RESEARCH ARTICLE

Accuracy and Quality of Educational Videos for Elbow Physical Examination: A Search from the Earliest Year until October 2018

Elisa L. Zwerus, MD[1,2]; Rik J. Molenaars, MD[1,3]; Michel P.J. van den Bekerom, MD, PhD[4]; Bertram The, MD, PhD[5]; Denise Eygendaal, MD, PhD[1,5]

Research performed at Department of Orthopedic Surgery, Amsterdam UMC, Location AMC Meibergdreef , Amsterdam, The Netherlands

Received: 25 March 2020 *Accepted: 07 May 2020*

Abstract

Background: Medical students and residents rely increasingly on web-based education. Online videos provide unique opportunities to share knowledge. The objective of this study was to investigate the accuracy and quality of instructional videos on the physical examination of the elbow and identify factors influencing educational usefulness.

Methods: On October 7, 2018, a search on YouTube, VuMedi, Orthobullets, and G9MD was performed. Videos were rated for accuracy and quality by two independent authors using a modified version of a validated scoring system for the nervous and cardiopulmonary system. Inter-rater reliability was analysed.

Results: The 126 included videos were uploaded between June 2007 and February 2018. Twenty-three videos were indicated as useful for educational purposes. Accuracy, quality and overall scores were significantly higher for videos from specialized platforms (VuMedi, Orthobullets, G6MD) compared to YouTube. Video accuracy and quality varied widely and were not correlated. Number of days online, views, and likes showed no or weak correlation with accuracy and quality. For the overall score, our assessment tool showed excellent inter-rater reliability.

Conclusion: There is considerable variation in accuracy and quality of currently available online videos on the physical examination of the elbow. We identified 23 educationally useful videos and provided an assessment method for the quality of educational videos. In educational settings, this method may help students to assess video reliability and aid educators in the development of high-quality instructional online content.

Level of evidence: III

Keywords: Education, Educational video, Elbow, Physical examination, YouTube

Introduction

Driven by increasing emphasis on problem-based and self-directed learning, medical students and doctors in orthopaedic specialty training rely increasingly on the internet (Google, YouTube) as learning resource (1-3). As students' or residents' performance on physical examination may be less supervised in comparison to other clinical skills, such as surgical competence, online videos may provide a valuable source for education of physical examination skills. Cognitive psychological research has shown that videos can help viewers to understand techniques and manage the sequential steps of physical examination and approach of patients (4-9). YouTube is the largest open-access video sharing platform available with over four

Corresponding Author: Elisa L. Zwerus, Department of Orthopaedic Surgery, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands
Email: elisazwerus@gmail.com

AB&JS THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

billion videos watched every day. YouTube offers a wide variety of user-generated and corporate media videos, including video clips, TV show clips, music videos, short and documentary films, audio recordings, movie trailers, live streams, and other content such as video blogging, short original videos, and educational videos. In this latter category, YouTube provides access to educational videos on a wide variety of orthopaedics-related topics. Compared to YouTube, VuMedi, G9MD, and Orthobullets are online platforms with video content that is more directly focused on orthopaedic topics, requiring user-registration to obtain access.The accuracy and quality of educational videos for health care providers have been studied for physical examination of the shoulder, the nervous system, cardiovascular and respiratory systems and direct ophthalmoscopy (10-14). Overall, quality and accuracy showed a wide variety among videos, not related to the amount of views or likes. The same conclusion was reached by studies focusing on educational videos for medical and nursing students about subjects such as anatomy, electrocardiography, and pharmacokinetics (15-21). This variability in quality and accuracy makes it challenging for students and residents to identify educationally valuable videos for self-learning. A guideline on instruction videos for laparoscopic cholecystectomy was published in 2018, including a list with 45 statements (22). Unfortunately, most of the statements in this list are not applicable for videos on physical examination skills. Currently, there are no studies on the accuracy and quality of educational videos on physical examination of the elbow.

In this study, the accuracy and quality of videos on general and specific physical examination of the elbow available through YouTube, VuMedi, Orthobullets, and G9MD were assessed using a standardized scoring system. We hypothesized considerable variability in quality and accuracy of the physical examination of the elbow in currently available online videos, with higher quality and accuracy of videos available through specialized platforms (VuMedi, Orthobullets, and G9MD) compared to YouTube, and high inter-rater reliability of quality and accuracy assessment of online videos using the modified scoring system.

Materials and Methods

Search

A YouTube search was performed on October 7, 2018 using key words aiming at general physical examination, namely "elbow exam" and "OSCE" (objective structured clinical examination), and key words aiming at specific tests [Table 1]. YouTube (https://www.youtube.com/) search settings were standard (sorted by relevance) and filtered for individual videos. In addition, all available videos in VuMedi (https://www.vumedi.com/), Orthobullets (https://www.orthobullets.com/) and G9MD (https://g9md.tv/) in the elbow and upper extremity sections were reviewed.

