A Review on Recent Advances in Video-based Learning Research: Video Features, Interaction, Tools, and Technologies

Evelyn Navarrete[1], Anett Hoppe[1,2] and Ralph Ewerth[1,2]

[1]L3S Research Center, Leibniz University Hannover, Germany
[2]TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

Abstract

Human learning shifts stronger than ever towards online settings, and especially towards video platforms. There is an abundance of tutorials and lectures covering diverse topics, from fixing a bike to particle physics. While it is advantageous that learning resources are freely available on the Web, the quality of the resources varies a lot. Given the number of available videos, users need algorithmic support in finding helpful and entertaining learning resources.

In this paper, we present a review of the recent research literature (2020-2021) on video-based learning. We focus on publications that examine the characteristics of video content, analyze frequently used features and technologies, and, finally, derive conclusions on trends and possible future research directions.

Keywords

Video-based Learning, web-based learning, literature review, video features

1. Introduction

The Web has fundamentally changed the way we learn. Especially video platforms, such as YouTube, play an increasing role here – from the about 30 million video views a day, about 50% relate to some kind of learning content [1]. Another YouTube-related statistic states that as of May 2019, about 500 hours of new content were uploaded to the platform every minute [2], implying that users are reliant on effective search and recommendation algorithms to find content that is relevant to their learning needs.

These algorithms need to consider multiple factors to provide suitable rankings and recommendations. Especially in learning contexts, an open question is: When is a video the best way to learn, depending on the individual user, the learning objective, and context factors? This question has been investigated in several studies in recent years, examining characteristics of the videos and the surrounding platforms to determine when video-based learning (VBL) processes are especially successful. The resulting insights are of interest for all the involved stakeholders: (a) Platform providers who wish to provide their users with the best possible content from their database, but also (b) content producers who aim to develop efficient and entertaining learning resources, (c) teachers who seek to enhance their teaching by integrating adapted multimedia resources, and finally, (d) students who pursue a more effective learning process and experience.

In this paper, we provide a review of recent research on the analysis of video-based learning, with a focus on studies that examine video characteristics. For this purpose, we performed a systematic literature search. Here, we present a preliminary summary of our findings from the most recent publications, covering the years 2020 and 2021. We systematize the contents of 41 reviewed papers and provide an overview on (a) the chosen research approach (e.g., empirical study, tool prototype, production guidelines), (b) considered video characteristics, and (c) tasks and technologies used to develop VBL tools and frameworks. From these dimensions, we derive recent research trends and identify gaps that provide directions for future research.

The rest of this paper is structured as follows: Section 2 presents the related work; Section 3 describes our methodology; Section 4 provides an overview of the video-based learning field, its benefits, potentials, and the current challenges; and explains why identifying relevant features in videos is important. Section 5 presents the results of the review. Finally, we summarize our findings, point out the limitations, and conclude with indications for future research in Section 6.

2. Related Work

We found four survey articles that examine video-based learning. Their core characteristics are summed up in

Proceedings of the CIKM 2021 Workshops, November 1–5, Gold Coast, Queensland, Australia
navarrete@l3s.de (E. Navarrete); anett.hoppe@tib.eu (A. Hoppe); ralph.ewerth@tib.eu (R. Ewerth)
0000-0002-5610-9908 (E. Navarrete); 0000-0002-1452-9509 (A. Hoppe); 0000-0003-0918-6297 (R. Ewerth)

Table 1
Core information of related survey articles Reference & publication year, temporal scope of the survey, number of reviewed papers and reviewed characteristics

Ref.	Review period	N° papers	Review dimensions	
[3] (2013)	2000-2012	166	1) Type of research 2) Sample 3) Subject area	4) Technology type 5) Style of use
[4] (2014)	2003-2014	76	1) Learning effectiveness 2) Teaching methods	3) Design 4) Reflection
[5] (2018)	2007-2017	178	1) Video type 2) Population 3) Control of prior knowledge	3) Sample size 5) Video features
[6] (2020)	2008-2019	39	Teacher perception and reflection	

Table 1. All of them reference the significant development in research on VBL [4] or the increase in publications [5, 3] as a motivation for their work. Giannakos [3] and Poquet et al. [5] find that there is an increase specifically in empirical studies.

All four surveys focus on different key characteristics (see Table 1, column "Dimensions"); and derive common themes and research trends based on these. For instance, Giannakos [3] centers on the development of the field and shows a shift in the used technologies. Yousef et al. [4] concentrate on the effectiveness of VBL and provide a review of teaching methods using video, design features, and how video-based contents and tools are integrated. Poquet et al. [5], spotlight the video characteristics that have been analyzed in research with specific regard to their influence on learning effectiveness. As stated by the authors, the most often used metrics to qualify the effectiveness are recall (remembering information) and transfer (applying what was learned to different scenarios), followed by motivation, cognitive load, mental effort, attention, and affect. The most analyzed features are text, audio, animations, and the video production style. Table 2 summarizes the features derived in their review. It is also found that when data sources are available, the most common sources are eye-tracking and click-stream data. A very recent work by Sablic et al. [6] reviews approaches to support teachers' self-reflection and professional development with video recordings of teaching sessions.

Giannakos [3] discovers a shift in the focus of research from social science disciplines to more technological domains. In accordance to these findings, Poquet et al. [5] state that more often than other subjects, the reviewed studies focus on STEM topics. The authors further find that these are followed by psychology, humanities, and social domains. The identified research trends include collaborative video-based learning [4], the effect of annotation and authoring features [4], and the use of videos as an instrument for reflective teacher education [6].

The present review follows a similar format to the one presented by Poquet et al. [5], such that our results are focused around the video characteristics that are analyzed or experimented with, and collect additional context information such as study sample sizes and covered subject areas. Further, this review is more comprehensive in that it does not include only papers that studied the effects on the learning effectiveness but any paper that examines features of video-based learning materials regardless of the ultimate goal. For example, papers whose aim is to propose tools, analyze data from e-learning platforms or suggest features that should be considered in the design of educational videos.Also, in contrast to Giannakos [3] and Yousef et al. [4], this review deepens into the technological trends of VBL by detailing not only the developed tools but the technological tasks carried out to develop such tools. Moreover, our review complements previous studies by focusing on the most recent publications (2020-2021).

Inductively, we deduce a taxonomy of researched video features. This is used to structure our account of research in the domain; and will guide the successive extension of our review towards earlier works.

3. Methodology

This literature review covers studies extracted from three academic databases: (a) the Digital Bibliography and Library Project (DBLP), (b) the Association for Computing Machinery (ACM), (c) and SpringerLink databases.

