Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Content and educational quality of videos on ossification of the posterior longitudinal ligament across Bilibili, TikTok, and Kwai: a cross-sectional study

Xuerui Zhu
Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine
Shaoguang Zhou
Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine
Ruilong Chen
Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine
Weiqing Qian
qwq_68@163.com

Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine

Article

Keywords: ossification of the posterior longitudinal ligament, short videos, social media, patient education, information quality, China

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10961314/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

The educational value of online information about ossification of the posterior longitudinal ligament (OPLL) remains uncertain. We aimed to characterize and compare OPLL-related video content and educational quality across Bilibili, TikTok, and Kwai. On August 29, 2026, the first 100 results per platform were screened. Two researchers independently evaluated eligible videos using GQS, mDISCERN, JAMA, PEMAT-A/V, and a study-specific content checklist. Comparisons used nonparametric tests and Spearman correlations. Among 265 included videos (Bilibili, 68; TikTok, 99; Kwai, 98), 95.5% originated from professional healthcare sources. Median GQS, mDISCERN, JAMA, and OPLL-specific content scores were 3, 2, 2, and 8/15, respectively. Median understandability and actionability were 70.0% and 33.3%; 58.5% and 3.0% met the respective 70% thresholds. TikTok had the highest observed interaction counts and higher PEMAT scores than both comparators. Bilibili videos were longest and had lower JAMA scores; Kwai videos had been online longest. GQS did not differ across platforms. One high-volume tertiary spine center contributed 26.4% of videos, but excluding these videos did not materially alter the principal platform differences. Interaction-quality associations varied within platforms. These findings identify moderate educational quality but limited actionable guidance, supporting clearer sourcing and context-appropriate next steps without equating popularity with quality.

1 Introduction

Ossification of the posterior longitudinal ligament (OPLL) involves ectopic bone formation within a spinal ligament and may cause spinal-cord compression and myelopathy (1, 2). A Chinese hospital-based imaging study identified cervical OPLL in 4.1% of 2,000 individuals, although this selected cohort does not provide a population-wide prevalence estimate (1). Long-term follow-up indicates that neurological deterioration is not inevitable in patients without myelopathy at presentation (2). Conversely, a cohort of patients hospitalized with OPLL and managed without surgery showed an elevated subsequent risk of spinal cord injury (3). Patient information should therefore communicate potential neurological risks without implying that every radiographic diagnosis requires surgery.

OPLL education also requires distinctions between diagnosis, neurological impairment, and treatment selection. Computed tomography characterizes ossified lesions, whereas magnetic resonance imaging helps assess spinal-cord compression and intramedullary changes (4). Clinical consensus emphasizes that management depends on symptoms and imaging findings, with surgical selection informed by disease extent and individual circumstances (5). Accessible explanations should help viewers understand why a brief symptom description or an isolated imaging finding cannot determine an appropriate treatment pathway.

Video-sharing platforms provide accessible channels for discussing spinal disorders, but professional presentation and online popularity do not necessarily indicate educational quality. Studies of spine tumors and cervical spondylosis on YouTube have reported limited or variable educational quality (6, 7). Research on Chinese platforms has extended these assessments to cervical spondylosis, spinal cord

injury, and lumbar disc herniation (8, 9, 10). Across these studies, platform rankings and associations with popularity differed by condition and assessment domain. These findings cannot establish whether OPLL-related videos adequately address its variable course, diagnostic requirements, and individualized treatment considerations.

General quality ratings do not replace disease-specific content assessment or separate evaluation of understandability and actionability (11). This study aimed to characterize and compare OPLL-related information on Bilibili, TikTok, and Kwai across content coverage, quality, reliability, transparency, understandability, and actionability. We also examined uploader sources, presentation formats, and associations between observed interaction counts and educational-quality measures. The study assessed the online information environment rather than treatment efficacy or actual patient learning.

2 Materials and methods

2.1 Study design and ethical considerations

This descriptive cross-sectional content analysis examined publicly accessible videos about OPLL on Bilibili, TikTok, and Kwai. TikTok and Kwai refer to the Chinese domestic platforms Douyin and Kuaishou, respectively. The three platforms were selected because the study was designed around public, video-first search results and a common set of visible interaction indicators suitable for cross-platform comparison. Xiaohongshu was not included because its mixed image-text and video content structure would require a different sampling frame and coding scheme; its exclusion was therefore a scope decision rather than an indication that the platform is unimportant. Reporting followed the Strengthening the Reporting of Observational Studies in Epidemiology statement where applicable (12). The study did not recruit or interact with human participants, and no private or personally identifiable information was collected. Ethics committee approval and informed consent were therefore not applicable.

2.2 Search strategy and video selection

All searches were conducted on August 29, 2026, using the Chinese term “□□□□□□” (ossification of the posterior longitudinal ligament). A newly created account without previous searches, viewing history, follows, likes, or other interactions was used on each platform to reduce personalization bias. Results were retained in the default comprehensive ranking, and the first 100 results from each platform were screened. Sampling the first 100 algorithm-ranked videos is consistent with previous cross-platform studies of medical video content (13).

Videos were included when OPLL was the principal topic, the complete audiovisual content was publicly accessible and sufficiently clear for assessment, and the video contained substantive educational or clinical information. Videos were excluded if they were irrelevant or mentioned OPLL only superficially, duplicated or substantially overlapped another result, lacked substantive audiovisual content, were inaccessible or of insufficient audiovisual quality, or had been online for fewer than seven complete

days. The seven-day minimum reduced instability in interaction measurements for very recent uploads. Eligible videos were therefore published on or before August 22, 2026.

2.3 Data extraction and classification

On the retrieval date, two researchers independently recorded platform, video identifier, publication date, duration, likes, comments, favorites, shares, uploader category, presentation format, and all quality scores in a standardized spreadsheet. Days online were calculated as the interval from publication to August 29, 2026. View counts were not publicly and consistently available across the three platforms. Likes, comments, favorites, and shares were therefore analyzed as observed interaction counts rather than viewer-normalized engagement rates. Institutional affiliation was coded from explicit uploader-profile and video-level identifiers. Videos attributed to the same tertiary spine center were grouped as a single institutional source for the source-concentration sensitivity analysis; institution and account names were not reported.

Uploaders were classified as professional physicians, medical institutions, independent users, patients or family members, or other sources. Presentation formats were classified as clinician-patient question-and-answer sessions, scientific popularization, lectures, surgical demonstrations, patient experiences, or other formats. Sparse categories were retained for descriptive reporting. For prespecified subgroup analyses, professional physicians and medical institutions were combined as professional healthcare sources, and the two common presentation formats were compared directly.

