BMC Oral Health

RESEARCH

Open Access

Quality and reliability evaluation of YouTube® exercises content for temporomandibular disorders

Halime Arikan[1*] and Erkan Erol[1]

Abstract

Objectives YouTube is a popular source for health-related content, including exercises for temporomandibular disorders (TMD). This study aimed to evaluate the content and quality of YouTube videos on TMD exercises.

Methods A systematic search on YouTube using keywords related to TMD exercises and rehabilitation identified relevant videos. Two researchers independently assessed these videos. Content was evaluated using the Video Information and Quality Index (VIQI), reliability with modified DISCERN, and quality with the Global Quality Scale (GQS).

Results One hundred twenty one exercise videos were included in the study. Of these, 47 (38.3%) were helpful, and 89 (73.6%) were uploaded by healthcare professionals. Useful and misleading videos had similar metrics for views, likes, comments, and interaction ($p > 0.05$), but useful videos had longer durations and older upload dates ($p < 0.05$). Cohen's Kappa scores showed substantial agreement between raters (mDISCERN = 0.709, GQS = 0.753). The Intraclass Correlation Coefficient for VIQI was 0.907. Significant differences were found in mDISCERN, GQS, and VIQI scores based on the source and usefulness of videos ($p < 0.05$), with high correlations among the instruments ($r = 0.740$ to 0.909). The area under the curve for the instruments ranged from 0.785 to 0.876.

Conclusions The majority of YouTube videos on TMD exercises were deemed useful, primarily uploaded by healthcare professionals, and contained reasonably safe, accurate, and quality information. Although this may seem like a positive outcome, unfortunately there was no explanation of TMD diagnoses in the videos. Therefore, it was unclear which types of TMD pathologies the videos were addressing.

Keywords Exercise, Rehabilitation, Physical therapy, Temporomandibular disorders, Video, YouTube

Background

Temporomandibular disorders (TMD) encompass a range of conditions involving intra-articular disc-related pathologies, structural abnormalities, and dysfunction within the associated muscular system [1]. Clinical manifestations commonly entail orofacial pain, joint noises, movement limitations or deviations, and cranial and/or myalgia symptoms [2]. TMD are the most prevalent pain conditions in the oral and facial regions. According to a 2024 meta-analysis, they affect 34% of the global population, with regional prevalence rates of 26% in North America, 47% in South America, 33% in Asia, and 29% in Europe. Studies also reveal that TMDs are more common in women, with the latest meta-analysis reporting female-to-male prevalence ratios of 1.26 in North America, 1.56 in South America, 1.26 in Asia, and 1.09 in Europe [3].

While many patients recover with simple measures like jaw rest and a soft diet, others require professional care, including occlusal splint therapy, physiotherapy, medications, or even surgical interventions. Procedures like

*Correspondence:
Halime Arikan
halimearikan92@gmail.com
[1] Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Tokat Gaziosmanpaşa University, Tokat 60250, Turkey

temporomandibular joint arthrocentesis and arthroscopy effectively manage "stuck" joints by lubricating the superior joint space, introducing medicaments like steroids or hyaluronic acid, and enabling disc mobilization [4]. Given the uncertain etiology of TMD, conservative approaches are typically favored as the initial treatment option [5]. Physical agents, manual techniques, self-management programs, cognitive behavioral therapy, pain neuroscience education, patient education, relaxation techniques, intraoral appliances, splints, mobilization, acupuncture, dry needling, electrophysical agents (such as iontophoresis, ultrasound, transcutaneous electrical nerve stimulation (TENS), and high voltage electrical stimulation (HVES)), and therapeutic exercises (including coordination, stretching, and postural exercises) are commonly utilized physiotherapy interventions [6–8]. Exercise is a physiotherapeutic approach employed to enhance strength, coordination, mobility, and alleviate muscular and articular pain [9]. Stretching exercises, muscle-strengthening exercises, jaw-opening exercises, mobilization exercises, and postural exercises are regarded as preferred modalities for managing TMD [5]. Numerous studies investigating the effects of exercise interventions on TMD can be found in the literature [10, 11]. Furthermore, exercise-related content for TMD is also increasingly prevalent on online platforms such as the internet, social media, and YouTube.

YouTube is a widely-used social media platform and ranks as the second most popular website globally [12]. Individuals affected by TMD often resort to the internet to understand the origins of their condition, assess the necessity of medical consultation, explore treatment options, and gather information about medical and home remedies. Given the prevalence of oversimplified diagnostic and therapeutic content regarding TMD, coupled with the unregulated nature of YouTube, it is inevitable that numerous videos containing misinformation exist on the internet [13]. Health misinformation is prevalent on YouTube; however, content produced by governmental and professional organizations typically offers high-quality, reliable information to the public [14–16].

Medical and dental professionals recognize YouTube's significance as a source of information for patients and the public. Numerous studies on TMD pathologies have assessed the quality of available materials [17]. According to the authors' research, no previous study has analyzed the quality and reliability of YouTube videos specifically focusing on exercises for TMD. The aim of this study was to evaluate the content, information delivery, and overall quality of YouTube videos pertaining to TMD exercises. The primary objective was to examine the exercise content presented in the videos without focusing on the subcategories of TMD. This approach was intended to reflect the general information available to the public and patients searching for TMD exercise guidance online. The study did not delve into the detailed kinesiology or biomechanics of specific exercises but aimed to assess the overall accuracy, comprehensibility, and potential implications of the information as presented to a general audience.

Methods

This study is a cross-sectional analysis. A cross-sectional study design was chosen, particularly given the dynamic nature of YouTube content. Recognizing the potential for changes in content quality over time will help us understand why this approach was adopted. All YouTube videos recommending exercises for TMJ dysfunctions that were accessible during the systematic search conducted between July 16–25, 2024, were examined. Ethics committee approval was not required because the study did not involve patients and only evaluated YouTube videos.

Search strategy

YouTube was systematically searched using the keywords 'temporomandibular disorders exercise,' 'temporomandibular disorders rehabilitation,' 'temporomandibular disorders physical therapy,' 'temporomandibular disorders physiotherapy,' 'temporomandibular joint exercise,' 'temporomandibular joint rehabilitation,' 'temporomandibular joint physical therapy,' and 'temporomandibular joint physiotherapy.'

