



Development and validation of a tool for evaluating YouTube-based medical videos

Mehmet Akif Guler¹ · Esref Orkun Aydın²

Received: 11 October 2021 / Accepted: 18 November 2021 / Published online: 25 November 2021
© The Author(s), under exclusive licence to Royal Academy of Medicine in Ireland 2021

Abstract

Background/aims Today, one of the ways to access medical information is the internet. Our objective was to develop a measurement tool to assess the quality of online medical videos.

Methods Online videos covering a variety of subjects (COVID-19, low back pain, weight loss, hypertension, cancer, chest pain, vaccination, asthma, allergy, and cataracts) were evaluated using our Medical Quality Video Evaluation Tool (MQ-VET) by 25 medical and 25 non-medical professionals. Exploratory factor analysis, Cronbach's alpha, and correlation coefficients were used to assess the validity and reliability of the MQ-VET.

Results The final MQ-VET consisted of 15 items and four sections. The Cronbach's alpha reliability coefficient for the full MQ-VET was 0.72, and the internal consistency for all factors was good (between 0.73 and 0.81). The correlation between the DISCERN questionnaire scores and MQ-VET scores was significant.

Conclusion Collectively, our findings indicated that the MQ-VET is a valid and reliable tool that will help to standardize future evaluations of online medical videos.

Keywords Medical videos · Questionnaire · Reliability · Validity · YouTube

Introduction

Eight out of ten internet users access medical information online [1]. The YouTube platform, in particular, allows users to create medical content, without any obligation to post verified information [2, 3]. In 2007, Keelan et al. first examined the quality of immunization-related online videos [4]. Many subsequent studies have further assessed the reliability of medical videos on YouTube. Presently, the search term “YouTube” returns more than 1,500 publications on PubMed and Scopus (accessed on 17 Jan 2021) [1]. However, a standardized tool for evaluating medical health videos is lacking. Most previous studies used novel, topic-specific scoring systems based on the literature and authors' own knowledge [5]. However, the generalizability of these scoring systems

is poor, and the results obtained using them are difficult to repeat. Moreover, their validity and reliability have not been adequately measured.

A variety of tools to evaluate the accuracy of medical information are available, including the DISCERN instrument, Health on the Net (HON) code, *Journal of the American Medical Association* (JAMA) evaluation system, brief DISCERN instrument, global quality score (GQS), and video power index (VPI); medical videos can also be evaluated subjectively [1, 5, 6]. The HON Foundation devised eight principles for websites to abide by, called the HONcode [7]. Certification by the HONcode Foundation is available for a fee, but the quality of medical information is not rated. Furthermore, the validity and reliability of this system for YouTube videos have not been confirmed. The JAMA scoring system was created to evaluate medical information on websites [8], but has not been validated for videos. The DISCERN instrument was created nearly 20 years ago for application to “written information about treatment choices” [9, 10]. Again, however, this instrument has not been validated for medical videos. In addition, the second part of the DISCERN questionnaire is focused on treatment information. Thus,

✉ Mehmet Akif Guler

¹ Department of Physical Medicine and Rehabilitation, Sultan II. Abdülhamid Han Training and Research Hospital, Selimiye, Tıbbiye Street Üsküdar, Istanbul 34668, Turkey

² Department of Physical Medicine and Rehabilitation, Gaziosmanpaşa Training and Research Hospital, Istanbul, Turkey

videos that exclude treatment information yield misleading results [10]. The VPI score, as a measure of audience approval, is calculated as the number of likes a video has divided by the number of likes and dislikes [11]. This scoring system, which is frequently used, is not suitable for evaluating the quality and reliability of medical videos. Given the lack of suitable instruments, we developed a reliable instrument, i.e., the Medical Quality Video Evaluation Tool (MQ-VET), for use by patients and healthcare professionals.

Materials and methods

Instrument development and item generation

Our original questionnaire included 42 novel items based on published evaluations of medical video quality. All questionnaires used in YouTube-related articles and questions used in subjective evaluations were examined by both authors [1, 5, 12–14]. Candidate items were rated by the authors (0 points, not applicable; 10 points, highly applicable). Duplicated questions and those with a score below the average were excluded, resulting in a total of 28 questions.

