Examining the use of text and video resources during web-search based learning—a new methodological approach

Georg Pardi[a]*, Daniel Hienert[b]* and Yvonne Kammerer[a,c]

[a]Leibniz-Institut für Wissensmedien, Tuebingen, Germany; [b]GESIS—Leibniz-Institute for the Social Sciences, Cologne, Germany; [c]Stuttgart Media University, Stuttgart, Germany

ABSTRACT

The present paper introduces a new methodological approach to capture and analyse the processing and use of text, images, and video content during web-search based learning on the free web. We asked 108 university students to search the web to learn about a natural science topic while recording their eye movements and navigation behaviour. Then, we used the 'reading protocol' software to automatically map participants' fixations to text, images, and video content that they had fixated upon on any information resource retrieved. Moreover, we retraced words from participants' post-search essays to words encountered in fixated text or in transcripts of viewed videos, in order to calculate the degree of overlap. Our results showed that the participants directed their attention significantly longer to text than to video or image resources. Nevertheless, multiple video resources were visited by the great majority of students, underlining the importance of videos in web-search based learning. Regarding the origin of learned concepts, more words included in the post-search essay could be retraced to fixated text than to words contained in transcripts of viewed videos. To conclude, we were able to retrace large parts of students' acquired knowledge to retrieved information resources with our approach.

KEYWORDS

Internet; learning; representation formats; text; video; eye tracking; essays

1. Introduction

The web has become a major knowledge resource, and thus, is also regularly used for learning purposes (e.g. Kammerer et al., 2018; Vakkari, 2016), with learning by searching the web being considered as "an active and critical process of knowledge building" (Mason et al., 2010, p. 629). To find

information online, individuals typically use search engines, such as Google, which has become a natural feature of the web (Hillis et al., 2012), providing easy access to vast amounts of information resources on almost any topic. Moreover, the web in general, and search engines in particular, no longer only provide access to textual webpages as part of the general search engine results pages (SERPs), but also to other representation formats, such as images and especially online videos, hereafter referred to as videos (Arguello, 2017; Azzopardi et al., 2018; Kammerer et al., 2018; Wopereis & van Merriënboer, 2011; also see Figure 1 for an example of a SERP from the present study). Therefore, it is not surprising that apart from using text-dominated webpages,

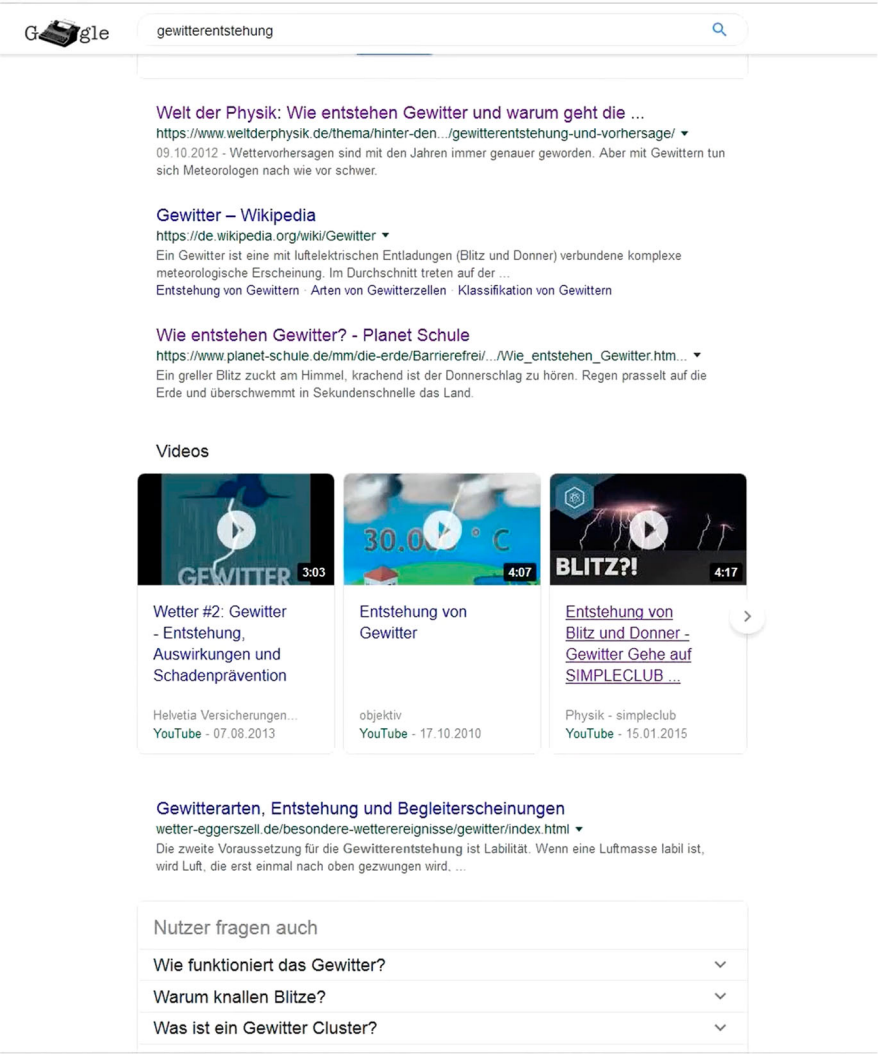

Figure 1. Example screenshot of a SERP for the query "gewitterentstehung" [thunderstorm formation] retrieved by a participant of the present study.

students report to increasingly use videos (e.g. from YouTube) for learning purposes (e.g. ACRL, 2015; Feierabend et al., 2020; Huang & Archer, 2017; Jebe et al., 2019; Koch & Beisch, 2020; Smith et al., 2018; for details see Section 1.1).

Theoretical models describing the process of web-search based learning (e.g. Brand-Gruwel et al., 2009; Frerejean et al., 2019; Gerjets et al., 2011; Kiili et al., 2018; Kuhlthau et al., 2008;) typically distinguish several iterative processing phases, such as: defining the information problem or learning goal (Phase 1); searching for and locating information, e.g. by using a search engine, and deciding which information resources to access (Phase 2); scanning and evaluating the information provided by the accessed resource (Phase 3); if deemed suitable processing the information more deeply and integrating it with prior knowledge and with information from other information resources (Phase 4); and, finally, synthesising the information and representing mentally or communicating in written or oral form what has been learned (Phase 5). However, little is known yet about how different representation formats, such as text, images, and video, contribute to this process of knowledge building while searching the web to learn about a particular topic.

The primary goal of the present study was to shed light on this issue to better understand how different representation formats are used for learning within the open web (cf. Garcia et al., 2021). To this end, 108 university students were asked to search the web freely in order to learn about the complex topic of how thunderstorms and lightning form. They were allowed to use any information resource they wanted. To analyse the degree of use of different kinds of representation formats, we recorded participants' eye movements, navigation logfiles, and HTML data of visited resources they wanted during their web search. We used a further refined version of the 'reading protocol' software (Hienert et al., 2019) that allows us to automatically assess fixation times on any text, image, or video content a participant retrieved. Thus, with our approach, we propose a possibility to automatically analyse areas of interest for web-search based learning sessions, as it recently has also been suggested by Schmidt et al. (2020). Furthermore, we were particularly interested in analysing from where the knowledge originated that participants acquired during web search. For this purpose, we mapped the textual content that participants processed on webpages (written text) and on videos (spoken text) to their essays that they composed about the inquired topic once before their web search (to assess their prior knowledge) and a second time after their web search (from memory).

In sum, with the present study, we aim to contribute a novel approach that allows to comprehensively analyse how learners use different representation formats during web-search based learning. With our approach, we take into account the request put forward by Wopereis and van Merriënboer (2011, p. 236; for a similar suggestion also see Greene et al., 2014) that "future research should consider the evolution of the web towards a predominantly multimedia-based information source."

