ORIGINAL ARTICLE
Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.530417

Educational Quality and Reliability of YouTube Videos on Hawley Retainer Fabrication: A Cross-sectional Content Analysis

Saira Tariq[a]*, Maria Shahwar[b], Abdullah Arslan[c], Zinat Rehana[d]

[a]Azra Naheed Dental College, Raiwind Road, Lahore, Pakistan
[b]Islamic International Dental College and Hospital, Islamabad, Pakistan
[c]University College of Medicine and Dentistry, Lahore, Pakistan
[d]Shaheed Suhrawardy Medical College, Dhaka, Bangladesh

*Corresponding address: *Azra Naheed Dental College, Raiwind Road, Lahore, Pakistan*
Email: sairat788@gmail.com

Received: 16 March 2026 / **Revised:** 28 May 2026 / **Accepted:** 05 June 2026 / **Available Online:** 26 June 2026

ABSTRACT

Objectives: To evaluate the reliability, educational quality, and viewer engagement of publicly available YouTube videos on Hawley retainer fabrication using three standardized assessment instruments.
Methods: A cross-sectional content analysis was conducted using five complementary search strings on YouTube. After deduplication and application of pre-specified eligibility criteria, 23 English-language videos demonstrating Hawley retainer fabrication were identified. Two calibrated reviewers independently scored each video using the Journal of the American Medical Association (JAMA) benchmarks, the Modified DISCERN tool (mDISCERN), and the Modified Global Quality Score (mGQS). Engagement metrics and an adapted Video Power Index (VPI) were extracted. Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). Pearson correlations and an exploratory multivariate linear regression were used to examine associations between engagement metrics and educational quality.
Results: Mean scores indicated low-to-moderate educational quality (mGQS 1.98/5; mDISCERN 0.52/5; JAMA 1.54/4), with 74% of videos classified as low quality on pre-specified mGQS bands. Likes correlated moderately with mGQS ($r = 0.46$) and mDISCERN ($r = 0.39$) and were the only significant predictor in the exploratory regression ($\beta = 1.6 \times 10^{-3}$, $p = 0.016$); however, the overall model did not reach significance ($R^2 = 0.434$, $p = 0.063$). Inter-rater reliability was moderate for mGQS (ICC = 0.443) and fair for mDISCERN (ICC = 0.323); JAMA returned a negative ICC (−0.257), indicating instrument–content mismatch.
Conclusion: YouTube content on Hawley retainer fabrication is predominantly of low educational quality. Viewer engagement partially tracks quality but cannot substitute for a structured evaluation. These findings highlight the need for professionally developed, peer-reviewed digital resources in orthodontic laboratory training.
Keywords: Dental education; Orthodontics; Hawley retainer; Social media; Video recording; Quality assessment

Citation: Tariq S, Shahwar M, Arslan A, Rehana Z. Educational Quality and Reliability of YouTube Videos on Hawley Retainer Fabrication: A Cross-sectional Content Analysis. J Sci Technol Educ Art Med. 2026;3(1):24-34

Introduction

Retention following active orthodontic treatment is essential for preserving tooth position and ensuring long-term occlusal stability. Without retention, periodontal remodeling and continued physiological growth predispose the dentition to relapse.[1] The Hawley retainer, first introduced by Charles Hawley in 1919, remains one of the most widely used removable retention appliances worldwide owing to its durability, adjustability, and ability to permit minor occlusal settling.[2]

The fabrication of a Hawley retainer requires precise wire bending and accurate adaptation of the acrylic baseplate, competencies that dental students and laboratory trainees are expected to acquire during pre-clinical and laboratory training.[3] Traditional instruction typically involves supervised hands-on practice; however, constraints such as limited laboratory time, increased material costs, reduced faculty contact, and ongoing effects of pandemic-related disruptions have collectively increased reliance on self-directed digital learning.[4] Recent surveys indicate that video-sharing platforms, particularly YouTube, are now routinely used to supplement formal curricula for procedural skills.[5,6] Despite this trend, the educational value of such resources remains largely unexamined for specific procedural competencies. Although students increasingly report supplementing laboratory training with YouTube content, few studies have assessed whether these materials meet the minimum educational standards required for instructional content in health professions curricula.

YouTube's accessibility and audiovisual format make it well-suited for procedural learning. However, the lack of peer review, editorial oversight, or structured quality control raises concerns regarding the accuracy, completeness, and reliability of instructional content available to trainees.[7,8] Existing research evaluating YouTube content across dental specialties generally reports low educational quality and considerable variability.[9-11] Studies of YouTube videos on topics such as vital pulp therapy, dental implant placement, and orthodontic bracket bonding have consistently found that most available content lacks clear procedural narration, evidence-based justification for technique choices, and adequate disclosure of authorship or conflicts of interest. Notably, most of these studies have focused on clinical chairside procedures rather than on laboratory-based fabrication skills. The assessment instruments employed, including the Modified DISCERN tool, the Modified Global Quality Score, and the JAMA benchmarks, have rarely been applied to benchwork content, where the absence of a patient context may influence the interpretation of reliability and transparency criteria. Evidence specific to orthodontic laboratory techniques, and to Hawley retainer fabrication in particular, remains limited, despite the procedural nature of the skill and the direct implications of inadequate technique for patient outcomes. Additionally, prior studies have seldom examined the relationship between viewer engagement metrics and educational quality. While likes, views, and comments are frequently reported descriptively, they are rarely evaluated as potential proxies or misleading surrogates for content quality. This represents a significant gap, as trainees who lack the expertise to independently assess instructional accuracy may rely on engagement signals, such as view counts or like ratios, as informal indicators of credibility. Whether these signals correspond to the actual educational value of procedural dental videos has not been established.