Participants

Videos were evaluated by 25 medical and 25 non-medical participants who have obtained sufficient points in any of the valid English language tests in the country and fluent in English. The questionnaire items were rated by participants in terms of quality and relevance (0–10 points for each item). The face and content validity of the questionnaire were also evaluated using the 10-point rating system. After excluding items with a score < 7, the final MQ-VET included 19 items. Ten unique videos (first appeared video for the popular topics from different medical subjects using YouTube's default setting) differing in terms of the uploader and medical topic were evaluated using a 5-point Likert scale (Table 1). The DISCERN instrument was also used to evaluate each video for concurrent validity.

Statistical analysis

Statistical analyses were performed using SPSS ver. 20.0 (IBM, Armonk, NY, USA). Data distribution was examined using the Shapiro–Wilk test and histograms. Continuous variables are expressed as means \pm standard deviation

Table 1 General information of the evaluated YouTube videos (updated on 12.02.2021)

Video link	Topic	Source	View count	Comment count	Like/dislike	Uploaded
https://www.youtube.com/watch?v=i0ZabxXmH4Y	COVID-19	Institution	517,921	536	5465/359	15 January 2020
https://www.youtube.com/watch?v=BOjTegn9RuY	Low back pain	Medical doctor	1,988,635	1237	16,390/1012	24 January 2014
https://www.youtube.com/watch?v=2MoGxae-zyo	How to lose weight	Private	139,422,338	116,586	2,801,454/18,000	8 August 2019
https://www.youtube.com/watch?v=X5TknCu3RV0	Hypertension	Internet source	157,110	72	2301/38	24 August 2019
https://www.youtube.com/watch?v=SGaQ0WwZ_0I	Cancer	Institution	2,423,944	2228	10,298/1067	31 October 2013
https://www.youtube.com/watch?v=vEQQidcJF1Q	Chest pain	Institution	96,213	172	398/33	15 October 2019
https://www.youtube.com/watch?v=Atrx1P2EkiQ	Vaccination	Company	364,676	75	3310/188	30 January 2020
https://www.youtube.com/watch?v=PzfLDi-sL3w	Asthma	Company	3,964,819	11,162	74,291/1009	11 May 2017
https://www.youtube.com/watch?v=IlZFx8n-WCQ	Allergy	Website/video channel	94,736	23	974/28	18 November 2019
https://www.youtube.com/watch?v=d5D0B2PoC7U	Cataract	Institution	207,946	41	1049/37	26 October 2015

Table 2 MQ-VET and DISCERN scores of the YouTube videos on different topics

Topic	MQ-VET scores					DISCERN scores			
	Part 1	Part 2	Part 3	Part 4	Total score	Section 1	Section 2	Section 3	Total score
COVID-19	18.12 ± 4.74	16.96 ± 2.39	13.24 ± 1.49	12.60 ± 2.08	60.92 ± 8.36	31.48 ± 4.02	15.06 ± 5.87	2.46 ± 12.65	49 ± 7.05
Low back pain	18.42 ± 4.22	16.72 ± 2.81	13 ± 1.85	12.04 ± 2.49	60.18 ± 9.16	30.54 ± 4.67	25.02 ± 4.11	3.9 ± 0.73	59.46 ± 6.4
How to lose weight	12.18 ± 3.70	10.60 ± 4.18	12.74 ± 2.25	11.24 ± 2.97	46.76 ± 8.61	22.66 ± 5.23	14.52 ± 6.58	2.12 ± 1.08	39.3 ± 11.06
Hypertension	13.06 ± 4.14	15.18 ± 2.89	12.56 ± 1.78	10.64 ± 2.57	51.44 ± 7.66	22.52 ± 5.26	16.48 ± 7.48	2.18 ± 1.28	41.18 ± 13.28
Cancer	14.10 ± 4.33	14.16 ± 2.66	12.14 ± 2.81	9.86 ± 2.49	50.26 ± 9.12	25.30 ± 6.76	21.18 ± 5.96	3.26 ± 1.08	49.74 ± 11.89
Chest pain	11.44 ± 2.89	12 ± 3.36	11.54 ± 9.04	9.04 ± 2.50	44.02 ± 7.18	19.24 ± 6.15	14.12 ± 5.99	1.96 ± 1.02	35.08 ± 11.86
Vaccination	13.96 ± 3.54	15.56 ± 2.74	12.36 ± 2.81	10.70 ± 2.43	52.58 ± 8.26	22.68 ± 6.39	18.20 ± 5.59	2.72 ± 1.19	43.60 ± 11.14
Asthma	14.08 ± 4.30	15.40 ± 2.51	12.34 ± 2.35	11.82 ± 2.45	53.64 ± 8.13	28.14 ± 3.30	23.62 ± 4.37	2.92 ± 1.24	54.68 ± 6.90
Allergy	11.08 ± 4.94	12.44 ± 3.16	9.34 ± 2.70	10.46 ± 2.08	43.32 ± 9.41	22.94 ± 4.91	16.48 ± 6.54	2.52 ± 1.11	41.94 ± 10.82
Cataract	10.80 ± 4.55	16.44 ± 2.77	12.40 ± 2.80	9.36 ± 2.21	49.00 ± 6.81	22.80 ± 5.17	13.50 ± 7.50	1.84 ± 1.09	38.14 ± 12.54