We selected 'sort by relevance' as the sorting option. English-language videos recommending exercises for TMJ dysfunctions were included in the study. Duplicate videos, non-English videos, videos without sound, videos that did not mention TMJ or jaw in the title or description, videos without exercises, promotional videos, and irrelevant videos were excluded. All remaining videos that appeared in the search results were watched.

Assessing the videos

The videos identified were assessed independently by two researchers. The Modified DISCERN (mDISCERN) tool was employed to assess reliability, while the Global Quality Scale (GQS) and the Video Information and Quality Index (VIQI) were utilized to evaluate quality. These categorization criteria for evaluating video content were developed based on established evidence and have been cross-verified to align with current clinical guidelines. This alignment ensures that the categorization process reflects contemporary clinical practices and supports the credibility and relevance of the assessments. Examples illustrating the application of these criteria have been provided to enhance transparency and context. The duration of the videos, number of views, likes, comments, upload date, and country of origin were recorded. Viewing rate and interaction index were calculated with the following formulas [18]:

$$\text{Viewing rate (\%)} = \frac{\text{Number of views}}{\text{Number of days since upload}} \times 100$$

$$\text{Interaction index (\%)} = \frac{\text{Number of likes} + \text{Number of comments}}{\text{Number of views}} \times 100$$

Videos were categorized based on their upload sources, including healthcare professionals, individual users, and trainers. The category of 'health professionals' includes individuals such as physiotherapists, physicians, dentists, and other licensed healthcare providers, while 'trainers' refers to individuals who specialize in fitness or exercise coaching without formal medical or healthcare training. This distinction ensures a clearer classification based on professional background and expertise. The usefulness of the video was scored as (1) misleading and potentially harmful, (2) misleading but not harmful, or (3) useful. The term 'usefulness' refers specifically to the clarity, comprehensibility, and informational quality of YouTube videos as they relate to viewer understanding. It does not imply or assess the clinical efficacy or validity of the treatments discussed within the videos. The purpose of evaluating 'usefulness' is to determine how well the content facilitates viewers' comprehension rather than the effectiveness of the health interventions themselves. If the exercises in a video contain clinically incorrect information along with incorrect explanations and applications, they are classified as misleading and potentially harmful. Misleading but not harmful videos include exercise content that lacks substantiated evidence based on current scientific knowledge. Videos considered useful contain clinically accurate exercise information with appropriate explanations and applications relevant to the disease [15]. In this study, the classification of videos as useful, harmful, or misleading was based on their overall clinical accuracy and applicability to TMD as a general condition. While we acknowledge that certain exercises may have different implications for specific subtypes of TMD, the objective of this research was to assess the content's suitability and potential impact from a broader perspective. This approach was taken to reflect the general public's experience when searching for TMD-related exercises online, without focusing on the specific needs or contraindications associated with individual TMD subtypes. While making this assessment, even though a specific search and examination was not made for each TMD subtypes, situations that could be misleading or harmful were taken into consideration. For example, in cases of unilateral intra-articular TMD (such as disc displacement without reduction), stretching exercises should be recommended with caution and with careful monitoring of symmetrical mouth opening. Since the body functions as a chain, exercises performed in proper posture or position can positively support symmetrical mouth opening [8]. At this point, if the description and application of the exercise could be misleading or harmful for different each TMD types, its usefulness was evaluated accordingly.

Modified DISCERN (mDISCERN): mDISCERN is a modification of the original DISCERN scale developed to evaluate videos. It rates five different parameters: 1. Are the videos clear, concise and understandable?; 2. Are valid sources cited? (From valid scientific studies, dentists, physiotherapists); 3. Is the information presented balanced and unbiased?; 4. Are additional sources of information listed for patient reference?; and 5. Does the video touch on controversial/ambiguous areas? (in summary clarity, reliability, bias/balance, provision of additional sources of information, and recognition of areas of uncertainty). Each of these 5 items is scored separately as present (score 1) or absent (score 0). The total score can range from 0 (poor quality) to 5 (excellent quality) [19].

Global Quality Scale (GQS): GQS is a 5-point Likert scale that evaluates the quality, flow, and ease of use of videos. Scores of 1 and 2 correspond to low quality, 3 to medium quality, and 4 to 5 to high quality. The scores mean: 1 = Bad quality, poor flow in video, missing a lot of information, no benefit for patients; 2 = Overall poor quality and poor flow, some information available but missing many important points, very limited use for patients; 3 = Fair quality, poor flow, some important information adequately discussed but others poorly discussed, somewhat useful for patients; 4 = Good quality and overall good flow. Most relevant information is available but some topics are not covered, useful for patients; and 5 = Excellent quality and flow, very helpful for patients. A higher score indicates a higher level of video quality [20].

Video Information and Quality Index (VIQI): VIQI was used to analyze the quality of the videos. It is a 5-point Likert scale that evaluates various aspects of the videos, including information flow, information accuracy, quality (one point each for the use of still images, animation, community interviews, video subtitles, and a report summary), and precision (the alignment between the video title and the content). Each of these parameters is scored from 1 to 5, and the total score is obtained by summing the scores of these four parameters [21].

The mDISCERN, GQS, and VIQI tools used in the study are based on Likert scales, which may allow for subjective evaluations. However, to minimize the potential impact of this, training and calibration sessions were conducted for the raters during the evaluation process. In these sessions, a common understanding was established among the researchers, and joint evaluations were performed on sample videos to ensure consistent scoring. Following this, the two researchers independently watched and scored the videos, remaining blind to each other's scores. Calibration sessions were conducted to align the raters' understanding and application of the evaluation criteria. These sessions were held daily for one hour over a week period prior to the formal data collection. During these sessions, raters discussed and practiced scoring with sample videos, refining their consistency and resolving discrepancies to ensure reliable assessments.

Statistical analysis

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS), version 22.0 computer software package for Windows. Kolmogorov Smirnov test was performed to show the parametric or nonparametric distribution of the data. To determine statistically significant differences between more than two groups of an independent variable, the Kruskal–Wallis test was used. Non-normally distributed variables were represented as medians (25th-75th interquartile range (IQR)) and compared using the Mann–Whitney U test.

The Kappa coefficient was utilized to assess inter-rater agreements of mDISCERN and GQS. Cohen's kappa coefficient values of ≤ 0, 0.01–0.2, 0.21–0.4, 0.41–0.6, 0.61–0.8, and 0.81–1 correspond to no agreement, slight, fair, moderate, substantial, and almost perfect agreement, respectively [22]. Addition, the Intraclass Correlation Coefficient (ICC) was calculated to assess inter-rater reliability of VIQI. ICC values of 0.80 and above are considered high [23].