2.4 Quality and educational-content assessment

The Global Quality Score (GQS) rated overall flow, completeness, usefulness, and educational value on a five-point ordinal scale (14). Scores of 1–2, 3, and 4–5 represented low, moderate, and high quality, respectively. The five-item modified DISCERN instrument assessed information reliability, with one point awarded for each criterion and total scores ranging from 0 to 5 (15, 16). The Journal of the American Medical Association (JAMA) benchmark criteria assessed authorship, attribution, disclosure, and currency, producing a total score from 0 to 4 (17).

A study-specific OPLL content score evaluated 15 prespecified information elements: definition, predilection sites, causes or risk factors, symptoms, spinal-cord manifestations, trauma-related risk, radiographic diagnosis, computed tomography, magnetic resonance imaging, non-surgical treatment, surgical indications, anterior and posterior strategies, individualized procedure selection, natural history, and indications for medical assessment. Each element was scored 0 or 1, yielding a total from 0 to 15. The score quantified disease-specific content coverage and was not treated as a validated measure of clinical accuracy. Detailed scoring criteria are provided in Supplementary Table S10.

The Patient Education Materials Assessment Tool for audiovisual materials (PEMAT-A/V) was used to assess understandability and actionability (11). Each domain score was calculated as the points obtained divided by the applicable points and multiplied by 100. Items rated not applicable were excluded from the denominator. Scores of at least 70% were classified as adequate for descriptive

comparisons; this analytical threshold was not considered a validated clinical cutoff (18). Full item-level criteria, response options, and scoring rules for GQS, mDISCERN, JAMA, PEMAT-A/V, and the OPLL-specific checklist are provided in Supplementary Tables S6–S10.

2.5 Rating procedure and inter-rater reliability

Two trained researchers independently screened and scored all eligible videos after standardized training and pilot calibration. Disagreements were reviewed with a third investigator and resolved by consensus. Reliability was calculated from the independent pre-consensus ratings for all 265 included videos. Quadratic-weighted Cohen's kappa was used for GQS, and two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficients [ICC(A,1), equivalent to ICC(2, 1)] were used for mDISCERN, JAMA, OPLL-specific content, PEMAT understandability, and PEMAT actionability scores (19). Ninety-five percent confidence intervals and exact percentage agreement for GQS were reported.

2.6 Statistical analysis

Analyses were performed using IBM SPSS Statistics version 32.0 (IBM Corp., Armonk, NY, United States) and R version 4.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Continuous and ordinal variables were summarized as medians and interquartile ranges (IQRs), and categorical variables as frequencies and percentages.

Platform differences in continuous or ordinal variables were assessed using Kruskal-Wallis tests. Significant overall tests were followed by Dunn pairwise comparisons with Bonferroni adjustment (20). Categorical distributions were compared using Monte Carlo contingency tests with 50,000 permutations. Professional healthcare sources were compared with all other sources using Mann-Whitney U tests. Presentation-format analyses were restricted to clinician-patient question-and-answer and scientific-popularization videos because the remaining categories were sparse.

Spearman rank correlations assessed associations among interaction indicators and between video duration, days online, interaction indicators, and quality scores. Correlations were examined in the overall sample and within each platform. The Benjamini-Hochberg procedure controlled the false-discovery rate within each correlation set (21). Raw interaction counts were the primary dissemination measures. A sensitivity analysis standardized interaction counts to 30 days by dividing the observed count by days online and multiplying by 30. These values standardized exposure time but were not interpreted as viewer-normalized engagement rates. A second sensitivity analysis repeated platform comparisons and within-platform correlations after excluding all 70 videos attributed to a single high-volume tertiary spine center, the most frequently represented institutional source. All tests were two-sided, and $P < 0.05$ was considered statistically significant.

3 Results

3.1 Video selection and overall characteristics

Of 300 screened records, 265 videos were included: 68 from Bilibili, 99 from TikTok, and 98 from Kwai. Bilibili excluded 30 irrelevant videos and two videos published fewer than seven days before retrieval. TikTok and Kwai excluded one and two recently published videos, respectively (Fig. 1).

The overall median duration was 91 seconds (IQR, 61–154), and the median time online was 245 days (IQR, 102–744). Professional physicians uploaded 246 videos (92.8%). Clinician-patient question-and-answer sessions (141/265, 53.2%) and scientific-popularization videos (116/265, 43.8%) were the predominant presentation formats.

A single high-volume tertiary spine center contributed 70 of 265 videos (26.4%). The center accounted for 24 of 68 Bilibili videos (35.3%), 38 of 99 TikTok videos (38.4%), and 8 of 98 Kwai videos (8.2%); the distribution differed across platforms ($\chi^2 = 26.85$, $P < 0.001$). Thirty-six of the 70 center-attributed videos addressed ACAF (51.4% of center videos; 13.6% of all videos).

The median GQS was 3 (IQR, 2–3), mDISCERN was 2 (IQR, 1–3), JAMA was 2 (IQR, 2–2), and OPLL-specific content score was 8 (IQR, 7–10). Median PEMAT understandability and actionability scores were 70.0% (IQR, 60.0–87.5) and 33.3% (IQR, 0–50.0), respectively. Overall, 155 videos (58.5%) met the PEMAT understandability threshold, whereas eight (3.0%) met the actionability threshold (Table 1).

Table 1
Overall video characteristics and quality scores

Characteristic	Overall sample
Videos included	265
Bilibili / TikTok / Kwai, n	68 / 99 / 98
Single high-volume tertiary spine center*	70 (26.4%)
Video duration, s	91 (61–154)
Days online	245 (102–744)
Likes	37 (6–191)
Comments	4 (0–20)
Favorites	8 (1–44)
Shares	4 (0–23)
Professional physicians	246 (92.8%)
Medical institutions	7 (2.6%)
Independent users	10 (3.8%)
Patients or family members	1 (0.4%)
Other uploaders	1 (0.4%)
Clinician-patient Q&A	141 (53.2%)
Scientific popularization	116 (43.8%)
Other presentation formats	8 (3.0%)
GQS	3 (2–3)
GQS quality: low / moderate / high	114 (43.0%) / 101 (38.1%) / 50 (18.9%)
mDISCERN	2 (1–3)
JAMA	2 (2–2)
OPLL-specific content score	8 (7–10)
PEMAT understandability, %	70 (60–87.5)
Adequate PEMAT understandability	155 (58.5%)
PEMAT actionability, %	33.3 (0–50)
Adequate PEMAT actionability	8 (3.0%)

Continuous and ordinal values are median (IQR); categorical values are n (%). Adequate PEMAT scores were defined as ≥ 70%. GQS, Global Quality Score; mDISCERN, modified DISCERN; JAMA, Journal of the American Medical Association benchmark criteria; Q&A, question and answer; OPLL, ossification of the posterior longitudinal ligament; PEMAT, Patient Education Materials Assessment Tool. *This overlapping source-concentration descriptor is not an additional uploader category.

