A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness

EVELYN NAVARRETE, L3S Research Center, Leibniz University, Germany

ANDREAS NEHRING, Institute for Science Education, Germany

SASCHA SCHANZE, Institute of Science Education, Germany

RALPH EWERTH, L3S Research Center, Leibniz University, Germany and TIB – Leibniz Information Centre for Science and Technology, Germany

ANETT HOPPE, L3S Research Center, Leibniz University, Germany and TIB – Leibniz Information Centre for Science and Technology, Germany

People increasingly use videos on the Web as a source for learning, be it for daily tasks or in formal as well as informal educational settings. To support this way of learning, researchers and developers are continuously developing tools, proposing guidelines, analyzing data, and conducting experiments. However, it is still not clear what characteristics a video should have to be an effective learning medium. In this paper, we present a comprehensive review of 257 articles on video-based learning from computer science databases for the period from 2016 to 2021 using the PRISMA guidelines. One of the aims of the review is to identify the video characteristics that have been explored by previous work. Based on our analysis, we suggest a taxonomy which organizes the video characteristics and contextual aspects into eight categories: (1) audio features, (2) visual features, (3) textual features, (4) instructor behavior, (5) learners' activities, (6) interactive features (quizzes, etc.), (7) production style, and (8) instructional design. Also, we identify four representative research directions: (1) proposals of tools to support video-based learning, (2) studies with controlled experiments, (3) data analysis studies, and (4) proposals of design guidelines for learning videos. We find that the most explored characteristics are textual features followed by visual features, learner's activities, and interactive features. Text of transcripts, video frames, and images (figures and illustrations) are most frequently used by tools that support learning through videos. The learner's activity is heavily explored through log files in data analysis studies, and interactive features have been frequently scrutinized in controlled experiments. We complement our review by contrasting research findings that investigate the impact of video characteristics on the learning effectiveness, report on tasks and technologies used to develop tools that support learning, and summarize trends in the development of design guidelines to produce learning videos.

Key Words and Phrases: Video-based Learning, web-based learning, literature review, video characteristics, video features

1 INTRODUCTION

Videos become increasingly a choice in the everyday life of a learner, facilitated not only by formal institutional platforms but also on commercial general-purpose video platforms such as YouTube. In May 2019 it was stated that about 500 hours of new content was uploaded every minute [39] on YouTube, where around 50% of the views have been associated with learning purposes [219]. Furthermore, it was predicted that 82% of all IP traffic in the world will be caused by video consumption by 2022 [37]. Learners who turn to video sources have to rely on the technologies offered by the platform for providing suitable learning material. Moreover, they need to trust the expertise (if any) of the content creators and instructors in designing and producing effective learning videos. However, even on dedicated

platforms videos are published with no guidelines or good practices in mind [85]. The reason is that, although efforts to better understand video-based learning (VBL) have increased [78, 196, 268], essential questions are still not answered, for example, what are the characteristics of a video[1] to support learning in the best way, according to individual needs, the learning goals, and other contextual aspects?

Previous VBL reviews have performed descriptive analysis on the growth tendency of the field, the research directions, the discipline addressed by the videos, the production style, the mode of use (i.e., main or supplementary learning medium), the context or setting (i.e., formal or informal learning), video characteristics, and other dimensions [78, 196, 268]. Nevertheless, they have put special emphasis on experimental research, and did not dive into different research directions.

In this paper, we present a comprehensive review of 257 articles on video-based learning from computer science databases for the period from 2016 to 2021 using the PRISMA guidelines [187]. Compared to other reviews, we have collected the largest number of articles and provide a comprehensive overview on VBL research. As a consequence, we can achieve a more robust analysis of the state of the art in the field. In contrast to previous reviews that focused on experimental research, we identify and consider four main types of research articles: (1) proposals of tools to support video-based learning, (2) studies with controlled experiments, (3) data analysis studies, and (4) proposals of design guidelines for learning videos. We present a close investigation of (a) video characteristics, (b) research approaches, (c) tasks and technologies used to develop tools, frameworks, and pipelines that support learning, (d) guidelines to design and create learning videos, as well as (e) findings from controlled experiments and data analysis. Furthermore, we derive a taxonomy that organizes video characteristics and contextual aspects into eight categories (and further sub-categories): (1) audio features, (2) visual features, (3) textual features, (4) instructor behavior, (5) learners' activities, (6) interactive features, (7) production style, and (8) instructional design. The presented results are relevant for different audiences: (a) platform providers who wish to support their users with the best possible content, (b) content producers who aim to develop effective learning resources, (c) teachers who want to enhance their teaching by adopting multimedia resources, and (d) researchers who seek to update on the current state of VBL, identify potential research gaps, and bring in ideas for further research.

This review is an extended version of our previous work [177], where we have presented a review of the literature on video-based learning (VBL) for only two years (2020 and 2021). Here, we expand the reviewed period to 2016-2021 and also include the most relevant results of the articles by contrasting, analyzing and summarizing their findings.

The rest of the paper continues with a discussion of related reviews (Section 2), the description of the methodology (Section 3), the analysis of the field of video-based learning, which provides an introduction to the importance of identifying the characteristics of learning videos (Section 4), and the results of the review (Section 5). After that, we discuss the impact of said video characteristics on learning effectiveness (Section 6). Finally, we draw conclusions and point out limitations and future work (Section 7).

2   RELATED WORK

To the best of our knowledge, four survey articles on video-based learning have been presented previously. A summary with the aspects of the VBL field they have reviewed is displayed in Table 1. Giannakos [78] focused increase in the number of publications on VBL, the shift of VBL research from theoretical to empirical, and the characteristics of research regarding the type, sample, discipline area, use of stable or mobile devices, and the context of the video usage

---

[1]In the context of this review, we understand video characteristics as specific attributes, features, or aspects of a video that might be relevant to VBL.

Table 1. Related survey articles: reference and publication year, temporal scope of the survey, number of reviewed papers, and reviewed characteristics

Reference	Review period	N° papers	Reviewed dimensions	
[78] (2013)	2000-2012	166	1) Type of research 2) Sample characteristics 3) Subject area	4) Technology type 5) Style of use
[268] (2014)	2003-2014	76	1) Learning effectiveness 2) Teaching methods	3) Design 4) Reflection
[196] (2018)	2007-2017	178	1) Video type 2) Population 3) Control of prior knowledge	3) Sample size 5) Video characteristics
[201] (2020)	2008-2019	39	Teacher perception and reflection	

(main, formal, and informal). Yousef et al. [268] summarized the findings of experimental research on the effectiveness of VBL, teaching methods when using videos, integration of tools with VBL content, and the use of videos to reflect on teaching methods. Poquet et al. [196] also concentrated on experimental studies. The authors present a descriptive analysis on the types of videos, the population to which the experiment was directed, if the experiment controlled the prior knowledge, the sample size, growth patterns in the number of publications, and the type of research (theoretical and empirical). In addition, the authors review the video characteristics that influence the learning effectiveness with special emphasis on the learning outcome. The most recent work was presented by Sablic et al. [201], who review research on the use of videos to support teacher's self-reflection and professional development. There are other reviews on learning from YouTube videos such as the recent work by Shoufan and Mohamed [215]. We have not considered this study or similar studies since they did not meet our criteria of having a computer science perspective and focusing on VBL in general, and not only on a specific platform.

All these reviews confirm the significant growth of research articles in the VBL field [78, 196, 268]. Giannakos [78] and Poquet et al. [196] reported a noticeable increase of empirical and quantitative research over theoretical and qualitative research. Furthermore, Giannakos [78] showed a shift in the research domain, from the social sciences to technological domains, which was also confirmed by the results of Poquet et al. [196] that science, technology, engineering, and mathematics (STEM) disciplines are currently a predominant research focus. Further important findings from Poquet et al. [196] suggested that the effectiveness of VBL has frequently been evaluated through two learning outcome metrics: recall (remembering information) and transfer (applying what was learned to a different scenario), which was also corroborated by Yousef et al. [268]. Other popular metrics of learning effectiveness were motivation, cognitive load, mental effort, attention, and affect [196]. Also, the most common sources to obtain data for analysis were clickstreams (e.g., views, likes, dislikes) and eye-tracking data. Regarding the video characteristics, Poquet et al. [196] found that the most examined characteristics were text, audio, animations, and voice over a PowerPoint presentation. Yousef et al. [268] found that trending research topics were the effects of annotations and authoring tools, collaborative VBL, and video usage for reflective teacher education.
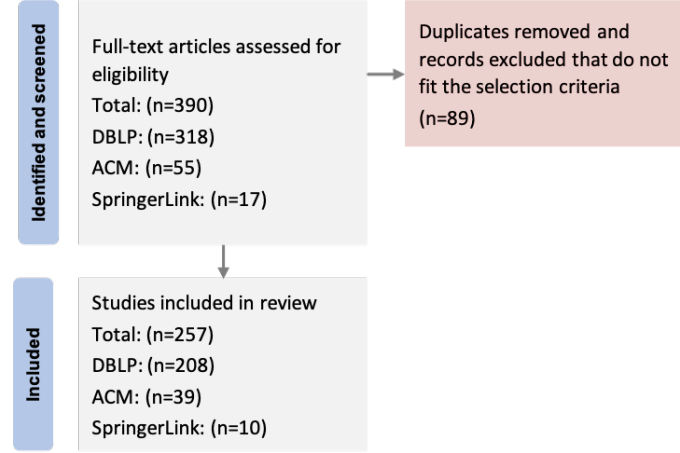
Fig. 1. The diagram shows the process of finding VBL-related articles in the databases. From an initial set of 390 papers, we removed duplicates and papers that did not fit the inclusion criteria. The final set consists of 257 articles from 2016 to 2021.

3 METHODOLOGY

The methodology for our systematic review was inspired by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [187]. They provide a guide to summarize results of reviewed research articles in an appropriate manner.

The papers covered by this review were found at SpringerLink [4], the Digital Library of the Association for Computing Machinery (ACM DL) [1], and the Digital Bibliography and Library Project (DBLP) [2]. These are focused in computer science (CS) domains, and therefore, fit well our task of reviewing the technological aspects of video-based learning. In each of these databases, we used advanced search options as a search strategy to refine our results. In ACM DL, we applied filters to retrieve publications that match the query terms only in the title and that fit the time frame. In SpringerLink, we applied the same filters complemented by an additional filter to obtain articles only in CS disciplines. In DBLP, however, we did not apply any filter since the search engine covers only CS and already retrieves articles that only match the query term in the title and sorts them all per year.

Table 2 shows the set of query combinations we used, which includes all the keywords from our previous work [177] plus new terms considered for this extended review (massive open online courses *MOOC* and *Lecture*). We followed an iterative process where the set of query terms was updated as new terms in the field of VBL were found (please see the search order in Table 2), and then performed a new search in the databases.

We included a paper in the final set only when the article: (a) considered videos in academic scenarios (this excludes video tutorials for everyday tasks such as cooking instructions), (b) focused on videos as learning resources (this excludes the use of classroom recordings for reflective tasks such as in teacher education), and (c) examined specific characteristics of video-based learning materials (this excludes, for instance, the use of videos to evaluate their viability in classroom). Figure 1 shows the obtained articles using the PRISMA-guided search process. The final set of articles consists of 257 articles published in the period from 2016 to 2021.

In our data collection process, we noted the article's title, authors, year, source database, matching terms of our query set, research direction, research objectives, video characteristics, and, if applicable, main findings, discipline of

Table 2. This table shows the set of query terms used to find relevant articles. The base term in the first column was combined with each of the extensions in the second column (using logical AND as a connector).

Base	Extension			
Video	1.1) Teach	1.2) Teaching	2.1) Education	2.2) Educational
	3.1) Instruction	3.2) Instructional	4.1) Learn	4.2) Learning
	5) -based Learning	6.1) Explanatory	6.2) Explaining	7) Tutorial
	8) Classroom	9) Student	10.1) Knowledge	10.2) Knowing
	11) Skills	12) University	13) School	14) Pupil
	15) MOOC	16) Lecture		

the video, study sample size, video production style(s), dependent variables, and target tasks and technologies. All the information was assembled in a data table to facilitate efficient data access and analysis.

4 VIDEO-BASED LEARNING

In this section, we discuss important aspects of video-based learning, potentials, challenges, and video characteristics. This lays the groundwork and motivates our literature review in the following section.

Potentials and Challenges: Video-based learning is a terminology that has already been adopted in several studies (e.g., [78, 196, 201, 268]), and can be considered as the task of learning using video sources, that is, acquiring knowledge and skills through video content. Videos have been implemented in classrooms as a teaching medium [196] and they form an important part of students' habits [219]. The trends show that videos might become the main learning medium in online education [86] due to its wide availability. Learning platforms, such as MOOCs, can disseminate videos in simple and scalable ways [78, 85].

The debate about the benefits of VBL continues. Yousef et al. [268], for instance, concluded from their results that learning videos are effective because they can improve the learning outcome, motivation, engagement, satisfaction, interaction, and communication between learners. They further argued that videos can support various learning styles and convey information otherwise hard to communicate by modalities such as plain text. However, the question of how to further support different learners' needs remains open, since, for example, VBL usually does not allow synchronous communication and feedback as in classic classroom settings.

Moreover, further issues have been found in the way videos are designed and created. Guo et al. [85] found in their interviews to video producers that "video editing was not done with any specific pedagogical "design patterns" in mind" [85, p. 45] and "MOOC video producers currently base their decisions on anecdotes, folk wisdom, and best practices distilled from studies with at most dozens of subjects and hundreds of video watching sessions" [85, p. 42]. Additionally, methods from typical classroom scenarios have been transferred to create learning videos, however, it is unclear whether this can be successful [85].

Examination of video characteristics: Learning videos can be a complex learning medium since audio, visual, textual, and a large set of derived features might collide when expressing a message. This complexity leads us to the following research question: Which of these characteristics make a video effective for learning? The analysis and exploration of characteristics in learning videos have been examined for different purposes, from experiments that try to understand the impact on the learning effectiveness to the construction of tools for supporting learning. Consequently, there is a multitude of video characteristics that have been extracted, studied, and manipulated. These can range from

low-level characteristics (e.g., text color, visual luminance, audio energy) to high-level characteristics such as models of a video's semantic content.

Video-based learning does not only involve the video and its instructors but also the surrounding hosting system or e-learning platform. Therefore, we analyzed characteristics related to the platform, for example, those related to the learner's activity, such as the usage of play and pause buttons, or number of views and likes. Also, more complex characteristics attempt to consider the human aspects in the video (e.g., instructor gestures and speech). In out analysis, we focused exclusively on characteristics related to the video content and its hosting platform while excluding learner-related characteristics, as for instance, previous knowledge, age, gender, etc.
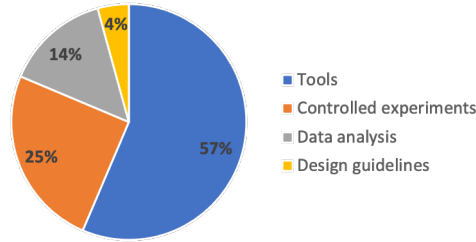
5   LITERATURE REVIEW

Fig. 2. The diagram shows the share of the different research directions identified in this review. Tools and studies that perform controlled experiments constituted the main research directions.

In this section, we present the results of our review and, briefly, the content of the relevant papers. Based on our analysis, we suggest to distinguish between four representative research directions: (1) tools, (2) controlled experiments, (3) data analysis, and (4) design guidelines. We found that the introduction and presentation of tools was present in more than the half of the articles, while reports about controlled experiments represented a quarter of the articles (see Figure 2).

In the following sub-sections, we report the results for each of these research directions, since this provides a natural way to analyze and contrast articles to draw conclusions. Thus, video characteristics are discussed differently in the context of each research direction. Table 3 summarizes the characteristics and contextual aspects identified in a single taxonomy, which we organized into eight main categories: (1) audio features, (2) visual features, (3) textual features, (4) instructor behavior, (5) learner's activities, (6) interactive features, (7) production style, and (8) instructional design. It is noticeable that specific characteristics fall naturally in specific categories of the taxonomy. For example, interactive features (e.g., quizzes and annotations) have been often studied in controlled experiments but little in data analysis studies (see Figure 3). Figure 4 depicts the structure of this section and the dimensions that were analyzed for each research type.

5.1   Tools, Pipelines, Frameworks, and Methods

In this section, we give an overview of the tools that the research community developed to assist and enhance video-based learning. We also report the automatic tasks carried out by the tools and the technologies that supported it.

Table 3. Taxonomy of video characteristics

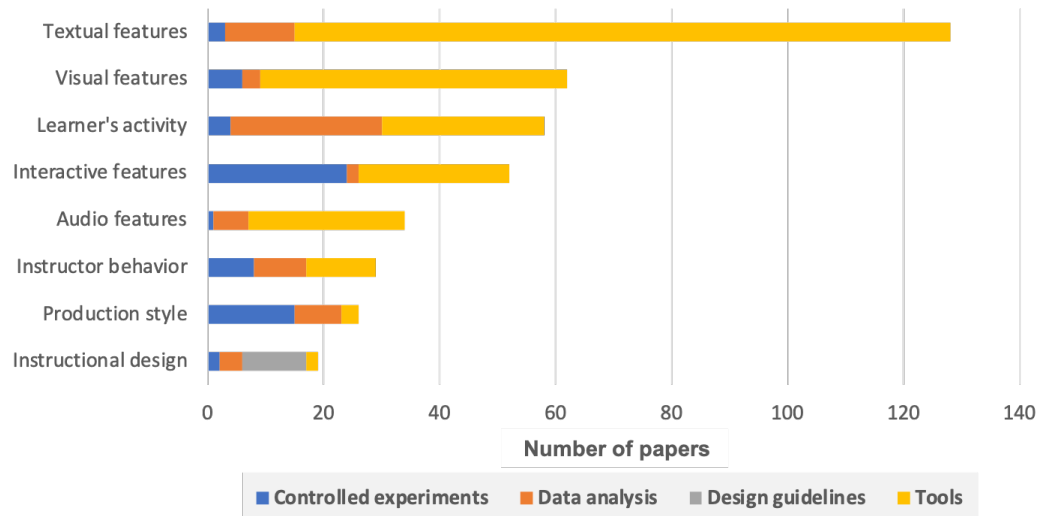| Category	Sub-category	Examples
Audio features	Audio records [14, 21, 31, 64, 65, 71, 81, 102, 109, 112, 131, 148, 160, 161, 182, 202, 207, 211, 220, 250, 273]	
	Sound-related [35, 56, 64, 76, 103, 112, 174, 211, 212, 214, 220]	Energy, entropy, spectral, jitter, shimmer, noise, Mel-frequency cepstral coefficients, perceived quality
	Representation characteristics (cues) [211, 253]	Instructor-related: verbal cues
Visual features	Video frames and images (figures and illustrations) [7, 14, 21, 29, 32, 38, 49, 50, 64, 65, 67, 72, 76, 80, 81, 83, 83, 102, 111, 111, 114, 115, 121, 123, 124, 131, 138, 148, 153, 158, 160, 161, 179, 182, 195, 198, 198, 207, 252, 255, 261–264, 264, 265, 273, 274, 274, 276]	
	Representation characteristics (cues) [110, 160, 170, 192, 193, 209, 209, 251, 253]	Instructor-related: pointing gestures, directed gaze. Content-related: highlighted content, arrows, moving objects
	Quality-related [56, 212, 214]	video quality, resolution in pixels
Textual features	transcripts [8, 12–14, 16, 17, 21, 22, 24, 31, 32, 40, 43, 48, 54, 61, 63, 64, 69, 71, 76, 92, 97, 102, 105, 107–111, 117, 127, 128, 128, 131, 152, 159, 160, 174, 176, 182, 195, 206, 211, 220, 221, 228, 233, 235, 243, 245, 250, 250, 252, 264, 267, 273, 275]	
	Metadata [16, 17, 25, 28, 32, 43, 48, 60, 61, 63, 70, 94, 108, 109, 128, 137, 140, 152, 167, 173, 183, 184, 194, 195, 203, 212, 216, 218, 225, 228, 231, 233, 235, 275]	Titles, keywords, tags, video length, comments, course syllabus, video url
	On-screen text [7, 14, 21, 29, 32, 38, 48–50, 52, 54, 67, 71, 72, 102, 114, 121, 124, 124, 131, 138, 152, 160, 160, 179, 182, 186, 195, 255, 257, 261, 262, 265, 273, 274, 276]	
	Representation characteristics [160]	Font size, boldeness
Instructor behavior	Body-related [18, 19, 35, 56, 80, 96, 98, 145, 214, 237, 254, 263]	Gestures: hand strokes and waving. Facial emotions, body language, pose, clothes, gender appearance
	Speech-related [12, 14, 28, 32, 56, 98, 105, 107, 111, 111, 160, 189, 212, 214, 221, 243]	Tone, speech rate, intonation, lexical diversity, syllable count, prosodic characteristics, readability, silence rate
	Attitudes [212]	Enthusiasm
Learner's activity	[9, 10, 12, 26–28, 36, 44, 45, 57, 60, 66, 74, 79, 89, 90, 99, 104, 122, 126, 127, 133, 134, 136, 137, 139, 140, 143, 146, 156, 164–168, 173, 184, 189, 200, 203, 205, 212–214, 227, 233, 235, 239, 243, 249, 250, 260, 266, 269, 270, 272]	Explicit: play, pause, speed rate, seek forward and backward. Implicit: viewing percentage, average of videos watched per week. Popularity-related: number of views, likes, dislikes, user ratings. Pace-related: automatic pauses, pauses controlled by the learner, type of pauses, fast pace.
Interactive features	[11, 15, 20, 33, 34, 41, 46, 47, 51, 53, 56, 62, 82, 84, 91, 100, 106, 110, 113, 114, 118–120, 125, 129, 130, 132, 142, 149–151, 154, 157, 162, 172, 175, 178, 188, 199, 208, 210, 217, 222, 234, 238, 240, 247–250, 259, 277]	Quizzes embedded in the video, virtual reality and 360-degree characteristics, in-video annotations, feedback to the learner, games, hyperlinks, interface elements for navigation
Production style	[6, 30, 42, 45, 49, 55, 56, 87, 88, 93, 94, 136, 140, 180, 181, 190, 191, 197, 207, 212, 214, 223, 224, 229, 230, 236]	Khan-style, with a talking head, dialogue-style or monologue-style, slide-based, with animations, paper-based explanations, classroom-style videos,
Instructional design and principles	[5, 58, 60, 75, 95, 101, 135, 140, 141, 169, 185, 212, 226, 241, 246, 256, 258, 270, 271]	Multimedia learning principles, guidelines for effective engaging, pedagogical characteristics of didactical mathematical knowledge, according to film theory, interaction theory, limited capacity theory
Others	[52, 116, 204, 211, 242]	Music, sign language, inclusion of a summary, content design, visual complexity, content design: text layout, image position, slide size, slide design, text ration, image ration

Fig. 3. The diagram shows the number of investigated video characteristics in the different research directions. Some video characteristics were more widely explored in certain research directions than in others.

Fig. 4. The diagram shows the dimensions we analyzed for every type of research.

We identified and classified the articles according to the following types of tools: (a) information extraction methods, (b) interactive tools, (c) recommendation systems, (d) classifiers, (e) video navigation and segmentation tools, (f) video analysis tools and methods, (g) video summarization, and (h) others, where the first five types represented about 80%
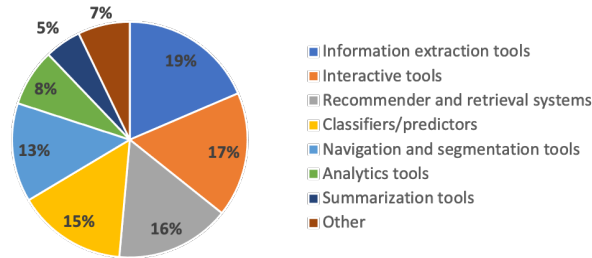
Fig. 5. types of tools that support video-based learning identified in this review

of all developed tools (see Figure 5). Also, please note that some articles proposed systems consisting of two or more tools [110, 114, 249, 250, 269].