The relationship among mDISCERN, GQS, and VIQI scores was determined using Spearman correlation analysis. A correlation coefficient between 0.00 and 0.30 indicates a negligible correlation, 0.30 to 0.50 indicates a low correlation, 0.50 to 0.70 indicates a moderate correlation, 0.70 to 0.90 indicates a high correlation, and 0.90 to 1.00 indicates a very high correlation [24].

Receiver operating characteristic (ROC) curve analysis was used to calculate cut-off values of mDISCERN, GQS, and VIQI instruments for qualified YouTube videos according to usefulness. For an instrument to be considered meaningful in measurement, the Area Under the Curve (AUC) should be greater than 0.5. If the AUC is between 0.50 and 0.60, the instrument is considered to fail; between 0.60 and 0.70, the instrument is considered weak; between 0.70 and 0.80, the instrument is considered moderate; between 0.80 and 0.90, the instrument is considered good; and between 0.90 and 1.00, the instrument is considered excellent [25].

A significance level of 0.05 was used.

Results

General characteristics of YouTube videos

A total of 553 videos were identified using the search keywords, of which 432 were excluded for various reasons, resulting in 121 videos being included in the study for analysis (Fig. 1). A significant proportion of the uploaded videos, approximately 66.2%, originated from the United States (Fig. 2). While exercise videos were included in the study, many also featured manual techniques such as massage and myofascial release, rather than solely focusing on pure exercise routines. Regarding the exercises, the videos covered active, isometric, strengthening, stretching, and eccentric exercises targeting the TMJ. Additionally, exercises like chin tuck and posture correction exercises were commonly featured. The videos also included tongue exercises such as sticking out the tongue, proper tongue posture, tongue slide, tongue curl, tongue push to the roof, tongue clock, and tongue strengthening exercises. Furthermore, exercises like jaw circular motion, puffing, monkey, paddle pop stick, reciprocal inhibition, half moon, jaw wiggle, roof scraping, and open books were also presented. Despite their frequent mention in the literature, the Rocabado 6×6 exercises were present in only two videos.

The descriptive information regarding the characteristics, reliability, and quality scores, source of upload and usefulness of the videos is presented in Table 1.

The mDISCERN, GQS, and VIQI scores of the videos, reviewed by rater 1 and rater 2, were compared based on the source of upload. For both rater 1 ($p=0.000$, $p=0.000$, and $p=0.001$, respectively) and rater 2 ($p=0.005$, $p=0.000$, and $p=0.000$, respectively), healthcare professionals scored higher than individual users across all three scales. Additionally, in rater 1's GQS evaluation, trainers also scored higher than individual users ($p=0.037$) (Table 2).

When analyzing video duration, time since upload, number of views, number of likes, number of comments, number of subscribers, viewing rate, and interaction index according to usefulness, it was observed that misleading, not harmful videos had a longer video duration ($p=0.019$) and misleading, potentially harmful videos were uploaded earlier ($p=0.004$) compared to misleading, not harmful videos. Additionally, useful videos had a longer duration compared to misleading, potentially harmful videos ($p=0.001$) (Table 3). When considering the mDISCERN, GQS, and VIQI scores, the reliability and quality scores of useful videos were significantly higher than those of misleading, not harmful,