Selection of videos

The first two-hundred videos on general physical examination and first fifty videos on specific elbow tests (sorted by relevance) on YouTube were included for initial screening. Because of the practically infinite output of the YouTube search engine, the authors decided on this arbitrary cut-off. All videos on overall and specific elbow examination on VuMedi, Orthobullets and G9MD were assessed. Videos had to be in English and in the format of an instructional video, i.e., videos in the form of a lecture (without moving video content), seminar, review, advertisement, and news or videos discussing history taking or symptoms were excluded. All videos were screened based on title and description on October 13, 2018. Videos were assessed and included in the study between October 13 and 31, 2018.

Data collection

For videos that met the inclusion criteria the following information was reported: title, duration of the video, URL, subject covered, days on YouTube, total number

Table 1. Search terms and hits

Search term	Hits
"Elbow exam"	101000
"Elbow OSCE"	45920
"Hook test biceps"	1690
"Biceps squeeze test"	2050
"Biceps crease interval test"	87
"Biceps crease ratio test"	7
"Triceps squeeze test rupture"	3560
"Valgus extension overload test"	57
"Posteromedial impingement test elbow"	265
"Arm bar test posteromedial impingement"	15
"Medial epicondylitis test"	355
"Valgus stress test elbow"	2790
"Moving valgus stress test elbow"	632
"Milking Maneuver test elbow"	70
"Mill's test elbow"	5630
"Maudsley's test elbow"	309
"Cozen's test elbow"	1870
"Grip and grind test elbow"	728
"Stand up test elbow"	10600
"Chair push-up test elbow"	3220
"Table-top relocation test elbow"	100
"Drawer sign elbow"	1630
"Lateral pivot shift test elbow"	1200
Elbow exam" section on VuMedi"	5
Elbow anatomy and evaluation" section on Orthobullets"	4
Elbow" search on G6MD"	8
Total	183802

of views, amount of likes and dislikes, and the name, profession and type of uploader/creating organization. The uploader type was categorized as follows: university/school, hospital, informative website, private, business/company, other or unknown. For the profession of the uploader the following categories were used: doctor (e.g., orthopaedic surgeon, sports physician, general practitioner or rheumatologist), physiotherapist, student, other or unknown.

Accuracy and quality assessment

Assessment of the accuracy and quality of the included videos was performed by two independent raters (ELZ and RJM) using a modified scoring tool based on previous studies in this field with excellent inter-rater reliability (12, 13, 15, 22). The scoring system is presented in [Figure 1]. Our modification aimed to increase focus on video accuracy instead of quality, including assessment of presentation of information on diagnostic accuracy. To rate the accuracy of the videos, a description of each test was provided to the raters as it was originally described in the literature (23). A total of 18 points could be obtained: eight for the accuracy assessment and ten for the quality assessment. Half points were not allowed. Both categories contained major (two points each) and minor (one point each) criteria. Videos that scored the maximum amount of points amongst the major criteria by both observers were defined as educational useful videos. To test the reliability of the modified scoring system, we performed reliability analysis for intra- and inter-rater reliability on the first 20 videos before proceeding with the full sample.

Statistical analysis

Data analysis was performed using Statistical Package for the Social Sciences 23 (IBM Corporation, Armonk, NY, USA). *P-values <0.05* were considered statistically significant.

Descriptive statistics and paired T-tests were used to examine demographics, accuracy, and quality of videos. 95% Confidence intervals (CI) were calculated using the group standard deviation in the total group and educationally non-useful videos. Due to the smaller sample size (<30), we used the total populations' standard deviation to calculate 95% CI for the educationally useful videos. Correlation between the accuracy and quality assessment was determined using Spearman's rho correlation coefficient. A correlation coefficient between 1.00 and 0.90 was considered very high, 0.90-0.70 high, 0.50-0.70 moderate, 0.50-0.30 low and 0.30-0.00 negligible, both in positive and negative direction(24). Pearson's correlation coefficient was used to assess the correlation between the numbers of days on YouTube, views, likes and dislikes using the same cut-off values for interpretation. Independent T-tests and contingency tables were used to evaluate the influence of uploader type and profession on scores and educational usefulness.

To evaluate the reliability of the scoring system, we calculated the intra-rater reliability of rater one (EZ) and the inter-rater reliability for rater one and two (EZ and RM) using the intra-class correlation coefficient (ICC) with a two-way random effects model for the first 20 videos. Inter-rater reliability was considered excellent for ICC values between 1.00 and 0.75, good for values between 0.74 and 0.60, fair for values between 0.59 and 0.40, and poor for values less than 0.40(25). It was not possible to calculate internal consistence using Cronbach's Alpha, due to the different scales for minor and major criteria and the small range (0-2 points and 0-1 points, respectively).

Finally, we calculated the inter-rater reliability of the scoring system for the complete dataset. Mean differences (MD) and the ICC with 95% CI were calculated using a two-way random effects model.