3.2 Inter-rater reliability

Inter-rater reliability was good or excellent for all assessed scores. Quadratic-weighted kappa for GQS was 0.901 (95% CI, 0.868–0.930), with 84.0% exact agreement. ICC(A,1) values ranged from 0.752 for JAMA to 0.970 for PEMAT understandability (Supplementary Table S1).

3.3 Comparison of video characteristics and quality across platforms

Video duration and days online differed across platforms (both $P < 0.001$). Bilibili had the longest videos, whereas Kwai had the longest median time online. Likes, comments, favorites, and shares also differed across platforms (all $P < 0.001$). TikTok had the highest median value for every observed interaction measure, and each value exceeded the corresponding Bilibili and Kwai value after Bonferroni adjustment (Table 2).

GQS did not differ across platforms ($P = 0.295$). TikTok scored higher than Kwai for mDISCERN and the OPLL-specific content score. Bilibili had lower JAMA scores than both TikTok and Kwai. TikTok also had higher PEMAT understandability and actionability scores than the other two platforms. Adequate understandability was observed in 76.8% of TikTok videos, compared with 44.1% of Bilibili and 50.0% of Kwai videos ($P < 0.001$). Adequate actionability remained uncommon and did not differ significantly by platform ($P = 0.104$; Table 2 and Fig. 2).

Table 2
Comparison of video characteristics and quality across platforms

Variable	Overall	Bilibili	TikTok	Kwai	*P* value
Video duration, s	91 (61–154)	151.5 (81.8–240.8)	92 (63.5–152.5)	75 (48–105.2)	< 0.001
Days online	245 (102–744)	224.5 (64–906)	158 (73–451.5)	464 (192.2–855.5)	< 0.001
Likes	37 (6–191)	2 (1–8.2)	210 (113–408.5)	15 (7.2–48)	< 0.001
Comments	4 (0–20)	0 (0–1)	20 (9.5–60)	2 (1–5.8)	< 0.001
Favorites	8 (1–44)	2 (1–5.5)	41 (23.5–102)	2 (0–10.8)	< 0.001
Shares	4 (0–23)	0 (0–2)	24 (11–67)	1 (0–8.5)	< 0.001
GQS	3 (2–3)	2.5 (2–3)	3 (2–3)	3 (2–3)	0.295
mDISCERN	2 (1–3)	2 (1–3)	2 (1–3)	2 (1–2)	0.026
JAMA	2 (2–2)	2 (2–2)	2 (2–2)	2 (2–2)	< 0.001
OPLL-specific content score	8 (7–10)	8 (6.8–10)	9 (7–11)	8 (6–10)	0.003
PEMAT understandability, %	70 (60–87.5)	65.2 (55.6–75.7)	87.5 (70–90)	68.3 (60–80)	< 0.001
PEMAT actionability, %	33.3 (0–50)	33.3 (0–33.3)	33.3 (25–50)	33.3 (0–33.3)	< 0.001
GQS quality: low / moderate / high	114 / 101 / 50	34 / 21 / 13	38 / 39 / 22	42 / 41 / 15	0.433
Adequate PEMAT understandability	155 (58.5%)	30 (44.1%)	76 (76.8%)	49 (50.0%)	< 0.001
Adequate PEMAT actionability	8 (3.0%)	1 (1.5%)	6 (6.1%)	1 (1.0%)	0.104

Continuous and ordinal values are median (IQR); categorical values are n (%) or category counts. *P* values were obtained using Kruskal-Wallis or Monte Carlo contingency tests, as appropriate. Pairwise results are described in the text. Sample sizes: overall, 265; Bilibili, 68; TikTok, 99; Kwai, 98. GQS quality categories were low (1–2), moderate (3), and high (4–5); adequate PEMAT scores were ≥ 70%. GQS, Global Quality Score; mDISCERN, modified DISCERN; JAMA, Journal of the American Medical

Association benchmark criteria; OPLL, ossification of the posterior longitudinal ligament; PEMAT, Patient Education Materials Assessment Tool.

3.4 Uploader and presentation-format characteristics

Professional healthcare sources accounted for 253 videos (95.5%). The complete uploader distribution differed across platforms ($P < 0.001$), mainly because Bilibili contained more independent-user videos and Kwai contained more medical-institution videos. Presentation-format distributions did not differ significantly across platforms ($P = 0.101$; Fig. 3 and Supplementary Table S2).

Professional healthcare sources had higher JAMA, PEMAT understandability, and PEMAT actionability scores than other sources, while duration and observed interaction counts did not differ. Scientific-popularization videos had more favorites than clinician-patient question-and-answer videos ($P = 0.029$) and scored higher for GQS, mDISCERN, OPLL-specific content, and PEMAT understandability. The remaining interaction and quality measures did not differ between the two common formats (Table 3).

Table 3
Comparison by uploader source and common presentation format

Variable	Professional healthcare source (n = 253)	Other source (n = 12)	*P*	Clinician-patient Q&A (n = 141)	Scientific popularization (n = 116)	*P*
Video duration, s	91 (61–154)	102.5 (41.5–199.5)	0.920	85 (58–148)	98 (63–155.2)	0.291
Likes	37 (6–200)	24.5 (6.5–49.8)	0.244	27 (5–200)	50 (9–182.2)	0.288
Comments	4 (0–20)	1 (0–4.2)	0.107	3 (0–19)	4 (1–20)	0.725
Favorites	8 (1–44)	13.5 (5.5–33.8)	0.682	4 (1–39)	12 (2.8–51)	0.029
Shares	4 (0–25)	3.5 (1.5–7)	0.517	3 (0–25)	6 (1–20)	0.136
GQS	3 (2–3)	2 (1.8–3.2)	0.262	2 (2–3)	3 (2–4)	< 0.001
mDISCERN	2 (1–3)	1.5 (0–3)	0.368	2 (1–2)	2 (1–3)	0.002
JAMA	2 (2–2)	1 (1–1.2)	< 0.001	2 (2–2)	2 (2–2)	0.894
OPLL-specific content score	8 (7–10)	6.5 (4–10)	0.126	8 (7–10)	9 (7–11)	0.026
PEMAT understandability, %	70 (60–87.5)	60 (50–72.5)	0.035	66.7 (60–77.8)	80.9 (66.7–90)	< 0.001
PEMAT actionability, %	33.3 (0–50)	0 (0–0)	0.005	33.3 (0–33.3)	33.3 (0–50)	0.284

Values are median (IQR). *P* values were obtained using Mann-Whitney U tests. Professional healthcare sources comprise physicians and medical institutions; all remaining uploader categories were combined as other sources. Only the two common presentation formats were included in the format comparison. These exploratory subgroup *P* values are unadjusted. GQS, Global Quality Score; mDISCERN, modified DISCERN; JAMA, Journal of the American Medical Association benchmark criteria; OPLL, ossification of the posterior longitudinal ligament; PEMAT, Patient Education Materials Assessment Tool; Q&A, question and answer.