This study was intentionally designed as an observational analysis. Its objective was to characterize the educational resources available to trainees seeking guidance on Hawley retainer fabrication via YouTube, with a specific focus on evaluating their reliability, comprehensiveness, and the relationship between perceived quality and viewer engagement metrics. Establishing this descriptive baseline is essential for informing subsequent interventions by educators, professional bodies, or accreditation organizations seeking to curate, recommend, or supplement online instructional materials in this domain. Given the procedural nature of the skill and the patient-safety implications of inadequate technique, this baseline is directly relevant to curriculum planning and the broader digital health information ecosystem.

Accordingly, this study aimed to: (i) evaluate the reliability and educational quality of YouTube videos depicting Hawley retainer fabrication using three standardized instruments; (ii) describe viewer engagement on these videos using a transparently adapted Video Power Index; and (iii) explore the extent to which engagement metrics correspond to underlying educational quality.

Materials and Methods

Study Design and Setting

This was a cross-sectional content analysis of publicly available YouTube videos conducted over three months. The study is purely observational; no interventions with video content, creators, or the YouTube platform were performed. As the analysis used only publicly accessible material and did not involve human participants or identifiable personal data, formal ethical approval was not required; the study was nevertheless conducted in accordance with the principles of the Declaration of Helsinki. This study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies.

Search Strategy

To minimize the selection bias inherent in single-keyword searches of YouTube, five complementary search strings were used (Table 1). Searches were performed in a private browsing window with cleared history, watch history disabled, and default platform filters ("Relevance" sort) restored between queries. Searches were conducted between August 2025 and October 2025. The exact date of each search was logged to support reproducibility. Within each search string, results were screened in descending order of platform relevance until two consecutive screens contained no topically relevant videos.

Table 1: Search strings used to identify candidate videos and the rationale for each

No.	Search String	Rationale
1	"Hawley retainer fabrication"	Primary procedural term used in the indexed dental education literature.
2	"Hawley retainer step by step"	Captures instructional/tutorial-style uploads aimed at students.
3	"Making a Hawley retainer from start to finish"	Captures full-procedure videos that may not use the term "fabrication."
4	"Dental lab benchwork Hawley"	Captures laboratory-technician-oriented content.
5	"Orthodontic wire bending tutorial Hawley"	Captures videos focused on the labial bow and clasp components.

Eligibility Criteria

Videos were eligible for inclusion if they: (i) were presented in English; (ii) demonstrated the actual procedure for fabricating a Hawley retainer (laboratory workflow or combined chair-side workflow); and (iii) represented unique, non-duplicate content. Videos were excluded if they: (i) were presented in a language other than English; (ii) functioned primarily as advertisements, product showcases, or institutional promotional material without procedural demonstration; (iii) were duplicate uploads; (iv) demonstrated only the clinical insertion or adjustment of an already-fabricated retainer; or (v) contained insufficient audio or video quality to permit reliable scoring. The full screening and selection process is summarised in Figure 1.

Data Extraction

For each included video, the following descriptive and engagement variables were recorded on the day of scoring to minimize temporal drift: total number of views, number of likes, number of comments, video duration (seconds), days since upload, and uploader category (individual practitioner, dental laboratory, institutional channel, or unspecified). Where a video contained substantial non-instructional introductory content or filler material, this was noted qualitatively during scoring, but the full reported duration was retained for descriptive purposes.

Educational Quality Assessment

Three established instruments were used to assess reliability and educational quality. Each was applied strictly according to its published operational definition, without modification of scoring rules.

The JAMA benchmarks scored each video on four binary criteria, authorship disclosure, citation of references, disclosure of conflicts of interest, and visibility of upload or publication date, yielding a maximum score of 4.[12] JAMA was originally developed for written health information; its application to procedural video content is recognised in the literature as imperfect, and this point is discussed below in light of the observed inter-rater reliability.