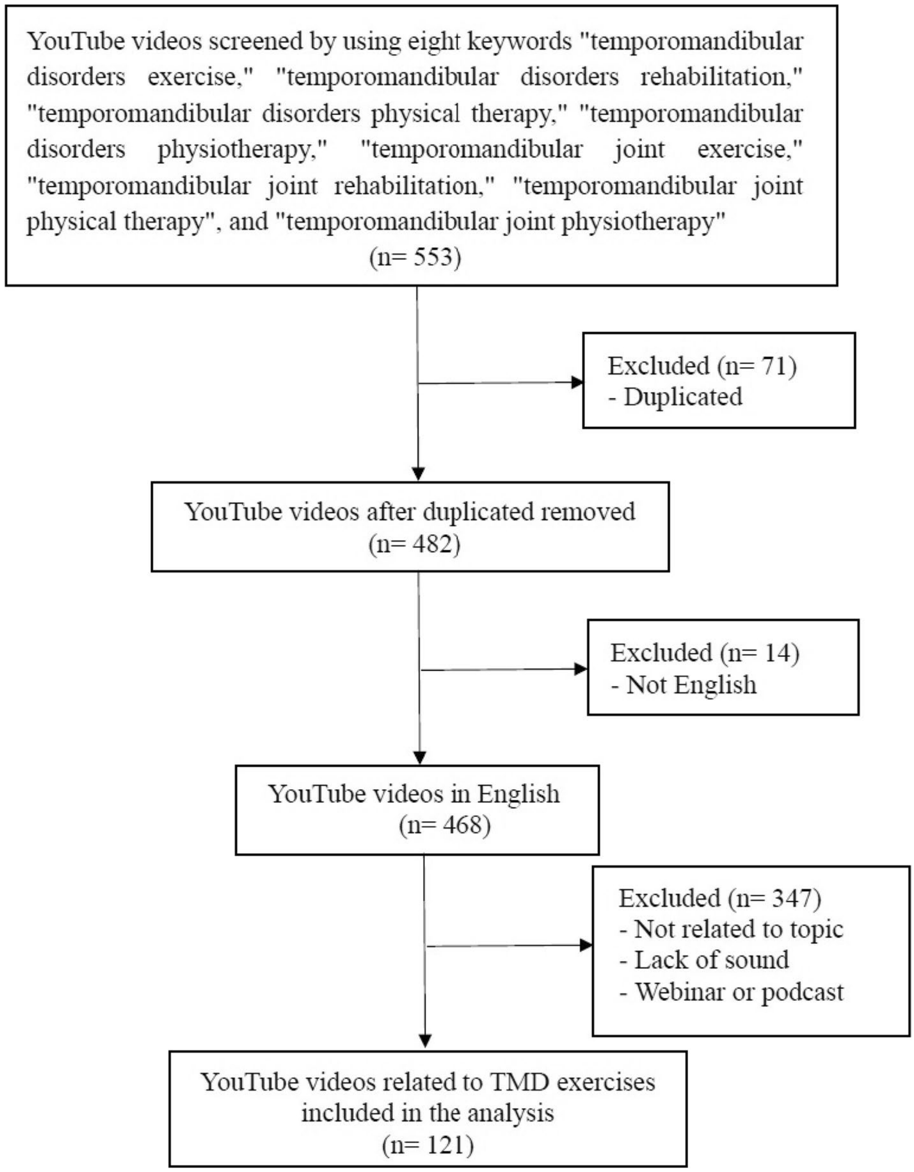


Fig. 1 Flow chart diagram of the included and excluded videos

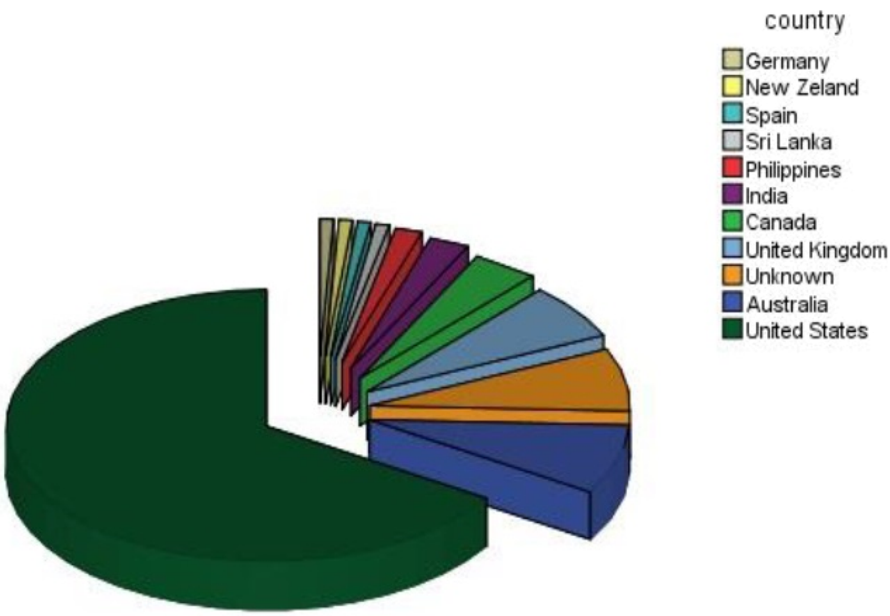
and misleading, potentially harmful videos ($p=0.000$). The mDISCERN rater 1 ($p=0.018$), mDISCERN rater 2 ($p=0.004$), and GQS rater 1 ($p=0.034$) scores of misleading, not harmful videos were higher than those of misleading, potentially harmful videos (Table 3).

Content and quality assessment of videos and correlation among evaluation scores of raters

The Cohen's kappa coefficient, which shows the inter-rater agreement, was calculated at 0.709 for the mDISCERN, and 0.753 for the GQS. Another measure of inter-rater reliability, the Intraclass Correlation Coefficient (ICC), was calculated for VIQI, resulting in a value of 0.907. Additionally, to emphasize the consistency of the observers and the reliability and validity of the evaluation criteria used, the relationships among these scales were also examined. All scales showed significant correlations with each other, with correlation coefficients ranging from 0.740 to 0.909, indicating very good to excellent agreement (Table 4).

Effectiveness of evaluation scores to predict high-quality videos

As in Table 5 and Fig. 3, ROC curve analyses for determining high-quality videos based on usefulness score



Country	n	%
Germany	1	0.8
New Zeland	1	0.8
Spain	1	0.8
Sri Lanka	1	0.8
Philippines	2	1.7
India	3	2.5
Canada	5	4.1
Unknown	9	7.4
United Kingdom	8	6.6
Australia	10	8.3
United States	80	66.2
Total	121	100

Fig. 2 Distribution of video uploads in countries

indicated that mDISCERN (cut-off 2.50, AUC 0.785), GQS (cut-off 2.50, AUC 0.851), and VIQI (cut-off 10.50, AUC 0.874) scores showed meaningful results for rater 1. Also, the analysis results for rater 2 were mDISCERN (cut-off 2.50, AUC 0.821), GQS (cut-off 2.50, AUC 0.876), and VIQI (cut-off 10.50, AUC 0.870) scores showed meaningful results. All three showed above excellent discrimination ability for high-quality videos.

Discussion

Whether individuals are patients or healthy, many turn to YouTube for information on various aspects of their lives. When this information pertains to health, the quality, reliability, and validity of the content on this platform become critically important. Despite being one of the most widely accessed sites, YouTube lacks standardization for the videos and information shared, which researchers must evaluate. To the authors' knowledge, this was the first study to evaluate and examine TMD exercise videos on YouTube. In the current study, exercise videos related to TMD were examined, focusing on those that patients can perform independently without needing direct guidance or specialized anatomical and physiological knowledge. Additionally, the reliability of instruments used to assess the quality of these videos was determined. Based on video usefulness, the videos were categorized into three groups: misleading, potentially harmful; misleading, not harmful; and useful videos.

Table 1 Video characteristics ($n = 121$)

Median (IQR 25–75), Min–Max	
Detailed features of videos	
Video duration, second	125 (78–346), 20–1778
Time since upload, day	1440 (720–2160), 15–5040
Number of views	246 (46–5804), 0–4300000
Number of likes	5 (0–78), 0–89000
Number of comments	0 (0–6), 0–7134
Number of subscribers	756 (100–13500), 0–9000000
Viewing rate	0.23 (0.05–3.61), 0–1990.74
Interaction index	0.01 (0–0.03), 0–2.56
Reliability and quality scores of videos	
mDISCERN rater 1	2 (2–3), 0–5
mDISCERN rater 2	2 (2–3), 0–5
GQS rater 1	2 (2–3), 1–5
GQS rater 2	2 (2–3), 1–5
VIQI rater 1	10 (8–12), 4–18
VIQI rater 2	10 (9–12), 4–18
Content of video	**n (%)**
Active exercises	57 (47.1)
Strengthening exercises	45 (37.2)
Stretching exercises	22 (19.9)
Chin tuck exercise	18 (14.9)
Posture or cervical exercises	27 (22.3)
Other exercises	16 (13.2)
Videos according source of upload	**n (%)**
Healthcare professionals	89 (73.6%)
Individual user	22 (18.2%)
Trainer	10 (8.3%)
Usefulness	**n (%)**
Misleading, potentially harmful	7 (5.8%)
Misleading, not harmful	67 (55.4%)
Useful	47 (38.8%)