Accuracy assessment (total 8 points)

- Major criteria (2 points each):
  1. Arm position
  2. Examiner action
  3. Interpretation
- Minor criteria (1 point each):
  1. Verbal description
  2. Information on diagnostic accuracy

Quality assessment (total 10 points)

- Major criteria (2 points each):
  1. The video uses (simulated) patients to demonstrate the examination
  2. Images are clear
  3. Sounds are clear
  4. Creator/organisation providing the video are mentioned
- Minor criteria (1 point each):
  1. The video covers the topic identified in the title
  2. Designed at a level of undergraduate medical students

Figure 1. Scoring system.

Results

Search

Our search resulted in 183,802 initial hits. One-thousand six-hundred and fifty video titles on YouTube were screened and 17 video titles were screened on VuMedi, Orthobullets, and G9MD. Of all screened videos, 7.4% (122/1,650) of the YouTube videos and 24% (4/17) of the videos from VuMedi, Orthobullets, and G9MD were eligible, resulting in a total of 126 videos (122/126 from YouTube, 96.8%) included in our study [Figure 2].

Characteristics of included videos

The 126 included videos were uploaded between June 2007 and February 2018. Video statistics for the total group and sorted by educational usefulness with mean, range and 95% CI are summarized in [Table 2]. Because of small number, dislikes per day online were not analysed.

The uploader type varied from videos from informative websites (mostly by physiotherapists) (34%), private individuals without a website (27%), universities or university hospitals (22%), general or private hospitals (9%), and commercial companies (8%). For more than half of the videos the profession of the uploader was unknown (56%), with the other half of videos predominantly produced by physiotherapists (21%) and

medical doctors (17%).

The content distribution of the included videos is summarized in [Table 3]. In 119 out of 126 videos (94%) a specific test was performed, with or without additional general examination. Stability tests of the medial collateral ligament (MCL) (n=46, 39% of the videos with a specific test) and lateral collateral ligament (LCL) (n=37, 31%), and specific tests for lateral epicondylitis (n=47, 39%) and medial epicondylitis (n=24, 20%) were covered most frequently.

Pilot: reliability of the scoring system

Intra-rater reliability analysis showed excellent ICC's of the modified scoring system for accuracy (0.97; 95% CI