These are specialized in the computer science field and, thus, provide suitable domain specificity not only for our review of technological systems surrounding video-based learning but also to identify the state of the art in this specific research area.

We iteratively refined our set of query terms, to include new vocabulary from the research papers. Whenever we encountered a new term to describe research around VBL in an article, we went back to the databases and re-ran

Table 2
Literature review findings of Poquet et al. [5]

Category	Features
Presentation features	Video modality (e.g., text, audio, animation, and voice over slides) Text customization (e.g., personal pronouns) Signaling (e.g., highlight relevant information)
Others	Video Usage (e.g., videos instead of face-to-face-lectures) Content (e.g., metaphors in contrast to descriptive language) Task (e.g., annotation features) Quizzes Learner characteristics Learner Control (e.g., pause, play) Distraction
Dependent variables	Recall (remembering information) Transfer (applying what was learned to different scenarios)

Table 3
Search queries: Base terms in the first column are joined (using logical AND) with extensions in the second column

Base	Extension		
Video-based	Learning		
Video	Education	Educational	Instructional
	Learn	Learning	Instruction
	Teach	Teaching	Explaining
	Knowing	Knowledge	Explanatory
	Tutorial	Student	Classroom
	Pupil	School	University
	Skills		

the search with the new terminology. The final set of query combinations can be found in Table 3.

The resulting set of papers was filtered based on the following criteria:

1. Videos are used in academic learning scenarios (this excludes tutorials for everyday tasks, such as cooking instructions).

2. The article considers videos as learning resources (this excludes the use of video recordings in reflective scenarios, such as in teacher education).

3. The article examines specific characteristics and features of the video-based learning materials.

Here, we report on a subset of 41 articles, focusing on very recent publications from the years 2020 and 2021. It is planned to extend the review in the future to cover a longer period time.

4. Video-based Learning

This section gives a short introduction to video-based learning, including a discussion of potentials and challenges as outlined in the reviewed papers, and motivates our specific interest in research works on the analysis of video characteristics. It lays the foundation for the following discussion of the literature review with a focus on video features.

The task of learning by using video sources has adopted the name of video-based learning (VBL). This terminology is used in several studies (e.g., [3, 5, 4, 6]). In the context of this review, the term "video-based learning" is employed in the same way, that is, as gaining knowledge and skills by using video material. This process is manifold and involves several actors beyond the learner and the video such as learning platforms. Consequently, we consider VBL in its wider context by including features of video platforms.

Potentials: It has been argued that VBL can be a powerful tool to enhance teaching and learning. Yousef et al. [4], for instance, state educational videos to be effective, able to increase motivation, engage the learner, and support diverse learning styles. The authors further underline the videos' potential to convey information that is hard to capture in text, e.g., to visualize procedural information. They cite several studies that present evidence about videos improving not only the learning outcome but also the satisfaction, interaction, and communication among learners.

Videos are already an integral part of students' learning habits [1] and common teaching practices [5]. Indeed, some even state that in online education, videos might become the main medium [7]. One reason for this development is certainly availability: Video production has become much easier [5]; and dedicated platforms provide simple, scalable ways for dissemination [3, 8]. This allows laypeople and professional teachers to make their materials available to the world.

Challenges: While offering many proven benefits, there are challenges involved in the production and dissemination of efficient video-based learning resources. Guo et al. [8] found in interviews that the video editing "was not done with any specific pedagogical "design patterns" in mind" [8, p. 45]. The authors also conclude that the production of videos is based on "decisions on anecdotes, folk wisdom, and best practices distilled from studies with at most dozens of subjects and hundreds of video watching sessions." [8, p. 42], which evidences the lack of understanding of how to produce effective videos. Furthermore, it is still unclear if techniques from classical classroom scenarios can be directly transferred to VBL environments, or if and to what degree VBL scenarios need customized didactic approaches and concepts.

Another challenge arises from the question of how a learner can be supported in her video-based learning trajectory. It is widely recognized that learners have diverse needs [8], depending on their current learning objective, context, and general preferences. How to analyze and index video resources to satisfy these individual requirements is still an open research question. As we will show below, exploring possible features for the automatic determination of video content and quality is one active area of research. These explorations are indispensable in the quest to allow diverse learners a satisfying video-based learning experience.

Lastly, MOOCs (Massive Open Online Courses) and similar educational courses are a central research area, mainly because they have enabled data analysis at scale. But it is precisely the large-scale nature of MOOCs that can carry issues in handling and processing such a massive amount of information. As stated by Guo et al. “The scale of data from MOOC interaction logs hundreds of thousands of students from around the world and millions of video watching sessions is four orders of magnitude larger than those available in prior studies” [8, p. 42]. Moreover, critics directed to the research community have been raised as well. For example, a challenge that needs to be confronted in these big scenarios is finding how to incorporate controlled experiments [4] and methods or instruments from other fields, such as the field of psychology, instead of relying only on data analysis [5].

Importance of features: Research on VBL poses a number of relevant questions: How can efficient learning be facilitated and improved? How to enable long-term learning success? How to ensure the learner’s engagement and satisfaction? The answer to these questions strongly relies on one base research question: What characterizes an effective learning video? [8, 4, 5] The search for relevant video features has been widely studied from the perspective of different research fields (e.g., [9]), and in an exploration of what is possible with current technologies (e.g., [10]). In consequence, there is a multitude of variables that have been extracted, studied, and manipulated. Video is a complex medium combining audio, text, and visual information from all of which features can be extracted and combined. These can range from low-level features like the use of audio quality features (e.g.,[11]), to high-level features where semantics try to be captured (e.g., [12]). Features can also not necessarily come directly from the video but can result from the interaction with a video, e.g., views and likes (e.g., [13]). Additional to the characteristics of the video, it is vital to study its environment. The success of a VBL process is further influenced by functionalities of the surrounding platform, the learner’s context, and the instructor captured in the video.

In our review, we focus on those features directly related to the learning video, with the goal to outline current insights and research trends. The discussion of learner-related features, such as previous knowledge and skills, is out of the scope of this review (see, for instance, [14] for a recent review).

5. Literature Review

This section summarizes our findings from the analysis of 41 publications on video-based learning in the period 2020-2021. We analyze (a) the principal research approach (Section 5.1); (b) the target task and the used technologies (Section 5.2); (c) the video features explored in the studies (Section 5.3); (d) and, finally, verify former surveys’ finding that there is a specific focus on STEM subjects in VBL studies (Section 5.4).