Figure 6 shows the video characteristics that have been used in articles to develop VBL tools. It can be seen that text and visual features were the most used characteristics, mainly by recommender systems, information extraction and summarization methods, as well as navigational tools. Regarding tasks and technologies, we identified some low-level tasks performed as basic steps, of which, many rely on deep learning approaches. For example, for textual video elements, several studies employed optical character recognition [14, 21, 29, 32, 49, 50, 54, 64, 67, 102, 109, 114, 123, 138, 152, 160, 161, 179, 195, 255, 261, 262, 265, 273, 276], keyword extraction [14, 40, 43, 61, 92, 105, 106, 109, 112, 121, 128, 131, 161, 206, 255, 264, 271], generic natural language processing methods (e.g., [29, 127, 128, 194, 243]), or utilized word embeddings (e.g., [54, 69, 76, 92, 176, 238]). For image-related data, many studies suggested approaches for slide transition detection and video segmentation [17, 29, 32, 64, 67, 69, 71, 72, 77, 83, 110, 115, 153, 160, 178, 205, 220, 262, 265, 273]. For audio tracks, various studies carried out speech recognition to generate transcripts [14, 16, 21, 24, 29, 31, 40, 54, 64, 69, 71, 102, 106, 109, 142, 161, 182, 202, 250, 267, 273]. In the next subsections, we discuss the tasks and technologies in detail according to each type of tool. Within each subsection, we have grouped the tool proposals according to their purpose and the video characteristics that they have employed.

5.1.1 Information Extraction Tools.

The reviewed articles in this section focused mainly on the extraction of text from videos (handwritten or machine printed) and metadata, as well as the extraction of key information such as the extraction of key video segments.

Extraction of text: We found seven proposals that extracted handwritten text [49, 72, 123, 124, 138, 262] and two papers on extracting machine-printed text (e.g., from slide-based videos) [115, 276]. Davila et al. [49] retrieved handwriting using a convolutional neural network (CNN). Kota et al. [115] used multiscale histogram of gradients for feature representation and deep neural networks to extract handwriting. In another article, Kota et al. [123] detected bounding boxes with handwriting and used clustering algorithms to group them, which could identify also content that was occluded by the instructor when she was standing in front of the whiteboard. The handwriting was then extracted by enclosing it in bounding boxes and binarizing it. Lee et al. [138] detected changes in the the handwriting pixels and extracted them using an adaptive threshold and time-series denoising techniques. Davila and Zanibbi [50] extracted handwritten formulas using a random forest classifier trained on handwritten symbols. Xu et al. [262] identified the main instructor's actions such as writing and erasing as clues to identify the video content. The actions were detected using pose estimation technologies. In the same way, Garber et al. [72] identified handwriting through instructor's
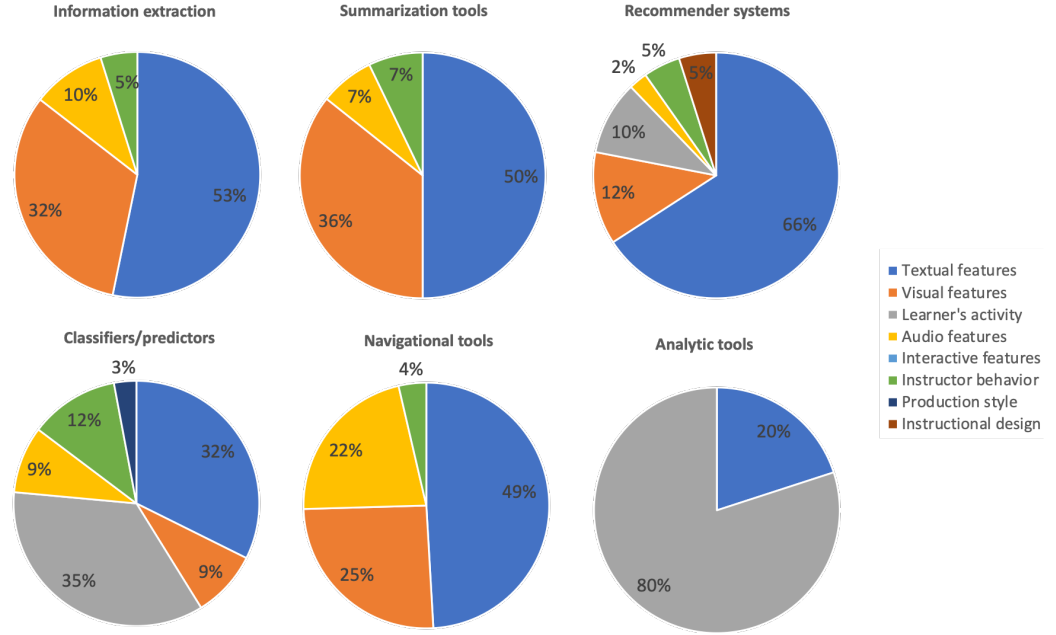
Fig. 6. The diagram displays the shares for targeted video characteristics used to develop tools for video-based learning. Textual and visual features were the mostly used characteristics, mainly by recommendation systems, information extraction, summarization, and navigational tools.

actions of writing and erasing. The instructor could be removed by masking the regions where there is a change among video frames. Zhong and Ma [276] extracted machine printed text using the maximally stable extremal regions method to detect candidate regions and clustering techniques were used to ensemble lines of text. Finally, Khattabi et al. [115] detected parts of the video that contained text using an AdaBoost meta-classifier.

Extraction of key information: Ghauri et al. [76] classified key segments in videos using different fusion schemes and bi-long-short-term memory network modules. Che et al. [32] detected the most important video segments using the slides structure and speech characteristics. Kravvaris and Kermanidis [128] identified important topics in the transcripts using Latent Dirichlet Allocation. They contrasted them with relevant terms from the learners' comments which were found using the Maximum Likelihood algorithm. Lin et al. [148] classified knowledge from non-knowledge by using a CNN that processes image and sound simultaneously. The classifier is trained on public datasets containing audio and visual human conversations, cooking events, and also collected video lectures; however, it was not clarified what the criteria were to classify something as "knowledge". Similarly, Liu et al. [152] extracted medical information according to their level of knowledge (low or high) by building a Logistic Regression classifier using the text from transcripts. Finally, for computer programming video tutorials, Ponzanelli et al. [195] classified segments according to the kind of content, for example, theoretical concepts, software source code, etc. The classifier employs temporal, structural, semantic, and code-related characteristics.

Extraction of metadata: Jung et al. [111] extracted keywords from transcripts using the TextRank algorithm. Tags were extracted using singular value decomposition techniques and other relevant information was extracted using entity recognition techniques. Escobar-Avila et al. [61] compared three algorithms (term frequency–inverse document

frequency, Latent Dirichlet Allocation using genetic algorithms, and the BM25F ranking function) to find the best technique to extract tags from video titles, description, and transcripts. Gomes et al. [109] extracted related topics with the help of a knowledge base, where a knowledge graph was built measuring the semantic similarity between concepts. The similarity was based on the number of resources (Uniform Resource Identifiers of instances in the knowledge base) the videos share and the categories that the resources share, and was calculated with the Sorensen-Dice coefficient. Furini [64] extracted tags by measuring the most frequent words. Ghosh et al. [77] detected concepts that are not familiar or do not belong to the domain of the video content but could be relevant to better understand the video. The concepts were linked to Wikipedia articles, which were used to build a concept similarity network. Unfamiliar topics were extracted using community detection techniques. Similarly, Nangi et al. [176] extracted unfamiliar topics, which were considered as cross-discipline concepts, also using community analysis in a concept graph.

Other: Part of the studies presented tools to extract key-phrases [8, 265], slides [83, 153], and software source code (from video tutorials to learn computer programming) [7, 114, 265].

5.1.2 Interactive Tools.

The development of interactive tools (tools that implement features allowing for communication and interaction between the video and the learner) mainly targeted in-video annotations, quizzes embedded in the video, and, interestingly, feedback to the learner for preventing mind wandering.

In-video Annotations: These tools offered a feature that allowed the learner to take notes while watching the video, and also were accompanied with other supporting characteristics. Fong et al. [62] suggested a tool in which the learner can interact not only through annotations but also interface elements for navigation. Liu et al. [151] allowed the learner to create annotations and highlight parts of the transcript. Malchow et al. [162] let the learner create annotations for each slide presented in the video. Lai et al. [132] did not only offer text-based annotations but also sketch-based annotations where the learner could draw into the video frame. Similarly, Singh et al. [217] created a tool for text-based and sketch-based annotations, and supported the student in comment creation, color-coding, and interface elements for navigation. Nguyen and Liu [178] automated the note taking process by tracking eye gaze. The tracking system automatically extracted relevant parts as candidate notes. Deb et al. [51] allowed the viewer to create annotations but did not introduce another supporting feature.

Quizzes: Herault et al. [91] proposed a 360-degree interactive environment with quizzes embedded in the video and feedback to support the learner. Kohen-Vacs et al. [120] introduced a tool to embed quizzes and announcements along the playback into the video. Walk et al. [250] included quizzes in the video with the aim of enhancing active learning. Wachtler et al. [249] presented an application that can include several types of questions such as single- and multiple-choice questions, as well as questions to the instructor. Ma and Ma [159] developed an automatic question generation method using the video transcript and an external knowledge graph. Wachtler and Ebner [247] created an application to schedule quizzes and other interactive features along the playing video.

Feedback to learner for preventing mind wandering: Lin et al. [149] introduced feedback in learning videos to alert the viewers in mind-wandering states. The tool kept track of brainwave signals to detect loss of focus and recommended to review again parts of the video. Jo and Lim [106] instead used a game that identified high-frequency words. To detect mind wandering, the game consisted in presenting a word to the learner and ask if that word was present previously in the video. Robal et al. [199] made use of web page, mouse, and face tracking technologies to determine whether the viewer was paying attention. If the user was inattentive, the system rang a bell, displayed a red frame around the video, and paused it. Sharma et al. [208] employed eye gaze tracking to identify mind wandering and

displayed red marks when distraction was detected. In a similar way, Llanda [155] offered a tool that provided feedback to the learner, although, not with regard to mind wandering but to the sentiment state. The tool tracked the viewers' emotions by analyzing face patterns and detecting negative sentiment to consequently suggest re-watching the video.

Other: Some studies suggested tools to create videos that integrate 360-degree characteristics [20, 91, 125]. Other papers introduced clickable objects [110, 238], interactive concept maps [150], feedback for the learners adapted to their previously identified proficiency level [277], and tools to interact directly with the video content for practicing, for example, image editing [118] and computer programming [15, 114].

5.1.3 Recommender and Retrieval Systems.

Recommender systems have mainly considered the similarity between transcripts and metadata to suggest new videos. There were, however, other interesting characteristics that have been exploited, for instance, video popularity (e.g., likes and views), learning style, teaching style, etc.

Based on the similarity of transcripts and metadata: Zhao et al. [275] suggested recommendations based on topic similarity, whereby the video is represented by means of its latent topics using the Latent Dirichlet Allocation algorithm. Waykar and Bharathi [255] retrieved videos based on keywords extracted from on-screen text and also the texture of keyframes using a texture histogram. Then, the videos were arranged with the help of Probability Extended Neighbor Classification. Bleoanca et al. [22] retrieved the top-N videos based on topic similarity using Latent Semantic Indexing to extract them, clustering techniques to group videos by discipline, and Wikipedia articles to augment the query context. Basu et al. [16] recommended videos when the learner was navigating Wikipedia pages. They used the Latent Dirichlet Allocation algorithm to extract relevant topics from Wikipedia content and video transcripts. Jung et al. [111] recommended similar videos based on keywords, tags, and the readability of the video. In a similar manner, Ji et al. [105] generated recommendations using metrics of linguistic difficulty and similarity of keywords identified by the TextRank algorithm. Jordán et al. [108] built a system that considers not only the characteristics of the current video but also previous videos watched by the learner. Walk et al. [250] made use of word embeddings constructed from the video transcript and compared them against the user search query to retrieve the most similar video segment. Tavakoli et al. [235] built a system to recommend videos for learning skills required in the labor market. The authors use the skills that can be learned from the video and the skills demanded by the market together with video popularity features (e.g., likes and views) and length.

Based on ontologies and external knowledge bases: Schulten et al. [206] determined related concepts through external ontologies based on Wikipedia articles and extracted keywords using the term frequency–inverse document frequency algorithm to discover new learning resources through a web search API. In the same way, Borges and Reis [24] extracted concepts and their relationships from an existing knowledge base to recommend videos. Coccoli and Vercelli [40] extracted concepts and also related concepts by linking the content to Wikipedia articles. The learners could receive recommendations of videos and also Web pages. Borges and Silveira [25] retrieved videos by analyzing the subject-object-predicate construction of the user query and searching keywords in a semantic graphic database.

Playlists: Aytekin et al. [13] assembled a set of videos for recommendations based on prerequisite relationship detection techniques, that is, identifying when a concept refers to another concept. Likewise, Tang et al. [233] recommended a playlist by building a concept map that showed the most relevant concepts in videos and how they are linked to each other. The authors also considered the sentiment analyzed from likes, dislikes, and comments. Furini et al. [67] recommended a playlist composed of video segments that covered the keywords given by the learner. For segmenting the video, silence periods and also visual transitions were detected using histogram-based techniques. Techniques for

optical character recognition were employed to understand the video content and match it to the user query. Cooper et al. [43] recommended videos based on the main topics, which were found by using the Latent Dirichlet Allocation algorithm and inter-topic relationships from the course syllabus. For creating the playlist, the authors used sequential pattern mining techniques to discover relationships.

Other: Further studies presented recommender systems based on video popularity [127, 270], videos delivered by the same instructor [83, 103], viewer's learning style [38], the current software application that the learner is using [63], and multiple pedagogical criteria [5].

5.1.4 Classifiers and Predictors.

Classifiers and predictors were mainly used to estimate learning outcome, while some other noticeable tasks were discipline classification, instructor behavior prediction, and dropout prediction.

Learning outcome prediction: Lemay and Doleck [143] used a repeated incremental pruning algorithm to predict the learners' grades based on the frequency of video viewing. Mbouzao et al. [165] predicted whether a learner will pass or fail a MOOC course using attendance rate (number of videos the student played divided by the number of videos mandatory to watch), utilization rate (percentage of video play time activity divided by the sum of video lengths of all videos), and watching index (attendance rate multiplied by utilization rate) metrics in a straightforward rule that compared a single student against the average. Mubarak et al. [173] trained a long-short-term memory network to predict the learning outcome as pass or fail based on learner's activity data (e.g., play, pause, speed, etc.). Furukawa et al. [68] used the number of views, interval of video views (e.g., 0 to 19 days), and the first access to videos in a multiple regression predictor to evaluate whether the learner will pass or fail a MOOC course. Similarly, Korosi et al. [122] predicted whether a learner will pass or fail a MOOC course using random forests and bagging based on characteristics related to platform usage (e.g., scrolling actions) and not representing the video itself. In a different way to determine learning success, Lallé and Conati [133] clustered learners into low and high achievers using characteristics such as the average proportion of videos watched per week. Aouifi et al. [10], instead, predicted learning outcome in terms of passing or failing the next quiz. The authors take as characteristics the sequence of previous quiz results to feed them into a recurrent neural network. Brinton et al. [27] also predicted whether the learner will pass a quiz successfully but using characteristics such as play and pause, revisiting, etc.

Discipline classification: Stoica et al. [228] built a support vector machine trained on a Wikipedia dataset to classify video transcripts and keywords into the discipline to which they belong. Dessi et al. [54] introduced a model based on variants of support vector machine algorithms to classify videos into their corresponding disciplines using transcripts, on-screen text features, and a relevance score for each feature. Basu et al. [17] clustered videos according to their disciplines by analyzing the topic similarity between the transcript and Wikipedia articles. Othman et al. [184] classified videos into categories such as Education and Science and Technology but also to rather different types including How-to and Style. The authors used decision trees and a Naive Bayes classifier for classification based on a various video metadata features. In a different approach, Chatbri et al. [31] built a CNN fed with image-like representations that characterize the transcript to then classify the video into its discipline. The images are generated using statistical co-occurrence transformation techniques, which arranges the transcript into an array of characters and uses ASCII codes of the characters to create gray scale images.

Instructor behavior prediction: Chen et al. [35] built a neural network to predict teachers' enthusiasm using facial expressions, voice, volume, and pose characteristics. Godavarthi [80] also used facial landmarks characteristics but to predict instructor emotions using a CNN. Sharma et al. [207] predicted video liveliness (energetic or active) using

two different long-short-term memory networks fed with audio features, visual features from the instructor actions, and production setup (e.g., instructor sitting or standing, slide-based video or whiteboard-based video). Xu et al. [263] predicted the instructor's actions (write, pick eraser, erase, etc.) based on pose characteristics using a random forest and adaptive CNN classifiers.

Dropout prediction: Jeon and Park [104] used learner's activity data to predict weekly course dropout in a MOOC platform by embedding click data sequences and feeding them into a gate recurrent unit. Furini et al. [66] investigated dropout prediction in an e-learning platform relying on characteristics related to video access, and using random forest and K-nearest neighbor classification.

Other: Further articles proposed models to predict how engaging a video was [28], when a learner will jump back [269], and to classify learners' comments [194].

5.1.5   Navigation and Segmentation Tools.

Many articles proposed navigation in the style of a table of contents, but also navigation using keywords and visual elements has been presented. The suggested tools commonly use segmentation techniques to match the extracted information with the corresponding video segment to allow user navigation.

Table of contents navigation: Such tools consist of a table of contents which the user can click on and jump to the corresponding part in the video. These approaches usually allow users to find segments where a certain sub-topic or keyword is presented. The segments to which the sub-topic or keyword belong are identified and indexed to allow the learner to click on them and to jump there directly [21, 48, 71, 102, 128, 131, 160, 174, 264, 273]. Sub-topics and keywords have been extracted with technologies such as the Latent Dirichlet Allocation algorithm [21, 102, 128, 174], TextRank algorithm [161, 264], Markov chain Monte Carlo method [21], and concept extraction algorithms [161].

Keyword navigation: Another frequent kind of proposal for navigation extracted keywords and indexed them according to specific segments of the video [112, 121, 128, 131, 182, 273].

Navigation using visual elements: This kind of approach enabled exploration by presenting visual parts of the video, for example, main slides, figures, tables, charts, etc., which the viewer clicked and then jumped to the corresponding segment in the video [131, 158, 264, 273, 274].

Video segmentation: One of the natural ways of splitting a video is to detect slide changes, which can be used to extract key video frames. For video segmentation, TextTiling [174, 273] and dynamic programming [71] were suggested. For the detection of slide changes, edge-based differences between video frames [273] and overlapping patches [160] were expoited. Also, we found studies dedicated to exclusively find better segmentation methods for learning videos (e.g.[69, 220, 221]). These approaches used natural language processing for phrase extraction, named entity recognition, and part-of-speech-tagging [69] to segment the video. Other studies used audio features to segment according to sound characteristics (e.g., power spectral density, fundamental frequencies, low level acoustic characteristics) [220, 221].

5.1.6   Analytic Tools.

Analytic tools relied heavily in the learner's activity data to understand video usage, and in some cases, contrasted this usage with the learners' performance. Another small group of analytic tools used characteristics such as metadata and eye gaze.

Based on learner's activity: General analytic tools exploited log data of learner activity from e-learning platforms to show how videos are used. Walk et al. [250] developed a tool that shows the audience retention for specific video segments, the number of times the segment was re-watched, and how much time learners spent on it. Wachtler et

al. [249] presented a tool that provided an analysis about the video (e.g., number of views) and also about the student (e.g., extent to which the video was watched). Long et al. [156] proposed a framework to analyze patterns regarding video usage, for instance, usage of a single video, specific video segments, all videos in a course, etc.

Other tools used log data to understand the learners' behavior in relation to their academic performance. Min et al. [168] suggested an analytic system that contrasted the log data and the student performance on the quizzes with the watching behavior. He et al. [89] analyzed the viewing behavior using two metrics, where the first one records if the learner watched the video and the second one records the amount of video viewing. Later, He at al. [90] showed how these metrics can be adopted to understand the results in academic performance.

In addition, learner activity data has been used in other interesting analytic tools for diverse purposes. Mbouzao et al. [164] presented a method to embed the viewer's sequence of activities into a vector and found if they characterize a specific video, this is, if a certain video elicits particular behavior. Bothe and Meinel [26] developed a metric to track viewers' visits to previous MOOC videos, considering that learners do not always strictly follow the path suggested by a platform. Choi et al. [36] included smartwatch data in their system log data to improve the measurement of interest and difficulty of a video.

Based on video metadata: Othman et al. [183] mined MOOC videos by exploiting their metadata (e.g., video length, comments, title, country). Stepanek and Dorn [225] presented a framework to analyze viewers' comments to identify and predict the engagement generated by a video. Sung et al. [231] analyzed viewers' comments considering the sentiment, its relevance, its type, and its topic for displaying this information in the video timeline.

Other: Zhang et al. [272] analyzed learning videos using data from eye tracking systems.

5.1.7 Summarization Tools.

Summarization approaches aim at generating a concise and shorter representation of the entire video. Text summarization was the most applied approach, however, also summarization of visual elements was suggested.

Text-based summarization: This kind of summarization was achieved by sentence segmentation, named entity recognition, extraction of key phrase extract, subtitles, and other metadata using transcripts as a main source [14, 117, 245] and on-screen text in a few cases [29, 261]. Methods that supported this task were term frequency–inverse document frequency [245], Naive Bayes [14], and Latent Semantic Analysis [267] for identifying the relevant information.

Image-based summarization: Rahman et al. [198] proposed a visual summary by extracting a group of relevant images (figures or illustrations) from slide-based videos. The image was included or not in the summary according to its importance, which was computed based on its size, unique aspects in the image, and the duration of its occurrence in the video.

5.1.8 Other.

Further interesting articles proposed methods to create videos automatically [65, 87, 88], alleviate the split attention problem [81, 209], enhance speech recognition [92, 202], and find content that is similar not in other videos but in other media such as in textbooks [252] and books [97].

5.2 Controlled Experiments

In this section, we report general findings regarding the evolution of experimental research, main disciplines, production style, and sample sizes. We also present the video characteristics that have been tested in controlled experiments and

the dependent variables used to measure the impact of those characteristics (e.g., on engagement, learning outcome, etc.).

5.2.1 Trends, Disciplines, Production Style, and Sample Size.

Trends: There is an increase of articles on controlled experiment for the last three years in comparison to the previous three years. More than 60% of articles in this category have been published in the last three years [6, 11, 19, 33, 34, 42, 46, 47, 49, 74, 84, 96, 98, 100, 113, 129, 130, 134, 136, 142, 157, 170, 175, 180, 181, 190, 193, 205, 222–224, 229, 234, 240, 241, 253, 254, 257, 259, 271].

Disciplines: Most of the experiments were performed using learning videos in STEM disciplines (around 80%) [6, 18, 19, 30, 34, 41, 46, 52, 55, 73, 74, 84, 93, 96, 98, 113, 119, 129, 134, 136, 142, 145, 157, 166, 170, 175, 180, 181, 186, 188, 190–193, 204, 205, 209, 210, 222–224, 230, 241, 242, 248, 251, 257, 259, 271] and some studies experimented with videos in STEM and non-STEM disciplines (e.g., [19]). Experiments with STEM topics have primarily considered physics, mathematics, and computer science subjects. In general, we identified the following STEM disciplines: (1) physics and mathematics (19% of controlled experiments) [6, 34, 93, 96, 98, 129, 142, 166, 188, 190, 223, 224, 248], (2) computer science (19%) [34, 46, 134, 136, 145, 175, 180, 181, 204, 222, 251, 257, 259], (3) biology, chemistry, and medicine (14%) [18, 52, 55, 84, 157, 193, 223, 224, 230, 271], (4) software use, digital media and communication (13%) [73, 74, 113, 119, 186, 191, 192, 241, 242], and (5) environmental science and meteorology (11%) [19, 30, 34, 41, 73, 74, 170, 205, 209, 210, 241, 242] Experiments on non-STEM topics focused mainly on language education (6%) [33, 49, 100, 130], history and geography (6%) [19, 47, 240, 253], and psychology (4%) [34, 42, 172].

Production style: Most of the experiments reported the production style of the videos (69% of controlled experiments). The results showed that most of them used slide-based videos (39% of the experiments) [30, 46, 93, 96, 98, 116, 136, 142, 172, 181, 192, 193, 209, 223, 224, 251, 254]. The presence of an instructor was also very common among the experiments (36%) [18, 19, 30, 46, 55, 93, 96, 98, 116, 180, 192, 193, 223, 229, 230, 254]. Less frequent production styles included handwriting (e.g., Khan-style, classroom-style videos) (16%), [93, 136, 190, 204, 209, 223, 230], talking-head (16%) [30, 49, 93, 145, 190–192], screencasts (14%) [73, 74, 113, 186, 241, 242], productions with virtual reality and 360-degree characteristics (14%) [11, 33, 47, 84, 100, 175], and animations in videos (content created entirely with drawings, illustrations, or computer-generated effects) (14%) [6, 42, 145, 170, 205, 223, 224]. Interestingly, we noticed that animations were considered more frequently in the last three years [6, 42, 170, 205, 223, 224]. Please note that several studies used more than one production style (e.g., [93, 192, 223]).

