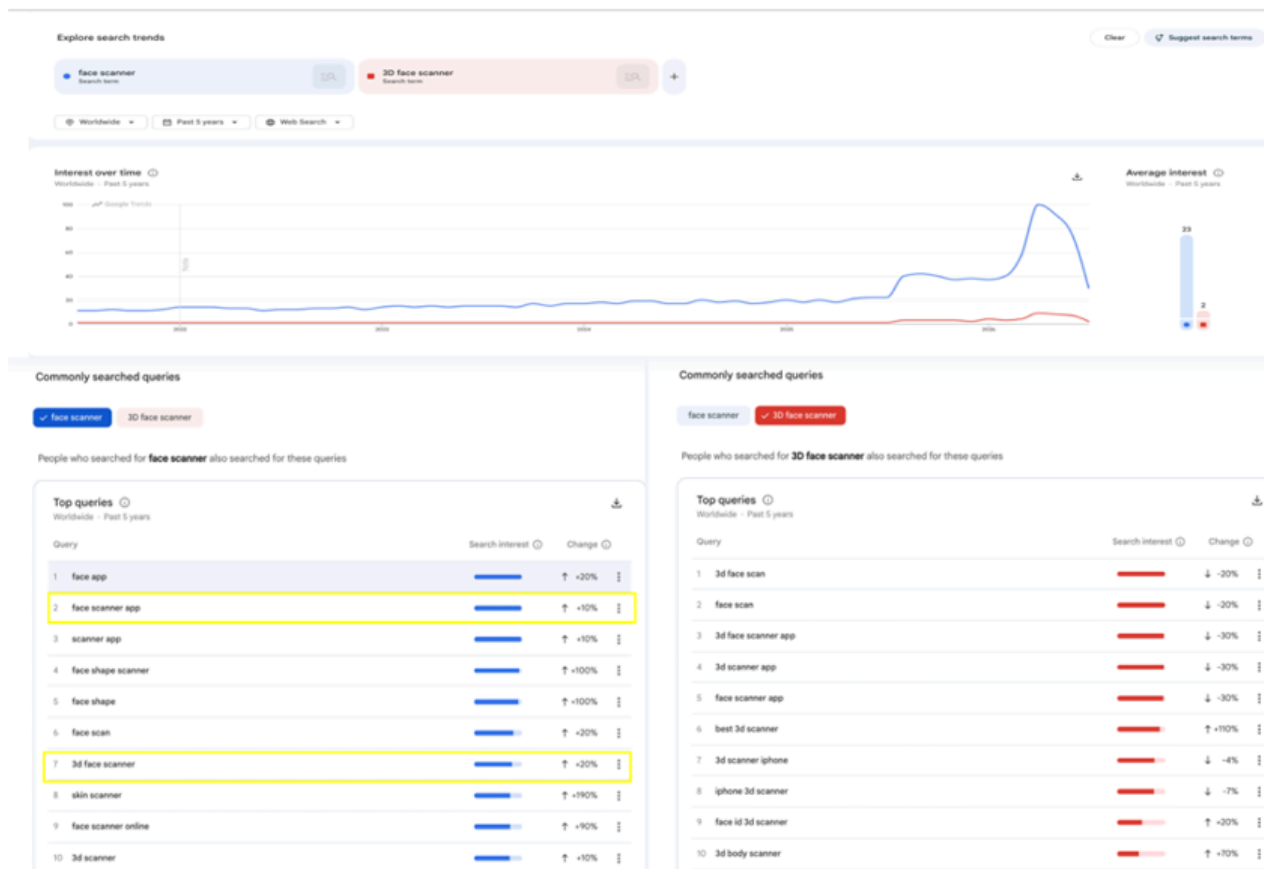


Figure 1

Determination of search queries

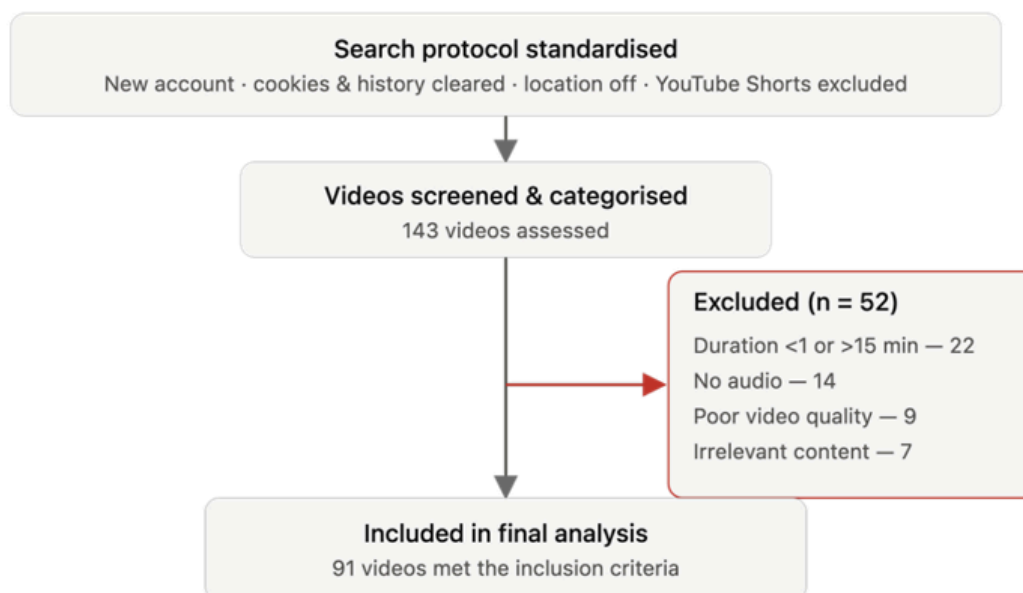


Figure 2

Flowchart diagram of the video selection process.