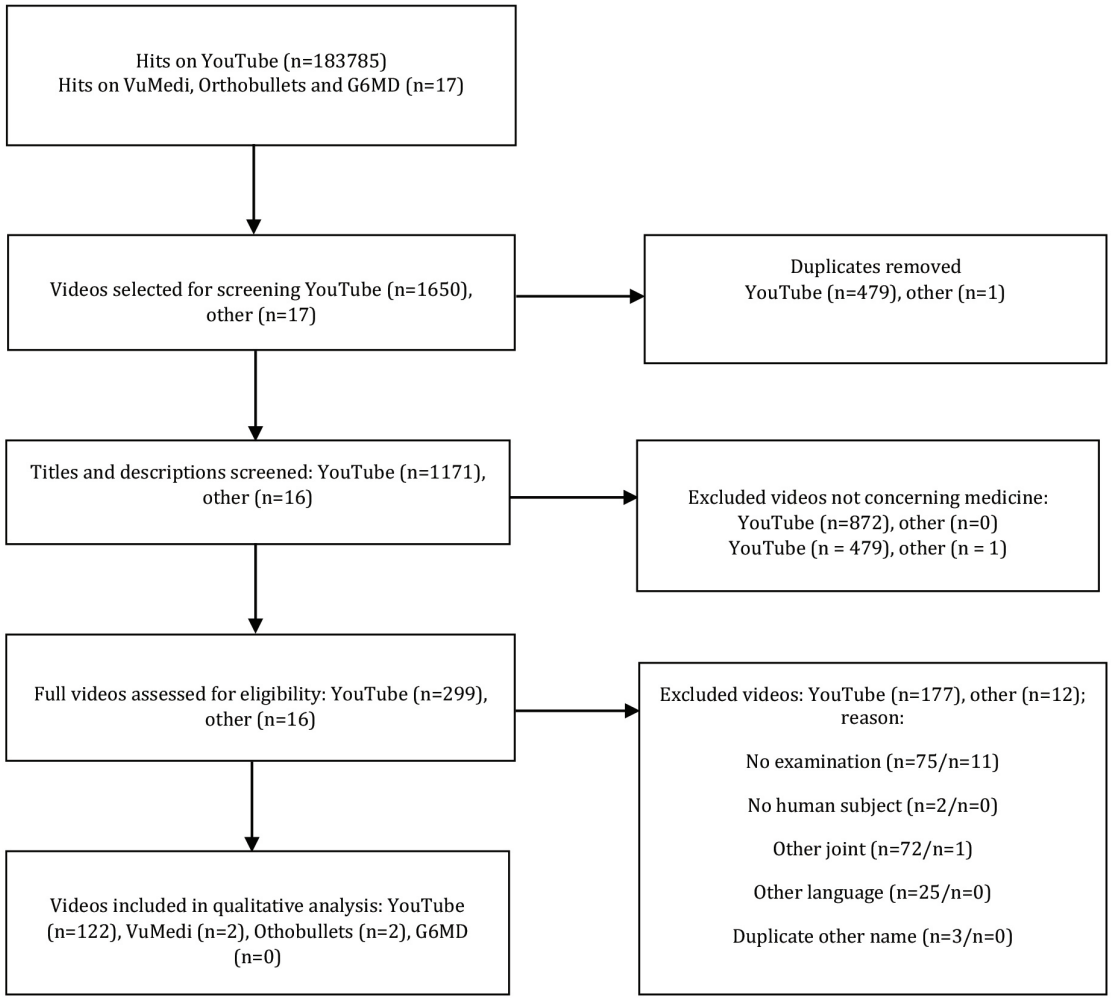

Figure 2. Flow chart of video selection.

Table 2. Characteristics of included videos

	Educational useful (n=23)	Educational not useful (n=103)	Total (n=126)	*P-value*
Days online (mean (95%CI))	1017.04 (842 to 1190)	1935.23 (1750 to 2120)	1624.25 (1450 to 1800)	*P<0.05*
Views total (mean (95%CI))	6168.30 (1650 to 10700)	20211.94 (14600 to 25800)	16954.79 (12400 to 21500)	*P<0.05*
Views per day (mean (95% CI))	5.69 (2.14 to 7.86)	8.47 (5.88 to 10.1)	7.96 (5.25 to 8.75)	*P<0.05*
Likes (mean (95%CI))	32.35 (21.1 to 43.7)	32.56 (19.3 to 45.9)	31.44 (20.0 to 42.8)	*P=0.989*
Likes per day (mean (95% CI)	0.03 (0 to 0)	0.02 (0 to 0)	0.02 (0 to 0)	*P=0.218*
Dislikes (mean (95%CI))	0.30 (-0.22 to 0.82)	1.99 (1.3 to 2.68)	1.68 (0.48 to 1.52)	*P=0.27*

Table 3. Content distribution of included videos

Content	Videos (n)	Percentage
History taking	6	4.8%
Inspection	22	17.5%
Anatomy	15	11.9%
Carrying angle	10	7.9%
Range of motion	23	18.3%
Specific test	119	94.4%
Biceps	15	11.9%
Triceps	3	2.4%
VEOS	7	5.6%
ME	24	19.0%
MCL	46	36.5%
LE	47	37.3%
LCL	37	29.4%
OCD	2	1.6%

0.93-0.99), quality (0.97; 95% CI 0.93-0.99) and overall score (0.98; 95% CI 0.94-0.99). The inter-rater reliability analysis of the pilot sample showed excellent ICC's for accuracy (0.87; 95% CI 0.70-0.95), quality (0.92; 95% CI 0.79-0.97) and overall score (0.94; 95% CI 0.86-0.98) as well. Therefore, we further analysed the full sample of 126 videos with the scoring system as described in the methods section.

Accuracy and quality assessment

The mean accuracy assessment score of the total sample was 5.6/8 points (95% CI 5.3 to 5.9), ranging from 0 to 8 points. For videos on VuMedi, Orthobullets and G9MD, the mean accuracy assessment score was significantly higher (mean 6.5; 95% CI 6.2 to 6.8) compared to videos on YouTube (mean 5.6; 95% CI 5.3 to 5.9) (*P<0.001*).

Out of the maximum 10 points, the mean score for quality assessment was 7.3 points (95% CI 7.0 to 7.7), with a range from 4 to 10 points. For videos posted on the specific platforms, the mean quality assessment score was significantly higher (mean 10; 95% CI 9.6 to 10.0) compared to videos available on YouTube (mean 7.2; 95% CI 6.9 to 7.6) (*P<0.001*).

Table 4. Educational useful videos

Title	Duration	Link	Content	Mean overall score
Orthopaedics Video 6 - examination of the Elbow	06:57	https://www.youtube.com/watch?v=qJRCP67NG9c	General, medial/lateral epicondylitis and LCL stability tests	17
Wrist and Elbow Examination	11:26	https://www.youtube.com/watch?v=V6tgYwatAYU	General, medial/lateral epicondylitis, LCL/MCL stability tests	17
Elbow Exam Tests - Sugar Land Houston - Dr. J. Michael Benett	09:28	https://www.youtube.com/watch?v=2IoopfRaR6o	General, biceps/triceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
The Reumatological examination of Elbows	02:18	https://www.youtube.com/watch?v=fYoeX5ZpC2w	General	17
Elbow exam	14:59	https://www.youtube.com/watch?v=Q-ef6inJ95E	General, biceps, medial/lateral epicondylitis, LCL/MCL stability tests	17,5
Musculoskeletal Examination and Joint injection Series_Examination of the Elbow	03:38	https://www.youtube.com/watch?v=LweM-dl6P2E	General	17
Physical Examination of the Elbow	11:43	https://www.orthobullets.com/video/view.aspx?id=1948	General, biceps/triceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
Elbow exam	13:33	https://www.vumedi.com/video/elbow-exam/	General, medial/lateral epicondylitis, LCL/MCL stability tests	17
Physical Examination of the Elbow	10:28	https://www.vumedi.com/video/physical-examination-of-the-elbow/	General, biceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
Biceps Squeeze Test	00:58	https://www.youtube.com/watch?v=HbOwTtbvrWM	Biceps test	17
LTA Diagnostic Tool: Biceps interval crease test	01:32	https://www.youtube.com/watch?v=OKL0PW5FD74	Biceps test	18