1.1 The increasing use of videos for web-search based learning

A representative survey by Feierabend et al. (2020) about information-related Internet activities of German adolescents showed for the age group of 18–19 year olds that 62% indicated to use videos on YouTube "daily" or "at least several times a week" to inform themselves about a topic. Besides, 36% indicated to inform themselves "daily" or "at least several times a week" through Wikipedia and comparable websites, 30% through Twitter or Facebook, and 27% through news portals of online newspapers. Focusing specifically on the usage motives of online videos, Koch and Beisch (2020) found for a German sample between 14 and 29 years that 72% of those participants who reported to use YouTube at least once a month indicated to use YouTube "occasionally" to "frequently" to watch explanatory videos and tutorials. Similar results were obtained, for instance, in a recent U.S. representative survey, with 53% of 18–29 years old having reported in 2018 that YouTube was "very important" to them to figure out how to do things they have not done before (Smith et al., 2018).

To conclude, considering these survey results, the importance of videos for learning is clearly recognisable. A potential reason for using online videos for learning purposes is that learning with videos is perceived as easier and less demanding than learning with text materials (e.g. Salomon, 1984). At the same time, this, however, bears the risk of overestimating one's learning performance (e.g. Kardas & O'Brien, 2018). Yet, as we will outline in the following, empirical research on the actual use of online videos compared to other, mostly text-based information resources during web-search based learning is still scarce.

1.2 Learning with textual and video materials

Information resources on the web, such as webpages and videos, often comprise combinations of verbal (written or spoken) and pictorial (static or dynamic) representations (e.g. Mayer, 2017), with different representations being distributed across multiple information resources (e.g. Rouet & Britt, 2014). In his cognitive theory of multimedia learning, Mayer (e.g. 2014) describes how learners select, organise, and integrate verbal and pictorial information during learning. Based on Mayer's work, numerous studies have investigated in controlled experiments whether and, if so, how different representation formats, such as textual as compared to video representations, affect learning when the amount and structure of information is kept equal across representation formats. While some studies found that one format benefitted learning more than the other (e.g. Salmerón, Sampietro, et al., 2020; Schmidt-Weigand & Scheiter, 2011), other research did not find differences between videos and text-based materials regarding learning outcomes (e.g. Delgado et al., 2022; Gerjets et al., 2009; List, 2018; List & Ballenger, 2019; Merkt et al., 2011; Tarchi et al., 2021).

For instance, Schmidt-Weigand and Scheiter (2011) found that university students who were asked to learn with on-screen text perceived learning as more cognitively demanding than students who were provided with an animated video accompanied by on-screen text for their learning. In addition, when the on-screen text did not convey any spatial information, it resulted in inferior retention than when also having the video available. Salmerón, Sampietro et al. (2020) compared secondary-school students' comprehension and integration of information when learning with two textual webpages or with two "talking head" videos. Results showed that the videos were more persuasive than the textual webpages, such that after learning, students defended the views presented in the videos more than those presented on the webpages. Furthermore, participants who learned with the textual webpages better integrated information from the two information resources than those who learned with the videos.

In contrast, List (2018) found university students' comprehension and integration of information to be comparable regardless of whether they learned with two textual webpages or with two animated videos. However, the representation format influenced students' processing strategies. For example, students more frequently reported to consciously direct their attention towards the videos than towards the textual webpages, while they more frequently reported to identify the meaning of vocabulary in textual webpages than in videos.

Similarly, a recent study by Delgado et al. (2022) found no differences in secondary school students' metacognitive calibration and comprehension when learning with video blogs or text-based blogs. Tarchi et al. (2021) investigated undergraduate students' (immediate and delayed) learning outcomes after learning with a text, a video, or a subtitled video. They found no differences between representation formats for immediate testing. However, for a delayed transfer task (six weeks after the learning phase) that required solving tasks about different topics based on the learned content, students who had learned with the text outperformed those who had learned with the subtitled video.

To conclude, even in controlled settings, no clear advantage of one representation format over the other has been shown, and effects of textual as compared to video materials on learning might also depend on the concrete design and content of the learning materials. Thus, in the present research in which we examined web-search based learning in a natural setting with authentic information resources, it was not our goal to investigate which kinds of representation formats would be better or worse for learning. Instead, the main question was to explore to what extent learners *accessed* and *processed* different representation formats when they could choose between a large and heterogeneous set of webpages and videos, and from which information resources the knowledge originated that participants *acquired* during web search.

1.3 The role of different representation formats in web-search based learning

As outlined above, the process of learning with different representation formats has been investigated substantially in the research area of multimedia learning. Yet, the focus within research investigating students' learning with online information so far has been on textual resources (for recent overviews, see e.g. Kammerer et al., 2018; Zlatkin-Troitschanskaia et al., 2021). In contrast, the use of different representation formats has been only rarely addressed in prior scientific research on web-search based learning.

One recent study bringing together web-search based learning and different representation formats is a study by Andresen, Anmarkrud, and Bråten (2019). In their study, secondary-school students were provided with three webpages, each comprising a text, an image, and a video, to learn about the potential health effects of UV radiation. Learning outcomes were assessed as oral responses. The different representation formats (text, images, video) provided complementary information, which allowed the researchers to identify to what extent learners drew on information from the different representation formats in their oral responses. While the focus of the study was to examine differences between students with and without dyslexia, the results for students without dyslexia showed that most information reported in their oral responses originated from the texts, followed by information from the videos. Least information was drawn from the images. Linking this to the abovementioned phases of web-search based learning the study by Andresen, Anmarkrud, and Bråten (2019) provides first insights into Phase 5 (i.e. regarding the origin of the communicated learning outcome).

Furthermore, in a case study with four dyslexic students, Andresen, Anmarkrud, Salmerón et al. (2019) used the same web materials to explore in greater detail how (i.e. in what sequence and to what extent) the four students processed the different representation formats (text, image, video) on the three webpages. Analyses of eye-tracking and logfile data provided insights into Phase 4 of web-search based learning, that is, into the processing of information on the three webpages: The usage patterns differed across learners, with half of the learners first reading text, then viewing the image, and finally watching the video, while the other half of learners followed the linear structure of each webpage starting with watching a video (that was presented at top) followed by reading text and ending with inspecting the image.

The studies mentioned here, investigating the role of different representation formats, had in common that they were conducted with predefined materials instead of analysing search behaviour in an open, authentic setting. The same applies to most previous studies that focused on web-search based learning with (text-based) websites only, usually providing a set of up to 10 preselected and experimentally controlled websites (e.g. Brand-Gruwel et al., 2017; Mason

et al., 2018; Salmerón, Delgado et al., 2020; only to mention a few recent examples). In the following, in contrast, we want to elaborate shortly on methodological approaches of investigating learners' web-search based learning in open, authentic web environments.

1.4 Methods of investigating learning in open web search environments

Methodologically, research investigating user behaviour during web-search based learning on the open web has often focused on capturing learners' interaction with search results and webpages by collecting logfiles (e.g. Câmara et al., 2021; Kalyani & Gadiraju, 2019; Kammerer et al., 2021; Knight et al., 2017; Liu et al., 2010; Marenzi & Zerr, 2012; Roy et al., 2020; Tibau et al., 2018; Yu et al., 2018) or recording gaze behaviour (e.g. Bhattacharya & Gwizdka, 2019; Gwizdka & Chang, 2020; Lewandowski & Kammerer, 2021).

An example of using logfiles for investigating web-search based learning is the work of Yu et al. (2018) which proposed a machine learning model to predict a user's prior knowledge and knowledge gain from 70 specific features, classifiable into session features, query features, SERP features, browsing features, and mouse movement features. Among the most promising features for predicting learning were time-based browsing features, such as maximum or average visit time per page. An example for using logfiles in actual teaching and learning contexts is the LearnWeb (Marenzi & Zerr, 2012), which is designed as a collaborative learning platform that allows monitoring learners' search activities and learning success through learning dashboards based on explicit (e.g. a glossary tool filled by the learner) and implicit measurements (e.g. tracking of queries and search activities). This monitoring allows to individually support learners during the web-search based learning process (Jaakonmäki et al., 2020).