3.5 Associations between interaction indicators and video quality

The four observed interaction indicators were strongly intercorrelated (Spearman's rho = 0.81–0.91; all Benjamini-Hochberg-adjusted $P < 0.001$). Video duration was not associated with any interaction indicator after adjustment. Days online showed weak positive associations with favorites and shares, but not with likes or comments.

Observed interaction counts showed their strongest quality associations with PEMAT understandability. Correlation coefficients ranged from 0.40 for likes and comments to 0.44 for favorites and shares (all adjusted $P < 0.001$). Associations with the remaining quality measures were weaker, and no interaction indicator was associated with PEMAT actionability after adjustment. Platform-stratified analyses showed heterogeneous directions and magnitudes (Fig. 4 and Supplementary Table S3).

3.6 Sensitivity analyses

Excluding the 70 videos from the single high-volume tertiary spine center left 195 videos: 44 from Bilibili, 61 from TikTok, and 90 from Kwai. Platform differences remained significant for all four observed interaction indicators and for mDISCERN ($P = 0.020$), JAMA ($P < 0.001$), OPLL-specific content ($P = 0.001$), PEMAT understandability ($P < 0.001$), and PEMAT actionability ($P = 0.001$); GQS remained similar across platforms ($P = 0.546$). Although institutional concentration was evident, this sensitivity analysis did not materially alter the principal findings. Platform-stratified interaction-quality correlations remained heterogeneous (Supplementary Table S4).

After standardizing observed interactions to a 30-day period, platform differences remained significant for likes, comments, favorites, and shares (all $P < 0.001$), and TikTok retained the highest median value for every measure. Time-standardized interaction-quality correlations were again strongest for PEMAT understandability, although their direction and magnitude varied across platforms (Supplementary Table S5).

4 Discussion

This study identified an actionability gap as the principal educational finding in a predominantly professional sample of OPLL-related videos. Although professional healthcare sources produced 95.5% of included videos, median global quality was moderate, and only 3.0% met the prespecified actionability threshold, compared with 58.5% meeting the understandability threshold. Platform differences varied by assessment domain, while associations between interaction counts and quality were heterogeneous within platforms. These findings distinguish professional participation and visibility from the provision of usable patient education. Similar limitations were reported for spine-tumor videos despite the predominance of academic medical sources (6).

The difference between understandability and actionability was the clearest educational limitation. Median PEMAT scores were 70.0% and 33.3%, respectively, resembling the imbalance reported for stroke-related videos on Chinese platforms (22). A recent study of Meniere's disease videos on TikTok found that 91.6% were uploaded by physicians, yet mean PEMAT understandability was 73.73% and mean actionability was 51.00% (28). This comparison reinforces that professional authorship can coexist with incomplete next-step guidance. PEMAT evaluates features that support understanding and identification of actions; it does not establish clinical accuracy or actual changes in patient behavior (11, 18). Some clips may explain a condition without intending to guide action, so a low actionability score does not itself indicate incorrect advice. Nevertheless, patient-facing OPLL videos could more clearly describe when to seek assessment and how to discuss management options with clinicians. Such guidance should support individualized care rather than prescribe a universal treatment pathway. For example, explanations could connect progressive neurological symptoms with the need for clinical assessment and clarify the complementary purposes of computed tomography and magnetic resonance imaging (4, 5).

The median OPLL-specific content score of 8 out of 15 indicated partial coverage of the prespecified educational topics. Disease-specific checklists have also supplemented general quality tools in cervical spondylosis and lumbar disc herniation video research (7, 10). Our checklist distinguished topic coverage from presentation quality, but its items were equally weighted and the instrument has not undergone external validation. High inter-rater reliability demonstrates scoring reproducibility, not clinical validity. A focused video may appropriately cover only a subset of topics; therefore, a lower total score should not automatically be interpreted as misleading content.

TikTok showed higher understandability and actionability scores than the other platforms, but GQS did not differ significantly. TikTok also exceeded Kwai in reliability and disease-specific content coverage, whereas Bilibili had lower JAMA scores than both comparators. A recent cervical spondylosis study likewise reported platform-dependent differences, although its disease scope and comparator platform differed from ours (8). However, spinal cord injury videos on the same three platforms showed higher JAMA and GQS scores on Bilibili (9). A knee osteoarthritis study similarly identified different platform leaders for interaction counts and individual quality measures (23). Recent studies of ground-glass nodules, osteosarcoma, pediatric pneumonia, CPR, and Wilson disease also found condition- and platform-specific mixtures of uploader types, content depth, and quality (29, 30, 33–35). Our findings therefore describe different platform information ecologies in the retrieved samples, not intrinsic platform performance. Contributor mix, content selection, user behavior, and recommendation algorithms are plausible explanations, but they were not causally evaluated. The nonsignificant comparison of actionability-threshold proportions, despite differences in domain scores, also cautions against treating threshold classifications as interchangeable with score distributions.

Professional sources had higher transparency, understandability, and actionability scores, but not uniformly higher scores across all domains. Physicians specifically accounted for 92.8% of the sample, whereas videos from patients or family members accounted for only 0.4%. OPLL's reliance on CT or MRI

interpretation, neurological assessment, and complex surgical decision-making may help explain this professional dominance, but this is a speculative interpretation; platform algorithms and user posting or viewing behavior are alternative explanations. Professional dominance should not be equated with patient-centeredness. The scarcity of patient-created content may underrepresent concerns about decision-making, recovery, functional limitations, and lived experience. The professional-source pattern partly agrees with stroke-video research (22) and with recent studies showing that professional contributors often achieve higher information-quality scores (28, 29, 33, 35). However, our comparison involved 253 professional-source videos and only 12 from other sources, limiting precision and generalizability. Scientific-popularization videos scored higher than clinician-patient question-and-answer videos on several measures, but not actionability. Spinal cord injury research also found higher scores for disease-knowledge videos than for outpatient-scene content (9). One possible explanation is that general educational presentations provide more context than responses to an individual clinical question. The comparison remains observational and does not establish that changing presentation format would improve quality.

In the pooled sample, interaction counts were strongly correlated with one another and showed their strongest quality associations with understandability. They were not associated with actionability after multiplicity adjustment. Cervical spondylosis and knee osteoarthritis studies likewise found that popularity was not a consistent indicator of educational quality (7, 8, 23). In contrast, lumbar disc herniation research reported positive associations between likes and quality scores (10). That study selected videos by likes, underscoring the importance of considering sampling and measurement differences when interpreting cross-study contrasts. In our study, aggregate associations may partly reflect differences between platforms, because within-platform patterns varied. Moreover, counts depend on audience reach as well as viewers' propensity to interact. Without consistently available view counts, these components cannot be separated. Thirty-day standardization addresses time online only and assumes that accumulated interactions can be summarized as an average over that period. Its similar results do not establish comparability of audience reach or a uniform accumulation pattern.