Sample size: The experiment with the least number of participants included four learners [116], the one with most of participants included 398 learners [210], the average was 94 participants. There was one experiment with a very large sample size of more than 12,000 participants, which was performed on a MOOC platform [53].

5.2.2 Video Characteristics in Experiments.

Figure 7 summarizes and ranks the video characteristics studied in controlled experiments. Specifically, we found four categories of characteristics that have been examined the most: (1) interactive features (e.g., quizzes) (39% of controlled experiments) [11, 33, 34, 41, 46, 47, 53, 84, 100, 113, 119, 129, 130, 142, 157, 172, 175, 188, 210, 222, 234, 240, 248, 259], (2) production style (e.g., talking head) (25%) [6, 30, 42, 49, 55, 93, 136, 180, 181, 190, 191, 223, 224, 229, 230], (3) instructor's behavior (e.g., gestures) (10%) [18, 19, 96, 98, 145, 254], and (4) visual features (10%) [170, 192, 193, 209, 251, 253].

Interactive features: These features allow the communication and interaction between the video and the learner. The most used interactive features in experiments were multiple-choice quizzes and questions embedded in the
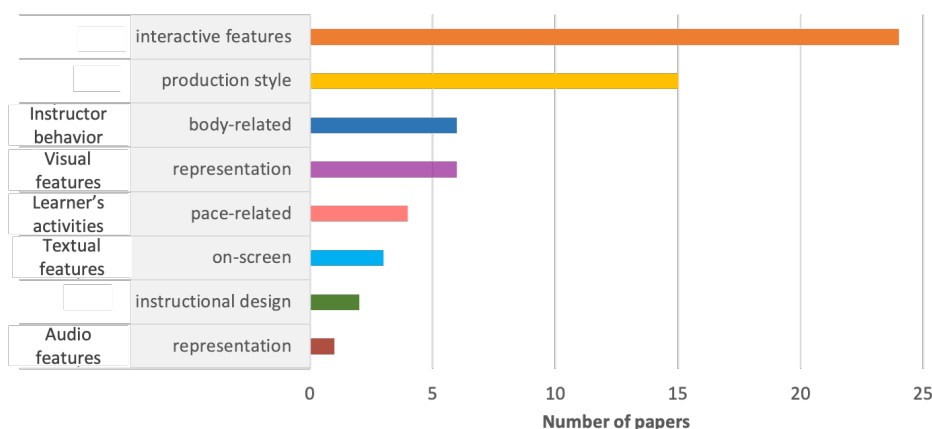
Fig. 7. The diagram shows the number of papers for each video feature type studied in controlled experiments. More than 80% of these studies focused on interactive features, production style, instructor behavior, and visual features

.

video [46, 119, 142, 188, 222, 248], virtual reality and 360-degree environments [11, 33, 47, 84, 100, 175], in-video annotations (feature that allow the learner to take notes while watching the video) [34, 130, 142, 234], and navigational elements (e.g., table of content navigation, timelines with clickable points of interest) [41, 46, 130]. Examples of less researched interactive features were games (e.g., crossword puzzles) [41], human-like conversational agents [259], and self-regulated learning activities [172, 240].

Production style: In general, the production style refers to the scenery, the elements used to convey the information, and also the physical arrangements considered to record the video. Examples for production style settings are Khan-style video, classroom-style videos, with talking head, etc. Specifically, we use the typology of video production styles of Hansch et al. [86]. We found that more than a third of the experiments that examined production styles addressed the instructor's presence [30, 49, 93, 190, 191]. Another third of the experiments evaluated dialogue-style in contrast to the commonly used monologue-style [55, 136, 180, 181, 229]. The last third of the experiments in this category investigated animations [6, 42, 223, 224].

Visual and audio features: This kind of experiments primarily addressed characteristics that attract the attention of the viewer, here called representation characteristics [170, 192, 193, 209, 251, 253]. Experiments on representation characteristics have clearly focused on content-related cues that direct attention (e.g., arrows, moving elements, color cues, gaze overlay) [170, 193, 209, 253] and instructor-related elements that direct attention (e.g., pointing gestures, spoken words, directed eye gaze) [170, 192, 193, 209, 251, 253]. Other experiments studied the impact of the complexity of visual content [84] and the gender appearance of the instructor [96].

Instructor behavior characteristics: In controlled experiments, the behavior of the instructors has been studied principally with respect to their gestures [18, 19, 98, 145, 254]. Less studied characteristics were related to the speech (e.g., tone of voice [98]), other visual aspects such as pose (e.g., frontal, lateral), and dressing style [18].

Other characteristics: Some articles studied the pace at which videos are viewed [73, 74, 134, 166, 205]. Among the less popular researched topics, we found experiments with captioned and non-captioned videos [186, 257], incorporating music [204], summaries at the end of video [242], the inclusion of sign language [116], and experiments that test

Multimedia Learning Theory such as the effect of pre-training material and spatial contiguity between captions and visual elements [241, 271].

5.2.3 Dependent Variables.

In controlled experiments, the dependent variable measured the impact of specific video characteristics on different learning aspects. We have reviewed all the dependent variables and categorized them into different groups according to their purpose. In this way, we have obtained the following categories that are related to: (a) self-perceived effectiveness (e.g., self-efficacy, instructional efficiency) [41, 46, 93, 96, 100, 113, 136, 172, 205, 222, 241, 242], (b) difficulty (e.g., cognitive load, time effort) [18, 19, 30, 34, 41, 46, 49, 52, 74, 96, 100, 113, 136, 157, 166, 170, 172, 191, 192, 223, 224], (c) general learning experience (e.g., lecture experience, learning experience) [30, 33, 42, 100, 119, 145, 157, 188, 190, 191, 210, 230, 240, 254], (d) emotions and feelings of the viewer (e.g., satisfaction, enjoyment) [11, 41, 84, 96, 98, 204, 223, 224, 229, 254], (e) instructor (e.g., enthusiasm, presentation skills) [6, 30, 41, 49, 93, 96, 136, 145, 191, 251], (f) video and its content (e.g., quality, usability) [47, 84, 113, 116, 181, 188, 210, 242] (g) engagement with the video (e.g., motivation, engagement) [11, 33, 42, 98, 100, 113, 136, 142, 157, 175, 186, 209, 210, 224, 251, 259], (h) learning performance (e.g., learning quality, learning outcome) [6, 11, 18, 30, 33, 41, 41, 42, 46, 47, 49, 52, 53, 55, 73, 74, 93, 96, 98, 100, 113, 119, 129, 130, 134, 136, 142, 145, 147, 154, 157, 166, 170, 172, 180, 186, 188, 191–193, 204, 205, 209, 210, 222, 223, 230, 234, 240–242, 248, 251, 253, 254, 257, 271], and (i) other (e.g., cybersickness, perceived similarity of the learner to the instructor) [18, 41, 84, 93, 96, 100, 181].

Specifically, within these categories, the evaluated variable in almost every experiment was the learning outcome (86% of controlled experiments), followed by cognitive load (20%), satisfaction (19%), mental effort (11%), social presence (learners' perception of their own social presence) (11%), motivation (9%), and self-efficacy (9%). Please note that several studies measured the effects on more than one dependent variable.

The learning outcome has been studied under different names such as learning gain, academic achievement, etc. Lately, this metric has been measured not only by testing how much the learner can remember (*recall*) but also to what extent the learner can apply the gained knowledge to new problems (*transfer*). Our results showed that transfer tests have been more frequently used in the last four years than in the previous years [18, 41, 55, 73, 74, 93, 129, 130, 157, 166, 193, 205, 240, 253, 271]. However, we found that most of the articles did not report explicitly if the type of test was recall or transfer, only 38% of experiments reported it [11, 18, 41, 49, 55, 73, 74, 93, 119, 129, 130, 145, 157, 166, 180, 193, 205, 240, 253, 257, 271].

In addition, we found that dependent variables were derived from tracking facial expressions [224], neurophysiological aspects (e.g., galvanic skin) [259], thermal aspects [223, 224], log data [46, 134], and eye gaze [116, 136, 170, 180, 192, 193, 209, 224, 230, 251, 271]. The last one was the most used technique, from which popular metrics such as dwell time have been derived to measure, for example, attention.

5.3 Data Analysis Studies

The majority of data analysis studies focused on the understanding of patterns in learning behavior. Furthermore, around half of the data analysis papers have also tried to understand the impact of those patterns on a specific variable that measures learning effectiveness [9, 12, 27, 44, 45, 60, 70, 82, 94, 126, 137, 146, 200, 213, 227, 236, 260]. In the same way as in the controlled experiment papers, the impact of video characteristics has been studied mostly on the learning outcome [9, 27, 45, 70, 94, 126, 146, 200, 213, 227, 236, 260] and on engagement [12, 45, 70, 82, 94, 146, 236].

Figure 8 sorts the characteristics studied in data analysis articles according to their frequency. We found that data analysis studies usually made use of learner's activity characteristics by taking advantage of system logs (e.g., play, pause) [9, 12, 44, 45, 57, 60, 79, 99, 126, 137, 139, 140, 146, 167, 189, 200, 203, 212, 213, 227, 239, 243, 260, 266]. Other
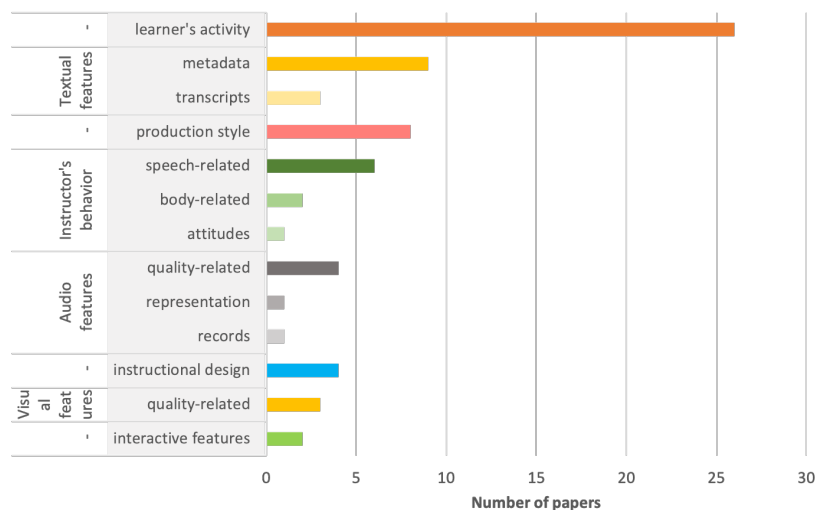
Fig. 8. The diagram displays the number of papers with regard to video characteristics explored in their data analysis studies. These studies usually made use of log data of learner's activities.

characteristics often analyzed were metadata and video transcripts [12, 56, 60, 70, 94, 137, 167, 203, 211, 212, 216, 218, 243]. Moreover, data analysis studies have inspected production style characteristics [45, 56, 94, 140, 197, 212, 214, 236], audio features [56, 211, 212, 214] and visual features [56, 212, 214], instructor's behavior characteristics [12, 56, 189, 212, 214, 237, 243], design principles and guidelines [60, 140, 185, 212], and interactive features [56, 82]. The corresponding findings are presented in the next section (Section 6).

5.4 Design Guidelines

General guidelines: Wijnker et al. [258] developed a framework that applies a film theory perspective to learning videos. The authors provided guidance to analyze and categorize these videos according to their design, aim and type in film theory [23]. Voronkin [246] proposed universal guidelines focusing on methodological, psychological, didactic, ergonomic, and technical aspects. Lange and Costley [135] considered media delivery problems and addressed them by proposing guidelines that aim at improving the pace, intelligibility, quality, media diversity, and congruence. The authors grounded their guidelines in the limited capacity theory in multimedia learning [163], generative theory [163], and extraneous processing research [144]. Dodson et al. [58] introduced a framework of active viewing that categorized learner's watching behavior. The authors stated that VBL can be designed in a way to promote this behavior.

    Guidelines for specific disciplines: Minnes et al. [169] proposed design guidelines for producing short learning podcast videos for computer science courses and emphasized the use of representation characteristics, especially call-out boxes, zooming, and animations. Hodges [95] also collected several recommendations for creating videos in CS. The authors emphasized the need of captions, descriptions of animations as they occur, zooming in relevant code, and for video systems that allow learners to pause, seek, and speed up or slow down the video. Stephenson [226] concentrated on the development of programming videos following a specific content structure. Lei et al. [141] discussed an approach to produce videos for Humanities disciplines. They suggested to favor short videos, use animations, highlight key concepts and terminologies, include an introduction and conclusions, employ consistent topography and graphic styles,

and to add music in specific parts. Ge and Li [75] identified beneficial characteristics to learn English as a foreign language. They recommended to take care of the speech rate, sentences and vocabulary complexity, use of closely related examples, use of realistic visuals when higher levels of context are demanded, include the instructor, and use keyword subtitles instead of complete subtitles.

Guidelines for videos with interactive features Weinert et al. [256] introduced reference guidelines to design Interaction Theory [171], Cognitive Load Theory [232], and own research. The authors recommended to include elements for interacting not only with the content but also with other learners and the instructor. They also discussed recommendations for the best positions to place specific types of interactive features. Hung et al. [101] proposed guidelines to design interactive tasks for learning video content based on literature on embodied cognition. The authors propose that the video should include engaging, prompting, experiencing, facilitating, demonstrating, and questioning tasks.

6 IMPACT OF VIDEO CHARACTERISTICS ON THE LEARNING EFFECTIVENESS

In this section, we contrast, analyze, and summarize the findings of studies from controlled experiments and data analysis research. These are organized according to the different categories and subcategories of video characteristics.

6.1 Instructor Behavior Characteristics

Gestures: Study results showed that gestures do not have a significant impact on the learning outcome [19, 98, 254]. One study even found that gestures can cause negative comments [56]. However, learners still care about gestures, facial expressions, and body language [212]. Among the benefits, gestures were found to reduce mental effort [19], increase presence [19, 145] and satisfaction [254]. Moreover, studies have obtained a positive effect on delayed learning outcome tests and a higher impact of human gestures compared to gestures of instructors as animations (human-like artificial agents) [98, 254].

Speech: Data analysis studies revealed that characteristics related to the speech such as speaking rate, lexical diversity, etc., affected how learners interacted with the video [12]. Speech characteristics were also found to correlate with sound characteristics [211]. Further findings showed that speech characteristics had an impact on dwell time [243] and the level of understanding [212]. For example, whether the instructor's language was native or not significantly affected the level of understanding in the learner, and the number of likes in the video [212, 214]. Also, pronunciation and intonation was found to affect the learner's attention. Based on these discoveries, it has been recommended to reduce lexical diversity, long sentences, and speech rate [12].

Other: Other characteristics related to instructor behavior which the learners cared about and had an impact were the appearance, clothes, and enthusiasm [56, 212]. Interestingly, the gender did not have a significant impact [212], it did not affect the learning outcome, the perception of self-efficacy [96], and learners liking or disliking a video [214]. Yet, it was observed that learners perceived themselves more similar to the instructor when the gender was the same [96]. With regard to instructors as animations, these were considered interesting and attractive [56].

6.2 Interactive Features

Quizzes and questions: Most of the studies that experimented with quizzes evidenced that the learning outcome was positively affected [46, 119, 142, 188, 248]. For Palaigeorgiou and Papadopoulou [188], the positive effect was found when quizzes were included together with other interactive features. For Wachtler et al. [248], the positive effect was found when quizzes were placed at the right time.

Furthermore, positive impacts were found on other aspects such as satisfaction [119, 188] and engagement [46]. Data analysis studies have also confirmed these findings. Kovacs [126] found that most viewers engaged in quizzes, which seemed to reduce dropout. On the contrary, Ding [56] showed that quizzes improved understanding of the content. But they did not draw the learner's attention in cases where the quizzes were complex and at the end of the video.

Virtual reality and 360-degree characteristics: The studies in this categories commonly reported an improved learning outcome, however, not to a significant degree [11, 33, 47, 100]. Huang et al. [100] found positive effects only in certain parts of the post test. Although statistically insignificant, Araiza-Alba et al. [11] reported short- and long-term improvements in learning outcome. Studies also showed differing impact of virtual reality and 360-degree environments on learner motivation. While Huang et al. [100] found no positive effect on the motivation, Muñoz-Carpio et al. [175] and Chen [33] found that viewers felt indeed more motivated. Regarding satisfaction, several findings confirmed an increase [33, 100, 175]. Furthermore, Huang et al. [100] also found a positive impact on self-efficacy. Our analysis over time shows that these kind of characteristics have been investigated more frequently in the last years [11, 33, 47, 100, 175]. Naturally, experiments in this category evaluated the impact in other non-common dependent variables such as sense of "being there" [11, 84] and cybersickness [84]. Interestingly, most of these experiments were conducted with non-STEM videos [11, 33, 47, 100].

Other: Most of the studies that experimented with in-video annotations found an increase in the learning outcome [130, 142, 234], whereas neutral effects were found for cognitive load [34]. Regarding navigational features, experiments using table content navigation [41, 130] and timeslines with clickable points [41, 46] found a positive effect on the learning outcome [41, 46, 130]. However, this was not always significant [46], or depended on the combination with other features to be more beneficial such as table content navigation together with timeslines with markers [41]. According to Granjo and Rasteiro [82], videos with interactive features needed to be used with other pedagogical techniques to improve the learning outcome.

6.3 Visual and Audio Features

Instructor-related representational characteristics: Tian and Bourguet [237] observed that the most common gestures that instructors used as cues were pointing and extending (extend the arms away from the body with palms up) cues. Results in evaluating the impact of these types of cues have been shown to increase the learning outcome in most cases [192, 193, 251, 253]. However, no significant effects [209] or even negative effects where found when the instructor was an animation [170].

Content-related representational characteristics: Examples of these characteristics are arrows, color-coded cues, highlighted content, etc. Experiments showed that these cues improve the learning outcome in most cases [192, 209, 253], but had no significant effect on cognitive load [209] or increased it [170]. When comparing content-related cues with instructor-related ones, the results were contradictory. In one case, instructor-related cues had a higher positive impact on the learning outcome than content-related cues when hand-pointing gestures were compared to arrows on slide-based videos [192]. In another case, contrary results were found when comparing hand-pointing gestures (a pencil following the handwriting) with gaze overlay defined by an expert (explicit drawing of the path that learners should follow with their eyes) [209].

Other visual and audio elements: General visual and audio features such as the image and audio quality seemed to matter to the learners and caused positive effects on the learning outcome [56, 204]. In the experiments of Uchidiuno et al. [239], English language students often avoided videos that did not have visual support (figures, charts, equations,

or on-screen text) but presented only an instructor talking. Interestingly, Ding [56] found that music in the background helped to keep students focused.

6.4 Production Style

Instructor presence: Experiments have shown that viewers preferred a video in which the instructor appears [49, 190]. Hew and Lo [93] observed that the three most widely used styles in a sample of YouTube videos in computer science had an instructor: (1) slide-based videos, (2) classroom-style videos, and (3) Khan-style videos with talking head. In addition, viewers voiced their preference for videos that show the instructor's hands [190]. However, although the presence of an instructor was preferred, it did not bring notable advantages [49, 93]. Davila et al. [49] and Hew and Lo [93] found that a talking head did not show a significant effect on the learning outcome, the social presence [93] and the cognitive load [49]. Experiments have also tested the effects of the instructor's size in the video. A small size seemed to cause higher learning gains [191] and satisfaction [30] when the topic was not complex. However, the size of the instructor had no significant effect on cognitive load, mental effort [30, 191] and social presence [30, 191].

Dialogue-style vs. Monologue-style: The experiments did not find a significant effect of dialogue-style over the monologue-style on learning gain [55, 136, 180]. In addition, no effect on the social presence was found [136]. Ohashi et al. [181] conducted an experiment using videos in computer programming topics, one using slides only, another one with slides accompanied by two puppets in a dialogue. They found that the number of viewers of the video including the dialogue decreased over time (throughout a compulsory course). The authors argued that the puppets' dialogue was effective at the beginning, but not when the learners became increasingly familiar with the course content. Nevertheless, the dialogue style seemed to offer benefits with regard to self-efficacy [136] and positive sentiment [55].

Animations: Animations in videos caused higher learning outcomes when compared to (a) text-only videos (i.e., slide-based videos only with text, without illustrations, figures, charts, and similar), [223] (b) textbooks [42], and (c) traditional live classroom lectures [6]. Animations also affected other aspects such as participation [42], cognitive load [136], mental effort [223], and perceived difficulty [224].

Video tutorials and video lectures: In several studies, learners believed that video tutorials were more interesting, attractive, kept them engaged, and helped them understand better [56, 185]. It was also found that learners tended to skip theoretical discussions, which were usually found in video lectures [140]. However, other studies showed that learners not only watched lectures significantly more often than tutorials [94] but also watched them significantly more often to completion [236]. In any case, Hildebrand and Ahn [94] reported that both styles had a similar positive correlation with the learning outcome.

Other: Other studies found that media diversity correlated positively with learning outcome, satisfaction, engagement, and interest [45]. Shoufan and Mohamed [214] reported that the most predominant styles in CS videos were paper-based explanations, whiteboard-based explanations, slide-based videos, and Khan-style videos.

6.5 Pace of the Video

Automatic pauses vs. non-automatic pauses: No significant positive effects of automatic pauses in videos were found compared to non-automatic pauses (where learners can pause the video, also called self-pacing) [205]. Indeed, systems with automatic pauses in videos were ineffective [74] or even increased cognitive load [73, 74]. Merkt et al. [166] did not find advantages of automatic pauses even in long lasting videos. On the contrary, viewers in the non-automatic pauses condition were more successful and reported a lower mental effort [73, 205].

Video acceleration: The acceleration rate at which the video is watched was found to be important for learners [56]. Furthermore, there was evidence that students tend to accelerate the videos at least once [266]. However, the impact reported on the learning outcome has been diverse. On the one hand, Lang et al. [134] showed that when the learner accelerated the video, the learning outcome improved for learners in Stanford Online [244] (a platform of open online courses) in diverse disciplines such as CS and mathematics. On the other hand, Garrett [74] reported that learners who watched accelerated CS videos did not outperform those who watched non-accelerated videos. However, in this experiment the learner could not decide when to play, pause or accelerate the video but this was given automatically by the system.

6.6 Beginners, Specialists, High and Low Performers

Learners show specific behaviors depending on their performance level with respect to the learning topic. High achievers engaged in more diverse activity patterns (e.g., play, pause seek) [200], interacted more intensely with quizzes embedded in the video [126], and watched videos more constantly every week [200]. Galster et al. [70] reported that students who were more active through in-video annotations, also watched more videos. One study suggested that if low-performance learners imitate this behavior, advantages could be attained [9]. However, Galster et al. [70] did not find differences in learning outcomes between more engaged and less engaged learners.

Regarding the differences between beginners and specialists, the findings showed that specialists engaged more often in different video activities than beginners [146] and achieved a higher learning outcome [227]. Nevertheless, the same level of engagement was found in both groups and they also consumed the same amount of videos [146, 227].

6.7 Video Completion and Length

Video completion: It was reported that learners usually watched 30% to 50% of the total video length (on a YouTube channel [203] and on the MOOC platform edX [3, 227]) and that skipping through videos was a very common behavior [94]. However, studies in other learning contexts obtained different results. If the videos were part of a university course, most of the learners watched the full video [79, 266]. Regarding the gain in the learning outcome, studies showed that the extent to which a video was watched impacted the student's grades [9, 94] and that viewers who watched the video to completion were more likely to watch other videos [260]. Regarding dropping behavior, the findings show that this occurred within the first five seconds. On the other hand, when the learners watched the video for a longer time, dropping behavior seemed to be reduced [260]. Wynn et al. found that when viewers surpassed the first five minutes, they were likely to watch the video to completion [260]. Strategies such as including in-video quizzes seemed to reduce video dropout [126].