Beyond the usage of log data, several researchers investigated different aspects of web-search based learning with the help of eye-tracking. Lewandowski and Kammerer (2021) provided a comprehensive review of previous research that used eye-tracking to investigate the viewing behaviour on SERPs (in controlled or authentic settings), which falls into the phase of searching for and locating information (i.e. Phase 2 of the web-search based learning process). An example of eye-tracking research that investigated how learners scanned and processed information in websites (Phases 3 and 4) they accessed during learning in an open web search context, is the work of Bhattacharya and Gwizdka (2019). They investigated in detail the reading behaviour of 30 participants performing web search tasks on several health-related topics. Their results showed that participants with higher knowledge gain had read significantly less on webpages but had entered more sophisticated queries than participants with lower knowledge gain.

In contrast to research especially investigating learning with textual and video materials, research investigating free web-search based learning has mostly neglected to consider the type of resources (text, video, or image) learners consult for learning. Moreover, the actual content of the visited web resources has also been neglected. In the present work, we argue that collecting and combining eye-tracking and logfile data, resource data (i.e. the accessed web contents, such as text and video transcripts), and essay data allow to investigate (1) to which extent learners use different representation formats (such as text and video) and (2) how different resources contribute to learning. This can be achieved by mapping and analysing the overlap between the content of visited web resources and participants' newly acquired knowledge as recalled in their post-search essays. We will elaborate on how we implemented this within our approach in the following.

1.5 The present study

In the present research, we tracked and analysed the resource usage (based on eye-tracking data and logfiles) and essay data of 108 university students learning about a complex natural science topic on the web. Generally, our approach (see Figure 2) includes the three steps of (1) data logging, (2) data processing, and (3) mapping. One main difference of our approach compared to most existing work is that beyond logfile and eye-tracking data, we also tracked the data of all visited resources, which enabled us to map newly learned knowledge (Phase 5 of the web-search based learning process) to the processing of the resources (Phase 4).

Specifically, by processing eye-tracking and resource data through a refined version of the 'reading protocol' software (Hienert et al., 2019), we generated a corpus of words that participants had read on websites. Additionally, we traced the words encountered in videos through video transcripts. Subsequently, we

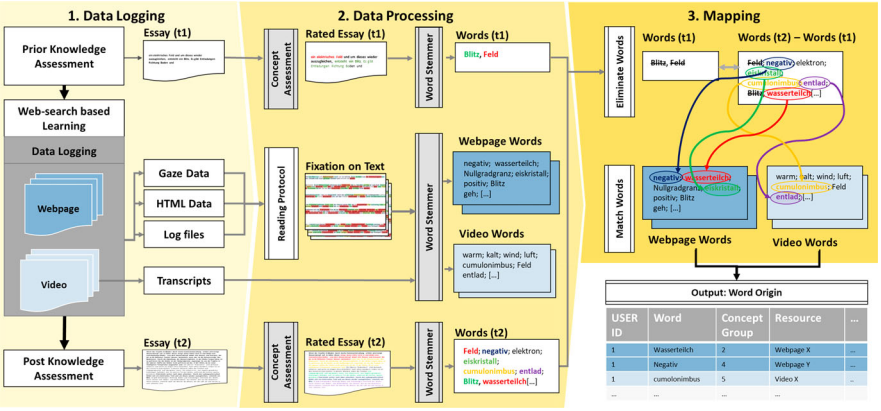

Figure 2. Procedure of retracing words from essay (t2) to visited webpage and video resources.

analysed the overlap between the corpus of encountered words and information recalled in the essays. This enabled us to determine which resources and words processed during web-search based learning participants subsequently also used (i.e. recalled from memory) in their post-search essays.

Further, as our first research question (**RQ1**), we explored to what extent (and how many different) webpages (with text and images) and videos students accessed during web search (Phase 2 of the models outlined in the introduction). As our second research question (**RQ2**), we analysed the extent to which students processed text, images, and video content during web search (Phase 4). Finally, as a third research question (**RQ3**), we investigated to what extent (and from which information resources) students incorporated (i.e. recalled from memory) information from text and from video content in their final essay, in which they summarised what they had learned about the topic (Phase 5).

2. Method

2.1 Participants

Participants were 130 university students from different majors at a large German university, who were compensated with 16€ for their participation. Due to technical problems during data recordings and other issues (e.g. misunderstanding the instructions), data from 15 participants had to be excluded from the dataset. Additionally, we excluded the data of another seven participants due to insufficient tracking ratios (< 80%) of the eye-tracking recordings. The final dataset for the analyses consisted of 108 participants (85.19% female; $M = 22.81$ years; $SD = 2.83$). Fifty-eight participants studied a social science major (e.g. educational science), 30 participants were from a humanities major (e.g. language studies, literature studies), and 20 participants from a natural science major (e.g. physics, medicine). Participants indicated to use the internet on average 32.02 h per week ($SD = 14.90$, scale from 1 to 70 h). Regarding participants' familiarity with search engines, they indicated on a scale ranging from "1 = not at all" to "5 = totally", that they felt quite proficient in using search engines to find suitable information ($M = 3.86$, $SD = 0.88$; "I know how to use search engines to find suitable information"). Participants' prior knowledge on the formation of thunderstorms and lightning was rather low, as indicated by the low number of correct concepts ($M = 1.75$, $SD = 1.80$, out of 20 concepts) included in their (t1) essay written before starting their web search (also see Section 3.3). Students from natural science majors reached a significantly ($F(2, 105) = 4.50$, $p = .009$) higher prior knowledge score ($M = 2.85$ concepts, $SD = 2.94$) compared to students from humanities ($M = 1.50$ concepts, $SD = 1.01$) or social science majors ($M = 1.50$, $SD = 1.48$). Beyond that, however, no significant differences were found between natural

science, social science, and humanities majors, regarding any of the assessed measures (such as, the number of concepts in essay (t2), the number of words in essay (t1) and essay (t2), total session time, or fixation times on text or videos).

2.2 Task

Participants were asked to learn with the help of the web as much as possible about the formation of thunderstorms and lightning. This topic is complex and requires knowledge about different physical and meteorological concepts and their interactions. Participants had a maximum of 30 min for their web-search based learning but could also quit the task earlier. Before and after this learning phase, participants were asked to write an essay in which they were asked to explain how thunderstorms and lightning form as detailed as possible.

2.3 Data logging and data processing

For the first step of data logging (see Figure 2) we used the SMI Experiment-Center 3.7 to record participants' eye movements. The software records gaze data and navigation logfiles. Collected raw data was exported with SMI BeGaze 3.59. To capture all webpages that participants visited in an HTML format, we installed the plugins "ScrapbookX" (1.5.14)[1] and "ScrapbookXAutosave" (1.4.3)[2] within the given Mozilla Firefox Browser (ESR 45.6.0). Each visit of a webpage automatically triggered an imperceptible download process of the necessary files to reconstruct the webpage. For the step of data processing, a refined version of the 'reading protocol' software (Hienert et al., 2019) was used to connect and analyse the collected gaze data and resource data (i.e. the HTML-files). In the reading protocol, raw eye-tracking data (x and y coordinates) are defined as fixations or saccades based on an ID-T algorithm (Salvucci & Goldberg, 2000). The software allows to analyse on any HTML page which parts of the page have been viewed and read by a participant and to calculate participants' total fixation times on text parts, such as words, sentences, or paragraphs, and in the refined version, now also on images and videos. For fixation times on the text and images, we added up all fixation durations on words or images, respectively, across all webpages. The time spent viewing videos comprised the time participants fixated HTML-video elements on YouTube videos and other embedded videos.