5.1. Research Approaches

Four main research approaches have been identified in the reviewed literature: (1) **Controlled experiments**, which in this study are defined as experiments where one or a few variables are manipulated (according to the learning context, video characteristics, or content) and where the impact of this manipulation is measured by some target metric. This group includes experiments run in laboratory settings, but also those performed in realistic academic environments such as university courses. This category accounts for 22 instances (54%) in the reviewed papers. (2) Novel **tools, architectures, or analysis pipelines** surrounding video-based learning scenarios are covered in 16 (39%) articles. (3) **Design principles and guidelines** for the creation of educational videos are presented in two publications (5%). (4) Results of the **analysis of data** generated by users in a video-based online learning platform are presented in one paper (2%).

Table 4 lists all the publications in each category. The high number of empirical studies matches Poquet et al.’s [5] statement that, especially after 2016, empirical works are on the rise. However, an extension of the reviewed period time is necessary to see if our findings concord with the authors’ temporal placement of the change. Within this category (controlled experiments), 21 out of 22 studies reported the number of participants in the experiment (sample size). The smallest study reported 12 participants [15] and the biggest acquired a group of 229 learners [16]. The average number of participants was 85 learners (SD=56). Studies that focused on the development of tools (group (2)) or data analysis of learning platforms such as MOOCs (group (4)) often do not report a sample size.

Some of the subsequent sections focus on a subset of the groups since not all dimensions can be meaningfully extracted from all research objectives. The following review of used technologies, for instance, only makes sense in the context of works that *have* a technology component, and will, consequently, mostly cover works of the group (2).

Table 4
Papers according to the research approach

Research Approach	Publications
Controlled Experiments	[16, 15, 17, 18, 19, 20, 10, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]
Tool, architecture, pipeline	[11, 12, 36, 37, 38, 13, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]
Design principles	[49, 50]
Data analysis	[9]

5.2. Target Task and Used Technologies

This section focuses on the 16 studies which present a technology component. From these, roughly a third of the studies considered in this section (N=5, 31%) present **recommender systems** that aim to support learners in finding suitable learning material. One of the systems does not only recommend video resources but also learning materials in other modalities (e.g., text articles such as Wikipedia pages [39]). Another two systems provide the user with a *video sequence* (playlist) instead of single learning contents [41, 42].

A popular approach to recommend the material is based on the extraction of topical keywords from the video's speech transcript [39, 40, 42], especially, by using the tf-idf algorithm. Those are then used to compute the similarity of the currently viewed video to the other ones in the database [39, 40]. Tavakoli et al. [13] suggest not only use the current video for this but all resources previously viewed by the user. They combine this information with a profile of the learner's skill set and predict relevant videos using a Random Forest classifier. Another technique is to detect prerequisite relations between the videos to generate sequences organized by difficulty and complexity [41, 42]. Lastly, Tang et al. [42] additionally analyzes viewer comments with sentiment analysis to generate a playlist, assuming that positive-sounding comments point to engaging learning material.

Other common use cases are systems for the **forecasting of learning success** (N=3, 19%). "Success" can be operationalized in different ways: In two studies, the objective is to predict the final score of the learner as "pass" or "fail". The first case is in the context of passing a university MOOC course [36], and the second case is about passing video quizzes presented by videos in an e-learning platform [38]. In a third study, the target is to forecast the probability of a student dropping out of a Coursera course (a MOOC provider) [37]. All these recent examples make use of deep learning technologies, specifically, neural networks with gated recurrent units [37], recurrent neural networks (RNN) [38], or long short-term memory neural networks (LSTM) [36]. These classifiers have been mainly fed with real-time features [36, 37], especially play, pause, and speed rate, however, aggregated data [37] and results of quizzes have also been explored [38].

Two of the 16 reviewed papers in this section (13%) focus on **video segmentation**. Video as a learning medium has the downside that all information is presented sequentially. Unlike text, videos cannot be easily skimmed to find relevant information. This can be alleviated by automatic video segmentation. In our sample, one paper aims to automatically determine important segments in the educational content using a bidirectional LSTM network classifier built on pre-trained models to extract text and visual features [11]. Das & Das [12] improve on topical segmentation of lectures videos by identifying concepts in the speech transcripts. To achieve this, speech-to-text-technologies are used, pre-trained neural network models are to analyze the resulting text, and the cluster centroid algorithm is used to group the resulting concepts.

The remaining six studies (38%) focus on diverse tasks such as video indexing [46], video ranking [43], video customization according to the type of learner [47], prediction of instructors' enthusiasm [48], matching videos with practical exercises [44] and helping educators with the creation of teaching materials [45].

5.3. Features

The core focus of this review is the collection of video characteristics that have been studied in the literature (2020 and 2021). The objective is to provide an account of how video features have been investigated, extracted, and varied in the context of studies on video-based learning in the computer science field. We developed a feature taxonomy that groups the video features into categories, which allows for a structured presentation of our findings.

We identified eight high-level categories: (a) audio features, (b) visual features, (c) textual features, (d) features related to the instructor's behavior, (e) features resulting from the interaction of the learner with the video (e.g., likes), (f) interactive features, (g) production style (e.g., Khan style), and (h) features related to instructional design principles. All of these will be defined and their appearances reviewed in separate paragraphs.

Audio features are all features that relate to the actual video's audio stream and also the features that can be extracted from it. This includes *audio records* (e.g., [39]), and *quality-related features*, such as energy, en-