Table 4. Continued

The valgus extension overload test	01:38	https://www.youtube.com/watch?v=HYG2spWs1qs	VEOS test	17,5
Elbow Valgus instability Stress test | Medial Collateral ligament	01:28	https://www.youtube.com/watch?v=3xF9_5fbJ8A	MCL stability test	17,5
The moving valgus stress test for MCL tears of the Elbow	02:08	https://www.youtube.com/watch?v=JIU_kv5VoQk	MCL stability test	18
The posterolateral rotatory drawer test for elbow instability	01:38	https://www.youtube.com/watch?v=y_Lm78EH-vkM	LCL stability test	17
The Lateral Pivot Shift Apprehension Test | Posterolateral Rotatory Instability of the Elbow	01:26	https://www.youtube.com/watch?v=A3zExo7cmSc	LCL stability test	17
Elbow Varus Instability Stress test | Lateral Collateral Ligament	01:21	https://www.youtube.com/watch?v=5zl8GsG3hR4	LCL stability test	17
Moving Valgus stress test	01:03	https://www.youtube.com/watch?v=vI9LPr1w-00	MCL stability test	17
Moving Valgus Stress test	01:44	https://www.youtube.com/watch?v=4KeoOJb3864	MCL stability test	16
Elbow Valgus Stress Test	01:07	https://www.youtube.com/watch?v=3d2H_BAMsCE	MCL stability test	17
Lateral Epicondylitis Tests	00:30	https://www.youtube.com/watch?v=xa4op1Hv-L8	Lateral epicondylitis test	17
Push-Up Sign	00:49	https://www.youtube.com/watch?v=3KZqQCQJ9ac	LCL stability test	17
Chair sign	00:51	https://www.youtube.com/watch?v=azxNAY6Sr-w	LCL stability test	17,5

The mean overall score was 12.9 points (95% CI 12.4 to 13.5) out of a maximum of 18 points, ranging from 5 to 18 points. Mean overall score for videos on specific platforms was 16.5 (95% CI 16.0 to 17.0) and for videos on YouTube 12.8 (95% CI 12.3 to 13.3) (*P<0.001*).

In total, two out of 126 videos (1.6%) achieved the maximum accuracy score (8 points) by both observers (both available on YouTube). Thirty-one videos (24.6%) had the maximum score (10 points) on quality assessment, including all the educational videos from the specific platforms. In total, six out of 126 (4.8%) videos were given a maximum score (18 points) by one of the observers (all on YouTube), but in only two videos both observers agreed on the maximum score. Twenty-three videos (18.3%) fulfilled all major criteria by both observers and were therefore determined to be educationally useful, this includes three out of the four included videos posted on specific platforms [Table 4].

Correlation between the accuracy and quality assessment scores was considered weak, based on a Spearman's rho of 0.26 (*P=0.003*). The number of days online showed a weak negative correlation with the quality assessment (Pearson's rho of -0.26, *P=0.003*), and no correlation was found between the number of days online and video accuracy and overall score (*P=0.090* and *P=0.175*). The total amount of views did not correlate with assessment, quality, or overall scores (P=0.938, *P=0.674* and *P=0.878* respectively). The number of likes showed a weak correlation with the quality score (r = 0.29, *P<0.001*) and overall score (r = 0.2, *P=0.026*), but not with the accuracy of the video (r = -0.07, *P=0.941*). The total number of dislikes was too small (n = 212) to draw conclusions on correlations.

Educational usefulness

The mean number of days online and viewers was significantly lower in educational useful videos compared to less useful educational videos (*P<0.01*; *P=0.02*). No significant difference was observed for the amount of likes between these two groups (*P=0.99*). There was a significant association between the type of uploader and usefulness of video content (*P<0.001*). Most educationally useful videos, originated from websites, universities/university hospitals, and general hospitals (48%, 26%, and 26% respectively). None of the videos uploaded by private persons, companies or other/unknown uploaders were classified as educational useful. Within the uploader type, general hospitals and websites had the highest scores (55% and 26% of videos were useful respectively, [Figure 3]). Additional analysis of all included videos showed no significant difference for accuracy scores in comparison to content from websites, universities/university hospitals and general hospitals (*P=0.534*), but around 0.5 to 1 point difference in quality and overall scores (*P<0.001*).