The graphical frontend of the 'reading protocol' software (see Figure 3) offers the possibility to illustrate the word-eye-fixations on all read webpages as a heat map. These heatmap visualisations can be displayed for individual participants or accumulated across all participants reading the particular webpage (example data: https://vizgr.org/nrhm_2021).

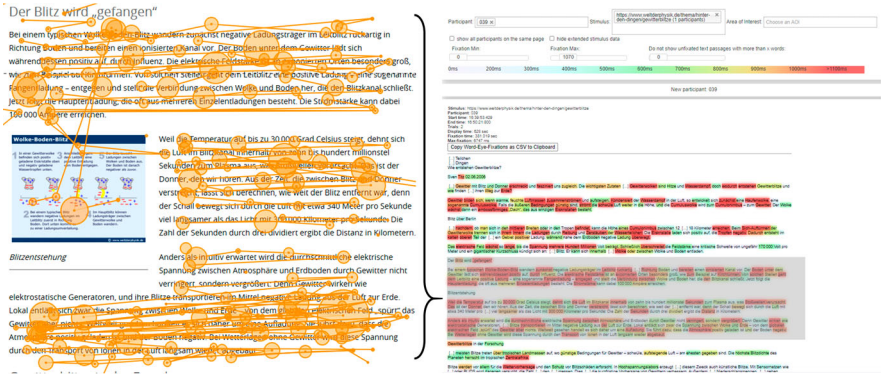

Figure 3. Scan path (from SMI BeGaze) of one participant for a section of the information resource https://www.weltderphysik.de/(...)/gewitterblitze-[…] (left) and the corresponding reading protocol data in the graphical frontend (right).

2.4 Coding system to assess pre- and post-knowledge

To assess participants' pre- and post-knowledge, both essay (t1) and essay (t2) were analysed based on a coding scheme which we developed in an iterative process based on a previous coding scheme by Schmidt-Weigand and Scheiter (2011). Our final coding scheme contained nine concept groups consisting of 20 concepts (see Table 1) that were all related to different aspects of the formation of thunderstorms and lightning. The more concepts a student had correctly addressed in their essay, the more comprehensive was their overall understanding of the formation of thunderstorms and lightning.

Each essay was scored according to the 20 concepts, which allowed to determine the respective concept group they belonged to. A concept was scored as present when a correct conceptual understanding of that concept could be noticed (e.g. "clouds are formed by condensation of humidity into water drops in the air"). Since concepts can be described within or across sentences, we coded on the level of idea units representing the concept, not on a sentence level. Two independent raters coded 55 essays. The overall agreement between the two raters was 95.8%, and the average Cohen's kappa across all concepts was κ .84. Disagreements were resolved through discussion between the raters. Subsequently, one rater coded the remaining essays.

2.5 Mapping essay data to information resources

Since we did not index images regarding their conceptual content, we were not able to map learned concepts to viewed images. For text and video contents, we applied the following steps for each participant: First, based on the coding of the individual concepts (see Table 1), all segments in participants' essay (t1) and

Table 1. Coding scheme with nine concept groups and 20 concepts to assess students' conceptual understanding of how thunderstorms and lightning form.

No.	Concept group	Concepts
1.	Convection	Heating of the soil
		Presence of moist warm air
		Ascent of air
2.	Condensation	Cloud formation due to condensation
		Additional ascent of air due to condensation
3.	Air circulation	Air flows within the cloud
4.	Cloud characteristics	Cloud shape and height
5.	Icing phase	Ice crystals formation and freezing zone within the cloud
6.	Thunderstorm electricity	Origin of electric charge
		Friction and collision of particles within the cloud (ice and water)
		Charge distribution within the cloud
		Electric potential between earth and cloud
		Electrostatic influence
7.	Pre-discharge	Pre-discharge
		Formation of ionised channel
		Upward streamers
8.	Main discharge	Charge equalisation
9.	Other aspects	Explanation flash of light
		Explanation of thunder
		Different lightning types

(t2) belonging to the same concept group were rated and marked with the same colour (see Figure 2, Data Processing). All words from the rated essays were then stemmed (Porter, 2001).

Next, as preparation for the mapping process, all words which had already been used in essay (t1), as well as stop words, words with less than four characters, and special characters, were eliminated from essay (t2). All remaining words then were registered in a list of words, together with information about the concept group they belonged to (see Figure 2, Mapping). This also informed about the most used words for each respective concept group.

Second, the reading protocol allowed us to create a list of all words that a participant had fixated on webpages (based on the extracted HTML-files). Only words fixated for at least 150 ms were included in our word analyses, which according to the E-Z reader model, represents the lower bound for lexical access (Reichle et al., 2009). For visited videos, the reading protocol provided us with the total fixation time on the video frame but not the concrete fixations on content. Therefore, instead, the transcripts offered by YouTube were crawled, checked, and corrected manually if necessary. For videos without available transcripts we manually generated the video transcripts. The encountered words across all visited resources (webpages and video transcripts of viewed videos) were then stemmed and summarised for each participant. Finally, we compared essay words (t2) with all stemmed words per information resource. As a result, we generated a list of word origins where we retraced the words and the associated concept groups of essay (t2) to particular information resources.

2.6 Procedure

Participants were tested in an eye-tracking lab in group sessions of up to four participants that lasted approximately one hour. Each participant had an individual workplace. The workplace consisted of a desk with a laptop connected to a 24-inch screen (1920 × 1080px), linked to a mouse and a keyboard. Below the screen, an SMI (Senso Motoric Instruments) RED250mobile eye-tracking device was attached. After participants were informed about the general procedure, they were positioned on a chin rest in front of the eye tracker. All subsequent steps in the experiment were displayed to the participants on the laptop and processed there.

First, they were asked to write an initial essay (t1) in which they should write down everything they knew about the formation of thunderstorms and lightning. There were no time restrictions or limitations on text length. After completing essay (t1), participants were informed that their task was to conduct a web search to learn about the formation of thunderstorms and lightning and that afterwards, they would have to explain everything they had learned about it to the other participants in the room. This was stated to generate a higher motivation to learn for the participants. In the debriefing, it was resolved that they did not need to explain the topic to other participants. Participants were also encouraged to use any kind of and as many information resources as they wanted.

Then, participants were calibrated on the eye-tracking system using a 9-point calibration, and subsequently started the web-search based learning phase. Participants were provided with access to the internet via the Mozilla Firefox browser (ESR 45.6.0), with the browser cache being cleared for each participant. The starting point for every participant was the Google search engine.

During the whole web-search based learning phase, the screen and eye movements of the participants were recorded with the SMI ExperimentCenter 3.7 software. After having terminated their web search, participants were asked to write another essay (t2) by writing down everything they now knew about the topic. Beyond the measurements reported above, a multiple-choice knowledge test (see von Hoyer et al., 2019; Otto et al., 2021) as well as participants' working memory capacity and reading comprehension skills (see Pardi et al., 2020) were also assessed in the course of this study, which, however, are beyond the scope of the present paper.