IQR Interquartile range, *Min* Minimum, *Max* Maximum, *mDISCERN* modified DISCERN, *GQS* Global Quality Scale, *VIQI* Video Information and Quality Index

There are existing studies that have examined and evaluated YouTube videos related to COVID-19 [26, 27], pelvic floor [28], ankylosing spondylitis [29], rheumatoid arthritis [30], migraine [31], and dysphagia [32]. Many of these studies [27–32] have employed assessment scales similar to those used in the present study. Similar to previous studies [30], this research found that most of the videos were shared by healthcare professionals (73.6%). However, despite the majority of the videos being uploaded by healthcare professionals, most were classified as misleading but not harmful (55.4%). This could be due to several factors, including the lack of patient education and information in many videos, the failure to recommend performing exercises in the correct posture and in front of a mirror, and the lack of guidance on selecting exercise. Regardless of video length or the variety of exercises presented, there was a noticeable absence of a scientific and literature-based approach. Additionally, factors such as the upload date of the videos, view counts, and subscriber numbers were not indicative of the quality or content of the videos. When examining the content of the videos, it becomes evident that they predominantly included a combination of various methods rather than specific exercise approaches. For instance, videos featured isometric strengthening exercises along with stretching exercises, as well as massages for the temporalis and masseter muscles. The majority of the videos (47.1%) included active mouth opening exercises. Controlled mouth opening exercise was observed in only 14 videos (11.6%). Of the videos containing strengthening exercises, 19 (15.7%) featured isometric exercises, 23 (19%) progressive resistive exercises, and 3 (2.5%) eccentric exercises. 14.9% of the videos included chin tuck exercises, 22.3% included posture or cervical exercises. Considering the anatomical, physiological and biomechanical connection of the TMJ with the cervical region, the evaluation and treatment of this region also gains importance. Rocabado's 6×6 exercises were included in 2 (1.7%) videos. It was seen that many methods and exercises were included without explaining the reason for their selection and application in videos. The importance and information of the pathology should also be included in the videos. In addition, one of the biggest deficiencies in the videos can be mentioned as follows: While patient education is very important for TMD, no one has addressed this in detail except for a few videos. On the other hand, all medical advice and treatments, including physiotherapy approaches, should be recommended appropriately and individually for a class of diagnosis, for example, pain-related TMD or intra-articular TMD pathology or both. Online assessment advice can be provided to observe the range of motion of the mandible and try to determine if there is a concern for disc displacement with or without reduction [33]. Stretching exercises aimed at increasing mouth opening should be recommended with caution, particularly in cases of unilateral intra-articular TMD (such as disc displacement without reduction) where symmetrical mouth opening has not been achieved, to prevent the persistence of asymmetrical mouth opening [8].

The Cohen's Kappa scores for the evaluators' GQS and mDISCERN assessments indicated substantial agreement between the evaluators' video evaluation results. Additionally, the ICC value for the VIQI scores among the evaluators was well above acceptable levels. ROC analysis also confirmed the accuracy of the methods used. When examining the exercise content videos watched by the evaluators, those uploaded by healthcare professionals

Table 2 Reliability and quality scores of videos according to the source of upload

Median (IQR 25–75)	Healthcare professionals (*n* = 89)	Individual user (*n* = 22)	Trainer (*n* = 10)	*p*
mDISCERN rater 1	2 (2–3)	1.50 (1–2)	2 (1–2)	**0.000**
Healthcare professionals > Individual user (0.000)				
mDISCERN rater 2	2 (2–3)	2 (1–2)	2 (1–2)	**0.000**
Healthcare professionals > Individual user (0.000)				
GQS rater 1	2 (2–3)	2 (1–2)	3 (2–3)	**0.001**
Healthcare professionals > Individual user (0.000), Trainer > Individual user (0.037)				
GQS rater2	3 (2–3)	2 (1–2)	3 (1–3)	**0.005**
Healthcare professionals > Individual user (0.003)				
VIQI rater 1	10 (9–12)	8 (5–9)	9 (7–13)	**0.000**
Healthcare professionals > Individual user (0.000)				
VIQI rater 2	11 (9–13)	9 (6–10)	10.50 (8–12)	**0.000**
Healthcare professionals > Individual user (0.000)				

IQR Interquartile range, *mDISCERN* modified DISCERN, *GQS* Global Quality Scale, *VIQI* Video Information and Quality Index

demonstrated higher quality and reliability compared to individual users. Furthermore, useful videos were of higher quality and reliability than misleading, potentially harmful, and misleading, not harmful videos. Consistent with previous studies, the present study found that videos deemed useful had significantly higher mDISCERN and GQS scores compared to misleading exercise videos [28, 30]. When examining high-quality and low-quality videos, the primary sources of high-quality videos were academics, non-physician healthcare personnel, universities/clinics/health centers, physicians, and physiotherapists. Low-quality videos were mostly from independent users. Aligning with our results, researchers have reported that high-quality videos predominantly originate from healthcare professionals or institutions, while the main sources of low-quality videos are medical advertisements, non-profit organizations, and independent users [29, 30]. Given the numerical abundance of online content, resources, and videos, it is essential for patients to be cautious and well-informed when utilizing these sources. However, no matter how knowledgeable they are, they cannot match the expertise of a healthcare professional. At this point, it might be crucial to first gain information about the channel, institution, or individuals uploading the video. The similarities in metrics such as likes, comments, subscribers, and view rates between misleading and useful videos pose a significant disadvantage as they facilitate access to incorrect exercise information. Individuals typically gravitate toward videos on the first page of search results, making it unclear whether they are exposed to misleading or useful content. However, we do not consider the evaluated parameters to be a reliable guide in this context. As stated, it is important to thoroughly investigate the source and origin of the content.