The Modified DISCERN tool consisted of five binary items appraising reliability and clarity,

Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.530417

yielding a maximum score of 5.[12]

The Modified Global Quality Score (mGQS) was a 5-point scale assessing overall educational quality, comprehensiveness, and usefulness, with 1 indicating poor quality and 5 indicating excellent quality.[13] To make the categorical interpretation of mGQS explicit and reproducible, and to address the question of how high- versus low-quality content was distinguished, cut-offs were specified before scoring began (Table 2) and applied uniformly by both reviewers

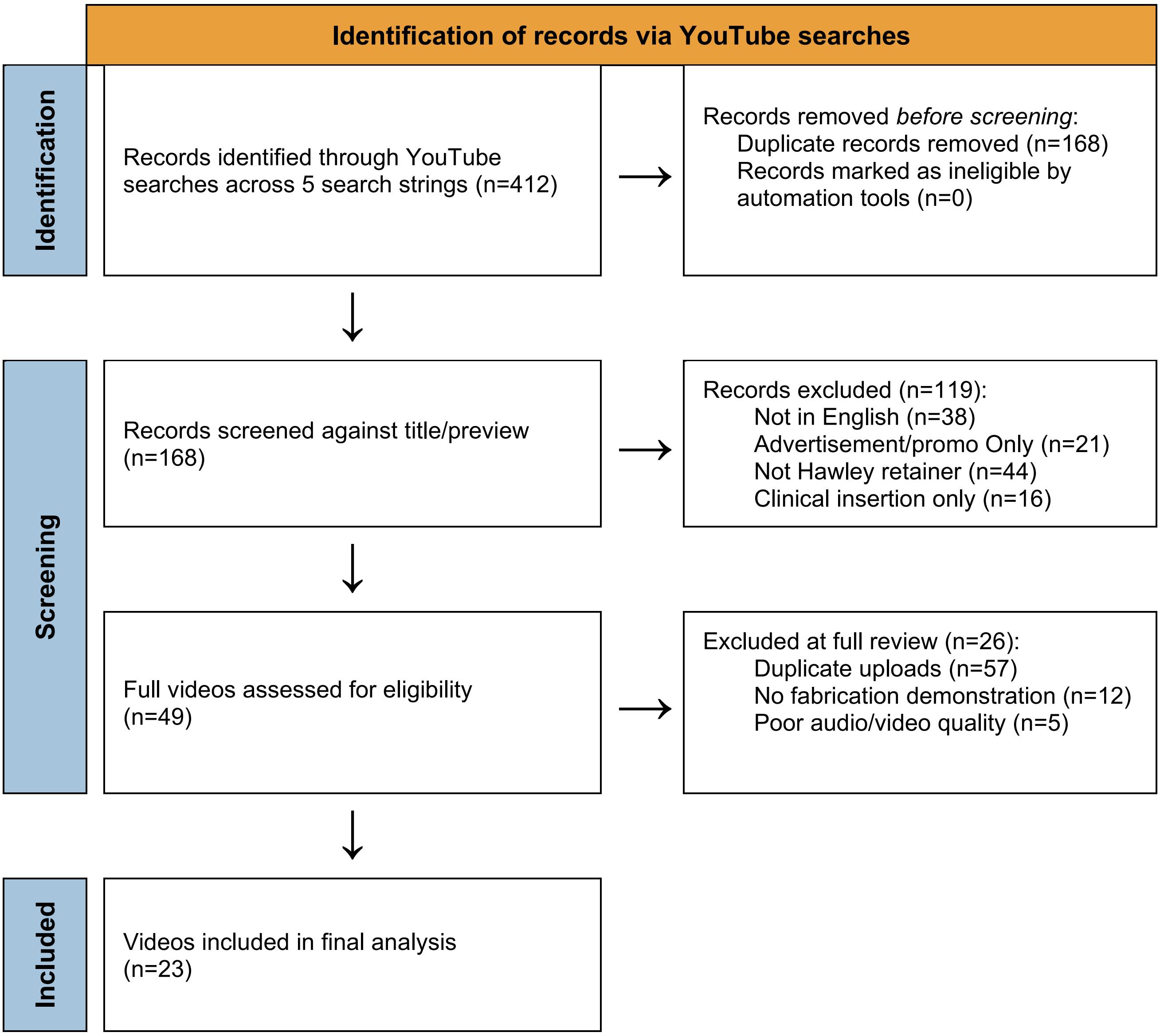

Figure 1: PRISMA-style flow diagram of YouTube video identification, screening, eligibility assessment, and inclusion. Counts are reported across the pooled, deduplicated results of all five search strings. Independently scored by two reviewers using JAMA benchmarks, mDISCERN and mGQS

Table 2: Pre-specified quality bands used to classify videos on the Modified Global Quality Score (mGQS)

Quality Band	mGQS Range	Operational Description
Low quality	1.0 – 2.0	Limited or incorrect procedural detail; critical steps missing; little or no educational structure or narration.
Moderate quality	2.1 – 3.5	Covers main steps but with notable omissions; partial demonstration of materials, wire-bending sequence, or acrylic adaptation.
High quality	3.6 – 5.0	Comprehensive, step-by-step demonstration with clear narration, correct technique, and accurate terminology.