3. Results

3.1 Extent of accessing webpages and videos

Overall, participants, on average, spent 25.47 min ($SD = 6.59$) on the web-search based learning task. All but one participant viewed Google SERPs ($M = 12.95$ SERPs, $SD = 10.34$) and spent an average of 2.53 min there ($SD =$

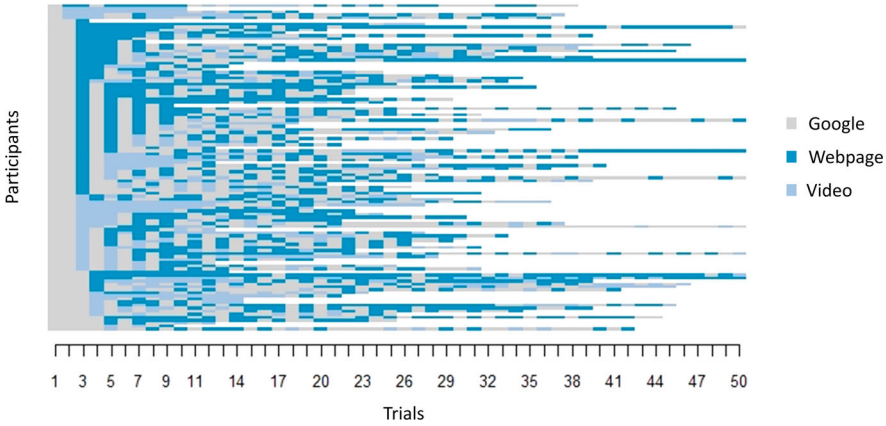

Figure 4. Sequence (trials 1–50) of visited resource types across the 108 participants (1 participant per row) clustered from the start.

2.34). Four participants went directly to YouTube. To address our RQ1, we analysed which content pages participants accessed during their web-search based learning session. In general, across the 108 participants, 239 distinct content resources were accessed. These consisted of 194 textual webpages (from 95 different website domains) and 46 different videos (41 from YouTube, 5 that were incorporated into textual webpages). Figure 4 illustrates the sequences 1–50 (participants had an average sequence length of $M = 29.89$, $SD = 15.92$) in which the 108 participants were accessing Google services, websites, and videos over time (1 participant per row).

Participants visited $M = 6.81$ ($SD = 3.97$, $min = 1$, $max = 22$) different textual webpages, from $M = 5.21$ ($SD = 2.74$) different website domains. For instance, 59 (54.63%) participants visited Wikipedia pages. Furthermore, the majority (91 participants, 84.26%) viewed at least one video. Those 91 participants viewed an average of $M = 3.29$ ($SD = 1.80$, $min = 1$, $max = 8$) different videos. Table A1 (Appendix) gives an overview of the 10 most visited information resources (SERPs were not classified as information resources). Noteworthy, all of those information resources addressed the formation of thunderstorms and lightning and had educational characteristics.

3.2 Extent of processing text, video, and image representations

Regarding RQ2, we analysed across all 108 participants how they devoted their reading and viewing times to text, video, and images. Only content-related webpages (Google-domains excluded) and videos were considered for the following analyses. Figure 5 shows the distribution for the 108 participants with the lengths of viewing times in a stocked bar chart. Each bar represents the measured total fixation time of a participant on text, video, and image representations, arranged in descending order of total fixation time on text.

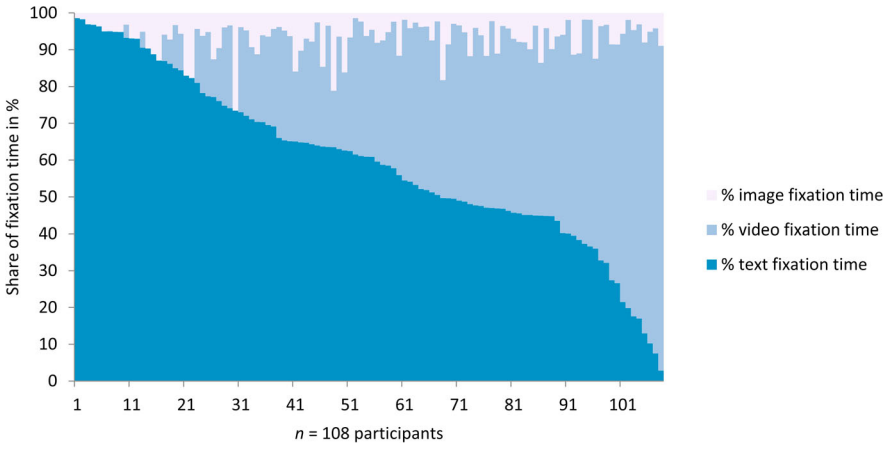

Figure 5. Share of fixation time (in %) for text, video, and images.

Averaged across all 108 participants, the total fixation time on text was dominant with $M = 441.27$ s per participant ($SD = 234.96$, $min = 6.27$ s, $max = 1035.70$ s) followed by $M = 342.61$ s ($SD = 258.69$, $min = 0$, $max = 1{,}202.59$) of video fixation time. Finally, images were fixated for $M = 72.13$ s ($SD = 55.94$, $min = 1.18$ s, $max = 313.20$ s). Pairwise comparisons (Bonferroni-corrected) showed that these time differences between the three representation formats all were significant (all $p < .001$). As shown in Figure 5, the largest share of fixation time was devoted to text ($M = 54.39\%$, $SD = 24.81$), followed by video fixation time ($M = 38.99\%$, $SD = 25.54$). In sum, 59 participants (54.63%) spent more than 50% of their fixation time on text, while only 43 participants (39.81%) spent more than 50% of their fixation time on video. The share of image fixation time ($M = 6.61\%$, $SD = 5.01$) played a minor role among participants.

3.3 Differences between essay (t1) and essay (t2)

Before addressing RQ3 concerning the origin of learned content, in the following, we will first report on the extent of knowledge that participants acquired through their web-search based learning. The essay written before their web-search based learning (t1) represented participants' prior knowledge and the essay written afterwards (t2) their acquired knowledge, respectively. As shown in Table 2, the number of written words (not stemmed), scored

Table 2. Means (and standard deviations) for number of words, scored concepts, and concept groups in essay (t1) and essay (t2) and inferential statistics.

	Essay (t1)	Essay (t2)	Inferential statistics
# words (not stemmed)	41.80 (28.98)	200.15 (72.83)	$t(107) = 23.94$, $p < .001$
# scored concepts	1.75 (1.80)	9.32 (2.60)	$t(107) = 27.73$, $p < .001$
# scored concept groups	1.54 (1.35)	5.08 (1.45)	$t(107) = 21.84$, $p < .001$

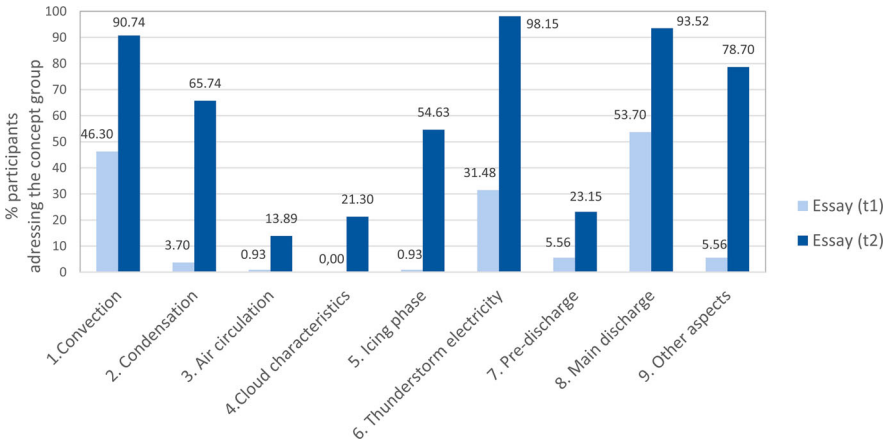

Figure 6. Percentage of participants addressing the nine concept groups in essay (t1) and essay (t2).

concepts, and concepts groups increased significantly from essay (t1) to essay (t2). Furthermore, Figure 6 shows the percentage of participants who referred to the different concept groups in essay (t1) and essay (t2), indicating substantially higher percentages for almost all concept groups in the post-search essay. At the same time, it can be seen that the concept groups addressing air circulation in clouds, cloud characteristics, or pre-discharge of lightning were only addressed by a few participants.