There was no significant impact of uploaders' profession on educational usefulness (*P=0.110*). Notable differences were observed between unknown professions (only 11% of videos in this group was useful) and medical doctors, physiotherapists and students (38%, 19% and 20% respectively) [Figure 4]. Comparison of videos developed by unknown professions with videos from known professions showed a significant difference in quality and overall scores (*P<0.001*), but not in accuracy scores (*P=0.220*). The mean quality score was 8.1 for videos from known professions and 6.7 for videos from unknown professions, with mean overall scores of 13.9 and 12.2, respectively.

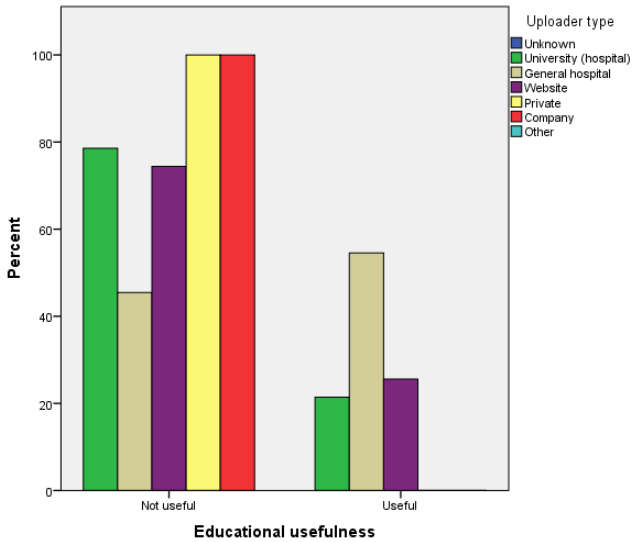

Figure 3. Bar chart showing influence of uploader type on educational usefulness.

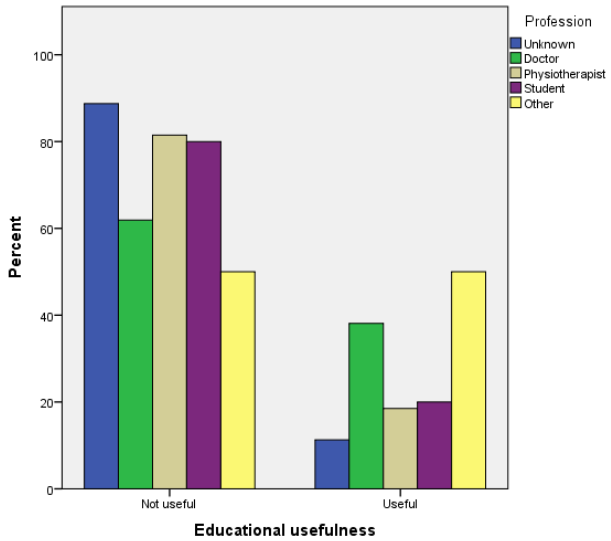

Figure 4. Bar chart showing influence of profession of uploader on educational usefulness.

Inter-rater reliability

For the accuracy and quality assessment, the mean difference between the two observers was significant (*P<0.001*), however, mean differences were less than 0.5 point. The mean difference for the overall score between rater one and rater two was not significant, with a mean difference of 0.024 (*P=0.871*).

Inter-rater reliability analysis for the accuracy assessment showed an excellent ICC of 0.80 (95% CI 0.70-0.86). Quality assessment also showed an excellent reliability with an ICC of 0.91 (95% CI 0.85-0.95). ICC's of the overall score was 0.93 (95% CI 0.09-0.95).

Discussion

This study assessed the accuracy and quality of currently available educational videos on physical examination of the elbow. The aim of this study was to provide characteristics of accurate and qualitative videos and a list of currently available educationally valuable videos. The findings of this study may guide students and residents in identifying educationally useful videos and to help health care providers with the development of future educational online content.

Main observations

A large variability of scores for both accuracy and quality were observed for the 126 videos reviewed in this study. Accuracy and quality assessment had a negligible correlation, suggesting that high-quality videos are not necessarily accurate in their presentation of physical examination and vice versa. Out of the included studies, 23 fulfilled all major criteria and were therefore classified as educationally useful. Notably, three out of the four included videos from specialized platforms (VuMedi and Orthobullets) were rated as educationally valuable and all showed higher accuracy, quality and overall scores compared to YouTube videos. Given the difficulty of regulating content on YouTube, these results are not surprising. This finding confirms our hypothesis that specialized platforms provide videos of higher quality and accuracy than videos from YouTube. Only two out of 126 videos achieved a maximum score by both observers. Overall, more videos received the maximum score for quality assessment than for accuracy assessment. One of the reasons for this is that only a few videos addressed diagnostic accuracy of the tests described. In previously described scoring systems, addressing the diagnostic accuracy was not a part of the accuracy assessment (12, 13, 15). However, to interpret the results of physical examination tests, knowledge on the diagnostic accuracy is of vital importance to rule in or out a diagnosis by physical examination. Notably, all examination videos on specialized platforms lacked information on the diagnostic accuracy. Interrater reliability of the modified scoring system showed excellent ICCs, especially when accuracy and quality assessment were combined. Mean scores for the combined accuracy and quality scores showed no significant difference.