Table 5
Features taxonomy

Category	Sub-category	Examples
Audio features	Audio records [39]	
	Quality-related features [11, 9, 45, 50]	Energy, entropy, spectral features
	Person-related audio features [24, 45, 48, 50]	Clarity of instructor's voice, speech rate, vocabulary, attention guiding emphasis
Visual features	Frames [11, 12, 44]	
	Representation features (directing attention) [24, 25, 26, 27, 19, 45]	Animations, realistic visuals, visual cues (e.g., color-coding, highlighting, arrows)
	Enhanced visuals [23, 20, 10, 21, 22]	Augmented/virtual reality, 360-degree features
	Quality-related features [9]	
Text	Transcripts [11, 12, 13, 39, 40, 41, 42, 43, 44, 46]	
	Metadata [13, 40, 42, 36, 9]	Titles, keywords, tags, video length
	Visual [24, 44]	Visual text (text extracted from frames)
		Textual cues (guiding attention to a key location in a slide)
Instructor's behavior	Gestures [16, 25, 48]	Beat gestures (e.g., hand strokes, natural hand-waving, and pointing gestures that aim at highlight the speech), facial emotions
	Body-related [48]	Volume (expansion of the teacher's body), pose
	Personality [45]	
Interaction between learners and the video	Real-time features [36, 37, 33, 26, 34, 35, 47]	Play, pause, rate-change (speed), seek forward, seek backward
		Percentage of the video that was watched
		Average proportion of videos watched per week
		Standard deviation of the proportion of videos watched per week
	Popularity [13, 9]	Number of views, user ratings, relevancy score (rank in search results)
Interactive features	[38, 33, 30, 28, 26, 32, 29, 31]	Quizzes, annotation, feedback, exchange (actions to interact with other learners)
Production style	[18, 32, 15, 17, 19]	Tutorial, lecture, Khan style, talking head, Dialogue and monologue, voice over slides, only-text, animations included
Instructional design principles	Principles of multimedia learning [9]	Coherence, signalling, spatial contiguity, segmentation (information in segments, rather than a long stream), pre-training (present first the basic information), modality (graphics and spoken words), multimedia (words together with pictures), personalization (informal conversational style), voice (human rather than a computer voice), image (animations)

tropy, and spectral features which allow making predictions on how well the auditory information can be perceived, e.g., [9, 11, 45]. A second big cluster is *person-related audio features* which groups characteristics directly related to the instructor's expression [24, 45, 48]. This includes, for instance, metrics for the teacher's voice [45, 48], speech rate [50], used vocabulary [50], emphasis on specific words that guide the listener's attention [24].

The category **visual features** contains analyses or adaptations related to the video's image information. This includes the analysis of information in *video frames* as images [11, 12, 44]. Particular to video, in contrast to text and still image, is that movement can guide the learner's attention. These types of features are captured in the category *visual representation features* where we can find, for example, highlighting and visual cues (e.g., animated and appearing elements) [24, 25, 26, 27, 19, 45]. There is also work investigating *visual quality features* [9]. Finally, some works explore *enhanced visual representations* by including 360-degree scene representations [20, 10, 21], augmented [23], and virtual reality features [22].

In the category **text features**, we group every text-based information which is available about the video or can be extracted from it. It is common practice to generate a *speech transcript* from the spoken content in a learning video. This pre-processing step transforms hard-to-analyze speech into written text, for which sophisticated and efficient analysis methods are available. Indeed, in our sample, speech transcripts are widely used [11, 12, 13, 39, 40, 41, 42, 43, 44, 46], and are referenced in all the works on video segmentation [11, 12], and on recommending subsequent videos [36, 37, 38].

Many video platforms provide *metadata* about the hosted videos, which are widely used as an easily available data source. These include structured information about the video's title, attributed keywords, tags, and similar [13, 40, 42]. Video length has been used, for instance, by Mubarak et al. [36] as a factor to predict the learning outcome, and by Tavakoli et al. [13] to suggest similar videos. Finally, text, especially in learning videos, also appears to support and visualize the speakers' message in the form of superimposed scene text, often in the form of presentation slides (*visual text*) [24, 44].

In every learning process, the teacher is a central facilitator, there is thus a number of features related to **instructor behavior** [16, 25, 45, 48]. This group captures *gesturing* [16, 25] and *facial emotions* [48], *features related to the body* such as volume and pose [48], and also information about the speaker's *personality* [45], which have been used as base techniques to determine higher-level information.

Several studies use **learner interactions with the video** as a feature. This includes *real-time features* of a user interacting with a certain instructional video [36, 37, 33, 26, 34, 35, 47], e.g., playing and pausing a video, skipping video content or interrupting; and aggregated measures, such as the number and ratio of videos watched in a time period, or the aggregated behavior of several users on a certain video (e.g., [47]). Besides, user interactions are used as indicators of *video popularity* [13, 9], in form of the number of views and ratings.

Interactive functionalities surrounding the video: A critical challenge in videos is that information is transient [51] and consumed in a rather passive way [52, 53]. This invites superficial processing and leads to challenges in knowledge acquisition and integration. In consequence, several studies investigate functionalities to actively involve the user [38, 33, 30, 28, 26, 32, 29, 31] by adding functionalities such as interactive quizzes, collaborative elements, or note-taking areas.

The **production style** category classifies characteristic types of video design [54]. It includes rough subcategories of how the learning setting is composed – e.g., if people are visible (talking-head video, lecture setting with or without an audience) or not (voice-over-slides, Khan-style lecture, animations, and films of real-world phenomena). If there is a single speaker (monologue) or several (dialogue, interview), and in which perspective the content is presented (upfront, or over-the-shoulder view in tutorials). Production style was mentioned several times as a feature in our sample [18, 32, 15, 17, 19].

There is a number of psychological and educational theories on how to efficiently support learners with multimedia resources, references to those are collected in the category **instructional design principles**. The only theoretical element referenced in our sample is the set of Multimedia Learning Principles introduced by Mayer [55], which was used by Eradze et al. [9]. They collected manual annotations that state whether a video follows those design principles (see Table 5 for examples). The aim was to find a correlation between the principles and students' perceptions regarding the quality of the video.

As shown by the taxonomy in Table 5, from all the studies reviewed in this research work (2020-2021), the most explored video features are text-related features, especially transcripts. The next most investigated are interactive features, which study additional functionalities (quizzes, note-taking) that impact the learning success. In third place, we find real-time features. Other usually explored features are visual features, especially features that direct attention (e.g., animation), enhanced visuals (e.g., 360-degree features), and production style (e.g., talking head). These findings are similar to the results of Poquet et al. [5] who show that text, animation, audio, and production style are the most explored features.

Lastly, we found that the impact of the above mentioned features has been measured mainly through the learning outcome using metrics such as recall and transfer [16, 33, 30, 23, 24, 25, 28, 26, 27, 10, 32, 15, 29, 22, 34, 17, 35, 19], which, again, conforms to the results of Poquet et al. [5]. Other frequently used metrics are: cognitive load [16, 25, 32, 22, 34, 17, 31, 19, 33, 23, 26, 18, 32, 22], motivation [30, 23, 20, 10, 22, 17], self-efficacy [28, 22, 17], mental effort [16, 32, 19], eye-gaze direction [25, 15, 17], engagement [20, 26], comprehension (understanding) [21, 29], and social presence [16, 17]. Of these, motivation, cognitive load, mental effort, and the results of eye-tracking were also included in the findings of Poquet et al. [5] as often used metrics. On the other hand, the less frequently used metrics encompass: enjoyment [10], agent-persona (credibility, engagement, and learning facilitating) [16], affective rating [16], para-social interaction [16], self-regulated learning [33], sense of presence (feeling of being present in the environment) [10], perceived benefits [21], confidence [19], creative thinking [22], application [21], analysis [21], synthesis [21], evaluation [21], attention span [32], views about the teaching technique [27], facial temperature [19], and challenge [19].