In the current study, most of the videos were sourced from the United States, which may limit the geographic generalizability of the findings. While the study primarily featured United States based videos, which may reflect certain healthcare delivery practices unique to that region, it is important to consider that cultural differences in healthcare approaches could influence how video content is produced, presented, and interpreted globally. This cultural variation may affect the generalizability of our findings to non-United States audiences. Future studies could benefit from analyzing video content from different countries and healthcare systems to better understand the impact of these cultural differences on the perceived quality and applicability of health information. Additionally, video length alone is not a definitive measure of quality or credibility; short but well-produced videos can also meet high-quality standards. Therefore, the evaluation criteria should be expanded to consider the content quality of both short and long videos. More research is needed to understand the reasons behind healthcare professionals producing misleading but harmless videos, especially in light of potential gaps in education and communication. Addressing these gaps could improve the quality of health-related content shared on platforms like YouTube. Furthermore, it is important to consider the impact of YouTube's algorithmic bias on the visibility of trustworthy versus untrustworthy content. The role that algorithms play in determining the types of content users are exposed to is a critical factor in health information acquisition. Understanding how these algorithms work can help identify ways to promote high-quality health information. Additionally, while video length is often considered an indicator of content credibility, producing shorter, high-quality, and

Table 3 Comparison of video characteristics and reliability and quality scores of videos according to the usefulness

Median (IQR 25–75)	Misleading, potentially harmful (*n* = 7)	Misleading, not harmful (*n* = 67)	Useful (*n* = 47)	*p*
Detailed features of videos (Median (IQR 25–75), min–max)				
Video duration, second	72 (22–82)	121 (66–310)	197 (92–362)	**0.002**
Misleading, not harmful > Misleading, potentially harmful (0.019), Useful > Misleading, potentially harmful (0.001)				
Time since upload, day	2520 (360–3600)	1440 (1080–2520)	1080 (360–1440)	**0.004**
Misleading, potentially harmful > Misleading, not harmful (0.004)				
Number of views	125 (72–1500)	676 (58–8000)	144 (34–5804)	0.367
Number of likes	0 (0–14)	8 (1–113)	2 (0–70)	0.102
Number of comments	0 (0–2)	0 (0–15)	0 (0–6)	0.457
Number of subscribers	624 (9–1190)	812 (103–34600)	635 (67–5890)	0.562
Viewing rate	0.11 (0.05–0.60)	0.46 (0.05–4.35)	0.19 (0.04–9.44)	0.518
Interaction index	0 (0–0.02)	0.02 (0–0.03)	0.01 (0–0.04)	0.136
Reliability and quality scores of videos				
mDISCERN rater 1	1 (0–1)	2 (2–2)	3 (2–3)	**0.000**
Misleading, not harmful > Misleading, potentially harmful (0.018), Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				
mDISCERN rater 2	0 (0–1)	2 (2–2)	3 (2–3)	**0.000**
Misleading, not harmful > Misleading, potentially harmful (0.004), Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				
GQS rater 1	1 (1–1)	2 (2–2)	3 (3–3)	**0.000**
Misleading, not harmful > Misleading, potentially harmful (0.034), Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				
GQS rater 2	1 (1–1)	2 (1–3)	3 (3–4)	**0.000**
Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				
VIQI rater 1	5 (4–8)	9 (8–10)	12 (11–14)	**0.000**
Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				
VIQI rater 2	6 (5–8)	9 (8–11)	13 (11–14)	**0.000**
Useful > Misleading, potentially harmful (0.000), Useful > Misleading, not harmful (0.000)				

IQR Interquartile range, *mDISCERN* modified DISCERN, *GQS* Global Quality Scale, *VIQI* Video Information and Quality Index

Table 4 Correlations among mDISCERN, GQS, and VIQI scores of raters

r p	mDISCERN rater 1	mDISCERN rater 2	GQS rater 1	GQS rater 2	VIQI rater 1	VIQI rater 2
mDISCERN rater 1	1.000					
mDISCERN rater 2	**0.853** **0.000**	1.000				
GQS rater 1	**0.793** **0.000**	**0.833** **0.000**	1.000			
GQS rater 2	**0.740** **0.000**	**0.839** **0.000**	**0.909** **0.000**	1.000		
VIQI rater 1	**0.743** **0.000**	**0.796** **0.000**	**0.807** **0.000**	**0.831** **0.000**	1.000	
VIQI rater 2	**0.760** **0.000**	**0.813** **0.000**	**0.816** **0.000**	**0.855** **0.000**	**0.889** **0.000**	1.000

mDISCERN modified DISCERN, *GQS* Global Quality Scale, *VIQI* Video Information and Quality Index

Table 5 Receiver operating characteristics (ROC) curve analyses for evaluation scores to discriminate qualified YouTube videos based on usefulness

	AUC (95% IC)	Cutt off value	Sensitivity	Specifity	Standart error	*p*
mDISCERN rater 1	0.785 (0.702–0.867)	2.50	0.553	0.108	0.042	0.000
GQS rater 1	0.851 (0.782–0.921)	2.50	0.809	0.176	0.035	0.000
VIQI rater 1	0.874 (0.811–0.937)	10.50	0.787	0.149	0.032	0.000
mDISCERN rater 2	0.821 (0.746–0.896)	2.50	0.660	0.135	0.038	0.000
GQS rater 2	0.876 (0.816–0.937)	2.50	0.872	0.230	0.031	0.000
VIQI rater 2	0.870 (0.806–0.934)	10.50	0.851	0.243	0.033	0.000

AUC Area under the curve, *CI* Confidence interval

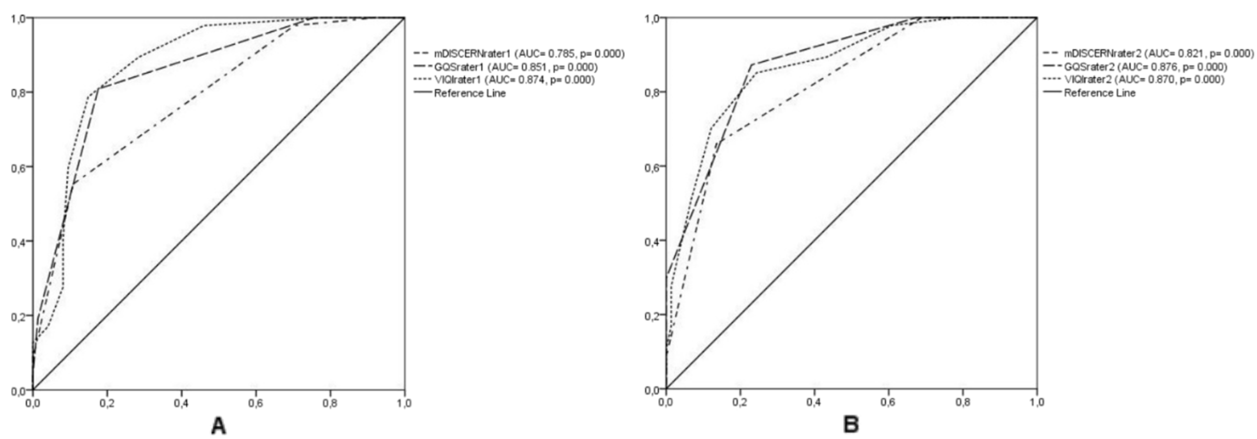


Fig. 3 Receiver operating characteristic (ROC) curve for evaluation scores to discriminate good quality YouTube videos based on usefullnes for rater 1 (**A**) and for rater 2 (**B**). mDISCERN, GQS and VIQI scores show good discrimination ability for high-quality YouTube videos