Video length: Several studies have concluded that short learning videos were preferred by learners [56, 59, 197, 203]. Videos within five to seven minutes were preferred in two studies [56, 59]. One of the studies experimented with videos with talking head, PowerPoint slides, drawings on a board, and animations for learning English as a foreign language by undergraduate students [56]. The other study experimented with videos of a MOOC on digital competence that presented pedagogical aspects and tutorials for teachers [59]. Further findings in both studies showed that although longer videos up to 10 minutes were not preferred, they were still well received [56, 59]. Saurabh and Gautam [203] found that the most popular learning videos lasted 10 to 15 minutes. Nevertheless, contrary findings presented by Shoufan [213] showed that videos perceived as too short were not well received. In this study, the videos were selected from YouTube with different liking ratios (number of likes of a video divided by the sum of likes and dislikes), different instructors, different speaking accents, and covering different topics in digital logic design and embedded systems

targeted to students of an engineering program. For lecture styles, Giannakos et al. [79] found that the preferred length of videos in software engineering was close to the traditional 45-minute classroom lectures for university students of a computer science degree. In any case, the context is an important factor to consider to select an appropriate video length [213].

6.8 Video popularity

Saurabh and Gautam [203] found that viewers tend to like videos that helped them to learn and dislike the ones that did not. Shoufan and Mohamed [214] discovered that the number of likes for learning videos increase (a) with higher video resolution, (b) with higher talking rates (words per second), (c) when English was the native language of the instructor, (d) when the production style included paper-based or a whiteboard-based explanations, and (e) when more than one production style was used. Later, Shoufan [213], found in a survey that learners liked a video when they perceived that (a) the explanation seemed appropriate (b) the presentation used animations, diverse colors, understandable handwriting, and appropriate usage of supporting technology and software, (c) the video seemed interesting and easy to focus on, (d) practical examples were included (e) the instructor was concise, used a simple vocabulary, and (f) had a confident and clear voice. On the other hand, in the same study, the authors [213] found that learners perceived that they disliked a video when the explanation was inappropriate, there were no practical examples, the problem was solved in non-common ways, unneeded details were included, the voice of the instructor seemed dull or annoying, or the content was boring and difficult to focus on. Additionally, in a different study, Shoufan [212] analyzed a survey and found that learners perceived that they disliked a video when they felt like they did not understand the video content, when the technical quality of audio or video was insufficient, when the instructor seemed incompetent, not friendly, or not enthusiastic, or the length of the video was not appropriate. Curiously, it was found that lengthy videos appeared to generate slightly less dislikes [214].

Lee and Osop [137] detected that popular videos have a high amount of comments with strong negative sentiment. Moreover, Lei et al. [140] found that popular videos had a high number of views but lower retention rates. The authors found that popular videos usually provided a general background, while less popular videos dived deeper into the topic, which might be the reason why they had higher retention rates. Indeed, it was warned that higher amount of views and comments might not necessarily entail a good video quality [137].

7 CONCLUSIONS

In this paper, we have reviewed 257 articles in the field of video-based learning (2016-2021) following the PRISMA guidelines [187]. Throughout the review, we (a) identified major research directions (b) surveyed researched video characteristics and organized them into a taxonomy, (c) reported on target tasks and technologies, (d) revealed derived trends in instructional design guidelines to produce videos, and (e) contrasted research findings and summarized the impact of the features on the learning effectiveness.

Research directions: We have identified four research directions in the field of VBL: (1) proposals of tools, (2) reports on controlled experiments, (3) data analysis studies, and (4) design guidelines proposals.

The majority of papers proposed tools for the support of VBL, while a moderate proportion of papers presented studies on experiments in controlled settings. A small part of articles were dedicated to data analysis studies and design guidelines.

Video characteristics: Table 3 presents the taxonomy we have built based on the investigated characteristics, topics, modalities, etc. identified in the research articles and their corresponding references. Figure 9 shows a ranking
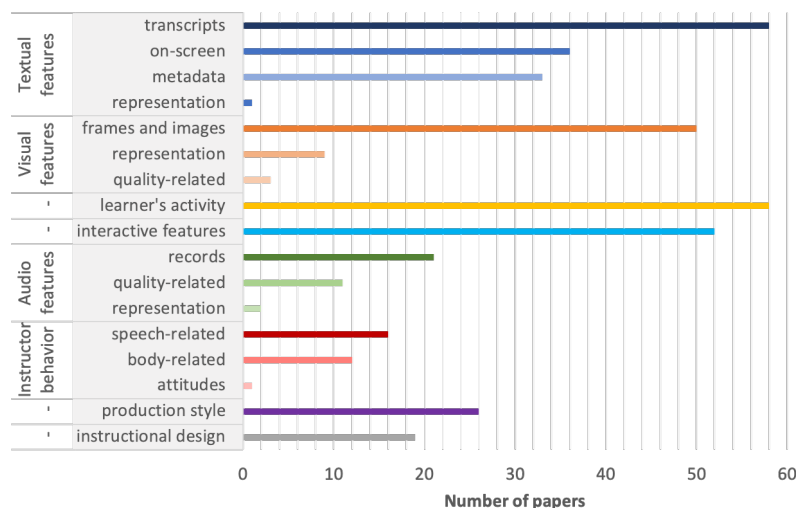
Fig. 9. the diagram shows all video characteristics identified in this review

of these video characteristics according to the number of related publications. The most explored characteristics were textual features followed by visual features, learner activity, interactive features. The most used textual features were transcripts, which were mainly employed to develop recommendation systems (see Section 5.1.3) or navigation tools (See Section 5.1.5), as a source to extract information (see Section 5.1.1), and for summarization tools (see Section 5.1.7 and Figure 6). The most explored visual features were video frames, essentially used to extract information from them. Characteristics related to the learner activity took advantage of log data and examined, for instance, play and pause activities. These were heavily explored in data analysis articles (see Section 5.3 and Figure 3) and in some tool proposals for developing predictors (see Section 5.1.4) and analytic tools (see Section 5.1.6 and Figure 6). Interactive features (e.g., quizzes embedded in the video, in-video annotations) were popularly scrutinized in experiments to evaluate their effectivity [11, 33, 34, 41, 46, 47, 53, 84, 100, 113, 119, 129, 130, 142, 157, 172, 175, 188, 210, 222, 234, 240, 248, 259] but also were integrated in tools for creating videos with interactive elements (see Section 5.1.2 Figure 3).

Moderately studied characteristics were related to audio features, production style, and instructor behavior. Audio features specifically have been used to develop tools [14, 21, 31, 35, 64, 64, 65, 71, 76, 81, 102, 103, 109, 112, 112, 131, 148, 160, 161, 174, 182, 202, 207, 220, 220, 250, 273]. Different types of production style have been tested mainly in controlled experiments [6, 30, 42, 49, 55, 93, 136, 180, 181, 190, 191, 223, 224, 229, 230] (see Figure 3). Regarding instructor behavior, speech-related characteristics [12, 14, 28, 32, 56, 98, 105, 107, 111, 111, 160, 189, 212, 214, 221, 243] have been used more often than body-related characteristics [18, 19, 35, 56, 80, 96, 98, 145, 214, 237, 254, 263] (see Figure 9). Less investigated characteristics were instructional design characteristics, which were mainly treated in articles on guidelines for the design and creation of videos [58, 75, 95, 101, 135, 141, 169, 226, 246, 256, 258].

Tools, tasks and technologies: The main tools, pipelines, frameworks, and methods that have been proposed for supporting VBL were related to information extraction, integration of interactive features, recommendation systems, classification and prediction, video navigation, and summarization (see Section 5.1 Figure 5). Information extraction focused mainly on retrieving text [49, 49, 50, 72, 72, 115, 115, 115, 123, 123, 124, 138, 138, 262, 262, 276, 276], metadata [61,

64, 77, 109, 111, 176], and key information from videos such as key video segments and key topics [32, 76, 128, 148, 152, 195]. Some approaches used deep learning for information extraction [32, 49, 115, 148], but also other (shallow) machine learning methods based on hand-crafted features were still present [32, 152, 195]. Knowledge bases [77, 109, 176] and natural language [61, 64, 111] processing methods were also utilized.

The development of interactive tools mainly targeted the integration of in-video annotations (feature that allow the learner to take notes while watching the video) [51, 62, 132, 151, 162, 178, 217], quizzes [91, 120, 159, 247, 249, 250], and, interestingly, also feedback to the learner for preventing mind wandering [106, 149, 155, 199, 208]. A few proposals introduced tools to create videos with 360-degree characteristics [20, 91, 125].

Recommendation systems have mainly used the similarity between transcripts and metadata to suggest videos [16, 22, 105, 108, 111, 235, 250, 255, 275]. A popular approach was to extract topics, concepts, and keywords to compute the similarity to other videos. The use of ontologies and external knowledge bases was used particularly to extract concepts [24, 25, 40, 206]. Some interesting approaches considered other characteristics such as video readability [111], the skills demanded by the labor market [235], video popularity [127, 270], videos from the same instructor [83, 103], learning style [38], and pedagogical characteristics [5].

For classifiers and predictors the main task was to predict the learning outcome using learner's activity data [27, 68, 122, 133, 143, 165, 173]. Other, but less frequent tasks were discipline classification [17, 31, 54, 184, 228], instructor behavior prediction [35, 80, 207, 263], and dropout prediction [66, 104].

Navigation tools popularly proposed table content navigation styles [21, 48, 71, 102, 128, 131, 160, 174, 264, 273] but some interesting approaches have emerged such as navigation according to keywords [112, 121, 128, 131, 182, 273] or visual elements [131, 158, 264, 273, 274]. These tools have relied heavily in technologies to segment the video and to extract subtopics, concepts, and keywords.

Analytic tools relied usually on characteristics of the learner's activity to understand video usage [156, 249, 250] and contrasted it with the learners performance [89, 90, 168]. Some interesting tools analyzed the comments of videos to understand the learners' sentiment [231] and their engagement [225].

Summarization methods used commonly text transcripts [14, 117, 245], however, there were also approaches using (visual) frame-based video content [198]. Relevant methods were related to the extraction of keywords, relevant subtitles, entities, and metadata.

Design guidelines: Some of these studies proposed general guidelines [58, 135, 246, 258], other suggested guidelines for specific disciplines [75, 95, 141, 169, 226], and two cases suggested guidelines for creating interactive videos [101, 256] . They have leveraged on theories and frameworks, for example, film theory [23], limited capacity theory in multimedia learning [163], interaction theory [171], active viewing framework [58], etc. These guidelines aimed at the improvement of a wide variety of aspects such as pace, intelligibility, quality, ergonomics, interaction, visuals, etc.

Experiments, data analysis, and their results on the effectiveness of video characteristics: One of the major characteristics explored in controlled experiments were interactive features. Other characteristics that were experimented are related to production style, instructor behavior, and cues. Learner's activity characteristics have been specially inspected in data analysis studies.

Interactive features focused mainly on quizzes, virtual reality and 360-degree characteristics, and in-video annotations. Quizzes showed to be effective [46, 56, 119, 126, 142, 188, 248], but not in all cases – sometimes they were only effective when they were used together with other interactive features and placed at the right moment [56, 188, 248]. Virtual reality and 360-degree characteristics seemed to be effective; however, the impact was not significant [11, 33, 47, 100]. In-video annotations seemed to improve the learning outcome [130, 142, 234] and also navigational elements, specifically,

table content navigation and timelines with clickable points, impacted positively the learning outcome [41, 46, 130]; however, results were not always statistically significant [46] or they were more beneficial when they were together with other features [41]. Indeed, further claims state that videos with interactive features need to consider other pedagogical techniques to improve the learning outcome [82].

Experiments on production styles investigated mainly the videos with instructor's presence, videos with animations, video tutorials vs. video lectures, and dialogue-style vs. monologue-style. The learners preferred the presence of an instructor, although the benefits were not always significant [49, 49, 93, 93, 190, 190]. Animations caused higher learning outcomes when compared to text media and traditional class [6, 42, 223]. Tutorials and lecture styles correlated positively with the learning outcome [94], however, preferences for one style over the other were different between studies [56, 94, 140, 185, 236]. Dialogue-style instead of the typical monologue-style showed no significant effects on learning outcome [55, 136, 180]. In any case, the use of divers styles appeared to be effective [45].

Studies on instructor behavior examined particularly human gestures and speech-related characteristics, but very little other visual elements such as pose and gender. Speech characteristics seemed to impact certain aspects of the learning effectiveness such as understanding, attention, etc [12, 212, 243]. Gestures proved to benefit only certain aspects such as satisfaction and presence, but very little the learning outcome [19, 98, 254]. Instructor-related cues caused a positive effect on the learning outcome [193, 253]. However, this did not apply to animated instructors (human-like artificial agents), in those it even caused negative effects [170, 209]. Content-related cues have also positively affected the learning outcome [192, 209, 253]; however, they can also cause cognitive overload [170].

Regarding the choice of video length, several studies found that short videos were preferred by the learners [56, 59, 197, 203]. However, this was not true for every type of learning videos and context, in some cases longer videos were still preferred [79, 213]. Regarding video completion, experiments showed that the learning outcome improved when learners watched the video more than a certain proportion [9, 94]. However, video completion was affected by common behavior such as skipping and watching only parts of the video [79, 94, 266]. Studies on the impact of video pace showed no significant positive effects of automatic pauses, and even negative effects on the learning outcome were found [73, 74, 166, 205]. Interestingly, watching videos at faster speed was beneficial in some cases and counterproductive in others [74, 134].

Studies on experts and high performance learners revealed that these type of viewers engaged more actively with the video [126, 146, 200, 200] and achieved a higher learning outcome [227]. However, in some cases they achieved the same learning outcome and engagement as their counterparts [70, 146, 227]. Also, behavior might change according to the discipline, topics, and content of the video [94, 203, 239]. Also, YouTube videos might be preferred over institutional videos [79].

Regarding video popularity, studies identified that learners tended to like videos perceived as helpful for learning [203], however, a high popularity might not necessarily entail a good video quality [137].

Further trends and directions: Our analyses have confirmed several findings from previous reviews. We observed an increase in the number of VBL research articles [78, 196], and a significant emphasis on the use of videos in STEM disciplines [78, 196]. We also identified a tendency to use primarily the learning outcome as an indicator to measure learning effectiveness, traditionally in terms of recall tests and, more recently, through transfer tests [196, 268]. Moreover, similarly to Poquet et al. [196], we found that the next most popular indicators were cognitive load, satisfaction, mental effort, social presence, motivation, and self-efficacy. We could also confirm that eye-tracking [116, 136, 170, 180, 192, 193, 209, 224, 230, 251, 271] and log data [46, 134] characteristics were used; however, we also found studies that used thermal [223, 224], neurophysiological [259], and facial expressions-tracking [224] characteristics. Regarding production

styles, we could recognize slide-based productions as one of the most used production styles in controlled experiments. In addition, we found that the experiments frequently used videos with a visible instructor. However, contrary to Poquet et al. [196], we did not find that animations were popularly used. Among the styles that were less used, there were handwriting, screencast, and virtual reality styles. We conclude that VBL research in controlled experiments have not tested the diverse range of learning video styles, disciplines, and video sources. Thus, the field requires more research efforts in the areas not targeted yet.

Another tendency in the reviewed articles was that audio, visual, textual and further features are usually treated in a separated manner. However, learning through videos requires the fusion and interpretation of synchronous audio and visual sensory input. Thus, we believe that VBL could benefit from multimodal representations and analysis in order to better capture the relationships between the different modalities.

Research on video, text, and image analysis has been dominated by deep learning approaches in recent years. This is also the case for VBL, especially for the development of tools that support the learning process. We noticed that deep learning was not only used for basic tasks such as video segmentation, optical character recognition, neural word representations (embeddings), speech-to-text tasks, etc. but also in entire pipelines of recommendation systems, learning outcome predictors, navigational tools, etc.

Contradictory results were found about the impact of video characteristics on the learning effectiveness. We could observe that the discrepancies originate not only from the small or big differences in the test conditions but also from the production style of the video. In some cases it was not possible to know in detail these differences, since not all the studies reported accurately the characteristics of the investigated videos.

Future work: Experimental research should consider more the diversity in video styles and also the diverse disciplines apart from STEM subjects. An important gap we see is that research needs to move towards multimodal representations were video characteristics can be unified to approach to the way human beings learn. There are very few articles that propose design guidelines. We believe that more research in this matter, especially guidelines customized to specific disciplines, can be highly useful for practical scenarios such as content creation. Although articles that propose tools for supporting learning encompass most of VBL research in our sample, we think that more efforts are needed to bring new technologies to the field for understanding whether they can improve the way people learn through videos. Finally, finding what characteristics a video should have to be effective for learning remains ongoing work. Obviously, more research is needed to identify what makes us learn in a video and in what learning scenarios.

Limitations: In this review, we searched and retrieved articles exclusively from academic computer science data bases. Therefore, we did not consider other disciplines that could contribute to VBL advances, e.g., instructional insights from science education. We have researched in three dedicated databases, where one of them (DBLP) can be considered to track the vast majority of relevant CS venues, including journals and conference proceedings [2]. We retrieved articles with the query terms only in the title to make sure that the main focus was on VBL – however, there might be articles treating VBL-related topics without it being reflected in the title. Moreover, we restricted our review to articles that used and experimented with video characteristics, thus, advances in other use cases such as teacher reflection were not reported. Finally, it is important to recognize that an accurate comparison of the effects of video characteristics on the learning effectiveness is difficult due to the heterogeneity in the experimental settings and test conditions. This is the reason why we could not provide a meta-analysis about these effects.

ACKNOWLEDGMENTS

REFERENCES

[1] 2022. *Association for Computing Machinery.* Retrieved November 28, 2022 from https://www.acm.org/

[2] 2022. *dblp: computer science bibliography.* Retrieved November 18, 2022 from https://dblp.org/

[3] 2022. *edX. Free Online Courses by Harvard, MIT, more.* Retrieved December 14, 2022 from https://www.edx.org

[4] 2022. *Springer International Publisher Science, Technology, Medicine. Springer International Publisher.* Retrieved November 28, 2022 from https://www.springer.com/gp

[5] Claudia Margarita Acuña-Soto, Vicente Liern, and Blanca Pérez-Gladish. 2020. Multiple criteria performance evaluation of YouTube mathematical educational videos by IS-TOPSIS. *Oper. Res.* 20, 4 (2020), 2017–2039. https://doi.org/10.1007/s12351-018-0405-2

[6] Mohammad Ahmad Al-Khateeb and Ahmed Mohammad Alduwairi. 2020. Effect of Teaching Geometry by Slow-Motion Videos on the 8th Graders' Achievement. *Int. J. Interact. Mob. Technol.* 14, 18 (2020), 57–67. https://doi.org/10.3991/ijim.v14i18.12985

[7] Mohammad Alahmadi, Jonathan Hassel, Biswas Parajuli, Sonia Haiduc, and Piyush Kumar. 2018. Accurately Predicting the Location of Code Fragments in Programming Video Tutorials Using Deep Learning. In *Proceedings of the 14th International Conference on Predictive Models and Data Analytics in Software Engineering, PROMISE 2018, Oulu, Finland, October 10, 2018*, Burak Turhan, Ayse Tosun, and Shane McIntosh (Eds.). ACM, 2–11. https://doi.org/10.1145/3273934.3273935

[8] Abdulaziz Albahr, Dunren Che, and Marwan Ali Albahar. 2021. A novel cluster-based approach for keyphrase extraction from MOOC video lectures. *Knowl. Inf. Syst.* 63, 7 (2021), 1663–1686. https://doi.org/10.1007/s10115-021-01568-2

[9] Lawrence Angrave, Zhilin Zhang, Genevieve M. Henricks, and Chirantan Mahipal. 2020. Who Benefits?: Positive Learner Outcomes from Behavioral Analytics of Online Lecture Video Viewing Using ClassTranscribe. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE 2020, Portland, OR, USA, March 11-14, 2020*, Jian Zhang, Mark Sherriff, Sarah Heckman, Pamela A. Cutter, and Alvaro E. Monge (Eds.). ACM, 1193–1199. https://doi.org/10.1145/3328778.3366953

[10] Houssam El Aouifi, Youssef Es-Saady, Mohamed El Hajji, Mohamed Mimis, and Hassan Douzi. 2021. Toward Student Classification in Educational Video Courses Using Knowledge Tracing. In *Business Intelligence - 6th International Conference, CBI 2021, Beni Mellal, Morocco, May 27-29, 2021, Proceedings (Lecture Notes in Business Information Processing, Vol. 416)*, Mohamed Fakir, Mohamed Baslam, and Rachid El Ayachi (Eds.). Springer, 73–82. https://doi.org/10.1007/978-3-030-76508-8_6

[11] Paola Araiza-Alba, Therese Keane, Bernadette Matthews, Kate Simpson, Grace Strugnell, Won Sun Chen, and Jordy Kaufman. 2021. The potential of 360-degree virtual reality videos to teach water-safety skills to children. *Comput. Educ.* 163 (2021), 104096. https://doi.org/10.1016/j.compedu.2020.104096

[12] Thushari Atapattu and Katrina E. Falkner. 2017. Discourse analysis to improve the effective engagement of MOOC videos. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, March 13-17, 2017*, Marek Hatala, Alyssa Wis, Phil Winne, Grace Lynch, Xavier Ochoa, Inge Molenaar, Shane Dawson, Shady Shehata, and Jennifer Pei-Ling Tan (Eds.). ACM, 580–581. https://doi.org/10.1145/3027385.3029470

[13] Mehmet Cem Aytekin, Stefan Räbiger, and Yücel Saygin. 2020. Discovering the Prerequisite Relationships Among Instructional Videos From Subtitles. In *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020*, Anna N. Rafferty, Jacob Whitehill, Cristóbal Romero, and Violetta Cavalli-Sforza (Eds.). International Educational Data Mining Society. https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_99.pdf

[14] Vidhya Balasubramanian, Doraisamy Gobu Sooryanarayan, and Navaneeth Kumar Kanakarajan. 2016. A multimodal approach for extracting content descriptive metadata from lecture videos. *J. Intell. Inf. Syst.* 46, 1 (2016), 121–145. https://doi.org/10.1007/s10844-015-0356-5

[15] Lingfeng Bao, Zhenchang Xing, Xin Xia, and David Lo. 2019. VT-Revolution: Interactive Programming Video Tutorial Authoring and Watching System. *IEEE Trans. Software Eng.* 45, 8 (2019), 823–838. https://doi.org/10.1109/TSE.2018.2802916

[16] Subhasree Basu, Yi Yu, Vivek K. Singh, and Roger Zimmermann. 2016. Videopedia: Lecture Video Recommendation for Educational Blogs Using Topic Modeling. In *MultiMedia Modeling - 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 9516)*, Qi Tian, Nicu Sebe, Guo-Jun Qi, Benoit Huet, Richang Hong, and Xueliang Liu (Eds.). Springer, 238–250. https://doi.org/10.1007/978-3-319-27671-7_20

[17] Subhasree Basu, Yi Yu, and Roger Zimmermann. 2016. Fuzzy clustering of lecture videos based on topic modeling. In *14th International Workshop on Content-Based Multimedia Indexing, CBMI 2016, Bucharest, Romania, June 15-17, 2016.* IEEE, 1–6. https://doi.org/10.1109/CBMI.2016.7500264

[18] Maik Beege, Steve Nebel, Sascha Schneider, and Günter Daniel Rey. 2019. Social entities in educational videos: Combining the effects of addressing and professionalism. *Comput. Hum. Behav.* 93 (2019), 40–52. https://doi.org/10.1016/j.chb.2018.11.051

[19] Maik Beege, Manuel Ninaus, Sascha Schneider, Steve Nebel, Julia Schlemmel, Jasmin Weidenmüller, Korbinian Moeller, and Günter Daniel Rey. 2020. Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Comput. Educ.* 156 (2020), 103955. https://doi.org/10.1016/j.compedu.2020.103955

[20] Anke Berns, José Miguel Mota, Ivan Ruiz-Rube, and Juan Manuel Dodero. 2018. Exploring the potential of a 360° video application for foreign language learning. In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, October 24-26, 2018*, Francisco José García-Peñalvo (Ed.). ACM, 776–780. https://doi.org/10.1145/3284179.3284309

[21] Jiang Bian and Mao Lin Huang. 2019. Semantic Topic Discovery for Lecture Video. In *Intelligent Systems and Applications - Proceedings of the 2019 Intelligent Systems Conference, IntelliSys 2019, London, UK, September 5-6, 2019, Volume 1 (Advances in Intelligent Systems and Computing, Vol. 1037)*, Yaxin Bi, Rahul Bhatia, and Supriya Kapoor (Eds.). Springer, 457–466. https://doi.org/10.1007/978-3-030-29516-5_36

[22] Diana Iulia Bleoanca, Stella Heras, Javier Palanca, Vicente Julián, and Marian Cristian Mihaescu. 2020. LSI Based Mechanism for Educational Videos Retrieval by Transcripts Processing. In *Intelligent Data Engineering and Automated Learning - IDEAL 2020 - 21st International Conference, Guimaraes, Portugal, November 4-6, 2020, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12489)*, Cesar Analide, Paulo Novais, David Camacho, and Hujun Yin (Eds.). Springer, 88–100. https://doi.org/10.1007/978-3-030-62362-3_9

[23] David Bordwell, Kristin Thompson, and Jeff Smith. 1993. *Film art: An introduction.* Vol. 7. McGraw-Hill New York.