Adapted Video Power Index

Viewer engagement was further summarised using an adapted Video Power Index (VPI). The original VPI incorporates a like ratio defined as likes ÷ (likes + dislikes); however, YouTube removed publicly visible dislike counts in November 2021, rendering this denominator unobtainable for any post-2021 evaluation. To remain transparent about this platform-level constraint while preserving a within-study composite engagement measure, the VPI was adapted as follows:

- View ratio = total views ÷ days since upload.
- Adapted like ratio = (likes ÷ views) × 100.
- Adapted VPI = (view ratio × adapted like ratio) ÷ 100.

This adaptation is not numerically identical to the pre-2021 VPI used in earlier YouTube content-analysis studies and is therefore not directly comparable to those reported numerical values. It is intended to provide an internally consistent engagement composite under current platform conditions.

Reviewer Calibration and Inter-rater Reliability

Two reviewers independently scored all videos: Reviewer 1 was a postgraduate orthodontic trainee, and Reviewer 2 was a general dental practitioner with prior laboratory experience. Before formal scoring, both reviewers jointly scored three pilot videos (excluded from the final dataset) and resolved discrepancies through structured discussion to establish a shared operational understanding of each scoring item. During formal scoring, reviewers were blinded to one another's assessments and worked independently from a shared coding sheet. The Intraclass Correlation Coefficient (ICC, two-way random effects, absolute agreement) was used to quantify agreement for each instrument. No video was excluded on the basis of low inter-rater reliability; instead, the implications of low ICC for instrument suitability are discussed in the relevant Discussion section. Final scores used in all analyses are the mean of the two reviewers' independent ratings for each instrument.

Statistical Analysis

Continuous variables were summarised using means, standard deviations, medians, and interquartile range, and presented descriptively. Pearson correlation coefficients were used to describe bivariate relationships between engagement metrics and educational quality scores.

A multivariate linear regression model was fitted to explore which engagement variables, if any, were associated with educational quality. The dependent variable was a composite quality score defined as the mean of mDISCERN and mGQS; JAMA was excluded from this composite a priori on the basis of its observed instrument–content mismatch (see Inter-rater reliability, below). Independent variables were the number of views, likes, comments, video duration, and the adapted VPI. Because the available population ($n = 23$) corresponded to approximately 4.6 observations per predictor, below the conventionally recommended 10:1 ratio for confirmatory linear regression[14], the regression was pre-declared as exploratory and hypothesis-generating rather than confirmatory. The Discussion is anchored in the descriptive and correlational findings, with the regression interpreted only as a tentative signal warranting confirmatory replication in a pooled dataset. Statistical significance was set at $\alpha = 0.05$ for descriptive interpretation. Analyses were performed using Python 3.11 (pandas, scipy, statsmodels).

ORIGINAL ARTICLE
Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.530417

Ethical Considerations

This study used only publicly accessible YouTube videos and did not involve human participants, identifiable personal data, or animal subjects. Formal institutional ethics review was therefore not required. The study was nevertheless conducted in accordance with the ethical principles of the Declaration of Helsinki.

Results

Selection and Characteristics of Included Videos

A total of 23 YouTube videos satisfied all eligibility criteria after pooling and deduplication across the five search strings (Figure 1). Uploaders were heterogeneous, including individual practitioners, dental laboratories, and institutional channels; no single channel dominated the eligible set.

Descriptive Characteristics of Videos and Engagement

Descriptive statistics are presented in Table 3. View counts ranged from 70 to 157,000 (median 18,000), and as counts ranged from 0 to 989 (median 42), reflecting a positively skewed distribution typical of social media content. The typical video lasted approximately 506 s (median ≈ 8.4 minutes). However, a single outlier of 68,500 s (≈19 hours) substantially inflated the mean length. On manual inspection, this outlier was a continuous-stream upload in which the active procedural demonstration accounted for only a small fraction of the runtime; the remainder consisted of unrelated or filler material. The value is retained in Table 3 for completeness and is discussed below as illustrative of a recurring quality issue: total video duration on the platform is not a reliable proxy for the duration of substantive instructional content.

Table 3: Descriptive statistics for engagement metrics and educational quality scores (n = 23)

Variable	n	Mean	SD	Min	Q1	Median	Q3	Max
Number of views	23	29,845	38,435	70	779	18,000	46,000	157,000
Number of likes	23	183.3	290.1	0	4.5	42	183	989
Number of comments	23	19.4	49.3	0	0	2	12	226
Video length (s)	23	3,835	14,154	20	165	506	1,018	68,500
Adapted VPI	23	14.3	17.1	0.03	—	—	—	61.4
mDISCERN (0–5)	23	0.52	0.41	0	0	0.5	1	1
mGQS (1–5)	23	1.98	0.87	1	1	2	2.5	3.75
JAMA (0–4)	23	1.54	0.45	0	1.5	1.5	2	3

† Q1, median, and Q3 omitted for adapted VPI due to highly skewed distribution with limited interpretability of quartile values

Educational quality scores were low overall: mean mGQS 1.98/5 (SD 0.87), mean mDISCERN 0.52/5 (SD 0.41), and mean JAMA 1.54/4 (SD 0.45). Applying the pre-specified mGQS cut-offs (Table 2), 17 videos (74 %) fell into the low-quality band, 5 (22 %) into the moderate-quality band, and only 1 video (4 %) reached the high-quality threshold.