5.4. Covered Subject Domain

Previous literature reviews stated that STEM (Science, Technology, Engineering & Mathematics) areas are over-represented in the investigation of VBL [3, 5]. This is confirmed in our sample of research papers from 2020-2021. In our first category of controlled experiments, 14 of 22 studies (64%) use videos on some STEM domain. Specifically, computer science [28, 20, 17, 35] and physics [30, 18] are often targeted. In the remaining eight papers, teaching English as a foreign language is the most common discipline [23, 29], followed by other social and humanities domains.

In the research categories "Data analysis" and "Design principles & guidelines" the main focus is also on STEM video content. However, the small sample size in these categories does not lead to further interpretation. Papers in the "Tools" research category do not usually focus on specific disciplines but aim at offering general solutions for VBL. This is why no specific analysis regarding the subject domain is provided.

6. Conclusions

In this paper, we have presented a survey on 41 papers in the field of video-based learning (2020-2021) to provide a structured account of current tendencies in the field. Specifically, we identified (a) common research approaches, (b) target applications and the used technologies, (c) investigated video features and developed a taxonomy which allows structuring the related work, (d) and, finally, collected information on the covered subject domains of the learning environments.

Research approaches: The results show that more than half of the studies are dedicated to controlled experiments (54% of the studies). A significant proportion focused on tool development (39% of the studies), while a very small proportion of studies dedicated effort to analyze data from learning platforms such as MOOCs (2% of the studies) and proposed design principles and guidelines (5% of the studies).

Tools and technologies: The main applications in our sample are three: (1) recommender systems (35% of the studies in the "Tools" category), (2) predictors of learning outcome (19%), and (3) video segmentation techniques (13%). Technology-wise, keyword-based analyses (e.g., using tf-idf to detect pertinent keywords) are the most commonly adopted. Approaches that aim to provide video playlists (as opposed to singular video recommendations) often refer to techniques from the area of prerequisite detection. Recent approaches to forecasting learning success mainly build upon deep learning techniques (e.g., multilayer perceptrons, gated recurrent units, RNNs, and LSTMs). A similar tendency towards deep learning approaches can be seen in methods aiming to improve video segmentation for educational videos; the approaches widely rely on pre-trained models for the extraction of textual and visual features.

Video features: In general, the most explored video features are text-related, especially metadata and speech transcripts. The second most studied features are interactive (such as interactive quizzes and other additional functionalities), followed by click-stream features, visual features, and production style.

Specifically, in the controlled experiments category, the main features manipulated are related to interactive features (32% of the studies), principally quizzes and annotations. Other often explored features in this category are enhanced visual features, mainly 360-degree features, production style features (18% of the studies) with an special emphasis in dialogue and monologue, and features that direct the attention of the learner (18% of the studies) (e.g., animations.)

Subject domain: Based on our sample, we found that STEM domains are prevalent in the targeted subject areas.

Tendencies and directions: In comparison to previous reviews, the tendencies that could be reaffirmed in this study are mainly related to the type of research that was performed, the type of features that have been manipulated, and the subject domain in which the educational videos have focused. Specifically, the trends discovered by Poquet et al. [5] could be confirmed by our review: (1) It was observed that the research community-directed effort principally towards controlled experiments (54% of the reviewed papers). (2) Text, animation, audio, and production style are situated among the most explored features. (3) Recall and transfer are the most-used metrics to measure learning outcome and effectiveness. (4) Cognitive load, mental effort, motivation, and the results of eye-tracking are considered among the most popular metrics for measuring learning experience. (5) Target disciplines of the videos were primarily addressing STEM topics

Research on video, text, and image analysis has been dominated by *deep learning* approaches in recent years. Increasingly, these also find application in the analysis of educational data, exemplified by their dominance in learning outcome prediction, educational video segmentation, and feature extraction. In most cases, the developed systems make use of general-purpose pre-trained models, even though first examples of models specifically trained on educational data exist, e.g., EduBERT [56], a word embedding trained on educational materials. It will be exciting to see the potential of deep learning techniques properly adapted to educational media, and how they will enhance educational applications.

There are various studies that investigate learning on the web, mainly focusing on features of user behavior [57] or textual materials. Only recent publications such as [58] explore the role of videos in such web-based learning processes. However, as shown here, the analysis of video as a learning resource is a highly active research area. The related work outlined here can be a guide for future works which aim to integrate video-based resources into individualized online-learning trajectories, and will provide interested scientists with a starting point for their research.

The reviewed papers investigated a wide range of video features, including information from the visual content, be it textual or image, and audio information. However, the modalities were mainly considered separately, without regard to their interactions. Research on VBL efficiency could be brought forward by a truly *multi-modal analysis* of educational videos, which explores how spoken language, shown text and image, and speaker behavior exhibit a combined message as, for instance, explored by Shi et al. [59]. This is in line with psychological research on instructional design, and might bring new insights for educational research in quantifying formal relationships between image, text and speech in their impact on learning success.

Limitations and future work: This review, with a focus on publications from the years 2020 and 2021, only considers a short period time. It is planned to extend the study in the future, to provide a more thorough analysis of tendencies and directions in VBL. Moreover, although the keyword set used in the paper retrieval process is extensive, we plan to broaden this set to ensure a more comprehensive literature review. As pointed out by the reviewers of this paper, there is a surprisingly low number of MOOC-related studies included in our dataset. This will be considered in the extension of our keyword set.

Acknowledgments

This work has been partly supported by the Ministry of Science and Education of Lower Saxony, Germany, through the Graduate training network "LernMINT: Data-assisted classroom teaching in the MINT subjects". Also, this study has been partly supported by the Leibniz Association, Germany (Leibniz Competition 2018, funding line "Collaborative Excellence", project SALIENT [K68/2017]). Finally, we would like to thank the reviewers for their valuable feedback.

References

[1] A. Smith, S. Toor, P. Van Kessel, Many turn to youtube for children's content, news, how-to lessons, Pew Research Centre 7 (2018). URL: https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/.

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

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

[4] A. M. F. Yousef, M. A. Chatti, U. Schroeder, The state of video-based learning: A review and future perspectives, Int. J. Adv. Life Sci 6 (2014) 122–135.