3.4 Used representation formats and word origin

Concerning our RQ3, we analysed the percentage of words from essay (t2) that we could also find in the fixated text on webpages and/or in transcripts of viewed videos. In sum, for all 107 participants who visited at least one textual webpage, words from essay (t2) could be retraced to fixated text. Likewise, for all 91 participants who viewed at least one video, words from essay (t2) could be retraced to video content. Across all 108 participants, an average of 68.52 stemmed words ($SD = 22.20$) were extracted from participants essay (t2). From these words, $M = 10.36$ words ($SD = 5.88$), representing 15.54%, could be matched to stemmed words of essay (t1) and were excluded for further analysis. For the remaining stemmed words ($M = 58.16$, $SD = 19.73$), an average of 79.94% ($SD = 9.90$) could be retraced either to fixated text or to video transcripts of viewed videos, while only $M = 20.06\%$ ($SD = 9.89$) could not be retraced to visited resources.

Looking more closely at the match between stemmed words from essay (t2) and words found in fixated text or in video transcripts (across all 108 participants), a significantly higher percentage ($t(107) = 4.75$, $p < .001$) of words from participants' post-search essays (t2) could be retraced to fixated text (M

= 65.85%, $SD = 17.10$) than to video transcripts ($M = 50.56\%$, $SD = 25.40$). When only considering those 90 participants who had viewed at least one webpage and at least one video, an average of 65.41% of essay (t2) words ($SD = 16.16$) could be retraced to text and $M = 59.89\%$ ($SD = 14.13$) to video transcripts, which represents a significant difference ($t(89) = 2.28$, $p = .025$).

3.5 The origin of different concept groups

Finally, we analysed from which representation format (i.e. text, video) and from which information resources (i.e. concrete webpages and videos) the words connected to learned concept groups potentially originated. Therefore, we compared words falling into concept group annotations and compared them to fixated text and to the transcripts of viewed videos. As before, words that had already been included in essay (t1) were not considered in these analyses. The five most commonly used words per concept group that participants included in their essay (t2) are provided in Table A2 (Appendix). For each concept group, we calculated the average number of words fixated in text for at least 150 ms and/or included in transcripts of viewed videos (see Figure 7). As a result, we identified 106 (out of 107) participants who incorporated words in essay (t2) from fixated text, 91 (out of 91) participants from video content, and 89 (out of 90) participants from both.

As can be seen in Figure 7, when only considering those participants who had visited at least one webpage or at least one video, respectively, for most concept groups, a similar number of words could be retraced to text and to video content (as indicated by the 'text' bars and the 'video' bars). Moreover, a considerable amount of words could be retraced to both text and video

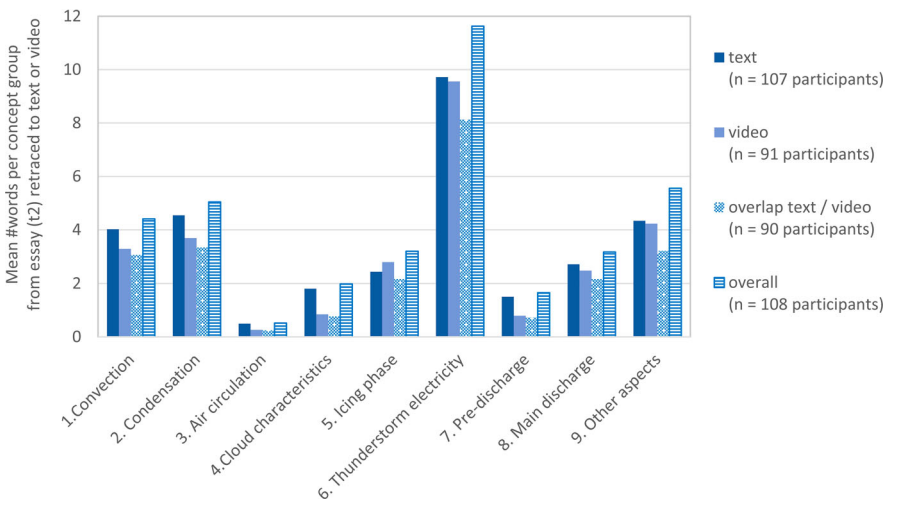

Figure 7. Average number of words from essay (t2) retraced to fixated text or video transcripts of viewed videos.

content (as indicated by the 'overlap' bars). However, to a smaller extent, text and video also contributed different words to the respective concept groups (as indicated, for instance, by the differences between the 'overall' bars and the 'text' bars or the 'video' bars, respectively).

In addition, our approach also allowed us to retrace words for each scored concept group to specific information resources (i.e. specific webpages and videos). Figure 8 shows the average number of words per concept group that participants had come across in the top three webpages (only considering participants who had visited the respective webpage) and subsequently used in their essay (t2). Among the nine concept groups, the highest number of matched words was found for concept group #6, "Thunderstorm Electricity", which was also the most complex group comprising five different concepts. The results further indicate that the three webpages contributed to a different extent to the acquisition of specific concept groups. For instance, the first 'world of physics' webpage seemed to contribute more to concept groups #2 and #4, whereas both the 'planet school' webpage and the second 'world of physics' webpage contributed more to concept group #6.

We used the video transcripts to analyse the overlap between words in essay (t2) related to the nine concept groups and words included in viewed videos. Figure 9 shows the average number of words per concept group that participants potentially had come across in the top three videos (only considering participants who visited the respective video). Again, the highest number of matched words among the nine concept groups was found for concept group #6 "Thunderstorm Electricity". Furthermore, as for the webpages, the results indicate that the three videos contributed to a different extent to the acquisition

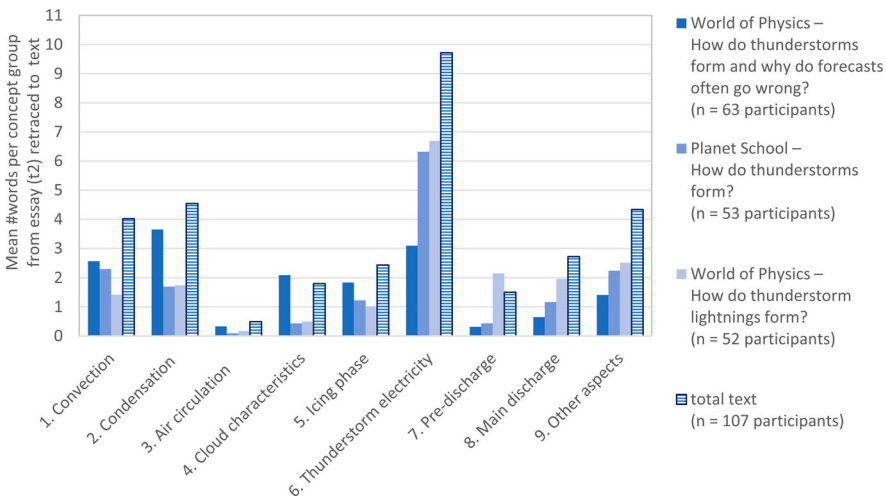

Figure 8. Average number of words from essay (t2) retraced to fixated text (overall or in the three most visited webpages, respectively) as a function of concept group.