(SD) with ranges, and categorical variables are expressed as numbers and percentages.

For item analysis, kurtoses, item-item correlations (IIC), and item-total correlations were calculated. Exploratory factor analysis (EFA) was conducted to verify construct validity. Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett’s sphere test were conducted to check whether the data were suitable for EFA. In general, KMO values between 0.8 and 1 indicate that the sampling is adequate, and KMO values less than 0.6 indicate that the sampling is not adequate [15]. Factors in the EFA were extracted using principal components analysis and the varimax kappa 4 of the rotation.

Reliability was assessed by analyzing Cronbach’s alpha value. Spearman’s correlation coefficients between MQ-VET and DISCERN scores were used for concurrent validity.

After the study was completed, the post hoc power analysis was performed using the G*Power version 3.1.9.2 software (Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany). For the bivariate normal model correlation from exact test family, the post hoc power was calculated as 0.81 in the power analysis using correlation between the MQ-VET and the DISCERN scores (Table 2).

Results

The mean age of the participants was 30.98 ± 4.38 years (range: 25–42 years). The professions of the participants were as follows: doctor (44%, $n = 22$), pharmacist (6%, $n = 3$), academic/teacher (20%, $n = 10$), and engineer (24%, $n = 12$). Profession data were missing in three cases. There were 23 (46%) participants with a bachelor’s degree, 6 (12%) with a masters, and 21 (42%) with a doctorate.

Exploratory factor analysis

The Kaiser–Meyer–Olkin (KMO) value for the final MQ-VET (19 items) was 0.83 and the Bartlett’s test statistic was $\chi^2 = 3920.72$ ($p < 0.001$). Thus, the data were suitable for further analysis. The first exploratory factor analysis (EFA) had five factors. The component correlation matrix was orthogonal, and varimax rotation with Kaiser normalization was applied. The factor loadings of the final EFA are displayed in Table 3. Ultimately, our questionnaire included 15 items across four factors (5, 4, 3, and 3 items for factors 1–4, respectively).

Concurrent validity

Correlation between the final form of the MQ-VET and DISCERN questionnaire used for concurrent validity. The scores of the questionnaires have shown in Table 2. The first part of the MQ-VET significantly correlated with all sections of DISCERN: Sect. 1 ($\rho = 0.617$, $p < 0.001$), Sect. 2 ($\rho = 0.508$, $p < 0.001$), Sect. 3 ($\rho = 0.436$, $p < 0.001$), and total score of DISCERN ($\rho = 0.640$, $p < 0.001$). The second part of the MQ-VET also significantly weakly correlated with all sections of DISCERN: Sect. 1 ($\rho = 0.456$, $p < 0.001$), Sect. 2 ($\rho = 0.167$, $p < 0.001$), Sect. 3 ($\rho = 0.123$, $p = 0.006$), and total score of DISCERN ($\rho = 0.326$, $p < 0.001$). The third part of the MQ-VET only significantly weakly correlated with Sect. 1 of the DISCERN scores ($\rho = 0.228$, $p < 0.001$), but not significantly correlated with Sect. 2, Sect. 3, and total scores of DISCERN ($p = 0.975$, $p = 0.578$, $p = 0.18$, respectively). The fourth part of the MQ-VET significantly correlated with all DISCERN scores: Sect. 1 ($\rho = 0.510$, $p < 0.001$), Sect. 2 ($\rho = 0.231$, $p < 0.001$), Sect. 3 ($\rho = 0.205$, $p < 0.001$), and total score ($\rho = 0.395$, $p < 0.001$). Total scores of the MQ-VET significantly correlated with all sections: Sect. 1