[24] Marcos Vinícius Macêdo Borges and Júlio Cesar dos Reis. 2019. Semantic-Enhanced Recommendation of Video Lectures. In *19th IEEE International Conference on Advanced Learning Technologies, ICALT 2019, Maceió, Brazil, July 15-18, 2019*, Maiga Chang, Demetrios G. Sampson, Ronghuai Huang, Alex Sandro Gomes, Nian-Shing Chen, Ig Ibert Bittencourt, Kinshuk, Diego Dermeval, and Ibsen Mateus Bittencourt (Eds.). IEEE, 42–46. https://doi.org/10.1109/ICALT.2019.00013

[25] Paulo Renato Soares Borges and Ismar Frango Silveira. 2019. Adding and Segmenting Educational Videos: Experiences of Teacher Users in an Educational Portal. *IEEE Access* 7 (2019), 87996–88011. https://doi.org/10.1109/ACCESS.2019.2924946

[26] Max Bothe and Christoph Meinel. 2020. When Do Learners Rewatch Videos in MOOCs?. In *IEEE Learning With MOOCS, LWMOOCS 2020, Antigua Guatemala, Guatemala, September 29 - October 2, 2020*. IEEE, 148–151. https://doi.org/10.1109/LWMOOCS50143.2020.9234368

[27] Christopher G. Brinton, Swapna Buccapatnam, Mung Chiang, and H. Vincent Poor. 2016. Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance. *IEEE Trans. Signal Process.* 64, 14 (2016), 3677–3692. https://doi.org/10.1109/TSP.2016.2546228

[28] Sahan Bulathwela, María Pérez-Ortiz, Aldo Lipani, Emine Yilmaz, and John Shawe-Taylor. 2020. Predicting Engagement in Video Lectures. In *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020*, Anna N. Rafferty, Jacob Whitehill, Cristóbal Romero, and Violetta Cavalli-Sforza (Eds.). International Educational Data Mining Society. https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_62.pdf

[29] Luca Cagliero, Lorenzo Canale, and Laura Farinetti. 2019. VISA: A Supervised Approach to Indexing Video Lectures with Semantic Annotations. In *43rd IEEE Annual Computer Software and Applications Conference, COMPSAC 2019, Milwaukee, WI, USA, July 15-19, 2019, Volume 1*, Vladimir Getov, Jean-Luc Gaudiot, Nariyoshi Yamai, Stelvio Cimato, J. Morris Chang, Yuuichi Teranishi, Ji-Jiang Yang, Hong Va Leong, Hossain Shahriar, Michiharu Takemoto, Dave Towey, Hiroki Takakura, Atilla Elçi, Susumu Takeuchi, and Satish Puri (Eds.). IEEE, 226–235. https://doi.org/10.1109/COMPSAC.2019.00041

[30] Jianxia Cao, Akinori Nishihara, and Shijuan Wang. 2018. Effect of instructor inclusion type and course complexity on students' learning in lecturer video. In *Proceedings of the 10th International Conference on Education Technology and Computers, ICETC 2018, Tokyo, Japan, October 26-28, 2018*, Minoru Nakayama, Fang Lou, Betsy J. Bannier, and Harumi Watanabe (Eds.). ACM, 181–188. https://doi.org/10.1145/3290511.3290544

[31] Houssem Chatbri, Kevin McGuinness, Suzanne Little, Jiang Zhou, Keisuke Kameyama, Paul Kwan, and Noel E. O'Connor. 2017. Automatic MOOC Video Classification using Transcript Features and Convolutional Neural Networks. In *Proceedings of the 2017 ACM Workshop on Multimedia-based Educational and Knowledge Technologies for Personalized and Social Online Training, MultiEdTech@MM 2017, Mountain View, CA, USA, October 27, 2017*, Ansgar Scherp, Vasileios Mezaris, Thomas Köhler, and Alexander G. Hauptmann (Eds.). ACM, 21–26. https://doi.org/10.1145/3132390.3132393

[32] Xiaoyin Che, Thomas Staubitz, Haojin Yang, and Christoph Meinel. 2016. Pre-Course Key Segment Analysis of Online Lecture Videos. In *16th IEEE International Conference on Advanced Learning Technologies, ICALT 2016, Austin, TX, USA, July 25-28, 2016*, J. Michael Spector, Chin-Chung Tsai, Demetrios G. Sampson, Kinshuk, Ronghuai Huang, Nian-Shing Chen, and Paul E. Resta (Eds.). IEEE Computer Society, 416–420. https://doi.org/10.1109/ICALT.2016.102

[33] Chih-Hung Chen. 2020. AR videos as scaffolding to foster students' learning achievements and motivation in EFL learning. *Br. J. Educ. Technol.* 51, 3 (2020), 657–672. https://doi.org/10.1111/bjet.12902

[34] Si Chen, Dennis Wang, and Yun Huang. 2021. Exploring the complementary features of audio and text notes for video-based learning in mobile settings. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing systems*. 1–7.

[35] Yujia Chen, Chongwen Wang, and Zefeng Jian. 2020. Research on Evaluation Algorithm of Teacher's Teaching Enthusiasm Based on Video. In *ICRAI 2020: 6th International Conference on Robotics and Artificial Intelligence, Singapore, November 20-22, 2020*. ACM, 184–191. https://doi.org/10.1145/3449301.3449333

[36] Jinhan Choi, Jeongyun Han, Woochang Hyun, Hyunchul Lim, Sun Young Huh, SoHyun Park, and Bongwon Suh. 2019. Leveraging Smartwatches to Estimate Students' Perceived Difficulty and Interest in Online Video Lectures. In *ICETC 2019, 11th International Conference on Education Technology and Computers, Amsterdam, The Netherlands, October 28-31, 2019*, Piet Kommers and Fang Lou (Eds.). ACM, 171–175. https://doi.org/10.1145/3369255.3369291

[37] V Cisco. 2018. Cisco visual networking index: Forecast and trends, 2017–2022. *White paper* 1, 1 (2018).

[38] Marius Andrei Ciurez, Marian Cristian Mihaescu, Maite Giménez, Stella Heras, Javier Palanca, and Vicente Julián. 2019. Automatic Categorization of Educational Videos According to Learning Styles. In *2019 International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2019, Split, Croatia, September 19-21, 2019*, Dinko Begusic, Nikola Rozic, Josko Radic, and Matko Saric (Eds.). IEEE, 1–6. https://doi.org/10.

23919/SOFTCOM.2019.8903601

[39] J Clement. 2019. Hours of video uploaded to YouTube every minute. *Statista.com* (2019). https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/

[40] Mauro Coccoli and Gianni Viardo Vercelli. 2016. A Tool for the Semantic Analysis and Recommendation of videos in e-learning. In *The 22nd International Conference on Distributed Multimedia Systems, DMS 2016, Salerno, Italy, November 25-26, 2016*, Giuseppe Polese and Vincenzo Deufemia (Eds.). KSI Research Inc. / Knowledge Systems Institute Graduate School, 153–157. https://doi.org/10.18293/DMS2016-018

[41] Salomé Cojean and Eric Jamet. 2017. Facilitating information-seeking activity in instructional videos: The combined effects of micro- and macroscaffolding. *Comput. Hum. Behav.* 74 (2017), 294–302. https://doi.org/10.1016/j.chb.2017.04.052

[42] April Cookson, Daesang Kim, and Taralynn Hartsell. 2020. Enhancing Student Achievement, Engagement, and Satisfaction Using Animated Instructional Videos. *Int. J. Inf. Commun. Technol. Educ.* 16, 3 (2020), 113–125. https://doi.org/10.4018/IJICTE.2020070108

[43] Matthew Cooper, Jian Zhao, Chidansh Amitkumar Bhatt, and David A. Shamma. 2018. MOOCex: Exploring Educational Video via Recommendation. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ICMR 2018, Yokohama, Japan, June 11-14, 2018*, Kiyoharu Aizawa, Michael S. Lew, and Shin'ichi Satoh (Eds.). ACM, 521–524. https://doi.org/10.1145/3206025.3206087

[44] Jamie Costley, Mik Fanguy, Chris Lange, and Matthew Baldwin. 2021. The effects of video lecture viewing strategies on cognitive load. *J. Comput. High. Educ.* 33, 1 (2021), 19–38. https://doi.org/10.1007/s12528-020-09254-y

[45] Jamie Costley and Christopher Henry Lange. 2017. Video lectures in e-learning: Effects of viewership and media diversity on learning, satisfaction, engagement, interest, and future behavioral intention. *Interact. Technol. Smart Educ.* 14, 1 (2017), 14–30. https://doi.org/10.1108/ITSE-08-2016-0025

[46] Maria da Graça Campos Pimentel, Cristiane A. Yaguinuma, Diogo Santana Martins, and Isabela Zaine. 2019. Anchoring interactive points of interest on web-based instructional video: effects on students' interaction behavior and perceived experience. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC 2019, Limassol, Cyprus, April 8-12, 2019*, Chih-Cheng Hung and George A. Papadopoulos (Eds.). ACM, 2445–2452. https://doi.org/10.1145/3297280.3297521

[47] Wajeeh Daher and Hiba Sleem. 2021. Middle School Students' Learning of Social Studies in the Video and 360-Degree Videos Contexts. *IEEE Access* 9 (2021), 78774–78783. https://doi.org/10.1109/ACCESS.2021.3083924

[48] Ananda Das and Partha Pratim Das. 2020. Incorporating Domain Knowledge To Improve Topic Segmentation Of Long MOOC Lecture Videos. *CoRR* abs/2012.07589 (2020). arXiv:2012.07589 https://arxiv.org/abs/2012.07589

[49] Kenny Davila, Fei Xu, Srirangaraj Setlur, and Venu Govindaraju. 2021. FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos. *IEEE Access* 9 (2021), 104469–104484. https://doi.org/10.1109/ACCESS.2021.3099427

[50] Kenny Davila and Richard Zanibbi. 2018. Visual Search Engine for Handwritten and Typeset Math in Lecture Videos and LATEX Notes. In *16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018*. IEEE Computer Society, 50–55. https://doi.org/10.1109/ICFHR-2018.2018.00018

[51] Suman Deb, Anindya Pal, and Paritosh Bhattacharya. 2017. Design Considerations for Self Paced Interactive Notes on Video Lectures - A Learner's Perspective and Enhancements of Learning Outcome. In *Intelligent Human Computer Interaction - 9th International Conference, IHCI 2017, Evry, France, December 11-13, 2017, Proceedings (Lecture Notes in Computer Science, Vol. 10688)*, Patrick Horain, Catherine Achard, and Malik Mallem (Eds.). Springer, 109–121. https://doi.org/10.1007/978-3-319-72038-8_9

[52] Tim Van der Zee, Wilfried Admiraal, Fred Paas, Nadira Saab, and Bas Giesbers. 2017. Effects of Subtitles, Complexity, and Language Proficiency on Learning From Online Education Videos. *J. Media Psychol. Theor. Methods Appl.* 29, 1 (2017), 18–30. https://doi.org/10.1027/1864-1105/a000208

[53] Tim Van der Zee, Dan Davis, Nadira Saab, Bas Giesbers, Jasper Ginn, Frans van der Sluis, Fred Paas, and Wilfried Admiraal. 2018. Evaluating retrieval practice in a MOOC: how writing and reading summaries of videos affects student learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge, LAK 2018, Sydney, NSW, Australia, March 07-09, 2018*, Abelardo Pardo, Kathryn Bartimote-Aufflick, Grace Lynch, Simon Buckingham Shum, Rebecca Ferguson, Agathe Merceron, and Xavier Ochoa (Eds.). ACM, 216–225. https://doi.org/10.1145/3170358.3170382

[54] Danilo Dessì, Gianni Fenu, Mirko Marras, and Diego Reforgiato Recupero. 2017. Leveraging Cognitive Computing for Multi-class Classification of E-learning Videos. In *The Semantic Web: ESWC 2017 Satellite Events - ESWC 2017 Satellite Events, Portorož, Slovenia, May 28 - June 1, 2017, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 10577)*, Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Lawrynowicz, Fabio Ciravegna, and Olaf Hartig (Eds.). Springer, 21–25. https://doi.org/10.1007/978-3-319-70407-4_5

[55] Lu Ding, Joshua Adams, Michelle D. Stephens, Sara Brownell, and Michelene T. H. Chi. 2018. Failure to Replicate Using Dialogue Videos in Learning: Lessons Learned from an Authentic Course. In *Rethinking learning in the digital age: Making the Learning Sciences count - Proceedings of the 13th International Conference of the Learning Sciences, ICLS 2018, London, UK, June 23-27, 2018*, Manolis Mavrikis and Kaska Porayska-Pomsta (Eds.). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2018.953

[56] Yan Ding. 2018. Exploring EFL Learners' Perspectives on Instructional Videos. *Int. J. Comput. Assist. Lang. Learn. Teach.* 8, 2 (2018), 60–74. https://doi.org/10.4018/IJCALLT.2018040104

[57] Samuel Dodson, Ido Roll, Matthew Fong, Dongwook Yoon, Negar M. Harandi, and Sidney S. Fels. 2018. Active Viewing: A Study of Video Highlighting in the Classroom. In *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, CHIIR 2018, New Brunswick, NJ, USA, March 11-15, 2018*, Chirag Shah, Nicholas J. Belkin, Katriina Byström, Jeff Huang, and Falk Scholer (Eds.). ACM, 237–240. https://doi.org/10.1145/3176349.3176889

[58] Samuel Dodson, Ido Roll, Matthew Fong, Dongwook Yoon, Negar M. Harandi, and Sidney S. Fels. 2018. An active viewing framework for video-based learning. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale, London, UK, June 26-28, 2018*, Rose Luckin, Scott R. Klemmer, and

Kenneth R. Koedinger (Eds.). ACM, 24:1–24:4. https://doi.org/10.1145/3231644.3231682

[59] Irina Engeness and Magnus Nohr. 2019. Videos as teaching and learning resources for developing of pre- and in-service teachers' professional digital competence in the ICTMOOC. In *Proceedings of Work in Progress Papers of the Research, Experience and Business Tracks at EMOOCs 2019, co-located with the European MOOCs Stakeholders Summit 2019 Conference, Naples, Italy, May 20-22, 2019 (CEUR Workshop Proceedings, Vol. 2356)*, Mauro Calise, Carlos Delgado Kloos, Catherine Mongenet, Justin Reich, José A. Ruipérez Valiente, Gideon Shimshon, Thomas Staubitz, and Martin Wirsing (Eds.). CEUR-WS.org, 128–133. http://ceur-ws.org/Vol-2356/experience_short8.pdf

[60] Maka Eradze, Anna Dipace, Bojan Fazlagic, and Anastasia Di Pietro. 2020. Semi-automated Student Feedback and Theory-Driven Video-Analytics: An Exploratory Study on Educational Value of Videos. In *Bridges and Mediation in Higher Distance Education - Second International Workshop, HELMeTO 2020, Bari, BA, Italy, September 17-18, 2020, Revised Selected Papers (Communications in Computer and Information Science, Vol. 1344)*, Laura Sara Agrati, Daniel Burgos, Pietro Ducange, Pierpaolo Limone, Loredana Perla, Pietro Picerno, Paolo Raviolo, and Christian M. Stracke (Eds.). Springer, 28–39. https://doi.org/10.1007/978-3-030-67435-9_3

[61] Javier Escobar-Avila, Esteban Parra, and Sonia Haiduc. 2017. Text retrieval-based tagging of software engineering video tutorials. In *Proceedings of the 39th International Conference on Software Engineering, ICSE 2017, Buenos Aires, Argentina, May 20-28, 2017 - Companion Volume*, Sebastián Uchitel, Alessandro Orso, and Martin P. Robillard (Eds.). IEEE Computer Society, 341–343. https://doi.org/10.1109/ICSE-C.2017.121

[62] Matthew Fong, Samuel Dodson, Xueqin Zhang, Ido Roll, and Sidney S. Fels. 2018. ViDeX: A Platform for Personalizing Educational Videos. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, JCDL 2018, Fort Worth, TX, USA, June 03-07, 2018*, Jiangping Chen, Marcos André Gonçalves, Jeff M. Allen, Edward A. Fox, Min-Yen Kan, and Vivien Petras (Eds.). ACM, 331–332. https://doi.org/10.1145/3197026.3203865

[63] C. Ailie Fraser, Tricia J. Ngoon, Mira Dontcheva, and Scott R. Klemmer. 2019. RePlay: Contextually Presenting Learning Videos Across Software Applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 297. https://doi.org/10.1145/3290605.3300527

[64] Marco Furini. 2018. On introducing timed tag-clouds in video lectures indexing. *Multim. Tools Appl.* 77, 1 (2018), 967–984. https://doi.org/10.1007/s11042-016-4282-5

[65] Marco Furini. 2021. Automatic and Smart Content Transformation of Video Lectures. In *18th IEEE Annual Consumer Communications & Networking Conference, CCNC 2021, Las Vegas, NV, USA, January 9-12, 2021*. IEEE, 1–6. https://doi.org/10.1109/CCNC49032.2021.9369592

[66] Marco Furini, Giovanna Galli, and Maria Cristiana Martini. 2021. On Using Video Lectures Data Usage to Predict University Students Dropout. In *GoodIT '21: Conference on Information Technology for Social Good, Roma, Italy, September 9-11, 2021*, Ombretta Gaggi, Pietro Manzoni, and Claudio E. Palazzi (Eds.). ACM, 313–316. https://doi.org/10.1145/3462203.3475890

[67] Marco Furini, Silvia Mirri, and Manuela Montangero. 2018. Topic-based playlist to improve video lecture accessibility. In *15th IEEE Annual Consumer Communications & Networking Conference, CCNC 2018, Las Vegas, NV, USA, January 12-15, 2018*. IEEE, 1–5. https://doi.org/10.1109/CCNC.2018.8319246

[68] Masako Furukawa, Hiroshi Itsumura, and Kazutsuna Yamaji. 2020. Estimation of Test Scores Based on Video Viewing Behavior in the Programming MOOC Course. In *9th International Congress on Advanced Applied Informatics, IIAI-AAI 2020, Kitakyushu, Japan, September 1-15, 2020*, Tokuro Matsuo, Kunihiko Takamatsu, Yuichi Ono, and Sachio Hirokawa (Eds.). IEEE, 155–162. https://doi.org/10.1109/IIAI-AAI50415.2020.00039

[69] Damianos Galanopoulos and Vasileios Mezaris. 2019. Temporal Lecture Video Fragmentation Using Word Embeddings. In *MultiMedia Modeling - 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 11296)*, Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis (Eds.). Springer, 254–265. https://doi.org/10.1007/978-3-030-05716-9_21

[70] Matthias Galster, Antonija Mitrovic, and Matthew Gordon. 2018. Toward enhancing the training of software engineering students and professionals using active video watching. In *Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training, ICSE (SEET) 2018, Gothenburg, Sweden, May 27 - June 03, 2018*, Patricia Lago and Michal Young (Eds.). ACM, 5–8. https://doi.org/10.1145/3183377.3183384

[71] Ankit Gandhi, Arijit Biswas, Kundan Shrivastava, Ranjeet Kumar, Sahil Loomba, and Om Deshmukh. 2016. Easy Navigation through Instructional Videos using Automatically Generated Table of Content. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces, IUI 2016, Sonoma, CA, USA, March 7-10, 2016*, Jeffrey Nichols, Jalal Mahmud, John O'Donovan, Cristina Conati, and Massimo Zancanaro (Eds.). ACM, 92–96. https://doi.org/10.1145/2876456.2879472

[72] Solomon E. Garber, Aaditya Prakash, Nick Moran, Maria Altebarmakian, Richard Alterman, Antonella Di Lillo, and James A. Storer. 2017. A two tier approach to chalkboard video lecture summary. In *2017 IEEE Frontiers in Education Conference, FIE 2017, Indianapolis, IN, USA, October 18-21, 2017*. IEEE Computer Society, 1–9. https://doi.org/10.1109/FIE.2017.8190536

[73] Nathan Garrett. 2018. Video Tutorials Improve Students' Ability to Recover from Errors. In *24th Americas Conference on Information Systems, AMCIS 2018, New Orleans, LA, USA, August 16-18, 2018*. Association for Information Systems. https://aisel.aisnet.org/amcis2018/Education/Presentations/19

[74] Nathan Garrett. 2021. Segmentation's failure to improve software video tutorials. *Br. J. Educ. Technol.* 52, 1 (2021), 318–336. https://doi.org/10.1111/bjet.13000

[75] Jun Ge and Xing Li. 2020. Design Strategies of EFL Learning Videos: Exampled by a China MOOC. In *ICEIT 2020, Proceedings of the 9th International Conference on Educational and Information Technology, Oxford, UK, February11-13, 2020*. ACM, 68–71. https://doi.org/10.1145/3383923.3383927

[76] Junaid Ahmed Ghauri, Sherzod Hakimov, and Ralph Ewerth. 2020. Classification of Important Segments in Educational Videos using Multimodal Features. In *Proceedings of the CIKM 2020 Workshops co-located with 29th ACM International Conference on Information and Knowledge Management (CIKM 2020), Galway, Ireland, October 19-23, 2020 (CEUR Workshop Proceedings, Vol. 2699)*, Stefan Conrad and Ilaria Tiddi (Eds.). CEUR-WS.org. http://ceur-ws.org/Vol-2699/paper15.pdf

[77] Krishnendu Ghosh, Sharmila Reddy Nangi, Yashasvi Kanchugantla, Pavan Gopal Rayapati, Plaban Kumar Bhowmick, and Pawan Goyal. 2021. Augmenting Video Lectures: Identifying Off-topic Concepts and Linking to Relevant Video Lecture Segments. *International Journal of Artificial Intelligence in Education* (2021), 1–31. https://doi.org/10.1007/s40593-021-00257-z

[78] Michail N. Giannakos. 2013. Exploring the video-based learning research: A review of the literature. *Br. J. Educ. Technol.* 44, 6 (2013), 191. https://doi.org/10.1111/bjet.12070

[79] Michail N. Giannakos, Letizia Jaccheri, and John Krogstie. 2016. Exploring the relationship between video lecture usage patterns and students' attitudes. *Br. J. Educ. Technol.* 47, 6 (2016), 1259–1275. https://doi.org/10.1111/bjet.12313

[80] Deekshith Godavarthi, Aswath S, Debanik Mishra, and Jayashree R. 2020. Analysing Emotions on Lecture Videos using CNN AND HOG (Workshop Paper). In *6th IEEE International Conference on Multimedia Big Data, BigMM 2020, New Delhi, India, September 24-26, 2020.* IEEE, 435–440. https://doi.org/10.1109/BigMM50055.2020.00072

[81] Uma Gopalakrishnan, P. Venkat Rangan, Ramkumar N, and Balaji Hariharan. 2017. Spatio-Temporal Compositing of Video Elements for Immersive eLearning Classrooms. In *19th IEEE International Symposium on Multimedia, ISM 2017, Taichung, Taiwan, December 11-13, 2017.* IEEE Computer Society, 138–145. https://doi.org/10.1109/ISM.2017.120

[82] José Granjo and Maria G. Rasteiro. 2018. LABVIRTUAL - A platform for the teaching of chemical engineering: The use of interactive videos. *Comput. Appl. Eng. Educ.* 26, 5 (2018), 1668–1676. https://doi.org/10.1002/cae.22007

[83] Minhuang Guan, Kai Li, Ran Ma, and Ping An. 2019. Convolutional-Block-Attention Dual Path Networks for Slide Transition Detection in Lecture Videos. In *Digital TV and Wireless Multimedia Communication - 16th International Forum, IFTC 2019, Shanghai, China, September 19-20, 2019, Revised Selected Papers (Communications in Computer and Information Science, Vol. 1181)*, Guangtao Zhai, Jun Zhou, Hua Yang, Ping An, and Xiaokang Yang (Eds.). Springer, 103–114. https://doi.org/10.1007/978-981-15-3341-9_9