[5] O. Poquet, L. Lim, N. Mirriahi, S. Dawson, Video and learning: a systematic review (2007-2017), in: A. Pardo, K. Bartimote-Aufflick, G. Lynch, S. B. Shum, R. Ferguson, A. Merceron, X. Ochoa (Eds.), Proceedings of the 8th International Conference on Learning Analytics and Knowledge, LAK 2018, Sydney, NSW, Australia, March 07-09, 2018, ACM, 2018, pp. 151–160. URL: https://doi.org/10.1145/3170358.3170376. doi:10.1145/3170358.3170376.

[6] M. Sablić, A. Mirosavljević, A. Škugor, Video-based learning (vbl)—past, present and future: An overview of the research published from 2008 to 2019, Technology, Knowledge and Learning (2020) 1–17.

[7] A. Hansch, L. Hillers, K. McConachie, C. Newman, T. Schildhauer, J. P. Schmidt, Video and online learning: Critical reflections and findings from the field (2015).

[8] P. J. Guo, J. Kim, R. Rubin, How video production affects student engagement: an empirical study of MOOC videos, in: M. Sahami, A. Fox, M. A. Hearst, M. T. H. Chi (Eds.), First (2014) ACM Conference on Learning @ Scale, L@S 2014, Atlanta, GA, USA, March 4-5, 2014, ACM, 2014, pp. 41–50. URL: https://doi.org/10.1145/2556325.2566239. doi:10.1145/2556325.2566239.

[9] M. Eradze, A. Dipace, B. Fazlagic, A. D. Pietro, Semi-automated student feedback and theory-driven video-analytics: An exploratory study on educational value of videos, in: L. S. Agrati, D. Burgos, P. Ducange, P. Limone, L. Perla, P. Picerno, P. Raviolo, C. M. Stracke (Eds.), Bridges and Mediation in Higher Distance Education - Second International Workshop, HELMeTO 2020, Bari, BA, Italy, September 17-18, 2020, Revised Selected Papers, volume 1344 of *Communications in Computer and Information Science*, Springer, 2020, pp. 28–39. URL: https://doi.org/10.1007/978-3-030-67435-9_3. doi:10.1007/978-3-030-67435-9_3.

[10] P. Araiza-Alba, T. Keane, B. Matthews, K. Simpson, G. Strugnell, W. S. Chen, J. Kaufman, The potential of 360-degree virtual reality videos to teach water-safety skills to children, Comput. Educ. 163 (2021) 104096. URL: https://doi.org/10.1016/j.

compedu.2020.104096. doi:10.1016/j.compedu.2020.104096.
[11] J. A. Ghauri, S. Hakimov, R. Ewerth, Classification of important segments in educational videos using multimodal features, in: S. Conrad, I. Tiddi (Eds.), 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, volume 2699 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2699/paper15.pdf.
[12] A. Das, P. P. Das, Incorporating domain knowledge to improve topic segmentation of long MOOC lecture videos, CoRR abs/2012.07589 (2020). URL: https://arxiv.org/abs/2012.07589. arXiv:2012.07589.
[13] M. Tavakoli, S. Hakimov, R. Ewerth, G. Kismihók, 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, 2020, pp. 1–5. URL: https://doi.org/10.1109/ICALT49669.2020.00008. doi:10.1109/ICALT49669.2020.00008.
[14] A. Abyaa, M. Khalidi Idrissi, S. Bennani, Learner modelling: systematic review of the literature from the last 5 years, Educational Technology Research and Development 67 (2019) 1105–1143. doi:10.1007/s11423-018-09644-1.
[15] A. Nugraha, I. A. Wahono, J. Zhanghe, T. Harada, T. Inoue, Creating dialogue between a tutee agent and a tutor in a lecture video improves students' attention, in: A. Nolte, C. Alvarez, R. Hishiyama, I. Chounta, M. J. Rodríguez-Triana, T. Inoue (Eds.), Collaboration Technologies and Social Computing - 26th International Conference, CollabTech 2020, Tartu, Estonia, September 8-11, 2020, Proceedings, volume 12324 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 96–111. URL: https://doi.org/10.1007/978-3-030-58157-2_7. doi:10.1007/978-3-030-58157-2_7.
[16] M. Beege, M. Ninaus, S. Schneider, S. Nebel, J. Schlemmel, J. Weidenmüller, K. Moeller, G. D. Rey, Investigating the effects of beat and deictic gestures of a lecturer in educational videos, Comput. Educ. 156 (2020) 103955. URL: https://doi.org/10.1016/j.compedu.2020.103955. doi:10.1016/j.compedu.2020.103955.
[17] B. Lee, K. Muldner, Instructional video design: Investigating the impact of monologue- and dialogue-style presentations, in: R. Bernhaupt, F. F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjøn, S. Zhao, B. P. Samson, R. Kocielnik (Eds.), CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020, ACM, 2020, pp. 1–12. URL: https://doi.org/10.1145/3313831.3376845. doi:10.1145/3313831.3376845.
[18] A. Pérez-Navarro, V. Garcia, J. Conesa, Students perception of videos in introductory physics courses of engineering in face-to-face and online environments, Multim. Tools Appl. 80 (2021) 1009–1028. URL: https://doi.org/10.1007/s11042-020-09665-0. doi:10.1007/s11042-020-09665-0.
[19] N. Srivastava, S. Nawaz, J. M. Lodge, E. Velloso, S. M. Erfani, J. Bailey, Exploring the usage of thermal imaging for understanding video lecture designs and students' experiences, in: C. Rensing, H. Drachsler (Eds.), LAK '20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23-27, 2020, ACM, 2020, pp. 250–259. URL: https://doi.org/10.1145/3375462.3375514. doi:10.1145/3375462.3375514.
[20] J. C. Muñoz-Carpio, M. A. Cowling, J. R. Birt, Doctoral colloquium - exploring the benefits of using 360^0 video immersion to enhance motivation and engagement in system modelling education, in: D. Economou, A. Klippel, H. Dodds, A. Peña-Ríos, M. J. W. Lee, D. Beck, J. Pirker, A. Dengel, T. M. Peres, J. Richter (Eds.), 6th International Conference of the Immersive Learning Research Network, iLRN 2020, San Luis Obispo, CA, USA, June 21-25, 2020, IEEE, 2020, pp. 403–406. URL: https://doi.org/10.23919/iLRN47897.2020.9155100. doi:10.23919/iLRN47897.2020.9155100.
[21] W. Daher, H. Sleem, Middle school students' learning of social studies in the video and 360-degree videos contexts, IEEE Access 9 (2021) 78774–78783. URL: https://doi.org/10.1109/ACCESS.2021.3083924. doi:10.1109/ACCESS.2021.3083924.
[22] H. Huang, G. Hwang, C. Chang, 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 (2020) 1386–1405. URL: https://doi.org/10.1111/bjet.12893. doi:10.1111/bjet.12893.
[23] C.-H. Chen, Ar videos as scaffolding to foster students' learning achievements and motivation in efl learning, British Journal of Educational Technology 51 (2020) 657–672.
[24] X. Wang, L. Lin, M. Han, J. M. Spector, 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. URL: https://doi.org/10.1016/j.chb.2020.106279. doi:10.1016/j.chb.2020.106279.
[25] J. Moon, J. Ryu, The effects of social and cognitive

cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction, J. Comput. High. Educ. 33 (2021) 39–63. URL: https://doi.org/10.1007/s12528-020-09255-x. doi:10.1007/s12528-020-09255-x.

[26] A. Cookson, D. Kim, T. Hartsell, Enhancing student achievement, engagement, and satisfaction using animated instructional videos, Int. J. Inf. Commun. Technol. Educ. 16 (2020) 113–125. URL: https://doi.org/10.4018/IJICTE.2020070108. doi:10.4018/IJICTE.2020070108.

[27] M. A. Al-Khateeb, A. M. Alduwairi, Effect of teaching geometry by slow-motion videos on the 8th graders' achievement, Int. J. Interact. Mob. Technol. 14 (2020) 57–67. URL: https://www.online-journals.org/index.php/i-jim/article/view/12985.

[28] M. C. Sözeri, S. B. Kert, Ineffectiveness of online interactive video content developed for programming education, Int. J. Comput. Sci. Educ. Sch. 4 (2021) 49–69. URL: https://doi.org/10.21585/ijcses.v4i3.99. doi:10.21585/ijcses.v4i3.99.

[29] A. A. Kuhail, M. S. Aqel, Interactive digital videos and their impact on sixth graders' english reading and vocabulary skills and retention, Int. J. Inf. Commun. Technol. Educ. 16 (2020) 42–56. URL: https://doi.org/10.4018/IJICTE.2020070104. doi:10.4018/IJICTE.2020070104.

[30] D. Leisner, C. G. Zahn, A. Ruf, A. A. P. Cattaneo, Different ways of interacting with videos during learning in secondary physics lessons, in: C. Stephanidis, M. Antona (Eds.), HCI International 2020 - Posters - 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part II, volume 1225 of *Communications in Computer and Information Science*, Springer, 2020, pp. 284–291. URL: https://doi.org/10.1007/978-3-030-50729-9_40. doi:10.1007/978-3-030-50729-9_40.

[31] S. Chen, D. Wang, Y. Huang, 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, 2021, pp. 1–7.

[32] X. Lu, Q. Li, X. Wang, Research on the impacts of feedback in instructional videos on college students' attention and learning effects, in: W. Shen, J. A. Barthès, J. Luo, Y. Shi, J. Zhang (Eds.), 24th IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021, Dalian, China, May 5-7, 2021, IEEE, 2021, pp. 513–516. URL: https://doi.org/10.1109/CSCWD49262.2021.9437774. doi:10.1109/CSCWD49262.2021.9437774.

[33] D. C. D. van Alten, C. Phielix, J. Janssen, L. Kester, Self-regulated learning support in flipped learning videos enhances learning outcomes, Comput. Educ. 158 (2020) 104000. URL: https://doi.org/10.1016/j.compedu.2020.104000. doi:10.1016/j.compedu.2020.104000.

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

[35] D. Lang, G. Chen, K. Mirzaei, A. Paepcke, Is faster better?: a study of video playback speed, in: C. Rensing, H. Drachsler (Eds.), LAK '20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23-27, 2020, ACM, 2020, pp. 260–269. URL: https://doi.org/10.1145/3375462.3375466. doi:10.1145/3375462.3375466.

[36] A. A. Mubarak, H. Cao, S. A. M. Ahmed, Predictive learning analytics using deep learning model in moocs' courses videos, Educ. Inf. Technol. 26 (2021) 371–392. URL: https://doi.org/10.1007/s10639-020-10273-6. doi:10.1007/s10639-020-10273-6.

[37] B. Jeon, N. Park, Dropout prediction over weeks in moocs by learning representations of clicks and videos, CoRR abs/2002.01955 (2020). URL: https://arxiv.org/abs/2002.01955. arXiv:2002.01955.

[38] H. E. Aouifi, Y. Es-Saady, M. E. Hajji, M. Mimis, H. Douzi, Toward student classification in educational video courses using knowledge tracing, in: M. Fakir, M. Baslam, R. E. Ayachi (Eds.), Business Intelligence - 6th International Conference, CBI 2021, Beni Mellal, Morocco, May 27-29, 2021, Proceedings, volume 416 of *Lecture Notes in Business Information Processing*, Springer, 2021, pp. 73–82. URL: https://doi.org/10.1007/978-3-030-76508-8_6. doi:10.1007/978-3-030-76508-8_6.

[39] C. Schulten, S. Manske, A. Langner-Thiele, H. U. Hoppe, Digital value-adding chains in vocational education: Automatic keyword extraction from learning videos to provide learning resource recommendations, in: C. Alario-Hoyos, M. J. Rodríguez-Triana, M. Scheffel, I. A. Sánchez, S. Dennerlein (Eds.), Addressing Global Challenges and Quality Education - 15th European Conference on Technology Enhanced Learning, EC-TEL 2020, Heidelberg, Germany, September 14-18, 2020, Proceedings, volume 12315 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 15–29. URL: https://doi.org/10.1007/978-3-030-57717-9_2. doi:10.1007/978-3-030-57717-9_2.