Table 3 Factor loadings from exploratory factor analysis

Items	Factor 1	Factor 2	Factor 3	Factor 4	Communality
13. Date and updates, if any, are clearly stated	0.822				0.724
15. The recording date of the video and date on which the information was accessed are mentioned	0.808				0.701
12. The resources and references used are clearly stated	0.764				0.606
14. Concerns about advertising and potential conflicts of interest have been resolved	0.731				0.670
23. Sufficient information was provided about the identity of the presenter in the video	0.571				0.578
9. The materials used in the video facilitated learning		0.783			0.729
10. The video covered the basic concepts of the subject		0.696			0.550
17. To explain the medical topic, visual resources were used sufficiently		0.671			0.672
19. The medical terms used were well-explained		0.456	0.444		0.418
22. The sound quality of the video was sufficient			0.857		0.764
21. The image quality of the video was sufficient			0.822		0.704
1. The information in the video is clear and understandable			0.496	0.459	0.488
4. The video generally met my expectations				0.774	0.698
5. Information about the video content was provided at the beginning				0.744	0.591
2. The video provided new knowledge and skills				0.702	0.577
Eigen value	4.817	2.151	1.303	1.201	
Explained variance (%)	32.112	14.340	8.684	8.004	
Total explained variance (%)	32.112	46.453	55.137	63.140	

($\rho = 0.654$, $p < 0.001$), Sect. 2 ($\rho = 0.377$, $p < 0.001$), Sect. 3 ($\rho = 0.320$, $p < 0.001$), and total scores of DISCERN ($\rho = 0.564$, $p < 0.001$).

Reliability

Regarding internal consistency, the Cronbach's alpha value was 0.81, 0.78, 0.75, and 0.73 for factors 1–4, respectively. The Cronbach's alpha reliability coefficient for the overall MQ-VET questionnaire was 0.72.

Discussion

Collectively, our results confirmed the validity and reliability of the MQ-VET questionnaire. Although previous publications have discussed YouTube medical video quality [16–18], standardized assessment tools were not utilized. Typically, de novo questionnaires were devised based on the literature and the authors' own knowledge [1, 5]. Several tools exist for the evaluation of written online information [5]. However, the applicability of these tools to videos is not known [7]. The DISCERN questionnaire

is designed to evaluate treatment options [10]. Thus, it is inappropriate for analyzing videos lacking treatment information. The JAMA questionnaire, the GQS, and the HON-code were also created for the evaluation of the medical sites and written information on the internet. The VPI was designed specifically for evaluating videos, but assesses popularity rather than quality and content. Scores based on popularity change over time, which impacts repeatability [11].

The MQ-VET resolves the aforementioned issues, and its validity and reliability have been demonstrated for a variety of medical topics. Also, the MQ-VET was designed for use by both medical professionals and the general population. Evaluation of additional medical topics by more reviewers will provide further support for the MQ-VET, while translation into other languages will increase its utility. This study was limited by the low number of participants and videos, and by the lack of the test–retest reliability of the MQ-VET. However, we believe that these problems will be resolved in future studies (Table 4).

In conclusion, we have developed a questionnaire to evaluate the quality of online medical videos posted by both medical professionals and members of the general public. We believe that this tool will help standardize evaluations of online videos.

Table 4 Final version of the Medical Quality Video Evaluation Tool

Item	Strongly disagree (1 point)	Disagree (2 points)	Neutral (3 points)	Agree (4 points)	Strongly agree (5 points)
Part 1					
1. Dates of updates, if any, are clearly stated					
2. The recording date of the video and date on which the information was accessed are mentioned					
3. The resources and references used are clearly stated					
4. Concerns about advertising and potential conflicts of interest have been resolved					
5. Sufficient information was provided about the identity of the presenter in the video					
Part 2					
6. The materials used in the video facilitated learning					
7. The video covered the basic concepts of the subject					
8. To explain the medical topic, visual resources were used sufficiently					
9. The medical terms used were well-explained					
Part 3					
10. The sound quality of the video was sufficient					
11. The image quality of the video was sufficient					
12. The information in the video is clear and understandable					
Part 4					
13. The video generally met my expectations					
14. Information about the video content was provided at the beginning					
15. The video provided new knowledge and skills					

Declarations

Ethics approval Not required. Since no human participants were used in our study, no ethical committee application was made.