Correlations between Engagement and Educational Quality

Pearson correlation coefficients are reported in Table 4. Views correlated strongly with likes (r = 0.67) and with the adapted VPI (r = 0.83), confirming internal consistency of the engagement measures. Likes correlated moderately and positively with mGQS (r = 0.46) and mDISCERN (r = 0.39), suggesting that videos perceived as more educationally valuable tended to attract more audience endorsement. Video duration correlated weakly but negatively with mGQS (r = −0.27) and mDISCERN (r = −0.28) — consistent with the qualitative observation that longer videos in this sample did not necessarily devote that additional runtime to substantive instruction. Correlations involving JAMA scores were weak across all variables, congruent with the inter-rater reliability findings reported below. At n = 23, correlations of approximately |r| ≥ 0.41 reached statistical significance at α = 0.05. Accordingly, the views–likes (r = 0.67), views–VPI (r = 0.83), likes–comments (r = 0.66), likes–VPI (r = 0.68), likes–mGQS (r = 0.46), and mDISCERN–mGQS (r =

0.48) correlations were statistically significant, while the remaining associations were not.

Table 4: Pearson correlation matrix for engagement metrics and educational quality scores. Values are correlation coefficients (r)

Item	Views	Likes	Comm.	Length	VPI	mDISC	mGQS	JAMA
Views	1.00	0.67	0.34	−0.19	0.83	0.25	0.39	−0.10
Likes	0.67	1.00	0.66	−0.14	0.68	0.39	0.46	0.16
Comments	0.34	0.66	1.00	−0.08	0.43	0.00	0.19	0.11
Length	−0.19	−0.14	−0.08	1.00	−0.18	−0.28	−0.27	0.24
VPI	0.83	0.68	0.43	−0.18	1.00	0.11	0.21	−0.11
mDISCERN	0.25	0.39	0.00	−0.28	0.11	1.00	0.48	0.00
mGQS	0.39	0.46	0.19	−0.27	0.21	0.48	1.00	−0.03
JAMA	−0.10	0.16	0.11	0.24	−0.11	0.00	−0.03	1.00

Inter-rater Reliability

Inter-rater reliability findings are summarised in Table 5. Reliability was moderate for mGQS (ICC = 0.443), fair for mDISCERN (ICC = 0.323), and notably negative for JAMA (ICC = −0.257). A negative ICC indicates that between-reviewer variance exceeded within-video variance; that is, the reviewers were not simply noisy but systematically diverged in how they applied the JAMA criteria. Qualitative review of the discrepant JAMA scores revealed that disagreement clustered around situations in which the four binary JAMA items (authorship, references, conflict-of-interest disclosure, currency) were either entirely absent from the video metadata or only partially inferable from it, a scenario for which JAMA, which was originally developed for written health information, was not designed to adjudicate. The JAMA scores are therefore retained in the descriptive analysis for transparency but were pre-emptively excluded from the composite outcome used in the regression model, on the grounds of instrument–content mismatch in this procedural setting rather than on the basis of the ICC alone.

Table 5: Inter-rater reliability of the three assessment instruments (n = 23 videos, 2 reviewers). Approximate 95 % confidence intervals based on two-way random-effects, absolute agreement ICC

Instrument	ICC	95% CI (approx.)	Interpretation
Modified DISCERN	0.323	[0.02, 0.61]	Fair agreement
Modified GQS	0.443	[0.10, 0.71]	Moderate agreement
JAMA benchmarks	−0.257	[−0.56, 0.14]	Systematic disagreement; instrument–content mismatch

Exploratory Regression Predicting Educational Quality

As previously stated, a multivariate linear regression was fitted with the composite quality score (mean of mDISCERN and mGQS) as the dependent variable, treating the analysis as exploratory given the 4.6:1 observation-to-predictor ratio. Results are shown in Table 6. The number of likes was the only individual predictor reaching conventional statistical significance ($\beta = 1.6 \times 10^{-3}$, $p = 0.016$). The number of views, comments, video length, and the adapted VPI did not reach significance. The model explained approximately 43 % of the variance in the composite quality outcome ($R^2 = 0.434$), but the overall model F-test did not reach conventional significance ($p = 0.063$). These results are interpreted as a tentative signal of a likes–quality association, consistent with the bivariate correlations reported in Table 4, rather than as a confirmatory predictive model. Replication in a substantially larger dataset, for example, by pooling videos across related orthodontic appliances and procedures, will be required to confirm the association.

Table 6: Exploratory multivariate linear regression predicting the composite educational quality score (mean of mDISCERN and mGQS). *Significant at α = 0.05.