Institutional concentration was also evident: a single high-volume tertiary spine center contributed 26.4% of all included videos. A scoliosis-exercise TikTok study similarly found that two institutional accounts contributed 83.7% of healthcare-organization videos (26). Of the 70 center-attributed videos in our sample, 36 (51.4%) addressed ACAF. This center was the original clinical development site for ACAF and has substantial clinical and academic experience in spine surgery (24), making sustained academic dissemination plausible; however, our design cannot determine promotional intent. ACAF is presented here only as contextual background: it is an emerging and increasingly investigated anterior surgical option, but comparative syntheses remain based predominantly on observational evidence (27, 31, 32). The present study neither compared surgical efficacy nor assessed the clinical correctness of individual videos. Procedure selection should remain individualized according to ossification extent, canal-occupying ratio, cervical alignment, K-line status, and other patient-specific factors (5, 31, 32). Excluding all 70 center-attributed videos did not materially alter the principal platform findings, although the analysis does not eliminate dependence among videos from other shared creators or institutions.

These findings favor improving the educational structure of professional content rather than merely increasing professional participation. Videos could state their intended audience, use plain language, display sources and dates, and distinguish general information from individualized advice. Concise action statements and clearly signposted explanations are consistent with PEMAT principles (11, 18). Their effectiveness should nevertheless be tested with viewers. Research on eHealth literacy distinguishes consumers' perceived information-seeking and appraisal skills from properties of the information itself (25). Future patient testing should therefore assess comprehension and appropriate next-step decisions alongside material scores, while accounting for differences in health literacy.

Several limitations should be considered. First, a single search term, retrieval date, and the first 100 ranked results per platform provide a snapshot rather than a representative platform-wide sample. New accounts reduce prior-history effects but cannot remove algorithmic selection, and the exclusion of Xiaohongshu limits generalizability to a mixed image-text/video ecosystem. Second, unequal platform sample sizes and sparse source categories limit some comparisons; nonsignificant findings should not be interpreted as equivalence. The very small patient/family subgroup also precluded a meaningful comparison of patient-generated and professional content. Third, multiple exploratory comparisons and unmodeled creator-level clustering warrant cautious interpretation despite the specified multiplicity corrections and sensitivity analyses. Finally, the study did not directly assess clinical accuracy, misinformation, patient comprehension, or health outcomes. Future work should combine guideline-based clinical-accuracy assessment with patient testing and external validation of the OPLL checklist. The 70% PEMAT threshold should remain an analytical convention rather than a validated clinical cutoff. The study also did not establish whether any observed video characteristic caused better educational outcomes.

5 Conclusion

OPLL-related videos on Bilibili, TikTok, and Kwai showed moderate median educational quality despite predominantly professional authorship. Features supporting understandability were substantially more common than explicit actionable guidance. Platform differences depended on the assessment domain, and interaction counts did not consistently identify better information within platforms. Clearer sourcing and context-appropriate next steps may strengthen patient-facing education, but clinical accuracy and actual patient benefit require separate evaluation.

Abbreviations

ACAF, anterior controllable antedisplacement and fusion; CI, confidence interval; CT, computed tomography; GQS, Global Quality Score; ICC, intraclass correlation coefficient; IQR, interquartile range; JAMA, Journal of the American Medical Association benchmark criteria; mDISCERN, modified DISCERN; MRI, magnetic resonance imaging; OPLL, ossification of the posterior longitudinal ligament; PEMAT-A, Patient Education Materials Assessment Tool actionability score; PEMAT-A/V, Patient Education Materials Assessment Tool for Audiovisual Materials; PEMAT-U, Patient Education Materials

Assessment Tool understandability score; Q&A, question and answer; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

Declarations

Ethics statement

This study analyzed publicly accessible online video content and did not recruit, interact with, or collect data from human participants. All individuals involved in video screening, rating, and adjudication were members of the study team and co-authors, rather than study participants. No external or non-author assessors were involved. No private information was accessed, and no usernames, account identifiers, or identifiable images are reproduced in the manuscript or Supplementary Information. Accordingly, ethics committee approval and informed consent were not applicable.

Author contributions

XZ and SZ independently screened and rated all eligible videos. RC adjudicated disagreements, and WQ supervised the study. XZ: Methodology, investigation, data curation, formal analysis, validation, and writing - original draft. SZ: Conceptualization, methodology, investigation, formal analysis, software, visualization, and writing - original draft. RC: Conceptualization, methodology, resources, validation, and writing - review and editing. WQ: Conceptualization, supervision, project administration, methodology, and writing - review and editing. All authors contributed to the article and approved the submitted version.

Data availability statement

The data analyzed in this study were derived from publicly accessible content on Bilibili, TikTok (Douyin), and Kwai (Kuaishou). Aggregated findings are provided in the article and Supplementary Information. The de-identified analytic dataset supporting the findings of this study is available from the corresponding author upon reasonable request. Links and archived records were retained for source verification. Copies of the original videos are not redistributed because they remain subject to platform availability and third-party rights.

Funding

This study was supported by the Nanjing Health Science and Technology Development Project of the Nanjing Municipal Health Commission, Jiangsu Province, China (Grant No. ZKX23050).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The authors thank the clinicians and science communicators who contribute to accessible public education on OPLL.