Conflict of interest The authors declare no competing interests.

References

- Madathil KC, Rivera-Rodriguez AJ, Greenstein JS et al (2015) Healthcare information on YouTube: a systematic review. *Health Informatics J* 21(3):173–194. <https://doi.org/10.1177/1460458213512220>
- Atkinson N, Saperstein S, Pleis J (2009) Using the internet for health-related activities: findings from a national probability sample. *J Med Internet Res* 11(1):e5. <https://doi.org/10.2196/jmir.1035>
- Burgess J, Green J (2009) *YouTube: online video and participatory culture*. Polity Press, Cambridge
- Keelan J, Pavri-Garcia V, Tomlinson G et al (2007) YouTube as a source of information on immunization: a content analysis. *JAMA* 298(21):2482–2484. <https://doi.org/10.1001/jama.298.21.2482>
- Drozdz B, Couvillonn E, Suarez A (2018) Medical YouTube videos and methods of evaluation: literature review. *JMIR Med Educ Jan-Jun* 4(1):e3. <https://doi.org/10.2196/mededu.8527>
- Bernard A, Langille M, Hughes S et al (2007) A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol* 102(9):2070. <https://doi.org/10.1111/j.1572-0241.2007.01325.x>
- Laversin S, Baujard V, Gaudinat A et al (2011) Improving the transparency of health information found on the internet through the honcode: a comparative study. *Stud Health Technol Inform* 654–658
- Silberg WM, Lundberg GD, Musacchio RA (1997) Assessing, controlling, and assuring the quality of medical information on the internet: Caveant lector et viewor—Let the reader and viewer beware. *JAMA* 277(15):1244–1245
- Khazaal Y, Chatton A, Cochand S et al (2009) Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. *Patient Educ Couns* 77(1):33–37
- Charnock D, Shepperd S, Needham G et al (1999) DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Commun H* 53(2):105–111. <https://doi.org/10.1136/jech.53.2.105>
- Erdem MN, Karaca S (2018) Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. *Spine* 43(22):E1334–E1339. <https://doi.org/10.1097/BRS.0000000000002691>
- Tolu S, Yurdakul OV, Basaran B et al (2018) (2018) English-language videos on YouTube as a source of information on self-administer subcutaneous anti-tumour necrosis factor agent injections. *Rheumatol Int* 38(7):1285–1292. <https://doi.org/10.1007/s00296-018-4047-8>
- Basch C, Hillyer G, Garcia P et al (2019) Content of widely viewed YouTube videos about celiac disease. *Public Health* 167:147–151. <https://doi.org/10.1016/j.puhe.2018.11.004>
- Basch CH, Fung I-H, Menafo A et al (2017) An exploratory assessment of weight loss videos on YouTube™. *Public Health* 151:31–38. <https://doi.org/10.1016/j.puhe.2017.06.016>
- Shrestha N (2021) Factor analysis as a tool for survey analysis. *Am J Appl Math Sta* 9(1):4–11. <https://doi.org/10.12691/ajams-9-1-2>
- Ozsoy-Unubol T, Alanbay-Yagci E (2020) YouTube as a source of information on fibromyalgia. *Int J Rheum Dis*. <https://doi.org/10.1111/1756-185X.14043>
- Kocuyigit BF, Nacitarhan V, Koca TT et al.(2019) YouTube as a source of patient information for ankylosing spondylitis exercises. *Clin Rheumatol Jun* 38(6):1747–1751. <https://doi.org/10.1007/s10067-018-04413-0>

18. Gokcen HB, Gumussuyu G (2019) A quality analysis of disc herniation videos on YouTube. *World Neurosurg* Feb 2. <https://doi.org/10.1016/j.wneu.2019.01.146>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.