Predictor	β (Coef.)	SE	t	p-value
Constant	1.198	0.140	8.56	<0.001
Views	3.61×10^{-6}	4.59×10^{-6}	0.79	0.443
Likes	1.6×10^{-3}	6.0×10^{-4}	2.67	0.016*
Comments	-3.4×10^{-3}	2.8×10^{-3}	−1.21	0.240
Video Length (s)	-1.07×10^{-5}	7.4×10^{-6}	−1.45	0.168
Adapted VPI	−0.016	0.011	−1.45	0.167

Discussion

This study assessed the educational quality and reliability of publicly available YouTube content on Hawley retainer fabrication using three standardized assessment instruments. The analysis was descriptive, documenting the resources a trainee would encounter on the platform rather than prescribing content-production standards. Interpreted in this context, the findings have several implications for orthodontic education, digital health information research, and methodologies for social media content analysis.

The principal finding is that the educational quality of available YouTube content on Hawley retainer fabrication is, on average, low. With 74% of videos falling into the pre-specified low-quality band on the mGQS and only one reaching the high-quality threshold, the dominant pattern is an incomplete or insufficient procedural demonstration. The low mDISCERN scores point to specific deficiencies in evidence-based justification of technique choices and in the clarity of procedural narration; the low JAMA scores reflect poor academic transparency, with authorship, citation, and conflict-of-interest disclosure rarely available. These findings are concordant with the broader literature evaluating YouTube content across dental and medical specialties.[15-17] These findings are consistent with Sezici et al., who reported a mean mGQS of 2.1 and similarly found that the majority of YouTube videos on orthodontic retainers fell below acceptable quality thresholds.[15] Alenezi and Jamleh reported comparable results for vital pulp therapy videos, with 68% classified as low quality on the mGQS.[16] The present study extends these findings to a specific laboratory fabrication procedure, suggesting that the pattern of low educational quality on YouTube is not confined to clinical chairside procedures but also applies to benchwork content, where procedural precision carries downstream implications for appliance fit and patient comfort.

Engagement metrics, particularly likes, were positively associated with educational quality both at the bivariate level ($r = 0.46$ with mGQS) and as the only individually significant predictor in the exploratory regression. It is not claimed, however, that engagement metrics can be used to predict educational quality in a clinically actionable sense. The overall regression model did not reach conventional significance ($p = 0.063$), and the 4.6:1 observation-to-predictor ratio is below conventional thresholds for confirmatory inference.[14] The regression should therefore be read as descriptive and hypothesis-generating; confirmatory replication would require either an extended observation window or a deliberately pooled dataset spanning related orthodontic procedures. The substantive interpretive weight of this study rests on the descriptive statistics and the bivariate correlation pattern, not on the regression's individual significant coefficient.

Video duration showed a weak but consistent negative correlation with both mDISCERN and mGQS. The 68,500 s outlier illustrates a broader phenomenon observed qualitatively across the sample: in several videos, the active instructional segment was preceded or interspersed with extended introductions, channel branding, advertising, unrelated commentary, or filler footage. As a result, total reported runtime on YouTube is an unreliable proxy for the duration of substantive educational content. This has two practical consequences. For trainees, selecting videos based on length is unlikely to identify the most informative resources. For researchers, it suggests that future content-analysis studies in this space may benefit from coding the proportion of runtime that constitutes genuinely instructional content. This has been flagged as a direction for future work.

The inter-rater reliability findings warrant particular methodological attention. The moderate ICC for mGQS (0.443) and the fair ICC for mDISCERN (0.323) are consistent with prior work using these instruments on procedural dental content[18] and indicate that, with appropriate reviewer calibration, both instruments perform acceptably in this setting. Ghai reported mGQS ICC values ranging from 0.40 to 0.55 across dental education YouTube studies, and Sezici et al. reported an ICC of 0.51 for mGQS in their analysis of orthodontic retainers.[15] The present values (mGQS ICC = 0.443, mDISCERN ICC = 0.323) fall within or slightly below this range, which may reflect the greater subjectivity involved in scoring laboratory fabrication videos where technique correctness requires domain-specific expertise that a general quality scale does not fully operationalize. The negative ICC for JAMA (−0.257) is more striking and merits explanation rather than mere exclusion. A negative ICC indicates systematic disagreement; the reviewers were not noisy but were applying the criteria differently. Qualitative inspection of the discrepant JAMA scores located the disagreement in cases where the four binary items were absent or only partially inferable from video metadata. This is a known limitation of applying JAMA, originally designed for written health information [12], to procedural video content. Rather than excluding without justification, the JAMA data were retained descriptively, excluded from the composite regression outcome due to instrument–content mismatch, and explicitly flagged. Future content-analysis research in this domain may benefit from either (i) operationalizing JAMA criteria for video formats prior to scoring, or (ii) relying on mDISCERN and mGQS as the primary instruments and treating JAMA as supplementary.