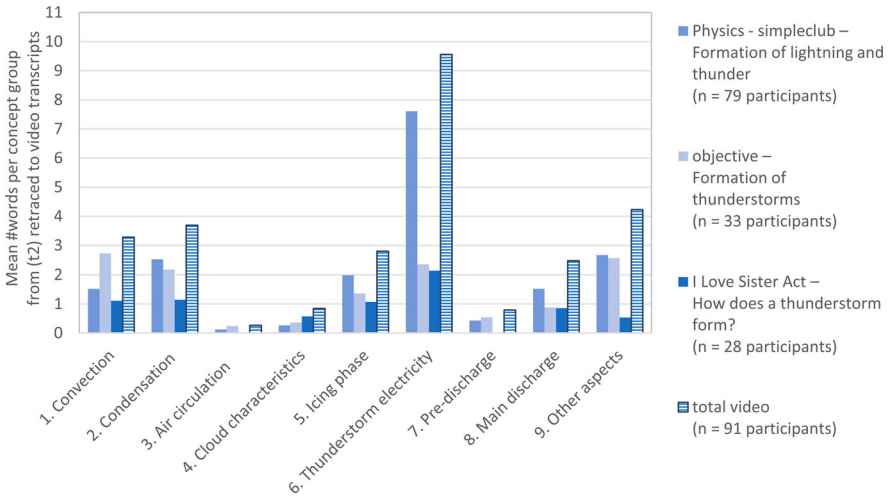

Figure 9. Average number of words from essay (t2) retraced to video transcripts (of viewed videos overall or the three most visited videos, respectively) as a function of concept group.

of specific concept groups. Particularly for concept group #6, the video from 'Physics—simpleclub' had the largest contribution.

4. Discussion and conclusion

To our knowledge, the present work is the first investigating in detail (a) the use of text and video resources and (b) how different resources contributed to knowledge construction in an open, authentic web environment. This could only be achieved through our new approach of combining logfile, eye-tracking, and resource data, which enabled us to map newly learned knowledge represented within essays (Phase 5) to information that was processed in different resources during web-search based learning (Phase 4). Thus, our approach contributes to the recently raised research question by Zlatkin-Troitschanskaia et al. (2021) regarding how students use online resources and information for domain-related learning on the web.

Concerning RQ1 (i.e. the extent to which webpages and videos were accessed; Phase 2 of the process of web-search based learning), we found that nearly twice as many webpages as videos were accessed during students' web-search based learning. Thus, for the students of the present study, webpages in general seemed to play a major role as information resources to learn about the formation of thunderstorms and lightning. Nevertheless, in line with the results of recent surveys (e.g. Feierabend et al., 2020; Koch & Beisch, 2020; Smith et al., 2018), videos, mainly from YouTube, also played a considerable role: Students, on average, viewed more than two videos and the list of the 10 most visited information resources was led by a YouTube video (and also included two more YouTube videos). Noteworthy, our

results also showed that with 194 distinct webpages and 46 different videos being visited by the 108 participants, learners seemed to select different resources rather than everyone using the same few information resources (e.g. Wikipedia).

Concerning RQ2 (i.e. the extent to which text, image, and video representations were processed; Phase 4), our results provide first insights into the actual extent of usage of different representation formats during web-search based learning. Taking the overall fixation time as an indicator for processing content, in line with the findings regarding the number of visited webpages, the results showed that participants fixated on average considerably longer on text (441.27 s) than on video content (342.61 s). Furthermore, images (72.13 s) played a rather subordinate role in students' web-search based learning.

With regard to RQ3 (i.e. the extent to which content from text and video was incorporated in students' final essay after learning; Phase 5), first of all, our results showed that participants included more correct concepts in their post-search essays than in their pre-search essays. Thus, by searching the web, learners managed to extend their conceptual knowledge about the respective topic and communicate it in their post-search essay. Furthermore, with our analyses, we were able to identify from which resources newly acquired knowledge potentially originated. Specifically, we retraced words from participants' post-search essays to words found in text or video transcripts of resources that participants had visited. These analyses revealed that a substantial degree of words addressing different concept groups could be retraced to the three most visited webpages and videos. Furthermore, we were able to map about 66% of the words learners used for the first time in their post-search essay to text they had processed during their web-search based learning and about 51% to videos. These findings align with the results of Andresen, Anmarkrud, and Bråten (2019), who also found an advantage for text as a source of information compared to video.

As a first limitation of the present work, however, it should be acknowledged that we based the matching of essay words to the contents of the viewed videos on the video transcripts of the complete videos. Thus, on the one hand, the number of matched words obtained from our analyses might overestimate the actual encountered words because we cannot exclude that some participants actually did not watch the complete videos. On the other hand, however, our analyses ignored any content that was only addressed visually in the videos but not in spoken text. Likewise, we did not analyse content encountered in images in the present work. Future research could combine our approach with automatic video (e.g. Ewerth et al., 2012) and image content (e.g. Otto et al., 2019) analysis methods.

As a second limitation it should be mentioned that for our analyses concerning concepts we only coded correct statements. Future research could expand

our approach by coding incorrect statements and analysing from which information resources those statements potentially originated. Furthermore, while we concentrated on analysing which resources contributed to the different concepts within the essays, future work could also analyse other aspects of the essays, such as argumentation quality (e.g. Brand-Gruwel et al., 2005; Whitelock-Wainwright et al., 2020) and how it relates to the extent of using websites or videos. However, we believe this would be more relevant when investigating web-search based learning regarding conflicting topics (e.g. Greene et al., 2014, 2018). Moreover, further research is needed to extend our findings to other learning topics and knowledge types (e.g. procedural knowledge). Indeed, it is reasonable to assume that learning with online videos will play an even more important role when learning how to perform a new sensorimotor procedure (e.g. Bétrancourt & Benetos, 2018).

Furthermore, our approach could also be applied to more prolonged and even multi-sessional endeavours of web-search based learning. In this context, longer time intervals between participants' web search and the assessment of their learning could be beneficial to analyse differences between immediate recall after the task and long-term learning (cf. Tarchi et al., 2021). Also, effects of note-taking could be examined (cf. Delgado et al., 2022). Further, by using our approach, future work could also gain additional insights into how learning evolves during a web-search-based learning session (Roy et al., 2020) and how different resources contributed to it. Likewise, our approach could also be applied to SERP viewing to examine which of the words or concepts they fixated in the search results (cf. Taibi et al., 2017) were later recalled in their essays.

A final limitation of the present work is that, even though students from natural science majors had a somewhat higher prior knowledge on the topic than students from humanities and social science majors, we examined a rather homogeneous sample of university students with rather low prior knowledge on the task at hand, but quite high experience in conducting web searches for learning purposes. In contrast, future studies could use our approach to investigate web-search based learning in more heterogeneous samples, for instance, to investigate differences in web search behaviour between domain experts and novices (cf. Brand-Gruwel et al., 2017) or between search experts and novices, respectively, in terms of their usage of different representation formats and how information from these resources was subsequently recalled in their learning products.

Notwithstanding the abovementioned limitations, we believe that our new methodological approach offers great potential to investigate web-search based learning processes, especially on the open, authentic web. Still, it could also be used in experimental environments with prepared sets of resources. It can provide valuable insights into which resources, and particularly which passages, were consulted and subsequently recalled by students in their writing.

While our approach is considered as a research tool in the first place, in the future, it might also help teachers to find out, for instance, which resources were consulted and which information was subsequently used or recalled by those students who performed best in the assigned learning task. Given this information, teachers could recommend struggling learners suitable resources or determine whether they had already consulted those resources but concentrated on the wrong passages or whether they had even read the correct passages but still did not use the information in their learning product. In addition, passages that contributed to learning for good learners could be automatically highlighted to guide poorer learners' attention to relevant content, for example, in learning environments such as the LearnWeb (Jaakonmäki et al., 2020).

In conclusion, to our knowledge, this is the first study that provides detailed insights about how and to what extent learners use textual and video resources in a free web environment and what they subsequently recall and understand from these resources. In sum, our results show that to learn about the formation of thunderstorms and lightning, the majority of the examined university students used both textual and video resources to a considerable extent.

Notes

1. https://github.com/danny0838/firefox-scrapbook
2. https://github.com/danny0838/firefox-scrapbook-autosave

Funding

The present research was funded by the Leibniz Association, Germany [Leibniz Competition 2018, funding line "Collaborative Excellence", project SALIENT (Grant Number K68/2017)].