[84] Esther Guervós, Jaime J. Ruiz, Pablo Pérez, Juan Alberto Muñoz, César Díaz, and Narciso García. 2019. Using 360 VR video to improve the learning experience in veterinary medicine university degree. In *Human Vision and Electronic Imaging 2019, Burlingame, CA, USA, 13-17 January 2019*, Damon M. Chandler, Mark E. McCourt, and Jeffrey B. Mulligan (Eds.). Ingenta. https://doi.org/10.2352/ISSN.2470-1173.2019.12.HVEI-217

[85] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How video production affects student engagement: an empirical study of MOOC videos. In *First (2014) ACM Conference on Learning @ Scale, L@S 2014, Atlanta, GA, USA, March 4-5, 2014*, Mehran Sahami, Armando Fox, Marti A. Hearst, and Michelene T. H. Chi (Eds.). ACM, 41–50. https://doi.org/10.1145/2556325.2566239

[86] Anna Hansch, Lisa Hillers, Katherine McConachie, Christopher Newman, Thomas Schildhauer, and J Philipp Schmidt. 2015. Video and online learning: Critical reflections and findings from the field. (2015). https://doi.org/10.2139/ssrn.2577882

[87] Arno Hartholt, Adam Reilly, Edward Fast, and Sharon Mozgai. 2020. Introducing Canvas: Combining Nonverbal Behavior Generation with User-Generated Content to Rapidly Create Educational Videos. In *IVA '20: ACM International Conference on Intelligent Virtual Agents, Virtual Event, Scotland, UK, October 20-22, 2020*, Stacy Marsella, Rachael Jack, Hannes Högni Vilhjálmsson, Pedro Sequeira, and Emily S. Cross (Eds.). ACM, 25:1–25:3. https://doi.org/10.1145/3383652.3423880

[88] Masaki Hayashi, Steven Bachelder, and Masayuki Nakajima. 2020. Vizclass: Making Computer Graphics Video Lecture from a Script. In *9th IEEE Global Conference on Consumer Electronics, GCCE 2020, Kobe, Japan, October 13-16, 2020.* IEEE, 688–689. https://doi.org/10.1109/GCCE50665.2020.9292069

[89] Huan He, Qinghua Zheng, and Bo Dong. 2018. VUSphere: Visual Analysis of Video Utilization in Online Distance Education. In *13th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2018, Berlin, Germany, October 21-26, 2018*, Remco Chang, Huamin Qu, and Tobias Schreck (Eds.). IEEE, 25–35. https://doi.org/10.1109/VAST.2018.8802383

[90] Huan He, Qinghua Zheng, Bo Dong, and Hongchao Yu. 2018. Measuring Student's Utilization of Video Resources and Its Effect on Academic Performance. In *18th IEEE International Conference on Advanced Learning Technologies, ICALT 2018, Mumbai, India, July 9-13, 2018*, Maiga Chang, Nian-Shing Chen, Ronghuai Huang, Kinshuk, Kannan M. Moudgalya, Sahana Murthy, and Demetrios G. Sampson (Eds.). IEEE Computer Society, 196–198. https://doi.org/10.1109/ICALT.2018.00053

[91] Romain Christian Herault, Alisa Lincke, Marcelo Milrad, Elin-Sofie Forsgärde, and Carina Elmqvist. 2018. Using 360-degrees interactive videos in patient trauma treatment education: design, development and evaluation aspects. *Smart Learn. Environ.* 5, 1 (2018), 26. https://doi.org/10.1186/s40561-018-0074-x

[92] Abner Hernandez and Seung Hee Yang. 2021. Multimodal Corpus Analysis of Autoblog 2020: Lecture Videos in Machine Learning. In *Speech and Computer - 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27-30, 2021, Proceedings (Lecture Notes in Computer Science, Vol. 12997)*, Alexey Karpov and Rodmonga Potapova (Eds.). Springer, 262–270. https://doi.org/10.1007/978-3-030-87802-3_24

[93] Khe Foon Hew and Chung Kwan Lo. 2020. Comparing video styles and study strategies during video-recorded lectures: effects on secondary school mathematics students' preference and learning. *Interact. Learn. Environ.* 28, 7 (2020), 847–864. https://doi.org/10.1080/10494820.2018.1545671

[94] Jordan Dale Hildebrand and Benjamin Ahn. 2018. Student Video Viewing Habits in an Online Mechanics of Materials Engineering Course. *Int. J. Eng. Pedagog.* 8, 3 (2018), 40–59. https://doi.org/10.3991/ijep.v8i3.7948

[95] Mark Hodges. 2018. Ensuring videos for a flipped classroom increase inclusivity. *Journal of computing Sciences in Colleges* 34, 1 (2018), 81–88.

[96] Vincent Hoogerheide, Margot van Wermeskerken, Hilke van Nassau, and Tamara van Gog. 2018. Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Comput. Hum. Behav.* 89 (2018), 457–464. https://doi.org/10.1016/j.chb.2017.11.012

[97] Shay Horovitz and Yaniv Ohayon. 2020. Boocture: Automatic Educational Videos Hierarchical Indexing with eBooks. In *IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020, Takamatsu, Japan, December 8-11, 2020*, Hiroyuki Mitsuhara, Yoshiko Goda, Yutato Ohashi, Ma. Mercedes T. Rodrigo, Jun Shen, Neelakantam Venkatarayalu, Gary Wong, Masanori Yamada, and Chi-Un Lei (Eds.). IEEE, 482–489. https://doi.org/10.1109/TALE48869.2020.9368461

[98] Tal Horovitz and Richard E. Mayer. 2021. Learning with human and virtual instructors who display happy or bored emotions in video lectures. *Comput. Hum. Behav.* 119 (2021), 106724. https://doi.org/10.1016/j.chb.2021.106724

[99] Hui Hu, Guofeng Zhang, Wanlin Gao, and Minjuan Wang. 2020. Big data analytics for MOOC video watching behavior based on Spark. *Neural Comput. Appl.* 32, 11 (2020), 6481–6489. https://doi.org/10.1007/s00521-018-03983-z

[100] Hsiu-Ling Huang, Gwo-Jen Hwang, and Ching-Yi Chang. 2020. Learning to be a writer: A spherical video-based virtual reality approach to supporting descriptive article writing in high school Chinese courses. *Br. J. Educ. Technol.* 51, 4 (2020), 1386–1405. https://doi.org/10.1111/bjet.12893

[101] I-Chun Hung, Kinshuk, and Nian-Shing Chen. 2018. Embodied interactive video lectures for improving learning comprehension and retention. *Comput. Educ.* 117 (2018), 116–131. https://doi.org/10.1016/j.compedu.2017.10.005

[102] Moula Husain and S. M. Meena. 2019. Multimodal Fusion of Speech and Text using Semi-supervised LDA for Indexing Lecture Videos. In *National Conference on Communications, NCC 2019, Bangalore, India, February 20-23, 2019*. IEEE, 1–6. https://doi.org/10.1109/NCC.2019.8732253

[103] Ali Shariq Imran, Zenun Kastrati, Torbjørn Karl Svendsen, and Arianit Kurti. 2019. Text-Independent Speaker ID Employing 2D-CNN for Automatic Video Lecture Categorization in a MOOC Setting. In *31st IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2019, Portland, OR, USA, November 4-6, 2019*. IEEE, 273–277. https://doi.org/10.1109/ICTAI.2019.00046

[104] Byungsoo Jeon and Namyong Park. 2020. Dropout Prediction over Weeks in MOOCs by Learning Representations of Clicks and Videos. *CoRR* abs/2002.01955 (2020). arXiv:2002.01955 https://arxiv.org/abs/2002.01955

[105] Seonmi Ji, Yongsung Kim, Seungwon Jung, and Eenjun Hwang. 2018. An Educational Video Scoring Scheme for Level and Topic-based Recommendation. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, IMCOM 2018, Langkawi, Malaysia, January 05-07, 2018*. ACM, 12:1–12:5. https://doi.org/10.1145/3164541.3164636

[106] Jaechoon Jo and Heuiseok Lim. 2017. A Study of Keywords Based on the Word Frequency Effect Theory in Video Lectures of Software Engineering Education for Detecting Mind. In *30th IEEE Conference on Software Engineering Education and Training, CSEE&T 2017, Savannah, GA, USA, November 7-9, 2017*, Hironori Washizaki and Nancy Mead (Eds.). IEEE, 241–245. https://doi.org/10.1109/CSEET.2017.46

[107] Kyung Je Jo, Hyeonggeun Yun, and Juho Kim. 2019. Supporting Instruction of Formulaic Sequences Using Videos at Scale. In *Proceedings of the Sixth ACM Conference on Learning @ Scale, L@S 2019, Chicago, IL, USA, June 24-25, 2019*. ACM, 57:1–57:4. https://doi.org/10.1145/3330430.3333671

[108] Jaume Jordán, Soledad Valero, Carlos Turró, and Vicent J. Botti. 2020. Recommending Learning Videos for MOOCs and Flipped Classrooms. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness. The PAAMS Collection - 18th International Conference, PAAMS 2020, L'Aquila, Italy, October 7-9, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12092)*, Yves Demazeau, Tom Holvoet, Juan M. Corchado, and Stefania Costantini (Eds.). Springer, 146–157. https://doi.org/10.1007/978-3-030-49778-1_12

[109] Jorão Gomes Jr., Laura Lima Dias, Eduardo Rocha Soares, Eduardo Barrére, and Jairo Francisco de Souza. 2019. Framework for Knowledge Discovery in Educational Video Repositories. *Comput. Informatics* 38, 6 (2019), 1375–1402. https://doi.org/10.31577/cai_2019_6_1375

[110] Hyeungshik Jung, Hijung Valentina Shin, and Juho Kim. 2018. DynamicSlide: Reference-based Interaction Techniques for Slide-based Lecture Videos. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings, UIST 2018, Berlin, Germany, October 14-17, 2018*, Patrick Baudisch, Albrecht Schmidt, and Andy Wilson (Eds.). ACM, 23–25. https://doi.org/10.1145/3266037.3266089

[111] Seungwon Jung, Minjae Son, Chung-Il Kim, Jehyeok Rew, and Eenjun Hwang. 2019. Video-based learning assistant scheme for sustainable education. *New Rev. Hypermedia Multim.* 25, 3 (2019), 161–181. https://doi.org/10.1080/13614568.2019.1678682

[112] Manish Kanadje, Zachary Miller, Anurag Agarwal, Roger S. Gaborski, Richard Zanibbi, and Stephanie Ludi. 2016. Assisted keyword indexing for lecture videos using unsupervised keyword spotting. *Pattern Recognit. Lett.* 71 (2016), 8–15. https://doi.org/10.1016/j.patrec.2015.11.012

[113] Alexander Keller, Michael Langbauer, Thomas Fritsch, and Franz Lehner. 2019. Interactive Videos vs. Hypertext Documents - The Effect on Learning Quality and Time Effort when Acquiring Procedural Knowledge. In *52nd Hawaii International Conference on System Sciences, HICSS 2019, Grand Wailea, Maui, Hawaii, USA, January 8-11, 2019*, Tung Bui (Ed.). ScholarSpace, 1–10. https://doi.org/10.24251/HICSS.2019.005

[114] Kandarp Khandwala and Philip J. Guo. 2018. Codemotion: expanding the design space of learner interactions with computer programming tutorial videos. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale, London, UK, June 26-28, 2018*, Rose Luckin, Scott R. Klemmer, and Kenneth R. Koedinger (Eds.). ACM, 57:1–57:10. https://doi.org/10.1145/3231644.3231652

[115] Zaynab El Khattabi, Youness Tabii, and Abdelhamid Benkaddour. 2019. Visual content summarisation for instructional videos using AdaBoost and SIFT. *Int. J. Knowl. Eng. Soft Data Paradigms* 6, 3/4 (2019), 207–218. https://doi.org/10.1504/IJKESDP.2019.103939

[116] Ritsuko Kikusawa, Motoko Okumoto, Takuya Kubo, and Laura Rodrigo. 2016. Developing eLecture Materials for Hearing Impaired Students and Researchers: Designing Multiple-Video Programs and Usability Assessment. In *Computers Helping People with Special Needs - 15th International Conference, ICCHP 2016, Linz, Austria, July 13-15, 2016, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 9759)*, Klaus Miesenberger, Christian Bühler, and Petr Penáz (Eds.). Springer, 415–418. https://doi.org/10.1007/978-3-319-41267-2_58

[117] Hyun Hee Kim and Yong Ho Kim. 2016. Generic speech summarization of transcribed lecture videos: Using tags and their semantic relations. *J. Assoc. Inf. Sci. Technol.* 67, 2 (2016), 366–379. https://doi.org/10.1002/asi.23391

[118] Alexandros Kleftodimos and Georgios Evangelidis. 2016. An Interactive Video-based Learning Environment that Supports Learning Analytics for Teaching 'Image Editing'. In *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning, Edinburgh, Scotland, April 26, 2016 (CEUR Workshop Proceedings, Vol. 1579)*, Michail N. Giannakos, Demetrios G. Sampson, Lukasz Kidzinski, and Abelardo Pardo (Eds.). CEUR-WS.org, 26–33. https://doi.org/10.1007/978-981-287-868-7_56

[119] Alexandros Kleftodimos and Georgios Evangelidis. 2018. Augmenting educational videos with interactive exercises and knowledge testing games. In *2018 IEEE Global Engineering Education Conference, EDUCON 2018, Santa Cruz de Tenerife, Tenerife, Islas Canarias, Spain, April 17-20, 2018*. IEEE, 872–877. https://doi.org/10.1109/EDUCON.2018.8363322

[120] Dan Kohen-Vacs, Marcelo Milrad, Miky Ronen, and Marc Jansen. 2016. Evaluation of enhanced educational experiences using interactive videos and web technologies: pedagogical and architectural considerations. *Smart Learn. Environ.* 3, 1 (2016), 6. https://doi.org/10.1186/s40561-016-0029-z

[121] Raga Shalini Koka, Farah Naz Chowdhury, Mohammad Rajiur Rahman, Thamar Solorio, and Jaspal Subhlok. 2020. Automatic Identification of Keywords in Lecture Video Segments. In *IEEE International Symposium on Multimedia, ISM 2020, Naples, Italy, December 2-4, 2020*. IEEE, 162–165. https://doi.org/10.1109/ISM.2020.00035

[122] Gábor Körösi, Peter Esztelecki, Richárd Farkas, and Krisztina Tóth. 2018. Clickstream-based outcome prediction in short video MOOCs. In *2018 International Conference on Computer, Information and Telecommunication Systems, CITS 2018, Alsace, Colmar, France, July 11-13, 2018*, Mohammad S. Obaidat, Pascal Lorenz, Kuei-Fang Hsiao, Petros Nicopolitidis, and Daniel Cascado Caballero (Eds.). IEEE, 1–5. https://doi.org/10.1109/CITS.2018.8440182

[123] Bhargava Urala Kota, Kenny Davila, Alexander Stone, Srirangaraj Setlur, and Venu Govindaraju. 2019. Generalized framework for summarization of fixed-camera lecture videos by detecting and binarizing handwritten content. *Int. J. Document Anal. Recognit.* 22, 3 (2019), 221–233. https://doi.org/10.1007/s10032-019-00327-y

[124] Bhargava Urala Kota, Alexander Stone, Kenny Davila, Srirangaraj Setlur, and Venu Govindaraju. 2020. Automated Whiteboard Lecture Video Summarization by Content Region Detection and Representation. In *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*. IEEE, 10704–10711. https://doi.org/10.1109/ICPR48806.2021.9412386

[125] Konstantinos Koumaditis and Francesco Chinello. 2019. Virtual Immersive Educational Systems: The case of 360° video and co-learning design. In *25th ACM Symposium on Virtual Reality Software and Technology, VRST 2019, Parramatta, NSW, Australia, November 12-15, 2019*, Tomas Trescak, Simeon Simoff, Deborah Richards, Anton Bogdanovych, Thierry Duval, Torsten W. Kuhlen, Huyen Nguyen, Shigeo Morishima, Yuichi Itoh, Richard Skarbez, and Martin Masek (Eds.). ACM, 99:1–99:2. https://doi.org/10.1145/3359996.3364714

[126] Geza Kovacs. 2016. Effects of In-Video Quizzes on MOOC Lecture Viewing. In *Proceedings of the Third ACM Conference on Learning @ Scale, L@S 2016, Edinburgh, Scotland, UK, April 25 - 26, 2016*, Jeff Haywood, Vincent Aleven, Judy Kay, and Ido Roll (Eds.). ACM, 31–40. https://doi.org/10.1145/2876034.2876041

[127] Dimitrios Kravvaris and Katia Lida Kermanidis. 2017. Advanced searching framework for open online educational video lectures. *Soc. Netw. Anal. Min.* 7, 1 (2017), 31:1–31:10. https://doi.org/10.1007/s13278-017-0452-3

[128] Dimitrios Kravvaris and Katia Lida Kermanidis. 2019. Automatic point of interest detection for open online educational video lectures. *Multim. Tools Appl.* 78, 2 (2019), 2465–2479. https://doi.org/10.1007/s11042-018-6372-z

[129] Wen-Ping Ku, Kai-Hsiang Yang, and Wan Ling Chang. 2019. The Design and Evaluation of Interactive Video-Based Flipped Classroom on Mathematics Learning. In *8th International Congress on Advanced Applied Informatics, IIAI-AAI 2019, Toyama, Japan, July 7-11, 2019*. IEEE, 1041–1042. https://doi.org/10.1109/IIAI-AAI.2019.00217

[130] Ashraf Ahmed Kuhail and Magdy Saeed Aqel. 2020. Interactive Digital Videos and Their Impact on Sixth Graders' English Reading and Vocabulary Skills and Retention. *Int. J. Inf. Commun. Technol. Educ.* 16, 3 (2020), 42–56. https://doi.org/10.4018/IJICTE.2020070104

[131] Abhishek Kumar, Kushal Srivastava, Kuldeep Yadav, and Om Deshmukh. 2017. Multi-faceted Index driven Navigation for Educational Videos in Mobile Phones. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI 2017, Limassol, Cyprus, March 13-16, 2017*, George A. Papadopoulos, Tsvi Kuflik, Fang Chen, Carlos Duarte, and Wai-Tat Fu (Eds.). ACM, 357–361. https://doi.org/10.1145/3025171.3025221

[132] Ah-Fur Lai, Wei-Hsin Li, and Horng-Yih Lai. 2018. A Study of Developing a Web-Based Video Annotation System and Evaluating Its Suitability on Learning. In *Proceedings of the 2nd International Conference on Education and Multimedia Technology* (Okinawa, Japan) *(ICEMT '18)*. Association for Computing Machinery, New York, NY, USA, 44–48. https://doi.org/10.1145/3206129.3239421

[133] Sébastien Lallé and Cristina Conati. 2020. A Data-Driven Student Model to Provide Adaptive Support During Video Watching Across MOOCs. In *Artificial Intelligence in Education - 21st International Conference, AIED 2020, Ifrane, Morocco, July 6-10, 2020, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12163)*, Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Millán (Eds.). Springer, 282–295. https://doi.org/10.1007/978-3-030-52237-7_23

[134] David Lang, Guanling Chen, Kathy Mirzaei, and Andreas Paepcke. 2020. Is faster better?: a study of video playback speed. In *LAK '20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23-27, 2020*, Christoph Rensing and Hendrik Drachsler (Eds.). ACM, 260–269. https://doi.org/10.1145/3375462.3375466

[135] Christopher Lange and Jamie Costley. 2020. Improving online video lectures: learning challenges created by media. *International Journal of Educational Technology in Higher Education* 17, 1 (2020), 1–18. https://doi.org/10.1186/s41239-020-00190-6

[136] Bridjet Lee and Kasia Muldner. 2020. Instructional Video Design: Investigating the Impact of Monologue- and Dialogue-style Presentations. In *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, 1–12. https://doi.org/10.1145/3313831.3376845

[137] Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, and Gani Kelni. 2017. Making sense of comments on YouTube educational videos: a self-directed learning perspective. *Online Inf. Rev.* 41, 5 (2017), 611–625. https://doi.org/10.1108/OIR-09-2016-0274

[138] Greg C. Lee, Fu-Hao Yeh, Ying-Ju Chen, and Tao-Ku Chang. 2017. Robust handwriting extraction and lecture video summarization. *Multim. Tools Appl.* 76, 5 (2017), 7067–7085. https://doi.org/10.1007/s11042-016-3353-y

[139] Chi-Un Lei, Donn Gonda, Xiangyu Hou, Elizabeth Oh, Xinyu Qi, Tyrone T. O. Kwok, Yip-Chun Au Yeung, and Ray Lau. 2017. Data-assisted instructional video revision via course-level exploratory video retention analysis. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, March 13-17, 2017*, Marek Hatala, Alyssa Wis, Phil Winne, Grace Lynch, Xavier Ochoa, Inge Molenaar, Shane Dawson, Shady Shehata, and Jennifer Pei-Ling Tan (Eds.). ACM, 554–555. https://doi.org/10.1145/3027385.3029454

[140] Chi-Un Lei, Cheuk-Wang Yau, King-Shan Lui, Vincent W. L. Tam, Allan Hoi Kau Yuen, and Edmund Y. Lam. 2019. Designing instructional videos and classwork activities: teaching internet of things via flipped classroom. *Int. J. Mob. Learn. Organisation* 13, 4 (2019), 392–411. https://doi.org/10.1504/IJMLO.2019.102611

[141] Chi-Un Lei, Yip-Chun Au Yeung, Tyrone T. O. Kwok, Ray Lau, and Andersen Man Shun Ang. 2016. Leveraging videos and forums for small-class learning experience in a MOOC environment. In *IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2016, Bangkok, Thailand, December 7-9, 2016*. IEEE, 409–411. https://doi.org/10.1109/TALE.2016.7851831

[142] David Leisner, Carmen G. Zahn, Alessia Ruf, and Alberto A. P. Cattaneo. 2020. Different Ways of Interacting with Videos During Learning in Secondary Physics Lessons. In *HCI International 2020 - Posters - 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part II (Communications in Computer and Information Science, Vol. 1225)*, Constantine Stephanidis and Margherita Antona (Eds.). Springer, 284–291. https://doi.org/10.1007/978-3-030-50729-9_40

[143] David John Lemay and Tenzin Doleck. 2020. Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior. *Educ. Inf. Technol.* 25, 2 (2020), 1333–1342. https://doi.org/10.1007/s10639-019-10022-4

[144] Jimmie Leppink, Fred Paas, Cees PM Van der Vleuten, Tamara Van Gog, and Jeroen JG Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behavior research methods* 45, 4 (2013), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

[145] Jamy Li, René F. Kizilcec, Jeremy N. Bailenson, and Wendy Ju. 2016. Social robots and virtual agents as lecturers for video instruction. *Comput. Hum. Behav.* 55 (2016), 1222–1230. https://doi.org/10.1016/j.chb.2015.04.005

[146] Liang-Yi Li. 2019. Effect of Prior Knowledge on Attitudes, Behavior, and Learning Performance in Video Lecture Viewing. *Int. J. Hum. Comput. Interact.* 35, 4-5 (2019), 415–426. https://doi.org/10.1080/10447318.2018.1543086

[147] Cheng-Ho Lin, Chih-Ming Chen, and Yong-Teng Lin. 2018. Improving Effectiveness of Learners' Review of Video Lectures by Using an Attention-Based Video Lecture Review Mechanism Based on Brainwave Signals. In *7th International Congress on Advanced Applied Informatics, IIAI-AAI 2018, Yonago, Japan, July 8-13, 2018*. IEEE, 152–157. https://doi.org/10.1109/IIAI-AAI.2018.00038

[148] Jinjiao Lin, Chunfang Liu, Yibin Li, Lizhen Cui, Rui Wang, Xudong Lu, Yan Zhang, and Jian Lian. 2019. Automatic Knowledge Discovery in Lecturing Videos via Deep Representation. *IEEE Access* 7 (2019), 33957–33963. https://doi.org/10.1109/ACCESS.2019.2904046

[149] Yong-Teng Lin and Chih-Ming Chen. 2019. Improving effectiveness of learners' review of video lectures by using an attention-based video lecture review mechanism based on brainwave signals. *Interact. Learn. Environ.* 27, 1 (2019), 86–102. https://doi.org/10.1080/10494820.2018.1451899