effective videos is equally important. Future research should focus on how these videos can effectively convey accurate health information. To address algorithmic bias and improve the visibility of credible health content, healthcare professionals could implement several optimization strategies when producing video content. These strategies may include using search engine optimization friendly titles and descriptions, incorporating trending and relevant keywords, and ensuring engaging thumbnails to attract a broader audience. Additionally, maintaining consistent video posting schedules and fostering viewer interaction through comments and community engagement can boost the algorithm's recognition of the content. By implementing these practices, healthcare professionals can increase their content's reach and contribute to a more balanced representation of reliable health information on platforms like YouTube.

There are numerous studies examining the content, quality, and reliability of YouTube exercise videos for different disease groups or pathologies [26–32]. Regarding TMD, studies have investigated videos on arthrocentesis, bruxism, and general TMD [15, 16, 18, 34]. To the best of our knowledge, this study is the first to examine the characteristics, content, quality, and reliability of exercise videos related to TMD. In light of this study, future research could evaluate specific exercise approaches or treatment videos such as manual therapy, electrotherapy, and taping for TMD. Considering the various sub-diagnostic categories of TMD, studies focusing on physiotherapy approaches for a single diagnostic group or specific approaches for a particular diagnostic group can also be planned.

This study has several limitations. Due to its cross-sectional nature, videos added after the analysis dates on this continuously updated social media platform were not considered. As a result, the dynamic nature of this platform could lead to different findings at different times among various researchers. While this study provides valuable insights through a cross-sectional approach, it is essential that future research revisits and re-evaluates newly published videos periodically, given the evolving nature of YouTube content. As a result, findings from future studies may differ from our own, which is based on content available up to the date of our study. Future research will contribute to a more comprehensive understanding of the platform's impact by capturing changes over time in the type, quality, and utility of the videos. Additionally, longitudinal comparisons of videos published in specific periods (e.g., comparing videos published within the last three years with those published three to six years ago) can be considered. Such studies could reveal whether newly added videos reflect any advancement by comparing them with older content. Therefore, the study is reproducible. The constant addition of new videos and the varying metrics of views, comments, likes, and dislikes necessitate continuous monitoring and evaluation. Additionally, the present study focused solely on English exercise videos. Lastly, the study was limited to YouTube®. Despite being the first study to examine exercise videos related to TMD, a strength of this study was its comprehensive analysis of all exercise videos. Two experienced researchers with backgrounds in physiotherapy and rehabilitation analyzed the exercise videos. Three different scales were used to assess the quality of the exercise videos. Moreover, the content, reliability, and quality of the exercise videos, along with the types of exercises demonstrated in the videos, were identified. However, in addition to these scales that evaluate videos from different perspectives, future studies should be conducted using more comprehensive scales developed for the usefulness and target audience of the videos. Another of the study's limitations was the difficulty in categorizing and quantifying the exercises shown in the videos due to their high variability and the frequent inclusion of non-exercise techniques. This complexity made it challenging to provide a structured table indicating the exact number of videos for each type of exercise. Future studies could aim to address this issue by focusing on a narrower scope of exercises or developing standardized methods for exercise categorization.

Conclusion

This study provides a comprehensive analysis of YouTube® videos related to TMD exercises, evaluating their content, quality, and reliability. Since exercises performed in uncontrolled and undesirable positions can be misleading and harmful for TMD, evaluations were made accordingly. Particular attention was paid to the way the exercises were explained and applied. Thus, misleading and harmful videos were identified. The majority of the videos were found to be useful, with healthcare professionals being the primary source of high-quality content. However, despite being uploaded by healthcare professionals, a significant portion of the videos were categorized as misleading but not harmful according to the usefulness classification. Common deficiencies identified in almost all videos included a lack of patient education and guidance on correct exercise practices, insufficient information on TMD. The study highlights the importance of critically assessing the source and content of online health information, given the ease of access to both high- and low-quality videos. The findings underscore the need for patients to be cautious and well-informed when using online resources for health-related information. The current studies and analyses addressed exercise approaches for general TMD rather than specific TMD diagnostic classifications. Therefore, it should be noted that individualized and specified exercise and physiotherapy approaches are needed for different subgroups of TMD (such as combining intra-articular TMD with pain-related TMD). Future research should focus on evaluating specific therapeutic approaches, such as manual therapy or electrotherapy, and their representation in online videos. Additionally, longitudinal studies monitoring the dynamic nature of video content on social media platforms could provide further insights into the evolving landscape of online health information.

Acknowledgements
None.

Authors' contributions
Conception and design: H.A. and E.E. Methodology: H.A. and E.E. Data acquisition: H.A. and E.E. Data analysis and interpretation: H.A. Writing of article: H.A. and E.E. Review of article: H.A. and E.E.

Funding
Funding for this study was not provided.

Data availability
Requests for data access can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The study does not involve human participants and therefore cannot be applicable.

Competing interests
The authors declare no competing interests.