References

1. Liang, H. et al. Epidemiology of ossification of the spinal ligaments and associated factors in the Chinese population: a cross-sectional study of 2000 consecutive individuals. *BMC Musculoskelet. Disord*. **20**, 253. 10.1186/s12891-019-2569-1 (2019).
2. Matsunaga, S., Sakou, T., Taketomi, E. & Komiya, S. Clinical course of patients with ossification of the posterior longitudinal ligament: a minimum 10-year cohort study. *J. Neurosurg.* **100** (3 Suppl Spine), 245–248. 10.3171/spi.2004.100.3.0245 (2004).
3. Wu, J. C. et al. Conservatively treated ossification of the posterior longitudinal ligament increases the risk of spinal cord injury: a nationwide cohort study. *J. Neurotrauma*. **29**, 462–468. 10.1089/neu.2011.2095 (2012).
4. Chikuda, H. The essence of clinical practice guidelines for ossification of spinal ligaments, 2019: 3. Diagnosis of OPLL. *Spine Surg. Relat. Res.* **5**, 325–327. 10.22603/ssrr.2021-0118 (2021).
5. Sun, X. F. et al. Consensus statement on diagnosis and treatment of cervical ossification of posterior longitudinal ligament from Asia Pacific Spine Society (APSS) 2020. J Orthop Surg (Hong Kong). **28**:2309499020975213. (2020). 10.1177/2309499020975213
6. Richardson, M. A., Park, W., Bernstein, D. N. & Mesfin, A. Analysis of the quality, reliability, and educational content of YouTube videos concerning spine tumors. *Int. J. Spine Surg.* **16**, 278–282. 10.14444/8215 (2022).
7. Wang, H. et al. YouTube online videos as a source for patient education of cervical spondylosis—a reliability and quality analysis. *BMC Public. Health*. **23**, 1831. 10.1186/s12889-023-16495-w (2023).
8. Lin, J. M., Li, K. H. & Lin, C. The quality and reliability of short videos about cervical spondylosis on TikTok and Redbook: a cross-sectional study. *BMC Musculoskelet. Disord*. **27**, 693. 10.1186/s12891-026-10105-7 (2026).
9. Zhang, Y. et al. Quality assessment of spinal cord injury-related health information on short-form video platforms: cross-sectional content analysis of TikTok, Kwai, and BiliBili. *Digit. Health*. **11**, 20552076251374226. 10.1177/20552076251374226 (2025).
10. Wang, Y., Lian, S., Liao, L. & Wang, W. Assessment of information quality and reliability of short videos related to lumbar disc herniation on selected video platforms. *Digit. Health*. **11**, 20552076251393278. 10.1177/20552076251393278 (2025).
11. Shoemaker, S. J., Wolf, M. S. & Brach, C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ. Couns.* **96**, 395–403. 10.1016/j.pec.2014.05.027 (2014).

12. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457. 10.1016/S0140-6736(07)61602-X (2007).
13. Liu, Z. Y. et al. YouTube/Bilibili/TikTok videos as sources of medical information on laryngeal carcinoma: cross-sectional content analysis study. *BMC Public. Health*. **24**, 1594. 10.1186/s12889-024-19077-6 (2024).
14. Bernard, A. et al. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am. J. Gastroenterol.* **102**, 2070–2077. 10.1111/j.1572-0241.2007.01325.x (2007).
15. 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*. **53**, 105–111. 10.1136/jech.53.2.105 (1999).
16. Singh, A. G., Singh, S. & Singh, P. P. YouTube for information on rheumatoid arthritis—a wakeup call? *J. Rheumatol.* **39**, 899–903. 10.3899/jrheum.111114 (2012).
17. Silberg, W. M., Lundberg, G. D. & Musacchio, R. A. Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor—let the reader and viewer beware. *JAMA* **277**, 1244–1245. 10.1001/jama.1997.03540390074039 (1997).
18. Agency for Healthcare Research and Quality. The Patient Education Materials Assessment Tool (PEMAT) and user's guide. Rockville, MD: AHRQ. (2013). https://www.ahrq.gov/health-literacy/patient-education/pemat.html (accessed August 30, 2026).
19. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. 10.1016/j.jcm.2016.02.012 (2016).
20. Dunn, O. J. Multiple comparisons using rank sums. *Technometrics* **6**, 241–252. 10.1080/00401706.1964.10490181 (1964).
21. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R Stat. Soc. Ser. B Methodol.* **57**, 289–300. 10.1111/j.2517-6161.1995.tb02031.x (1995).
22. Nie, S. J. et al. Quality analysis of stroke-related videos on video platforms: cross-sectional study. *JMIR Form. Res.* **9**, e80458. 10.2196/80458 (2025).
23. Duan, X. et al. Quality and reliability of knee osteoarthritis-related information on short video platforms in China: a multi-method cross-sectional study. *BMC Public. Health*. **26**, 770. 10.1186/s12889-026-26455-9 (2026).
24. Sun, J. et al. Anterior controllable antidisplacement and fusion surgery for the treatment of multilevel severe ossification of the posterior longitudinal ligament with myelopathy: preliminary clinical results of a novel technique. *Eur. Spine J.* **27**, 1469–1478. 10.1007/s00586-017-5437-4 (2018).
25. Norman, C. D. & Skinner, H. A. eHEALS: the eHealth Literacy Scale. *J. Med. Internet Res.* **8**, e27. 10.2196/jmir.8.4.e27 (2006).

26. Jang, C. W., Kim, M., Kang, S. W. & Cho, H. E. Reliability, quality, and educational suitability of TikTok videos as a source of information about scoliosis exercises: a cross-sectional study. *Healthc. (Basel)*. **10**, 1622. 10.3390/healthcare10091622 (2022).
27. Li, S. et al. Comparison of the surgeries for the ossification of the posterior longitudinal ligament-related cervical spondylosis: a PRISMA-compliant network meta-analysis and literature review. *Med. (Baltim).* **100**, e24900. 10.1097/MD.0000000000024900 (2021).
28. Wang, X., Lian, D. & Liu, Z. Physician-dominated yet suboptimal: evaluating the quality of Meniere's disease information on TikTok in China. *Digit. Health*. **12**, 20552076261418919. 10.1177/20552076261418919 (2026).
29. Qiu, X. et al. Quality and reliability of ground-glass nodule-related short Chinese videos on TikTok and Bilibili: a cross-sectional content analysis. *J. Thorac. Dis.* **18**, 138. 10.21037/jtd-2025-aw-2252 (2026).
30. Chen Jv, Luo, Z. et al. The quality and reliability of osteosarcoma information on TikTok and Bilibili: a cross-sectional study. *Digit. Health*. **12**, 20552076261433814. 10.1177/20552076261433814 (2026).
31. Wang, M. et al. Anterior cervical controllable antedisplacement and fusion versus anterior cervical corpectomy and fusion for ossification of the cervical posterior longitudinal ligament in the Chinese population: a systematic review and meta-analysis. *Neurosurg. Rev.* **47**, 783. 10.1007/s10143-024-02977-x (2024).
32. Wang, W., Xiang, C. H., Li, D., Wang, X. Z. & Xu, X. H. Comparative efficacy and safety of ACAF, ACCF, and laminoplasty for multilevel cervical OPLL: a network meta-analysis of observational studies. *J. Orthop. Surg. (Hong Kong)*. **34**, 10225536261432365. 10.1177/10225536261432365 (2026).
33. He, F. et al. Quality and reliability of pediatric pneumonia related short videos on mainstream platforms: cross-sectional study. *BMC Public. Health*. **25**, 1896. 10.1186/s12889-025-22963-2 (2025).
34. Pan, C. et al. Professional content analysis and quality assessment of cardiopulmonary resuscitation educational videos on social media platforms: a comparative study of YouTube, Bilibili, and TikTok. *Front. Public. Health*. **13**, 1657233. 10.3389/fpubh.2025.1657233 (2025).
35. Xu, Z. et al. A cross-sectional analysis of the quality and reliability of Wilson disease videos on Bilibili, Douyin, and Kuaishou. *Sci. Rep.* **16**, 16314. 10.1038/s41598-026-47222-1 (2026).