[150] Ching (Jean) Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative Concept Mapping for Video Learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 387. https://doi.org/10.1145/3173574.3173961

[151] Ching (Jean) Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. NoteStruct: Scaffolding Note-taking while Learning from Online Videos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Regan L. Mandryk, Stephen A. Brewster, Mark Hancock, Geraldine Fitzpatrick, Anna L. Cox, Vassilis Kostakos, and Mark Perry (Eds.). ACM. https://doi.org/10.1145/3290607.3312878

[152] Xiao Liu, Bin Zhang, Anjana Susarla, and Rema Padman. 2018. YouTube for Patient Education: A Deep Learning Approach for Understanding Medical Knowledge from User-Generated Videos. *CoRR* abs/1807.03179 (2018). arXiv:1807.03179 http://arxiv.org/abs/1807.03179

[153] Zhijin Liu, Kai Li, Liquan Shen, and Ping An. 2017. Sparse Time-Varying Graphs for Slide Transition Detection in Lecture Videos. In *Image and Graphics - 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part I (Lecture Notes in Computer Science, Vol. 10666)*, Yao Zhao, Xiangwei Kong, and David Taubman (Eds.). Springer, 567–576. https://doi.org/10.1007/978-3-319-71607-7_50

[154] Christopher John R. Llanda. 2019. Video Tutoring System with Automatic Facial Expression Recognition: An Enhancing Approach to E-Learning Environment. In *Proceedings of the 2019 4th International Conference on Intelligent Information Technology* (Da, Nang, Viet Nam) *(ICIIT '19)*. Association for Computing Machinery, New York, NY, USA, 5–9. https://doi.org/10.1145/3321454.3321462

[155] Christopher John R. Llanda. 2019. Video Tutoring System with Automatic Facial Expression Recognition: An Enhancing Approach to E-Learning Environment. In *Proceedings of the 2019 4th International Conference on Intelligent Information Technology, ICIIT '19, Da Nang, Vietnam, February 20-23, 2019*. ACM, 5–9. https://doi.org/10.1145/3321454.3321462

[156] Ryan Long, Tayfun Tuna, and Jaspal Subhlok. 2018. Lecture Video Analytics as an Instructional Resource. In *IEEE Frontiers in Education Conference, FIE 2018, San Jose, CA, USA, October 3-6, 2018*. IEEE, 1–7. https://doi.org/10.1109/FIE.2018.8659007

[157] Xin Lu, Qing Li, and Xue Wang. 2021. Research on the Impacts of Feedback in Instructional Videos on College Students' Attention and Learning Effects. In *24th IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021, Dalian, China, May 5-7, 2021*, Weiming Shen, Jean-Paul A. Barthès, Junzhou Luo, Yanjun Shi, and Jinghui Zhang (Eds.). IEEE, 513–516. https://doi.org/10.1109/CSCWD49262.2021.9437774

[158] Di Ma, Xi Zhang, Xu Ouyang, and Gady Agam. 2017. Lecture video indexing using boosted margin maximizing neural networks. *CoRR* abs/1712.00575 (2017). arXiv:1712.00575 http://arxiv.org/abs/1712.00575

[159] Lin Ma and Yuchun Ma. 2019. Automatic Question Generation Based on MOOC Video Subtitles and Knowledge Graph. In *Proceedings of the 2019 7th International Conference on Information and Education Technology* (Aizu-Wakamatsu, Japan) *(ICIET 2019)*. Association for Computing Machinery, New York, NY, USA, 49–53. https://doi.org/10.1145/3323771.3323820

[160] Debabrata Mahapatra, Ragunathan Mariappan, and Vaibhav Rajan. 2018. Automatic Hierarchical Table of Contents Generation for Educational Videos. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 267–274. https://doi.org/10.1145/3184558.3186336

[161] Debabrata Mahapatra, Ragunathan Mariappan, Vaibhav Rajan, Kuldeep Yadav, Seby A., and Sudeshna Roy. 2018. VideoKen: Automatic Video Summarization and Course Curation to Support Learning. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 239–242. https://doi.org/10.1145/3184558.3186988

[162] Martin Malchow, Matthias Bauer, and Christoph Meinel. 2018. Enhance learning in a video lecture archive with annotations. In *2018 IEEE Global Engineering Education Conference, EDUCON 2018, Santa Cruz de Tenerife, Tenerife, Islas Canarias, Spain, April 17-20, 2018*. IEEE, 849–856. https://doi.org/10.1109/EDUCON.2018.8363319

[163] Richard E Mayer. 2014. Multimedia instruction. In *Handbook of research on educational communications and technology*. Springer, 385–399.

[164] Boniface Mbouzao, Michel C. Desmarais, and Ian Shrier. 2019. A Methodology for Student Video Interaction Patterns Analysis and Classification. In *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019*, Michel C. Desmarais, Collin F. Lynch, Agathe Merceron, and Roger Nkambou (Eds.). International Educational Data Mining Society (IEDMS). https://drive.google.com/file/d/130CZcEi3W9WvYHanQ1n6Va3hTpnlmDha

[165] Boniface Mbouzao, Michel C. Desmarais, and Ian Shrier. 2020. Early Prediction of Success in MOOC from Video Interaction Features. In *Artificial Intelligence in Education - 21st International Conference, AIED 2020, Ifrane, Morocco, July 6-10, 2020, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 12164)*, Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Millán (Eds.). Springer, 191–196. https://doi.org/10.1007/978-3-030-52240-7_35

[166] Martin Merkt, Anne Ballmann, Julia Felfeli, and Stephan Schwan. 2018. Pauses in educational videos: Testing the transience explanation against the structuring explanation. *Comput. Hum. Behav.* 89 (2018), 399–410. https://doi.org/10.1016/j.chb.2018.01.013

[167] Angel Meseguer-Martinez, Alejandro Ros-Galvez, and Alfonso Rosa-García. 2019. Linking YouTube and university rankings: Research performance as predictor of online video impact. *Telematics Informatics* 43 (2019). https://doi.org/10.1016/j.tele.2019.101264

[168] Qiusha Min, Yating Chen, and Dan Xia. 2019. A Video Learning Analytics System for Students' Learning Behavior Collection and Visualization. *J. Comput.* 14, 6 (2019), 381–388. https://doi.org/10.17706/jcp.14.6.381-388

[169] Mia Minnes, Christine Alvarado, Max Geislinger, and Joyce Fang. 2019. Podcast Highlights: Targeted Educational Videos From Repurposed Lecture-capture Footage. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE 2019, Minneapolis, MN, USA, February 27 - March 02, 2019*, Elizabeth K. Hawthorne, Manuel A. Pérez-Quiñones, Sarah Heckman, and Jian Zhang (Eds.). ACM, 365–371. https://doi.org/10.1145/3287324.3287465

[170] Jewoong Moon and Jeeheon Ryu. 2021. The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction. *J. Comput. High. Educ.* 33, 1 (2021), 39–63. https://doi.org/10.1007/s12528-020-09255-x

[171] Michael G Moore. 1989. Three types of interaction. (1989).

[172] Daniel C. Moos and Caitlin Bonde. 2016. Flipping the Classroom: Embedding Self-Regulated Learning Prompts in Videos. *Technol. Knowl. Learn.* 21, 2 (2016), 225–242. https://doi.org/10.1007/s10758-015-9269-1

[173] Ahmed Ali Mubarak, Han Cao, and Salah A. M. Ahmed. 2021. Predictive learning analytics using deep learning model in MOOCs' courses videos. *Educ. Inf. Technol.* 26, 1 (2021), 371–392. https://doi.org/10.1007/s10639-020-10273-6

[174] Arpan Mukherjee, Shubhi Tiwari, Tanya Chowdhury, and Tanmoy Chakraborty. 2019. Automatic Curation of Content Tables for Educational Videos. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer (Eds.). ACM, 1329–1332. https://doi.org/10.1145/3331184.3331400

[175] Juan Carlos Muñoz-Carpio, Michael A. Cowling, and James R. Birt. 2020. Doctoral Colloquium - Exploring the Benefits of Using 360^0 Video Immersion to Enhance Motivation and Engagement in System Modelling Education. In *6th International Conference of the Immersive Learning Research Network, iLRN 2020, San Luis Obispo, CA, USA, June 21-25, 2020*, Daphne Economou, Alexander Klippel, Heather Dodds, Anasol Peña-Ríos, Mark J. W. Lee, Dennis Beck, Johanna Pirker, Andreas Dengel, Tiago M. Peres, and Jonathon Richter (Eds.). IEEE, 403–406. https://doi.org/10.23919/iLRN47897.2020.9155100

[176] Sharmila Reddy Nangi, Yashasvi Kanchugantla, Pavan Gopal Rayapati, and Plaban Kumar Bhowmik. 2019. OffVid: A System for Linking Off-Topic
        Concepts to Topically Relevant Video Lecture Segments. In *19th IEEE International Conference on Advanced Learning Technologies, ICALT 2019,
        Maceió, Brazil, July 15-18, 2019*, Maiga Chang, Demetrios G. Sampson, Ronghuai Huang, Alex Sandro Gomes, Nian-Shing Chen, Ig Ibert Bittencourt,
        Kinshuk, Diego Dermeval, and Ibsen Mateus Bittencourt (Eds.). IEEE, 37–41. https://doi.org/10.1109/ICALT.2019.00015

[177] Evelyn Navarrete, Anett Hoppe, and Ralph Ewerth. 2021. A Review on Recent Advances in Video-based Learning Research: Video Features,
        Interaction, Tools, and Technologies. In *Proceedings of the CIKM 2021 Workshops co-located with 30th ACM International Conference on Information
        and Knowledge Management (CIKM 2021), Gold Coast, Queensland, Australia, November 1-5, 2021 (CEUR Workshop Proceedings, Vol. 3052)*, Gao Cong
        and Maya Ramanath (Eds.). CEUR-WS.org. http://ceur-ws.org/Vol-3052/paper7.pdf

[178] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human
        Factors in Computing Systems, San Jose, CA, USA, May 7-12, 2016*, Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade
        (Eds.). ACM, 2093–2097. https://doi.org/10.1145/2858036.2858137

[179] Chunyin Nong, Qiao Zhang, Liguo Huang, Di Cui, Qinghua Zheng, and Ting Liu. 2019. FVT: a fragmented video tutor for "dubbing" software
        development tutorials. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering Education and Training, ICSE
        (SEET) 2019, Montreal, QC, Canada, May 25-31, 2019*, Sarah Beecham and Daniela E. Damian (Eds.). IEEE / ACM, 95–99. https://doi.org/10.1109/ICSE-
        SEET.2019.00018

[180] Ari Nugraha, Izhar Almizan Wahono, Jianpeng Zhanghe, Tomoyuki Harada, and Tomoo Inoue. 2020. Creating Dialogue Between a Tutee Agent
        and a Tutor in a Lecture Video Improves Students' Attention. In *Collaboration Technologies and Social Computing - 26th International Conference,
        CollabTech 2020, Tartu, Estonia, September 8-11, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12324)*, Alexander Nolte, Claudio Alvarez,
        Reiko Hishiyama, Irene-Angelica Chounta, María Jesús Rodríguez-Triana, and Tomoo Inoue (Eds.). Springer, 96–111. https://doi.org/10.1007/978-
        3-030-58157-2_7

[181] Yutaro Ohashi, Masashi Katsumata, Kazuhiro Nakamura, Hiroaki Hashiura, Takafumi Matsuura, Jiro Ishihara, Hidemi Yamachi, Fumihiro Kumeno,
        and Yasuhiro Tsujimura. 2019. Comparison of Educational Video Production Methods for Students Studying Computer Programming. In *Proceedings
        of the 2019 7th International Conference on Information and Education Technology* (Aizu-Wakamatsu, Japan) *(ICIET 2019)*. Association for Computing
        Machinery, New York, NY, USA, 105–110. https://doi.org/10.1145/3323771.3323814

[182] Yoshimasa Ohnishi, Shinnosuke Yamaguchi, Yoshiki Shimoikura, Kazunori Nishino, Hideki Kondo, and Akihiro Hayashi. 2019. Prototype
        Design of Playback and Search System for Lecture Video Content using Google Cloud API. In *Knowledge-Based and Intelligent Information
        & Engineering Systems: Proceedings of the 23rd International Conference KES-2019, Budapest, Hungary, 4-6 September 2019 (Procedia Computer
        Science, Vol. 159)*, Imre J. Rudas, János Csirik, Carlos Toro, János Botzheim, Robert J. Howlett, and Lakhmi C. Jain (Eds.). Elsevier, 1517–1526.
        https://doi.org/10.1016/j.procs.2019.09.322

[183] El Harrak Othman, Slimani Abdelali, and Jaber El Bouhdidi. 2016. Education data mining: Mining MOOCs videos using metadata based approach.
        In *4th IEEE International Colloquium on Information Science and Technology, CiSt 2016, Tangier, Morocco, October 24-26, 2016*, Mohammed El Mohajir,
        Mohamed Chahhou, Mohammed Al Achhab, and Badr Eddine El Mohajir (Eds.). IEEE, 531–534. https://doi.org/10.1109/CIST.2016.7805106

[184] El Harrak Othman, Abderrahim Ghadi, and Jaber El Bouhdidi. 2017. Mining Moocs Videos Metadata Using Classification Techniques. In *Proceedings
        of the 2nd international Conference on Big Data, Cloud and Applications, BDCA 2017, Tetouan, Morocco, March 29-30, 2017*, Mohamed Lazaar, Youness
        Tabii, Mohamed Chrayah, and Mohammed Al Achhab (Eds.). ACM, 94:1–94:6. https://doi.org/10.1145/3090354.3090450

[185] Chaohua Ou, Ashok K. Goel, David A. Joyner, and Daniel F. Haynes. 2016. Designing Videos with Pedagogical Strategies: Online Students'
        Perceptions of Their Effectiveness. In *Proceedings of the Third ACM Conference on Learning @ Scale, L@S 2016, Edinburgh, Scotland, UK, April 25 - 26,
        2016*, Jeff Haywood, Vincent Aleven, Judy Kay, and Ido Roll (Eds.). ACM, 141–144. https://doi.org/10.1145/2876034.2893391

[186] Muzaffer Ozdemir, Serkan Izmirli, and Ozden Sahin-Izmirli. 2016. The Effects of Captioning Videos on Academic Achievement and Motivation:
        Reconsideration of Redundancy Principle in Instructional Videos. *J. Educ. Technol. Soc.* 19, 4 (2016), 1–10. https://www.j-ets.net/ETS/journals/19_
        4/1.pdf

[187] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M
        Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W
        Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny
        Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* (March 2021), n71.
        https://doi.org/10.1136/bmj.n71

[188] George Palaigeorgiou and Anthea Papadopoulou. 2019. Promoting self-paced learning in the elementary classroom with interactive video, an
        online course platform and tablets. *Educ. Inf. Technol.* 24, 1 (2019), 805–823. https://doi.org/10.1007/s10639-018-9804-5

[189] Devangini Patel, Debjyoti Ghosh, and Shengdong Zhao. 2018. Teach Me Fast: How to Optimize Online Lecture Video Speeding for Learning in Less
        Time?. In *Proceedings of the Sixth International Symposium of Chinese CHI, ChineseCHI 2018, Montreal, QC, Canada, April 21-22, 2018.* ACM, 160–163.
        https://doi.org/10.1145/3202667.3202696

[190] Antonio Pérez-Navarro, Victor Garcia, and Jordi Conesa. 2021. Students perception of videos in introductory physics courses of engineering in
        face-to-face and online environments. *Multim. Tools Appl.* 80, 1 (2021), 1009–1028. https://doi.org/10.1007/s11042-020-09665-0

[191] Zhongling Pi, Jianzhong Hong, and Jiumin Yang. 2017. Does instructor's image size in video lectures affect learning outcomes? *J. Comput. Assist.
        Learn.* 33, 4 (2017), 347–354. https://doi.org/10.1111/jcal.12183

[192] Zhongling Pi, Jianzhong Hong, and Jiumin Yang. 2017. Effects of the instructor's pointing gestures on learning performance in video lectures. *Br. J. Educ. Technol.* 48, 4 (2017), 1020–1029. https://doi.org/10.1111/bjet.12471

[193] Zhongling Pi, Yi Zhang, Fangfang Zhu, Ke Xu, Jiumin Yang, and Weiping Hu. 2019. Instructors' pointing gestures improve learning regardless of their use of directed gaze in video lectures. *Comput. Educ.* 128 (2019), 345–352. https://doi.org/10.1016/j.compedu.2018.10.006

[194] Elizabeth Poché, Nishant Jha, Grant Williams, Jazmine Staten, Miles Vesper, and Anas Mahmoud. 2017. Analyzing user comments on YouTube coding tutorial videos. In *Proceedings of the 25th International Conference on Program Comprehension, ICPC 2017, Buenos Aires, Argentina, May 22-23, 2017*, Giuseppe Scanniello, David Lo, and Alexander Serebrenik (Eds.). IEEE Computer Society, 196–206. https://doi.org/10.1109/ICPC.2017.26

[195] Luca Ponzanelli, Gabriele Bavota, Andrea Mocci, Rocco Oliveto, Massimiliano Di Penta, Sonia Haiduc, Barbara Russo, and Michele Lanza. 2019. Automatic Identification and Classification of Software Development Video Tutorial Fragments. *IEEE Trans. Software Eng.* 45, 5 (2019), 464–488. https://doi.org/10.1109/TSE.2017.2779479

[196] Oleksandra Poquet, Lisa Lim, Negin Mirriahi, and Shane Dawson. 2018. Video and learning: a systematic review (2007-2017). In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge, LAK 2018, Sydney, NSW, Australia, March 07-09, 2018*, Abelardo Pardo, Kathryn Bartimote-Aufflick, Grace Lynch, Simon Buckingham Shum, Rebecca Ferguson, Agathe Merceron, and Xavier Ochoa (Eds.). ACM, 151–160. https://doi.org/10.1145/3170358.3170376

[197] Muhamad Izzat Rahim and Sarimah Shamsudin. 2019. Video Lecture Styles in MOOCs by Malaysian Polytechnics. In *Proceedings of the 3rd International Conference on Education and Multimedia Technology* (Nagoya, Japan) *(ICEMT '19)*. Association for Computing Machinery, New York, NY, USA, 64–68. https://doi.org/10.1145/3345120.3345169

[198] Mohammad Rajiur Rahman, Shishir K. Shah, and Jaspal Subhlok. 2020. Visual Summarization of Lecture Video Segments for Enhanced Navigation. In *IEEE International Symposium on Multimedia, ISM 2020, Naples, Italy, December 2-4, 2020*. IEEE, 154–157. https://doi.org/10.1109/ISM.2020.00033

[199] Tarmo Robal, Yue Zhao, Christoph Lofi, and Claudia Hauff. 2018. IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *Proceedings of the 29th on Hypertext and Social Media, HT 2018, Baltimore, MD, USA, July 09-12, 2018*, Dongwon Lee, Nishanth Sastry, and Ingmar Weber (Eds.). ACM, 106–114. https://doi.org/10.1145/3209542.3209547

[200] Manuel Caeiro Rodríguez, Martín Llamas Nistal, Fernando A. Mikic-Fonte, Manuel Lama Penín, and Manuel Mucientes Molina. 2018. Exploring the application of process mining to support self-regulated learning: An initial analysis with video lectures. In *2018 IEEE Global Engineering Education Conference, EDUCON 2018, Santa Cruz de Tenerife, Tenerife, Islas Canarias, Spain, April 17-20, 2018*. IEEE, 1766–1774. https://doi.org/10.1109/EDUCON.2018.8363448

[201] Marija Sablić, Ana Mirosavljević, and Alma Škugor. 2021. Video-Based Learning (VBL)—Past, Present and Future: an Overview of the Research Published from 2008 to 2019. *Technology, Knowledge and Learning* 26 (12 2021), 1–17. https://doi.org/10.1007/s10758-020-09455-5

[202] Isaias Sanchez-Cortina, Jesús Andrés-Ferrer, Alberto Sanchís, and Alfons Juan. 2016. Speaker-adapted confidence measures for speech recognition of video lectures. *Comput. Speech Lang.* 37 (2016), 11–23. https://doi.org/10.1016/j.csl.2015.10.003

[203] Samant Saurabh and Sanjana Gautam. 2019. Modelling and statistical analysis of YouTube's educational videos: A channel Owner's perspective. *Comput. Educ.* 128 (2019), 145–158. https://doi.org/10.1016/j.compedu.2018.09.003

[204] Benjamin J. Schreiber and John P. Dougherty. 2017. Assessment of Introducing Algorithms with Video Lectures and Pseudocode Rhymed to a Melody. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, SIGCSE 2017, Seattle, WA, USA, March 8-11, 2017*, Michael E. Caspersen, Stephen H. Edwards, Tiffany Barnes, and Daniel D. Garcia (Eds.). ACM, 519–524. https://doi.org/10.1145/3017680.3017789

[205] Noah L. Schroeder, Joshua Chin, and Scotty D. Craig. 2020. Learner Control Aids Learning from Instructional Videos with a Virtual Human. *Technol. Knowl. Learn.* 25, 4 (2020), 733–751. https://doi.org/10.1007/s10758-019-09417-6

[206] Cleo Schulten, Sven Manske, Angela Langner-Thiele, and Heinz Ulrich Hoppe. 2020. Digital Value-Adding Chains in Vocational Education: Automatic Keyword Extraction from Learning Videos to Provide Learning Resource Recommendations. In *Addressing Global Challenges and Quality Education - 15th European Conference on Technology Enhanced Learning, EC-TEL 2020, Heidelberg, Germany, September 14-18, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12315)*, Carlos Alario-Hoyos, María Jesús Rodríguez-Triana, Maren Scheffel, Inmaculada Arnedillo Sánchez, and Sebastian Dennerlein (Eds.). Springer, 15–29. https://doi.org/10.1007/978-3-030-57717-9_2

[207] Arjun Sharma, Arijit Biswas, Ankit Gandhi, Sonal Patil, and Om Deshmukh. 2016. LIVELINET: A Multimodal Deep Recurrent Neural Network to Predict Liveliness in Educational Videos. In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, Tiffany Barnes, Min Chi, and Mingyu Feng (Eds.). International Educational Data Mining Society (IEDMS), 215–222. http://www.educationaldatamining.org/EDM2016/proceedings/paper_64.pdf

[208] Kshitij Sharma, Hamed S. Alavi, Patrick Jermann, and Pierre Dillenbourg. 2016. A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, LAK 2016, Edinburgh, United Kingdom, April 25-29, 2016*, Dragan Gasevic, Grace Lynch, Shane Dawson, Hendrik Drachsler, and Carolyn Penstein Rosé (Eds.). ACM, 417–421. https://doi.org/10.1145/2883851.2883902

[209] Kshitij Sharma, Sarah D'Angelo, Darren Gergle, and Pierre Dillenbourg. 2016. Visual Augmentation of Deictic Gestures in MOOC Videos. In *Transforming Learning, Empowering Learners: Proceedings of the 12th International Conference of the Learning Sciences, ICLS 2016, Singapore, June 20-24, 2016*, Chee-Kit Looi, Joseph L. Polman, Ulrike Cress, and Peter Reimann (Eds.). International Society of the Learning Sciences. https://repository.isls.org/handle/1/117

[210] Catharyn Shelton, Annie Hale, and Leanna Archambault. 2016. Exploring the Use of Interactive Digital Storytelling Video: Promoting Student Engagement and Learning in a University Hybrid Course. *TechTrends* 60, 5 (06 2016), 465–474. https://doi.org/10.1007/s11528-016-0082-z

[211] Jianwei Shi, Christian Otto, Anett Hoppe, Peter Holtz, and Ralph Ewerth. 2019. Investigating Correlations of Automatically Extracted Multimodal Features and Lecture Video Quality. In *Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information* (Nice, France) *(SALMM '19)*. Association for Computing Machinery, New York, NY, USA, 11–19. https://doi.org/10.1145/3347451.3356731

[212] Abdulhadi Shoufan. 2019. Estimating the cognitive value of YouTube's educational videos: A learning analytics approach. *Comput. Hum. Behav.* 92 (2019), 450–458. https://doi.org/10.1016/j.chb.2018.03.036

[213] Abdulhadi Shoufan. 2019. What motivates university students to like or dislike an educational online video? A sentimental framework. *Comput. Educ.* 134 (2019), 132–144. https://doi.org/10.1016/j.compedu.2019.02.008