[40] J. Jordán, S. Valero, C. Turró, V. J. Botti, Recommending learning videos for moocs and flipped classrooms, in: Y. Demazeau, T. Holvoet, J. M. Corchado, S. Costantini (Eds.), 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, volume 12092

of *Lecture Notes in Computer Science*, Springer, 2020, pp. 146–157. URL: https://doi.org/10.1007/978-3-030-49778-1_12. doi:10.1007/978-3-030-49778-1_12.
[41] M. C. Aytekin, S. Räbiger, Y. Saygin, Discovering the prerequisite relationships among instructional videos from subtitles, in: A. N. Rafferty, J. Whitehill, C. Romero, V. Cavalli-Sforza (Eds.), Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020, International Educational Data Mining Society, 2020. URL: https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_99.pdf.
[42] C. Tang, J. Liao, H. Wang, C. Sung, W. Lin, Conceptguide: Supporting online video learning with concept map-based recommendation of learning path, in: J. Leskovec, M. Grobelnik, M. Najork, J. Tang, L. Zia (Eds.), WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021, ACM / IW3C2, 2021, pp. 2757–2768. URL: https://doi.org/10.1145/3442381.3449808. doi:10.1145/3442381.3449808.
[43] D. I. Bleoanca, S. Heras, J. Palanca, V. Julián, M. C. Mihaescu, LSI based mechanism for educational videos retrieval by transcripts processing, in: C. Analide, P. Novais, D. Camacho, H. Yin (Eds.), Intelligent Data Engineering and Automated Learning - IDEAL 2020 - 21st International Conference, Guimaraes, Portugal, November 4-6, 2020, Proceedings, Part I, volume 12489 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 88–100. URL: https://doi.org/10.1007/978-3-030-62362-3_9. doi:10.1007/978-3-030-62362-3_9.
[44] X. Wang, W. Huang, Q. Liu, Y. Yin, Z. Huang, L. Wu, J. Ma, X. Wang, Fine-grained similarity measurement between educational videos and exercises, in: C. W. Chen, R. Cucchiara, X. Hua, G. Qi, E. Ricci, Z. Zhang, R. Zimmermann (Eds.), MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020, ACM, 2020, pp. 331–339. URL: https://doi.org/10.1145/3394171.3413783. doi:10.1145/3394171.3413783.
[45] A. Hartholt, A. Reilly, E. Fast, S. Mozgai, Introducing canvas: Combining nonverbal behavior generation with user-generated content to rapidly create educational videos, in: S. Marsella, R. Jack, H. H. Vilhjálmsson, P. Sequeira, E. S. Cross (Eds.), IVA '20: ACM International Conference on Intelligent Virtual Agents, Virtual Event, Scotland, UK, October 20-22, 2020, ACM, 2020, pp. 25:1–25:3. URL: https://doi.org/10.1145/3383652.3423880. doi:10.1145/3383652.3423880.
[46] S. Horovitz, Y. Ohayon, Boocture: Automatic educational videos hierarchical indexing with ebooks, in: H. Mitsuhara, Y. Goda, Y. Ohashi, M. M. T. Rodrigo, J. Shen, N. Venkatarayalu, G. Wong, M. Yamada, C. Lei (Eds.), IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020, Takamatsu, Japan, December 8-11, 2020, IEEE, 2020, pp. 482–489. URL: https://doi.org/10.1109/TALE48869.2020.9368461. doi:10.1109/TALE48869.2020.9368461.
[47] S. Lallé, C. Conati, A data-driven student model to provide adaptive support during video watching across moocs, in: I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, E. Millán (Eds.), Artificial Intelligence in Education - 21st International Conference, AIED 2020, Ifrane, Morocco, July 6-10, 2020, Proceedings, Part I, volume 12163 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 282–295. URL: https://doi.org/10.1007/978-3-030-52237-7_23. doi:10.1007/978-3-030-52237-7_23.
[48] Y. Chen, C. Wang, Z. Jian, 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, 2020, pp. 184–191. URL: https://doi.org/10.1145/3449301.3449333. doi:10.1145/3449301.3449333.
[49] T. Weinert, M. T. de Gafenco, M. S. Billert, N. Boerner, Fostering interaction in higher education with deliberate design of interactive learning videos, in: J. F. George, S. Paul, R. De', E. Karahanna, S. Sarker, G. Oestreicher-Singer (Eds.), 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, Association for Information Systems, 2020. URL: https://aisel.aisnet.org/icis2020/digital_learning_env/digital_learning_env/12.
[50] J. Ge, X. Li, 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, 2020, pp. 68–71. URL: https://doi.org/10.1145/3383923.3383927. doi:10.1145/3383923.3383927.
[51] R. E. Mayer, C. Pilegard, Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles, The Cambridge handbook of multimedia learning (2005) 169–182.
[52] R. Ramachandran, E. M. Sparck, M. Levis-Fitzgerald, Investigating the effectiveness of using application-based science education videos in a general chemistry lecture course, J. Chem. Educ. 96 (2019) 479–485. URL: https://doi.org/10.1021/acs.jchemed.

8b00777. doi:10.1021/acs.jchemed.8b00777.
[53] G. Salomon, Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions., Journal of Educational Psychology 76 (1984) 647–658. doi:https://doi.org/10.1037/0022-0663.76.4.647.
[54] K. Chorianopoulos, A taxonomy of asynchronous instructional video styles, The International Review of Research in Open and Distributed Learning 19 (2018) 294–311.
[55] R. Mayer, R. E. Mayer, The Cambridge handbook of multimedia learning, Cambridge university press, 2005.
[56] B. Clavié, K. Gal, Edubert: Pretrained deep language models for learning analytics, CoRR abs/1912.00690 (2019). URL: http://arxiv.org/abs/1912.00690. arXiv:1912.00690.
[57] U. Gadiraju, R. Yu, S. Dietze, P. Holtz, Analyzing knowledge gain of users in informational search sessions on the web, in: C. Shah, N. J. Belkin, K. Byström, J. Huang, F. Scholer (Eds.), Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, CHIIR 2018, New Brunswick, NJ, USA, March 11-15, 2018, ACM, 2018, pp. 2–11. URL: https://doi.org/10.1145/3176349.3176381. doi:10.1145/3176349.3176381.
[58] C. Otto, R. Yu, G. Pardi, J. von Hoyer, M. Rokicki, A. Hoppe, P. Holtz, Y. Kammerer, S. Dietze, R. Ewerth, Predicting knowledge gain during web search based on multimedia resource consumption, in: I. Roll, D. S. McNamara, S. A. Sosnovsky, R. Luckin, V. Dimitrova (Eds.), Artificial Intelligence in Education - 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14-18, 2021, Proceedings, Part I, volume 12748 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 318–330. URL: https://doi.org/10.1007/978-3-030-78292-4_26. doi:10.1007/978-3-030-78292-4_26.
[59] J. Shi, C. Otto, A. Hoppe, P. Holtz, R. Ewerth, 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, SALMM '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 11–19. URL: https://doi.org/10.1145/3347451.3356731. doi:10.1145/3347451.3356731.