Regarding search breadth, the deliberate use of five complementary search strings (Table 1) was intended to minimize the selection bias that a single-term search would introduce. Despite this broadened strategy, only 23 unique English-language videos satisfied the inclusion criteria. This is interpreted as reflecting the true size of the publicly available educational pool for this procedure on this platform at the time of the search, rather than as undersampling. The small size of the available population is itself a substantive finding: trainees searching for guidance on Hawley retainer fabrication on the world's most-used video platform have a narrow set of resources to draw on, and the average educational quality of that set is low.

Two further methodological considerations are relevant to interpretation. First, the adapted VPI used here departs from the pre-2021 formulation because YouTube no longer publicly displays dislike counts; this loss of disapproval data is itself a constraint on YouTube's transparency as a source of health information, and the adapted formula cannot fully compensate for it. Future studies may consider supplementing engagement-based composites with structured expert review by clinical specialists, an approach recommended for any confirmatory study in this area. Second, although both reviewers had clinical and laboratory familiarity with Hawley retainer fabrication and were calibrated on a pilot set, expanding to a multi-reviewer expert panel (for example, two specialist orthodontists in addition to a generalist) would be expected to further improve reliability, particularly for mGQS.

Taken together, the present findings support the emerging consensus that ad hoc YouTube content cannot serve as an adequate substitute for structured procedural training in orthodontic laboratory skills.[19] They also support a more constructive corollary: that there is a clear and unmet need for professionally developed, peer-reviewed digital educational resources in this domain, ideally produced or endorsed by dental schools, professional bodies, or accreditation organizations, and accompanied by transparent authorship, citation, and disclosure information.[20] Based on these findings, three practical recommendations emerge. First, dental educators should not assume that directing students to YouTube for guidance on Hawley retainer fabrication will yield adequate instructional material; curated playlists or institutional video libraries with quality-assured content are preferable. Second, professional bodies such as the British Orthodontic Society or the American Association of Orthodontists could contribute by endorsing or producing standardized procedural videos that meet minimum quality benchmarks, including transparent authorship, evidence-based technique justification, and complete stepwise demonstration. Third, researchers conducting future YouTube content analyses in dental education should consider using mDISCERN and mGQS as primary instruments rather than JAMA benchmarks, given the instrument–content mismatch demonstrated here.

This study has several limitations that should be considered when interpreting the results. First, the analysis was confined to English-language videos on a single platform at a single time point; non-English content and content hosted on other platforms (e.g., institutional learning management systems or dedicated dental education sites) were not evaluated. Second, the available population of qualifying videos (n = 23) is small relative to the number of predictors in the regression model; the regression is presented as exploratory, and inferential claims are limited. Third, the adapted VPI used here departs from the pre-2021 formula owing to platform changes and is not directly comparable to numerical VPI values reported in earlier studies. Fourth, JAMA benchmarks were developed for written health information; their application to procedural video content is a recognized limitation reflected in the negative ICC observed here. Fifth, scoring was performed by two reviewers; expansion to a multi-reviewer expert panel would further test reliability. Sixth, total reported video duration is an imperfect proxy for instructional duration, and future work would benefit from coding the proportion of runtime that constitutes substantive procedural demonstration. Finally, YouTube's recommendation algorithm and platform indexing are dynamic; replication at different time points may yield different available populations.

Conclusion

Publicly accessible YouTube videos on Hawley retainer fabrication exhibit substantial variability and, on average, are of low educational quality. Viewer engagement, particularly likes, partially tracks educational quality, but popularity alone cannot substitute for the structured content evaluation. Modified DISCERN and mGQS appear more appropriate than JAMA benchmarks for evaluating procedural dental video content, and the JAMA application to this format requires methodological refinement. The findings point to a clear unmet need for professionally developed, peer-reviewed digital educational resources in orthodontic laboratory training, and to a methodological need for larger, pooled datasets that can support confirmatory analysis of the engagement–quality relationship.

Acknowledgments

The authors thank the calibration-round reviewers for their contribution to the development of the scoring protocol.

Author Contribution

ST conceived the research idea, developed the methodology, participated in the investigation, drafted the original manuscript, and supervised the study. MS contributed to the methodology, participated in the investigation, performed the formal analysis, and assisted in writing, reviewing, and editing the manuscript. AA participated in the investigation, performed data curation, and contributed to writing, reviewing, and editing the manuscript. MUR participated in the investigation, carried out validation, and contributed to writing, reviewing, and editing the manuscript. MR performed the formal analysis, prepared the visualizations, and contributed to writing, reviewing, and editing the manuscript. NN participated in the investigation, performed data curation, and contributed to writing, reviewing, and editing the manuscript. All authors read and approved the final manuscript.

Funding

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

Data Availability Statement

The de-identified dataset supporting the findings of this study, comprising video identifiers, engagement metrics, and reviewer scores, is available from the corresponding author on reasonable request. Video URLs are not republished to respect the platform's terms of service and the potential subsequent removal or modification of source content.

Conflict of Interest

The authors have no conflicts of interest to disclose.