With the abundance of online available content and increasing focus on self-learning, it is important for students and residents to identify educationally high quality videos. We observed that the majority of educational useful videos on elbow physical examination were provided by general hospitals and websites from physiotherapy practices or private individuals and to a lesser extent by universities and university hospitals. The limited contribution of university hospitals (as education institutions) may be explained by the fact that extensive physical examination of the elbow generally extends beyond the content of the general medical study curriculum. Between general hospitals and websites only quality scores differed (approximately 1 point higher score for websites compared to general hospitals), which may be explained by sponsorship or profit motives of more commercially oriented companies/websites. Considering

the profession of the uploader, our study show that when the profession is unknown, the quality and overall score of educational videos are lower compared to clips of known professions. Therefore, we recommend videos from providers of which the profession is presented for educational purposes. Comparing our findings with observations in previous literature on medical examination videos, only one study from Urch et al. (2016) compared videos about shoulder examination on YouTube with specialised platforms. Their conclusion supported our conclusion: videos on YouTube were less accurate compared to specialised platforms (VuMedi, G6MD and orthobullets) (10). All other studies available only included YouTube videos (11-15). Their results were mostly in accordance with our study. For example, Lee et al (2018) investigated the physical examination of the shoulder and found that videos originating from (known) physicians are more useful comparing to videos from (unknown) individuals (11). Also the limitations of the YouTube search algorithm and video regulations leading to a relative large number of screened videos and small number of educational useful videos were addressed in most studies. The studies by Azer (2012, 2012 and 2013) on the nervous system, surface anatomy and cardiorespiratory system used the scoring system we based our modification on and showed comparable scores for educational useful and non-useful videos (12, 13, 15).

Strengths and limitations

Because YouTube content changes continuously, potentially useful videos may be missed. However, search results often not match with search terms: our search yielded almost 200.000 hits, of which many did not focus on physical examination, medicine, and/or the elbow joint. Therefore we pragmatically decided to only screen titles of the first 1,650 videos. This problem is not present on video platforms that target medical professionals, however these platforms are often not known by non-specialized care providers such as medical students, general practitioners and physiotherapists and provide less videos. Improvement of YouTube's search algorithm system may lead to more accurate matching of search terms and resulting video content.

The inclusion of videos for this study was limited to videos in English, so that the information is understandable for most viewers. These factors might have led to a selection bias and a limited amount of educationally useful videos. Furthermore, the scoring system used in this study was not validated before, but limited modifications were made to the previously validated scoring system and our analysis showed excellent inter-rater reliability.

Recommandaties and future directions

Teachers and clinical supervisors should be aware that students and residents use open-access online learning platforms such as YouTube and recognize its pitfalls. Our study shows that numerous teachers and clinicians are creating online content, but that these videos frequently lack accuracy and/or quality. We advise content creators to film in a quiet room with good quality electronic video capture systems. Creators should provide step-by-step physical examination with clear verbal instructions and images. Furthermore, we advise the examiners to introduce themselves by providing information on their profession and institution. In order to enable the viewer to interpret the physical examination properly, information on diagnostic accuracy should be provided when available; in cases where there are no diagnostic accuracy studies available this should be indicated as well. Students and residents should be aware that not all online content is accurate and qualitative and that videos posted on specialized platforms such as VuMedi and Orthobullets are generally more educational useful compared to videos on YouTube. Viewership, likes/dislikes, and days online are not appropriate to determine whether or not a video is educationally useful.

For future research, it would be interesting to investigate the educational value of videos for students and residents in comparison to written content with or without static images and live education by a teacher or clinician. Furthermore, it may be useful to develop creator guidelines for physical examination videos for educational purpose such as the LAP-VEGaS guidelines for videos on laparoscopic cholecystectomy (22). The modified assessment tool used in this study might serve for this goal.

In this study, we indicated 23 educational useful videos on the physical examination of the elbow. Videos posted on specialized platforms, such as VuMedi and Orthobullets, by a creator of which the profession is known are generally more reliable compared to videos of which the creator's background and/or institution is unknown. Viewership, likes/dislikes, and days online do not indicate usefulness. The assessment tool used in this study for evaluating accuracy and quality of videos is easy to apply and covers key elements of good-quality educational videos. The tool can be used by students and residents to assess reliability of educational video content and aid educators in the development of new online content.