Figures

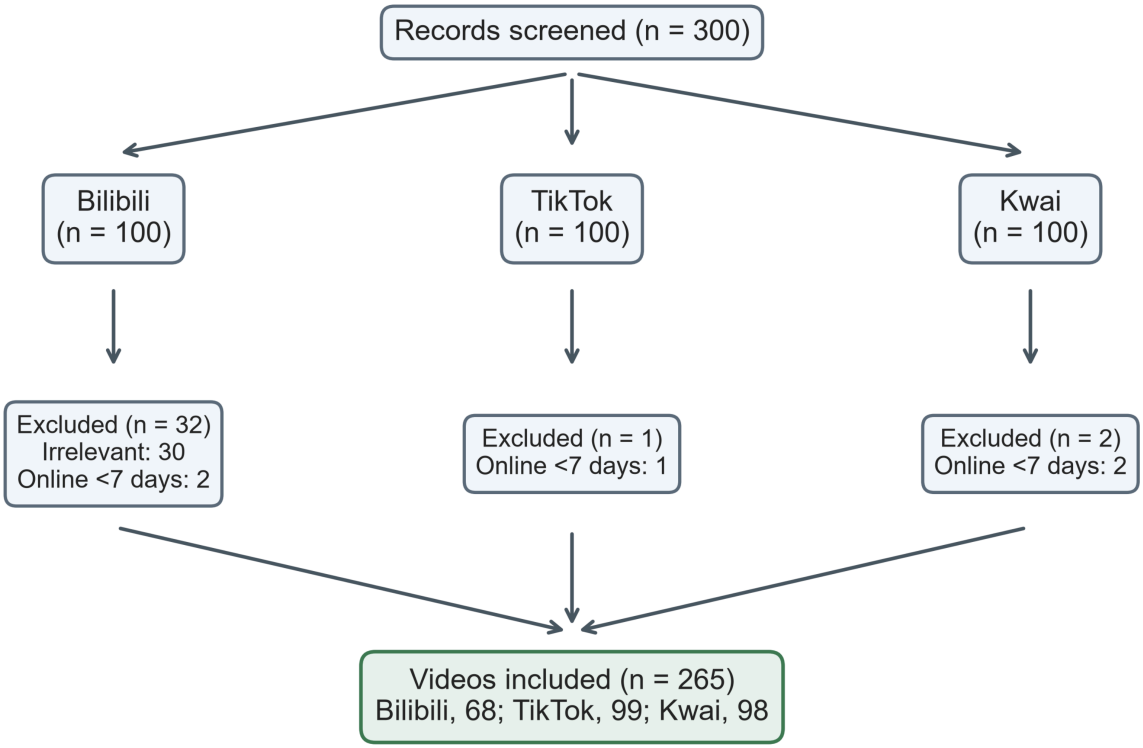


Figure 1

Video identification and inclusion. The first 100 records returned for the Chinese term "□□□□□□" were screened on each platform on August 29, 2026. Bilibili exclusions comprised 30 irrelevant videos and two videos online for fewer than seven complete days. TikTok and Kwai excluded one and two videos, respectively, for the same publication-age criterion. The final sample comprised 265 videos: Bilibili, n = 68; TikTok, n = 99; and Kwai, n = 98.

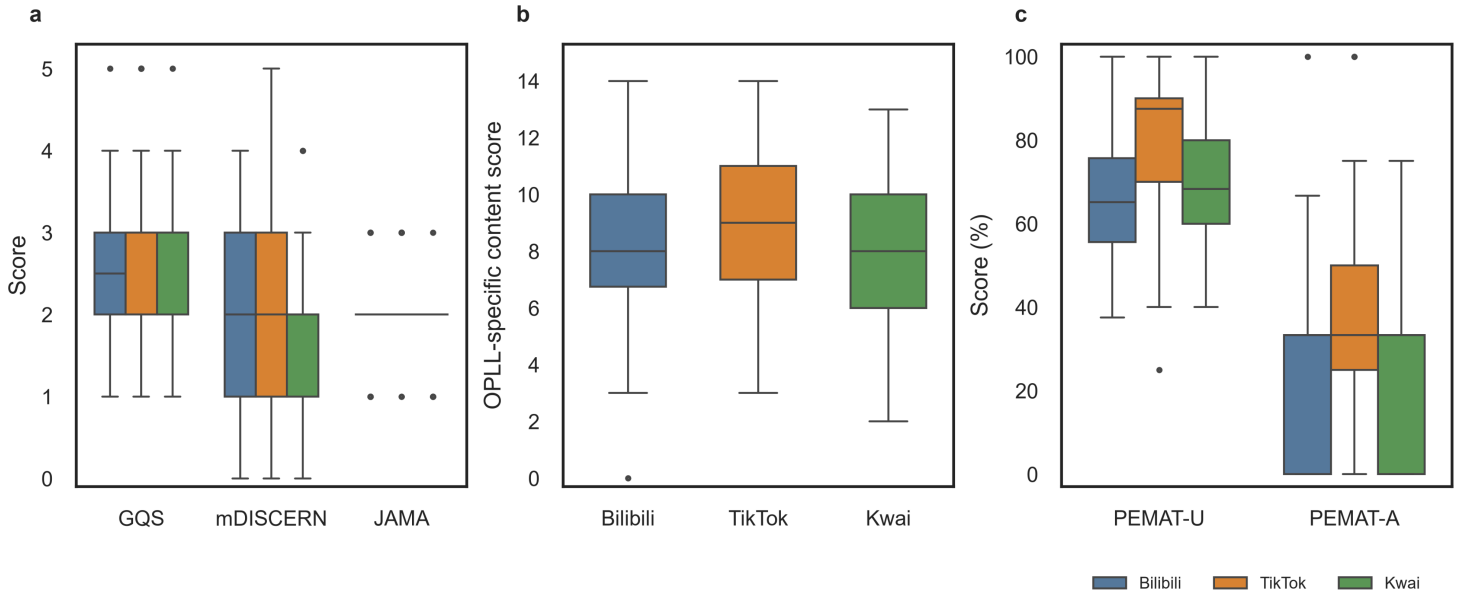


Figure 2

Quality, reliability, disease-specific content, and patient-education scores across platforms. a, GQS, mDISCERN, and JAMA scores. b, OPLL-specific content scores. c, PEMAT understandability and actionability scores. Boxes represent the interquartile range, central lines the median, and whiskers the most extreme observations within 1.5 interquartile ranges of each quartile. Points beyond the whiskers represent outliers. Each observation is one video (Bilibili, n = 68; TikTok, n = 99; Kwai, n = 98). Platform comparisons used Kruskal–Wallis tests, followed where significant by Dunn tests with Bonferroni adjustment; overall *P* values are reported in Table 2. GQS, Global Quality Score; mDISCERN, modified DISCERN; JAMA, Journal of the American Medical Association benchmark criteria; OPLL, ossification of the posterior longitudinal ligament; PEMAT, Patient Education Materials Assessment Tool.