References

1. Lyros I, Tsolakis IA, Maroulakos MP, Fora E, Lykogeorgos T, Dalampira M, et al. Orthodontic retainers-A critical review. Children (Basel). 2023;10(2):230. doi:10.3390/children10020230
2. Shoukat Ali U, Zafar K, Hoshang Sukhia R, Fida M, Ahmed A. Effect of bonded and removable retainers on occlusal settling after orthodontic treatment: a systematic review and meta-analysis. Dent Med Probl. 2023;60(2):327-34.

doi:10.17219/dmp/146194

3. Dervisbegovic S, Laky M, Tur D, Grundnig J, Rausch-Fan X, Moritz A, et al. Educational videos as a teaching approach to enhance dental students' practical skills in preclinical courses. BMC Med Educ. 2025;25(1):1299. doi:10.1186/s12909-025-07807-2
4. Schlenz MA, Wöstmann B, Krämer N, Schulz-Weidner N. Update of students' and lecturers' perspectives on online learning in dental education after a five-semester experience due to the SARS-CoV-2 (COVID-19) pandemic: insights for future curriculum reform. BMC Med Educ. 2023;23(1):556. doi:10.1186/s12909-023-04544-2
5. Khanagar SB, Alolayan LS, Alobaid TA, Alharbi AA, Alazaz NN, Alanazi MH. Use of YouTube as a learning modality for clinical procedures among dental students in Riyadh, Saudi Arabia-a cross-sectional study. Appl Sci. 2022;12(23):11977. doi:10.3390/app122311977
6. Shah SS, Jameel B, Ijaz W, Iqbal N, Adeeb H, Ahmad S. Effectiveness of YouTube as an educational tool for teaching orthodontics to undergraduate dental students. J Ayub Med Coll Abbottabad. 2024;36(4):716-20. doi:10.55519/JAMC-04-12740
7. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ. 2022;22(1):382. doi:10.1186/s12909-022-03446-z
8. Maganur PC, Hakami Z, Raghunath RG, Vundavalli S, Jeevanandan G, Almugla YM, et al. Reliability of educational content videos in YouTube about stainless steel crowns. Children (Basel). 2022;9(4):571. doi:10.3390/children9040571
9. Hakami Z, Maganur PC, Khanagar SB, Naik S, Alhakami K, Bawazeer OA, et al. Thumb-sucking habits and oral health: an analysis of YouTube content. Children (Basel). 2022;9(2):225. doi:10.3390/children9020225
10. Uzel İ, Ghabchi B, Akalın A, Eden E. YouTube as an information source in paediatric dentistry education: reliability and quality analysis. PLoS One. 2023;18(3):e0283300. doi:10.1371/journal.pone.0283300
11. Ozturk T, Gumus H. YouTube as an information and education source for early orthodontic treatment. Am J Orthod Dentofacial Orthop. 2022;162(3):e123-32. doi:10.1016/j.ajodo.2022.06.013
12. Jung MJ, Seo MS. Assessment of reliability and information quality of YouTube videos about root canal treatment after 2016. BMC Oral Health. 2022;22(1):494. doi:10.1186/s12903-022-02540-4
13. Cerci Akcay H, Kargu EC, Seker N. Assessing the reliability and educational value of YouTube videos on computer-controlled local anesthesia in dentistry. PLoS One. 2025;20(8):e0329291. doi:10.1371/journal.pone.0329291
14. Seabrook JA. Powering nutrition research: practical strategies for sample size in multiple regression. Nutrients. 2025;17(16):2668. doi:10.3390/nu17162668
15. 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 Dentofacial Orthop. 2021;161(1):e72-9. doi:10.1016/j.ajodo.2020.12.023
16. Jamleh A, Aljohani SM, Alzamil FF, Aljuhayyim SM, Alsubaei MN, Alali SR, et al. Assessment of the educational value of endodontic access cavity preparation YouTube video as a learning resource for students. PLoS One. 2022;17(8):e0272765. doi:10.1371/journal.pone.0272765
17. Alsulaiman AA, Alsulaiman OA, Alkhateeb RI, AlMuhaish L, Alghamdi M, Nassar EA, et al. Orthodontic elastics: a multivariable analysis of YouTube videos. Clin Cosmet Investig Dent. 2024;16:115-25. doi:10.2147/CCIDE.S460910
18. Aksoy M, Topsakal KG. YouTube for information on paediatric oral health instructions. Int J Dent Hyg. 2022;20(3):496-503. doi:10.1111/idh.12580
19. Yavan MA, Gökçe G. YouTube as a source of information on adult orthodontics: a video analysis study. J World Fed Orthod. 2022;11(1):41-6. doi:10.1016/j.ejwf.2021.09.001
20. Donnell CC, Bowers EJ, Monteiro J, Gilchrist F. Scrolls, screens and search bars: exploring family search behaviour and the quality of online information on molar incisor hypomineralisation. Eur Arch Paediatr Dent. 2026;27(2):1-12. doi:10.1007/s40368-026-01208-9