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

Elisa L. Zwerus MD[1,2]
Rik J. Molenaars MD[1,3]
Michel P.J. van den Bekerom MD PhD[4]
Bertram The MD PhD[5]
Denise Eygendaal MD PhD[1,5]
1 Department of Orthopaedic Surgery, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands
2 Department of Orthopaedic Surgery, Zuyderland MC, Sittard-Geleen, The Netherlands
3 Harvard Medical School at Massachusetts General Hospital, Sports Medicine Center, Boston, Massachusetts, USA
4 Shoulder and Elbow Unit, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands
5 Department of Orthopaedic Surgery, Amphia Hospital, Breda, The Netherlands

References

1. Keene AB, Shiloh AL, Dudaie R, Eisen LA, Savel RH. Online testing from Google Docs to enhance teaching of core topics in critical care: a pilot study. Med Teach. 2012;34(12):1075-7.
2. Kingsley K, Galbraith GM, Herring M, Stowers E, Stewart T, Kingsley KV. Why not just Google it? An assessment of information literacy skills in a biomedical science curriculum. BMC Med Educ. 2011;11:17.
3. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ. 2006;6:41.
4. Smith W, Rafeek R, Marchan S, Paryag A. The use of video-clips as a teaching aide. Eur J Dent Educ. 2012;16(2):91-6.
5. Casado MI, Castano G, Arraez-Aybar LA. Audiovisual material as educational innovation strategy to reduce anxiety response in students of human anatomy. Adv Health Sci Educ Theory Pract. 2012;17(3):431-40.
6. Srivastava G, Roddy M, Langsam D, Agrawal D. An educational video improves technique in performance of pediatric lumbar punctures. Pediatr Emerg Care. 2012;28(1):12-6.
7. Aronson ID, Plass JL, Bania TC. Optimizing educational video through comparative trials in clinical environments. Educ Technol Res Dev. 2012; 60(3):469-82.
8. Martens MJ, Duvivier RJ, van Dalen J, Verwijnen GM, Scherpbier AJ, van der Vleuten CP. Student views on the effective teaching of physical examination skills: a qualitative study. Med Educ. 2009;43(2):184-91.
9. Fincher RM, Simpson DE, Mennin SP, Rosenfeld GC, Rothman A, McGrew MC, et al. Scholarship in teaching: an imperative for the 21st century. Acad Med. 2000;75(9):887-94.
10. Urch E, Taylor SA, Cody E, Fabricant PD, Burket JC, O'Brien SJ, et al. The Quality of Open-Access Video-Based Orthopaedic Instructional Content for the Shoulder Physical Exam is Inconsistent. HSS J. 2016;12(3):209-15.
11. Lee H, Choi A, Jang Y, Lee JI. YouTube as a learning tool for four shoulder tests. Prim Health Care Res Dev. 2018:1-7.
12. Azer SA, Aleshaiwi SM, Algrain HA, Alkhelaif RA. Nervous system examination on YouTube. BMC Med Educ. 2012;12:126.
13. Azer SA, Algrain HA, AlKhelaif RA, AlEshaiwi SM. Evaluation of the educational value of YouTube videos about physical examination of the cardiovascular and respiratory systems. J Med Internet Res. 2013; 15(11):e241.
14. Borgersen NJ, Henriksen MJ, Konge L, Sorensen TL, Thomsen AS, Subhi Y. Direct ophthalmoscopy on YouTube: analysis of instructional YouTube videos' content and approach to visualization. Clin Ophthalmol. 2016;10:1535-41.
15. Azer SA. Can "YouTube" help students in learning surface anatomy? Surg Radiol Anat. 2012;34(5):465-8.
16. Azer SA. Understanding pharmacokinetics: are YouTube videos a useful learning resource? Eur Rev Med Pharmacol Sci. 2014;18(13):1957-67.
17. Akgun T, Karabay CY, Kocabay G, Kalayci A, Oduncu V, Guler A, et al. Learning electrocardiogram on YouTube: how useful is it? J Electrocardiol. 2014;47(1):113-7.
18. Burton A. YouTube-ing your way to neurological knowledge. Lancet Neurol. 2008;7(12):1086-7.
19. Raikos A, Waidyasekara P. How useful is YouTube in learning heart anatomy? Anat Sci Educ. 2014; 7(1):12-8.
20. Sharoff L. Integrating YouTube into the nursing curriculum. Online J Issues Nurs. 2011;16(3):6.
21. Cardoso AF, Moreli L, Braga FT, Vasques CI, Santos CB, Carvalho EC. Effect of a video on developing skills in undergraduate nursing students for the management of totally implantable central venous access ports. Nurse Educ Today. 2012;32(6):709-13.
22. Celentano V, Smart N, McGrath J, Cahill RA, Spinelli A, Obermair A, et al. LAP-VEGaS Practice Guidelines for Reporting of Educational Videos in Laparoscopic Surgery: A Joint Trainers and Trainees Consensus Statement. Ann Surg. 2018;268(6):920-6.
23. Zwerus EL, Somford MP, Maissan F, Heisen J, Eygendaal D, van den Bekerom MP. Physical examination of the elbow, what is the evidence? A systematic literature review. Br J Sports Med. 2017.
24. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71.
25. Cicchetti DV. Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research. J Clin Exp Neuropsychol. 1994;16(1):155-61.