Received: 13 December 2024 Accepted: 14 February 2025
Published online: 25 February 2025

References

1. Tanaka E, Detamore MS, Mercuri LG. Degenerative disorders of the temporomandibular joint: etiology, diagnosis, and treatment. J Dent Res. 2008;87(4):296–307.
2. Murphy MK, MacBarb RF, Wong ME, Athanasiou KA. Temporomandibular joint disorders: a review of etiology, clinical management, and tissue engineering strategies. Int J Oral Maxillofac Implants. 2013;28(6):e393-414.
3. Zieliński G, Pająk-Zielińska B, Ginszt M. A meta-analysis of the global prevalence of temporomandibular disorders. J Clin Med. 2024;13(5):1365.
4. Kumar MPS, Murugesan K, Hinaz N. Surgical treatment modalities in the management of temporomandibular joint disorders. Int J Dent Oral Sci. 2021;8(9):4168–79.
5. Zhang L, Xu L, Wu D, Yu C, Fan S, Cai B. Effectiveness of exercise therapy versus occlusal splint therapy for the treatment of painful

temporomandibular disorders: a systematic review and meta-analysis. Ann Palliat Med. 2021;10(6):6122–32.
6. Butts R, Dunning J, Pavkovich R, Mettille J, Mourad F. Conservative management of temporomandibular dysfunction: a literature review with implications for clinical practice guidelines (Narrative review part 2). J Bodyw Mov Ther. 2017;21(3):541–8.
7. Fernández-de-Las-Peñas C, Von Piekartz H. Clinical reasoning for the examination and physical therapy treatment of temporomandibular disorders (TMD): a narrative literature review. J Clin Med. 2020;9(11):3686.
8. Arikan H, Citaker S, Ucok C, Ucok O. Effect of high voltage electrical stimulation in temporomandibular disorders: a randomized controlled trial. Physiother Theory Pract. 2025;41(1):79–92.
9. Rashid A, Matthews NS, Cowgill H. Physiotherapy in the management of disorders of the temporomandibular joint—perceived effectiveness and access to services: a national United Kingdom survey. Br J Oral Maxillofac Surg. 2013;51(1):52–7.
10. Idáñez-Robles AM, Obrero-Gaitán E, Lomas-Vega R, Osuna-Pérez MC, Cortés-Pérez I, Zagalaz-Anula N. Exercise therapy improves pain and mouth opening in temporomandibular disorders: a systematic review with meta-analysis. Clin Rehabil. 2023;37(4):443–61.
11. Shimada A, Ogawa T, Sammour SR, Narihara T, Kinomura S, Koide R, et al. Effectiveness of exercise therapy on pain relief and jaw mobility in patients with pain-related temporomandibular disorders: a systematic review. Front Oral Heal. 2023;4:1170966.
12. Basch CH, Yin J, Kollia B, Adedokun A, Trusty S, Yeboah F, et al. Public online information about tinnitus: a cross-sectional study of YouTube videos. Noise Heal. 2018;20(92):1–8.
13. Basch CH, Yin J, Walker ND, De Leon AJ, Fung I. TMJ online: investigating temporomandibular disorders as "TMJ" on YouTube. J Oral Rehabil. 2018;45(1):34–40.
14. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J. 2015;21:173–94.
15. Kim MJ, Kim JR, Jo JH, Kim JS, Park JW. Temporomandibular disorders-related videos on YouTube are unreliable sources of medical information: a cross-sectional analysis of quality and content. Digit Health. 2023;9:20552076231154376.
16. Keskinrüzgar A, Yavuz GY, Acibadem E. Reliability of Information on Youtube™ regarding temporomandibular joint arthrocentesis: cross-sectional research. Turkiye Klinikleri J Dental Sci. 2023;29(2):234–42.
17. Özkan TH, Acar AH, Özkan E. Does YouTube provide high-quality information for patients regarding night guards, especially for over-the-counter ones? J Craniofac Surg. 2024;35(4):1157–9.
18. Vaira LA, Sergnese S, Salzano G, Maglitto F, Arena A, Carraturo E, et al. Are YouTube videos a useful and reliable source of information for patients with temporomandibular joint disorders? J Clin Med. 2023;12(3):817.
19. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis — a wakeup call? J Rheumatol. 2012;39(5):899–903.
20. Bernard A, Langille M, Hughes S, Rose C, Leddin D, van Zanten SV. A systematic review of patient inflammatory bowel disease information resources on the world wide web. Am J Gastroenterol. 2007;102(9):2070–7.
21. Lena Y, Dindaroğlu F. Lingual orthodontic treatment: a YouTube™ video analysis. Angle Orthod. 2018;88(2):208–14.
22. McHugh ML. Interrater reliability: the kappa statistic. Biochem Medica. 2012;22(3):276–82.
23. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.
24. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69–71.
25. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022;75(1):25–36.
26. Vancini RL, Borges Viana R, dos Santos Andrade M, Andre Barbosa de Lira C, Theodoros Nikolaidis P, Aparecido de Almeida A, et al. YouTube as a source of information about physical exercise during COVID-19 outbreak. Int J Sport Stud Heal. 2022;4(2):e123312.
27. Rodriguez-Rodriguez AM, Blanco-Diaz M, de la Fuente-Costa M, Hernandez-Sanchez S, Escobio-Prieto I, Casaña J. Review of the quality of YouTube videos recommending exercises for the COVID-19 lockdown. Int J Environ Res Public Health. 2022;19(13):8016.
28. Culha Y, Seyhan Ak E, Merder E, Ariman A, Culha MG. Analysis of the YouTube videos on pelvic floor muscle exercise training in terms of their reliability and quality. Int Urol Nephrol. 2021;53(1):1–6.
29. Kocyigit BF, Nacitarhan V, Koca TT, Berk E. YouTube as a source of patient information for ankylosing spondylitis exercises. Clin Rheumatol. 2019;38(6):1747–51.
30. Sari F, Bazancir Apaydin Z, Sari S. Assessment of reliability and quality of YouTube® exercise videos in people with rheumatoid arthritis. Physiother Theory Pract. 2025;41(2):362–9.
31. Reina-Varona Á, de Rivera-Romero BR, Cabrera-López CD, Fierro-Marrero J, Sánchez-Ruiz I, La Touche R. Exercise interventions in migraine patients: a YouTube content analysis study based on grades of recommendation. PeerJ. 2022;10:e14150.
32. Chang MC, Park D. Exercise interventions in migraine patients: a YouTube content analysis study based on grades of recommendation. Healthcare (Basel). 2021;9(8):1084.
33. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet J-P, et al. Diagnostic criteria for temporomandibular disorders (DC/TMD) for clinical and research applications: recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. J Oral Facial Pain Headache. 2014;28(1):6–27.
34. Bilir H, Yılancı H. Does YouTube™ give us accurate information about bruxism? J Heal Sci Med. 2023;6(2):380–4.

Publisher's Note