[214] Abdulhadi Shoufan and Fatma Mohamed. 2017. On the Likes and Dislikes of YouTube's Educational Videos: A Quantitative Study. In *Proceedings of the 18th Annual Conference on Information Technology Education and the 6th Annual Conference on Research in Information Technology, SIGITE/RIIT 2017, Rochester, New York, USA, October 4-7, 2017*, Stephen J. Zilora, Tom Ayers, and Daniel S. Bogaard (Eds.). ACM, 127–132. https://doi.org/10.1145/3125659.3125692

[215] Abdulhadi Shoufan and Fatma Mohamed. 2022. YouTube and Education: A Scoping Review. *IEEE Access* 10 (2022), 125576–125599. https://doi.org/10.1109/ACCESS.2022.3225419

[216] Hugo Silva and Isabel Azevedo. 2017. Instructional Videos and Others on Youtube - Similarities and Differences in Comments. In *Proceedings of the 9th International Conference on Computer Supported Education - Volume 1: CSEDU,*, Vol. 2. INSTICC, SciTePress, 418–425. https://doi.org/10.5220/0006333104180425

[217] Vikash Singh, Sarah Abdellahi, Mary Lou Maher, and Celine Latulipe. 2016. The Video Collaboratory as a Learning Environment. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education, SIGCSE 2016, Memphis, TN, USA, March 02 - 05, 2016*, Carl Alphonce, Jodi L. Tims, Michael E. Caspersen, and Stephen H. Edwards (Eds.). ACM, 352–357. https://doi.org/10.1145/2839509.2844588

[218] Björn Sjödén, Vania Dimitrova, and Antonija Mitrovic. 2018. Using Thematic Analysis to Understand Students' Learning of Soft Skills from Videos. In *Lifelong Technology-Enhanced Learning - 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018, Proceedings (Lecture Notes in Computer Science, Vol. 11082)*, Viktoria Pammer-Schindler, Mar Pérez-Sanagustín, Hendrik Drachsler, Raymond Elferink, and Maren Scheffel (Eds.). Springer, 656–659. https://doi.org/10.1007/978-3-319-98572-5_67

[219] Aaron Smith, Skye Toor, and Patrick Van Kessel. 2018. Many turn to YouTube for children's content, news, how-to lessons. *Pew Research Centre* 7 (2018). https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/

[220] Eduardo Rocha Soares and Eduardo Barrére. 2018. Automatic Topic Segmentation for Video Lectures Using Low and High-Level Audio Features. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web, WebMedia 2018, Salvador-BA, Brazil, October 16-19, 2018*, Manoel Carvalho Marques Neto, Renato Lima Novais, Carlos Ferraz, and Windson Viana (Eds.). ACM, 189–196. https://doi.org/10.1145/3243082.3243096

[221] Eduardo Rocha Soares and Eduardo Barrére. 2019. An optimization model for temporal video lecture segmentation using word2vec and acoustic features. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia 2019, Rio de Janeiro, Brazil, October 29 - November 01, 2019*, Joel A. F. dos Santos and Débora Christina Muchaluat-Saade (Eds.). ACM, 513–520. https://doi.org/10.1145/3323503.3349548

[222] Mahmut Can Sözeri and Serhat Bahadir Kert. 2021. Ineffectiveness Of Online Interactive Video Content Developed For Programming Education. *Int. J. Comput. Sci. Educ. Sch.* 4, 3 (2021), 49–69. https://doi.org/10.21585/ijcses.v4i3.99

[223] Namrata Srivastava, Sadia Nawaz, Jason M. Lodge, Eduardo Velloso, Sarah M. Erfani, and James Bailey. 2020. Exploring the usage of thermal imaging for understanding video lecture designs and students' experiences. In *LAK '20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23-27, 2020*, Christoph Rensing and Hendrik Drachsler (Eds.). ACM, 250–259. https://doi.org/10.1145/3375462.3375514

[224] Namrata Srivastava, Eduardo Velloso, Jason M. Lodge, Sarah M. Erfani, and James Bailey. 2019. Continuous Evaluation of Video Lectures from Real-Time Difficulty Self-Report. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 586. https://doi.org/10.1145/3290605.3300816

[225] Nicholas R. Stepanek and Brian Dorn. 2017. Automated Analysis of Lecture Video Engagement Using Student Posts. In *Artificial Intelligence in Education - 18th International Conference, AIED 2017, Wuhan, China, June 28 - July 1, 2017, Proceedings (Lecture Notes in Computer Science, Vol. 10331)*, Elisabeth André, Ryan S. Baker, Xiangen Hu, Ma. Mercedes T. Rodrigo, and Benedict du Boulay (Eds.). Springer, 565–569. https://doi.org/10.1007/978-3-319-61425-0_65

[226] Ben Stephenson. 2019. Coding Demonstration Videos for CS1. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE 2019, Minneapolis, MN, USA, February 27 - March 02, 2019*, Elizabeth K. Hawthorne, Manuel A. Pérez-Quiñones, Sarah Heckman, and Jian Zhang (Eds.). ACM, 105–111. https://doi.org/10.1145/3287324.3287445

[227] Christian Stöhr, Natalia Stathakarou, Franziska Mueller, Sokratis Nifakos, and Cormac McGrath. 2019. Videos as learning objects in MOOCs: A study of specialist and non-specialist participants' video activity in MOOCs. *Br. J. Educ. Technol.* 50, 1 (2019), 166–176. https://doi.org/10.1111/bjet.12623

[228] Alexandru Stefan Stoica, Stella Heras, Javier Palanca, Vicente Julián, and Marian Cristian Mihaescu. 2019. A Semi-supervised Method to Classify Educational Videos. In *Hybrid Artificial Intelligent Systems - 14th International Conference, HAIS 2019, León, Spain, September 4-6, 2019, Proceedings (Lecture Notes in Computer Science, Vol. 11734)*, Hilde Pérez García, Lidia Sánchez-González, Manuel Castejón Limas, Héctor Quintián-Pardo, and Emilio S. Corchado Rodríguez (Eds.). Springer, 218–228. https://doi.org/10.1007/978-3-030-29859-3_19

[229] Samantha Stranc and Kasia Muldner. 2019. Learning from Videos Showing a Dialog Fosters More Positive Affect Than Learning from a Monolog. In *Artificial Intelligence in Education - 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 11626)*, Seiji Isotani, Eva Millán, Amy Ogan, Peter M. Hastings, Bruce M. McLaren, and Rose Luckin (Eds.). Springer, 275–280. https://doi.org/10.1007/978-3-030-23207-8_51

[230] Andrew T. Stull, Logan Fiorella, and Richard E. Mayer. 2018. An eye-tracking analysis of instructor presence in video lectures. *Comput. Hum. Behav.* 88 (2018), 263–272. https://doi.org/10.1016/j.chb.2018.07.019

[231] Ching-Ying Sung, Xun-Yi Huang, Yicong Shen, Fu-Yin Cherng, Wen-Chieh Lin, and Hao-Chuan Wang. 2016. ToPIN: A Visual Analysis Tool for Time-anchored Comments in Online Educational Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, May 7-12, 2016, Extended Abstracts*, Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade (Eds.). ACM, 2185–2191. https://doi.org/10.1145/2851581.2892327

[232] John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction* 4, 4 (1994), 295–312.

[233] Chien-Lin Tang, Jingxian Liao, Hao-Chuan Wang, Ching-Ying Sung, and Wen-Chieh Lin. 2021. ConceptGuide: Supporting Online Video Learning with Concept Map-based Recommendation of Learning Path. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 2757–2768. https://doi.org/10.1145/3442381.3449808

[234] Yassin Taskin, Tobias Hecking, Heinz Ulrich Hoppe, Vania Dimitrova, and Antonija Mitrovic. 2019. Characterizing Comment Types and Levels of Engagement in Video-Based Learning as a Basis for Adaptive Nudging. In *Transforming Learning with Meaningful Technologies - 14th European Conference on Technology Enhanced Learning, EC-TEL 2019, Delft, The Netherlands, September 16-19, 2019, Proceedings (Lecture Notes in Computer Science, Vol. 11722)*, Maren Scheffel, Julien Broisin, Viktoria Pammer-Schindler, Andri Ioannou, and Jan Schneider (Eds.). Springer, 362–376. https://doi.org/10.1007/978-3-030-29736-7_27

[235] MohammadReza Tavakoli, Sherzod Hakimov, Ralph Ewerth, and Gábor Kismihók. 2020. A Recommender System For Open Educational Videos Based On Skill Requirements. In *20th IEEE International Conference on Advanced Learning Technologies, ICALT 2020, Tartu, Estonia, July 6-9, 2020*. IEEE, 1–5. https://doi.org/10.1109/ICALT49669.2020.00008

[236] Sera Thornton, Ceri Riley, and Mary Ellen Wiltrout. 2017. Criteria for Video Engagement in a Biology MOOC. In *Proceedings of the Fourth ACM Conference on Learning @ Scale, L@S 2017, Cambridge, MA, USA, April 20-21, 2017*, Claudia Urrea, Justin Reich, and Candace Thille (Eds.). ACM, 291–294. https://doi.org/10.1145/3051457.3054007

[237] Yi Tian and Marie-Luce Bourguet. 2016. Lecturers' Hand Gestures as Clues to Detect Pedagogical Significance in Video Lectures. In *Proceedings of the European Conference on Cognitive Ergonomics, ECCE 2016, Nottingham, United Kingdom, September 5-8, 2016*, Glyn Lawson, Tess Roper, and Geert de Haan (Eds.). ACM, 2:1–2:3. https://doi.org/10.1145/2970930.2970933

[238] Quan H. To, Ba-Huu Tran, and Minh-Triet Tran. 2016. How to Teach Young Kids New Concepts with Interactive Videos and Visual Recognition. In *HCI International 2016 - Posters' Extended Abstracts - 18th International Conference, HCI International 2016, Toronto, Canada, July 17-22, 2016, Proceedings, Part II (Communications in Computer and Information Science, Vol. 618)*, Constantine Stephanidis (Ed.). Springer, 282–286. https://doi.org/10.1007/978-3-319-40542-1_46

[239] Judith Uchidiuno, Ken Koedinger, Jessica Hammer, Evelyn Yarzebinski, and Amy Ogan. 2018. How Do English Language Learners Interact with Different Content Types in MOOC Videos? *Int. J. Artif. Intell. Educ.* 28, 4 (2018), 508–527. https://doi.org/10.1007/s40593-017-0156-x

[240] David C. D. van Alten, Chris Phielix, Jeroen Janssen, and Liesbeth Kester. 2020. Self-regulated learning support in flipped learning videos enhances learning outcomes. *Comput. Educ.* 158 (2020), 104000. https://doi.org/10.1016/j.compedu.2020.104000

[241] Hans van der Meij. 2019. Advance organizers in videos for software training of Chinese students. *Br. J. Educ. Technol.* 50, 3 (2019), 1368–1380. https://doi.org/10.1111/bjet.12619

[242] Hans van der Meij and Jan van der Meij. 2016. The effects of reviews in video tutorials. *J. Comput. Assist. Learn.* 32, 4 (2016), 332–344. https://doi.org/10.1111/jcal.12136

[243] Frans van der Sluis, Jasper Ginn, and Tim Van der Zee. 2016. Explaining Student Behavior at Scale: The Influence of Video Complexity on Student Dwelling Time. In *Proceedings of the Third ACM Conference on Learning @ Scale, L@S 2016, Edinburgh, Scotland, UK, April 25 - 26, 2016*, Jeff Haywood, Vincent Aleven, Judy Kay, and Ido Roll (Eds.). ACM, 51–60. https://doi.org/10.1145/2876034.2876051

[244] veena. 2020. STANFORD COURSES ON THE LAGUNITA LEARNING PLATFORM. https://online.stanford.edu/lagunita-learning-platform

[245] Anusha Vimalaksha, Siddarth Vinay, and N. S. Kumar. 2019. Hierarchical Mind Map Generation from Video Lectures. In *Tenth IEEE International Conference on Technology for Education, T4E 2019, Goa, India, December 9-11, 2019*, Maiga Chang, Ramkumar Rajendran, Kinshuk, Sahana Murthy, and Venkatesh Kamat (Eds.). IEEE, 110–113. https://doi.org/10.1109/T4E.2019.00-40

[246] Oleksii Voronkin. 2019. Educational Video in the University: Instruments, Technologies, Opportunities and Restrictions. In *Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference, Kherson, Ukraine, June 12-15, 2019 (CEUR Workshop Proceedings, Vol. 2387)*, Vadim Ermolayev, Frédéric Mallet, Vitaliy Yakovyna, Heinrich C. Mayr, and Aleksander Spivakovsky (Eds.). CEUR-WS.org, 302–317. http://ceur-ws.org/Vol-2387/20190302.pdf

[247] Josef Wachtler and Martin Ebner. 2019. Scheduling Interactions in Learning Videos: A State Machine Based Algorithm. *International Journal of Learning Analytics and Artificial Intelligence for Education* 1, 1 (29 July 2019), 58–76. https://doi.org/10.3991/ijai.v1i1.10995

[248] Josef Wachtler, Michael Hubmann, Helmut Zöhrer, and Martin Ebner. 2016. An analysis of the use and effect of questions in interactive learning-videos. *Smart Learn. Environ.* 3, 1 (2016), 13. https://doi.org/10.1186/s40561-016-0033-3

[249] Josef Wachtler, Mohammad Khalil, Behnam Taraghi, and Martin Ebner. 2016. On Using Learning Analytics to Track the Activity of Interactive MOOC Videos. In *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning, Edinburgh, Scotland, April 26, 2016 (CEUR Workshop Proceedings, Vol. 1579)*, Michail N. Giannakos, Demetrios G. Sampson, Lukasz Kidzinski, and Abelardo Pardo (Eds.). CEUR-WS.org, 8–17. http://ceur-ws.org/Vol-1579/paper3.pdf

[250] Ian Walk, Arnold Yim, Ed Novak, Charles Reiss, and Daniel Graham. 2020. Redesigning the Online Video Lecture Player to Promote Active Learning. In *IEEE Frontiers in Education Conference, FIE 2020, Uppsala, Sweden, October 21-24, 2020*. IEEE, 1–5. https://doi.org/10.1109/FIE44824.2020.9273898

[251] Hongyan Wang, Zhongling Pi, and Weiping Hu. 2019. The instructor's gaze guidance in video lectures improves learning. *J. Comput. Assist. Learn.* 35, 1 (2019), 42–50. https://doi.org/10.1111/jcal.12309

[252] Xin Wang, Wei Huang, Qi Liu, Yu Yin, Zhenya Huang, Le Wu, Jianhui Ma, and Xue Wang. 2020. Fine-Grained Similarity Measurement between Educational Videos and Exercises. In *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*, Chang Wen Chen, Rita Cucchiara, Xian-Sheng Hua, Guo-Jun Qi, Elisa Ricci, Zhengyou Zhang, and Roger Zimmermann (Eds.). ACM, 331–339. https://doi.org/10.1145/3394171.3413783

[253] Xue Wang, Lin Lin, Meiqi Han, and J. Michael Spector. 2020. Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos. *Comput. Hum. Behav.* 107 (2020), 106279. https://doi.org/10.1016/j.chb.2020.106279

[254] Yang Wang, Qingtang Liu, Wenli Chen, Qiyun Wang, and David Stein. 2019. Effects of instructor's facial expressions on students' learning with video lectures. *Br. J. Educ. Technol.* 50, 3 (2019), 1381–1395. https://doi.org/10.1111/bjet.12633

[255] B. Sanjay Waykar and C. R. Bharathi. 2017. Multimodal Features and Probability Extended Nearest Neighbor Classification for Content-Based Lecture Video Retrieval. *J. Intell. Syst.* 26, 3 (2017), 585. https://doi.org/10.1515/jisys-2016-0041

[256] Tim Weinert, Marian Thiel de Gafenco, Matthias Simon Billert, and Niklas Boerner. 2020. Fostering Interaction in Higher Education with Deliberate Design of Interactive Learning Videos. In *Proceedings of the 41st International Conference on Information Systems, ICIS 2020, Making Digital Inclusive: Blending the Locak and the Global, Hyderabad, India, December 13-16, 2020*, Joey F. George, Souren Paul, Rahul De', Elena Karahanna, Suprateek Sarker, and Gal Oestreicher-Singer (Eds.). Association for Information Systems. https://aisel.aisnet.org/icis2020/digital_learning_env/digital_learning_env/12

[257] Michael Whitney and Bryan Dallas. 2019. Captioning Online Course Videos: An Investigation into Knowledge Retention and Student Perception. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE 2019, Minneapolis, MN, USA, February 27 - March 02, 2019*, Elizabeth K. Hawthorne, Manuel A. Pérez-Quiñones, Sarah Heckman, and Jian Zhang (Eds.). ACM, 511–517. https://doi.org/10.1145/3287324.3287347

[258] Winnifred Wijnker, Arthur Bakker, Tamara van Gog, and Paul Drijvers. 2019. Educational videos from a film theory perspective: Relating teacher aims to video characteristics. *Br. J. Educ. Technol.* 50, 6 (2019), 3175–3197. https://doi.org/10.1111/bjet.12725

[259] Rainer Winkler, Sebastian Hobert, Tizian Fischer, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Engaging Learners in Online Video Lectures with Dynamically Scaffolding Conversational Agents. In *28th European Conference on Information Systems - Liberty, Equality, and Fraternity in a Digitizing World, ECIS 2020, Marrakech, Morocco, June 15-17, 2020*, Frantz Rowe, Redouane El Amrani, Moez Limayem, Sue Newell, Nancy Pouloudi, Eric van Heck, and Ali El Quammah (Eds.). https://aisel.aisnet.org/ecis2020_rp/97

[260] Julia Wynn, Wei Wei, Xinhang Li, Yat So, Suzanne R. Bakken, Chunhua Weng, and Wendy K. Chung. 2019. User engagement with web-based genomics education videos and implications for designing scalable patient education materials. In *AMIA 2019, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 16-20, 2019*. AMIA. https://knowledge.amia.org/69862-amia-1.4570936/t004-1.4574923/t004-1.4574924/3201690-1.4574943/3190483-1.4574940

[261] Chengpei Xu, Ruomei Wang, Shujin Lin, Xiaonan Luo, Baoquan Zhao, Lijie Shao, and Mengqiu Hu. 2019. Lecture2Note: Automatic Generation of Lecture Notes from Slide-Based Educational Videos. In *IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, July 8-12, 2019*. IEEE, 898–903. https://doi.org/10.1109/ICME.2019.00159

[262] Fei Xu, Kenny Davila, Srirangaraj Setlur, and Venu Govindaraju. 2019. Content Extraction from Lecture Video via Speaker Action Classification Based on Pose Information. In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*. IEEE, 1047–1054. https://doi.org/10.1109/ICDAR.2019.00171

[263] Fei Xu, Kenny Davila, Srirangaraj Setlur, and Venu Govindaraju. 2020. Skeleton-Based Methods for Speaker Action Classification on Lecture Videos. In *Pattern Recognition. ICPR International Workshops and Challenges - Virtual Event, January 10-15, 2021, Proceedings, Part IV (Lecture Notes in Computer Science, Vol. 12664)*, Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani (Eds.). Springer, 250–264. https://doi.org/10.1007/978-3-030-68799-1_18

[264] Kuldeep Yadav, Ankit Gandhi, Arijit Biswas, Kundan Shrivastava, Saurabh Srivastava, and Om Deshmukh. 2016. ViZig: Anchor Points based Non-Linear Navigation and Summarization in Educational Videos. In *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI 2016, Sonoma, CA, USA, March 7-10, 2016*, Jeffrey Nichols, Jalal Mahmud, John O'Donovan, Cristina Conati, and Massimo Zancanaro (Eds.). ACM, 407–418. https://doi.org/10.1145/2856767.2856788

[265] Shir Yadid and Eran Yahav. 2016. Extracting code from programming tutorial videos. In *2016 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, Onward! 2016, Amsterdam, The Netherlands, November 2-4, 2016*, Eelco Visser, Emerson R. Murphy-Hill, and Crista Lopes (Eds.). ACM, 98–111. https://doi.org/10.1145/2986012.2986021

[266] Shulong Yan and Emily Baxter. 2018. Learning Efficiency of Video-based Learning. *eLearn Mag.* 2018, 12 (2018), 12. https://doi.org/10.1145/3302261.3236701

[267] Taewon Yoo, Hyewon Jeong, Donghwan Lee, and Hyunggu Jung. 2021. LectYS: A System for Summarizing Lecture Videos on YouTube. In *IUI '21: 26th International Conference on Intelligent User Interfaces, College Station, TX, USA, April 13-17, 2021, Companion*, Tracy Hammond, Katrien Verbert, and Dennis Parra (Eds.). ACM, 90–92. https://doi.org/10.1145/3397482.3450722

[268] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, and Ulrik Schroeder. 2014. The state of video-based learning: A review and future perspectives. *Int. J. Adv. Life Sci* 6, 3/4 (2014), 122–135.

[269] Han Zhang. 2017. Smart Jump: Automated Navigation Suggestion for Videos in MOOCs. In *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*, Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich (Eds.). ACM, 1183–1184. https://doi.org/10.1145/3041021.3055364

[270] Jie Zhang, Yongsheng Zhang, Xiaolong Wu, and Guoyun Li. 2018. Teaching Video Recommendation Based on Student Evaluation. In *Cloud Computing and Security - 4th International Conference, ICCCS 2018, Haikou, China, June 8-10, 2018, Revised Selected Papers, Part II (Lecture Notes in Computer Science, Vol. 11064)*, Xingming Sun, Zhaoqing Pan, and Elisa Bertino (Eds.). Springer, 182–190. https://doi.org/10.1007/978-3-030-00009-7_17

[271] Xinyong Zhang. 2021. Using Eye Tracking to Analyze the Effects of Spatial Contiguity in MOOC Video Subtitles. In *Engineering Psychology and Cognitive Ergonomics - 18th International Conference, EPCE 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24-29, 2021, Proceedings (Lecture Notes in Computer Science, Vol. 12767)*, Don Harris and Wen-Chin Li (Eds.). Springer, 336–354. https://doi.org/10.1007/978-3-030-77932-0_27

[272] Xue-Bai Zhang, Shyan-Ming Yuan, Ming-Dao Chen, and Xiaolong Liu. 2018. A Complete System for Analysis of Video Lecture Based on Eye Tracking. *IEEE Access* 6 (2018), 49056–49066. https://doi.org/10.1109/ACCESS.2018.2865754

[273] Baoquan Zhao, Shujin Lin, Xiaonan Luo, Songhua Xu, and Ruomei Wang. 2017. A Novel System for Visual Navigation of Educational Videos Using Multimodal Cues. In *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017*, Qiong Liu, Rainer Lienhart, Haohong Wang, Sheng-Wei "Kuan-Ta" Chen, Susanne Boll, Yi-Ping Phoebe Chen, Gerald Friedland, Jia Li, and Shuicheng Yan (Eds.). ACM, 1680–1688. https://doi.org/10.1145/3123266.3123406

[274] Baoquan Zhao, Shujin Lin, Xin Qi, Ruomei Wang, and Xiaonan Luo. 2018. A novel approach to automatic detection of presentation slides in educational videos. *Neural Comput. Appl.* 29, 5 (2018), 1369–1382. https://doi.org/10.1007/s00521-017-3276-1

[275] Jian Zhao, Chidansh Amitkumar Bhatt, Matthew Cooper, and David A. Shamma. 2018. Flexible Learning with Semantic Visual Exploration and Sequence-Based Recommendation of MOOC Videos. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 329. https://doi.org/10.1145/3173574.3173903

[276] Huasong Zhong and Yuchun Ma. 2018. Fast and Robust Text Detection in MOOCs Videos. In *Proceedings of the 2018 International Conference on Distance Education and Learning* (Beijing, China) *(ICDEL '18)*. Association for Computing Machinery, New York, NY, USA, 98–102. https://doi.org/10.1145/3231848.3231856

[277] Yun Zhu, Leisi Pei, and Junjie J. Shang. 2017. Improving Video Engagement by Gamification: A Proposed Design of MOOC Videos. In *Blended Learning. New Challenges and Innovative Practices - 10th International Conference, ICBL 2017, Hong Kong, China, June 27-29, 2017, Proceedings (Lecture Notes in Computer Science, Vol. 10309)*, Simon K. S. Cheung, Lam-for Kwok, Will W. K. Ma, Lap-Kei Lee, and Harrison Hao Yang (Eds.). Springer, 433–444. https://doi.org/10.1007/978-3-319-59360-9_38