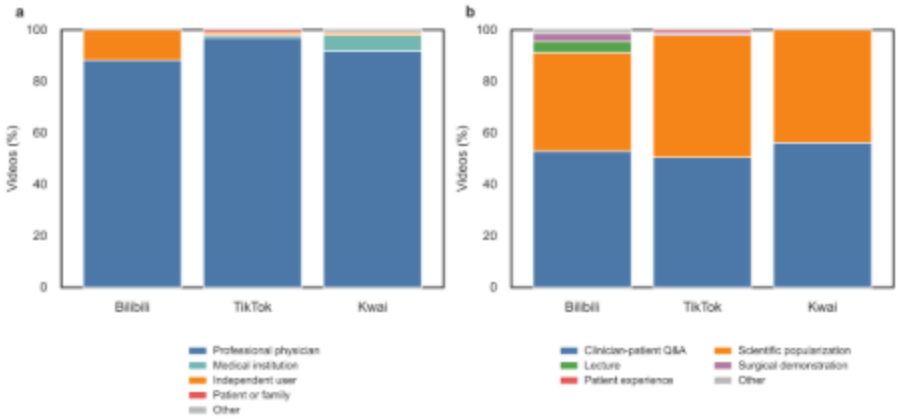


Figure 3

Uploader and presentation-format distributions across platforms. a, Proportions of videos in each uploader category. b, Proportions in each presentation-format category. Each stacked bar includes all eligible videos from that platform (Bilibili, n = 68; TikTok, n = 99; Kwai, n = 98). Categories within each panel are mutually exclusive. The uploader and presentation-format distributions were compared using Monte Carlo contingency tests with 50,000 permutations ($P < 0.001$ and $P = 0.101$, respectively). Detailed counts are provided in Supplementary Table S2. Q&A, question and answer.

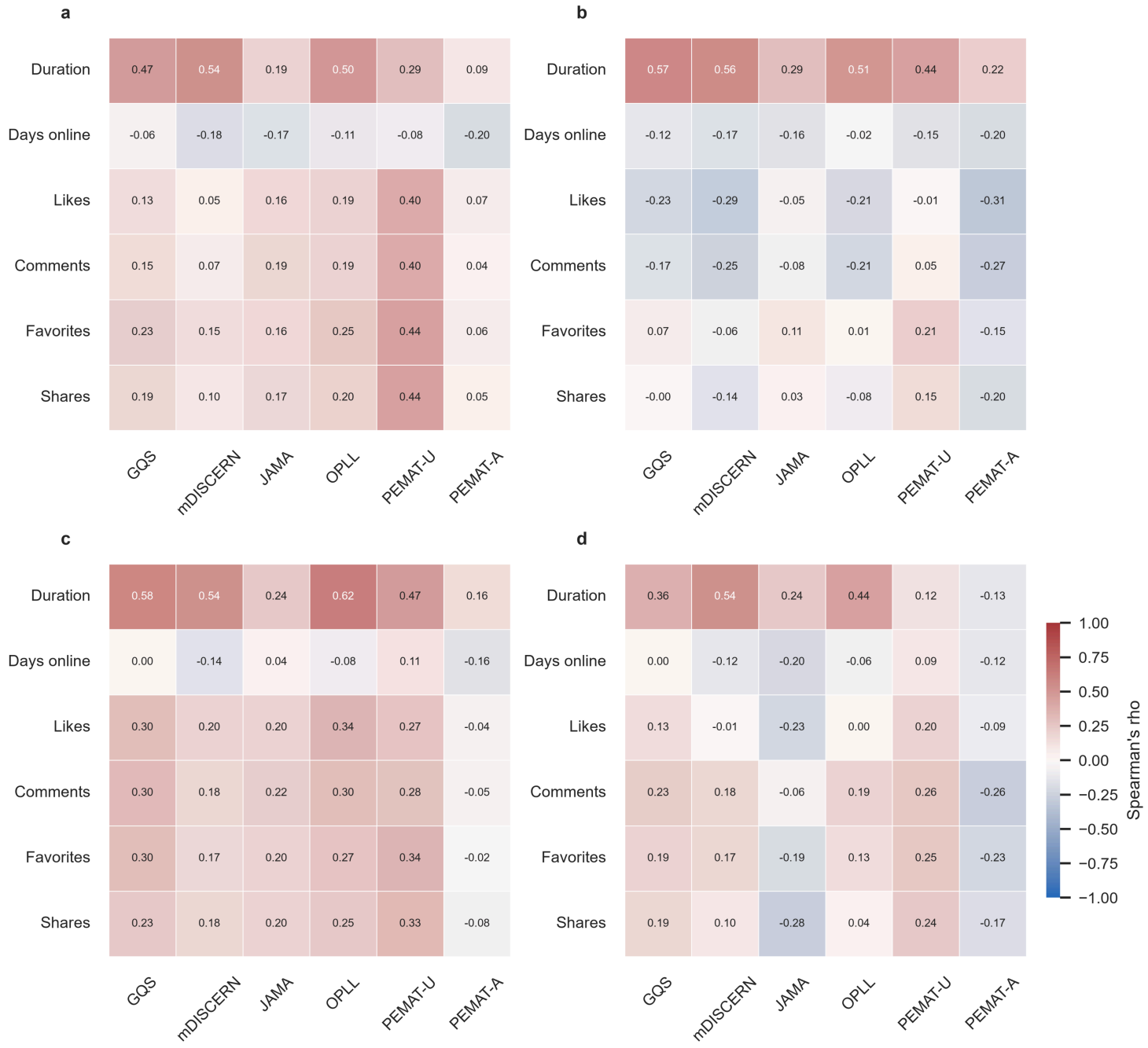


Figure 4

Associations between video characteristics, observed interaction counts, and quality measures. Spearman correlation coefficients are shown for a, the overall sample (n = 265); b, TikTok (n = 99); c, Kwai (n = 98); and d, Bilibili (n = 68). Rows represent video duration, days online, and observed interaction counts; columns represent the six quality and content scores. Red indicates positive and blue inverse associations on a common scale from −1 to 1. Numbers are correlation coefficients, not significance markers. Benjamini–Hochberg-adjusted *P* values are provided in Supplementary Table S3. GQS, Global Quality Score; mDISCERN, modified DISCERN; JAMA, Journal of the American Medical Association benchmark criteria; OPLL, OPLL-specific content score; PEMAT-U and PEMAT-A, Patient Education Materials Assessment Tool understandability and actionability scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OPLLSupplementaryTablesS1S5.docx
- OPLLSupplementaryScoringCriteriaS6S10.docx