References

ACRL. (2015). Framework for information literacy for higher education. The Association of College and Research Libraries. https://www.ala.org/acrl/standards/ilframework

Andresen, A., Anmarkrud, Ø, & Bråten, I. (2019). Investigating multiple source use among students with and without dyslexia. *Reading and Writing*, *32*(5), 1149–1174. https://doi.org/10.1007/s11145-018-9904-z

Andresen, A., Anmarkrud, Ø, Salmerón, L., & Bråten, I. (2019). Processing and learning from multiple sources: A comparative case study of students with dyslexia working in a multiple source multimedia context. *Frontline Learning Research*, *7*(3), 1–26. https://doi.org/10.14786/flr.v7i3.451

Arguello, J. (2017). Aggregated search. *Foundations and Trends in Information Retrieval*, *10*(5), 365–502. https://doi.org/10.1561/1500000052

Azzopardi, L., Thomas, P., & Craswell, N. (2018). Measuring the utility of search engine result pages: An information foraging based measure. In K. Collins-Thompson, & Q. Mei (Eds.), *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 605–614). ACM. https://doi.org/10.1145/3209978.3210027

Bétrancourt, M., & Benetos, K. (2018). Why and when does instructional video facilitate learning? A commentary to the special issue "developments and trends in learning with instructional video". *Computers in Human Behavior*, *89*, 471–475. https://doi.org/10.1016/j.chb.2018.08.035

Bhattacharya, N., & Gwizdka, J. (2019). Measuring learning during search: Differences in interactions, eye-gaze, and semantic similarity to expert knowledge. In L. Azzopardi, M. Halvey, I. Ruthven, H. Joho, V. Murdock, & P. Qvarfordt (Eds.), *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 63–71). ACM. https://doi.org/10.1145/3295750.3298926

Brand-Gruwel, S., Kammerer, Y., Van Meeuwen, L., & Van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning*, *33*(3), 234–251. https://doi.org/10.1111/jcal.12162

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. https://doi.org/10.1016/j.chb.2004.10.005

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, *53*(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004

Câmara, A., Roy, N., Maxwell, D., & Hauff, C. (2021). Searching to learn with instructional scaffolding. In F. Scholer, P. Thomas, D. Elsweiler, H. Joho, N. Kando, & C. Smith (Eds.), *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 209–218). ACM. https://doi.org/10.1145/3406522.3446012

Delgado, P., Anmarkrud, Ø, Avila, V., Altamura, L., Chireac, S. M., Pérez, A., & Salmerón, L. (2022). Learning from text and video blogs: Comprehension effects on secondary school students. *Education and Information Technologies*, *27*(4), 5249–5275. https://doi.org/10.1007/s10639-021-10819-2.

Ewerth, R., Mühling, M., & Freisleben, B. (2012). Robust video content analysis via transductive learning. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 1–26. https://doi.org/10.1145/2168752.2168755

Feierabend, S., Rathgeb, T., Kheredmand, H., & Glöckler, S. (2020). *JIM-Studie 2020: Jugend, Information, Medien: Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland [JIM-Study 2020: Youth, Information, Media: Basic Study on Media Use by 12- to 19-Year-Olds in Germany]*. Medienpädagogischer Forschungsverbund Südwest. https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf

Frerejean, J., Velthorst, G. J., van Strien, J. L., Kirschner, P. A., & Brand-Gruwel, S. (2019). Embedded instruction to learn information problem solving: Effects of a whole task

approach. *Computers in Human Behavior*, *90*, 117–130. https://doi.org/10.1016/j.chb.2018.08.043

Garcia, C., Argelagós, E., & Privado, J. (2021). Assessment of higher education students' information problem-solving skills in educational sciences. *Information Development*, *37*(3), 359–375. https://doi.org/10.1177/0266666920976189

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. https://doi.org/10.1016/j.learninstruc.2010.02.005

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, *25*(2), 360–370. https://doi.org/10.1016/j.chb.2008.12.015

Greene, J. A., Copeland, D. Z., Deekens, V. M., & Yu, S. B. (2018). Beyond knowledge: Examining digital literacy's role in the acquisition of understanding in science. *Computers & Education*, *117*, 141–159. https://doi.org/10.1016/j.compedu.2017.10.003

Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education*, *76*, 55–69. https://doi.org/10.1016/j.compedu.2014.03.008

Gwizdka, J., & Chang, Y. S. (2020). Search results viewing behavior vis-à-vis relevance criteria. In F. Davis, R. Riedl, J. vom Brocke, P. M. Léger, A. Randolph, & T. Fischer (Eds.), *Information Systems and neuroscience* (pp. 181–188). Springer. https://doi.org/10.1007/978-3-030-28144-1_20

Hienert, D., Kern, D., Mitsui, M., Shah, C., & Belkin, N. J. (2019). Reading protocol: Understanding what has been read in interactive information retrieval tasks. In L. Azzopardi, M. Halvey, & I. Ruthven (Eds.), *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 73–81). ACM. https://doi.org/10.1145/3295750.3298921

Hillis, K., Petit, M., & Jarrett, K. (2012). *Google and the culture of search*. Routledge. https://doi.org/10.4324/9780203846261

Huang, C.-W., & Archer, A. (2017). 'Academic literacies' as moving beyond writing: Investigating multimodal approaches to academic argument. *London Review of Education*, *15*(1), 63–72. https://doi.org/10.18546/LRE.15.1.06.

Jaakonmäki, R., vom Brocke, J., Dietze, S., Drachsler, H., Fortenbacher, A., Helbig, R., … & Yun, H. (2020). Understanding students' online behavior while they search on the internet: Searching as learning. In R. Jaakonmäki, J. vom Brocke, S. Dietze, H. Drachsler, A. Fortenbacher, R. Helbig, … , & H. Yun (Eds.), *Learning analytics cookbook* (pp. 75–88). Springer. https://doi.org/10.1007/978-3-030-43377-2_7

Jebe, F., Konietzko, S., Lichtschlag, M., & Liebau, E. (2019). *Jugend/YouTube/Kulturelle Bildung. Horizont 2019 [Youth/YouTube/Cultural Education. Horizon 2019]*. Rat für Kulturelle Bildung e. V. https://www.rat-kulturelle-bildung.de/fileadmin/user_upload/pdf/Studie_YouTube_Webversion_final_2.pdf

Kalyani, R., & Gadiraju, U. (2019). Understanding user search behavior across varying cognitive levels. In C. Atzenbeck, J. Rubart, & D. E. Millard (Eds.), *Proceedings of the 30th ACM Conference on Hypertext and Social Media* (pp. 123–132). ACM. https://doi.org/10.1145/3342220.3343643

Kammerer, Y., Brand-Gruwel, S., & Jarodzka, H. (2018). The future of learning by searching the Web: Mobile, social, and multimodal. *Frontline Learning Research*, *6*(2), 81–91. https://doi.org/10.14786/flr.v6i2.343

Kammerer, Y., Gottschling, S., & Bråten, I. (2021). The role of internet-specific justification beliefs in source evaluation and corroboration during web search on an unsettled socio-scientific issue. *Journal of Educational Computing Research*, *59*(2), 342–378. https://doi.org/10.1177/0735633120952731

Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science*, *29(4)*, 521–536. https://doi.org/10.1177/0956797617740646

Kiili, C., Leu, D. J., Utriainen, J., Coiro, J., Kanniainen, L., Tolvanen, A., Lohvansuu, K., & Leppänen, P. H. (2018). Reading to learn from online information: Modeling the factor structure. *Journal of Literacy Research*, *50*(3), 304–334. https://doi.org/10.1177/1086296X18784640

Knight, S., Rienties, B., Littleton, K., Mitsui, M., & Tempelaar, D. (2017). The relationship of (perceived) epistemic cognition to interaction with resources on the internet. *Computers in Human Behavior*, *73*, 507–518. https://doi.org/10.1016/j.chb.2017.04.014

Koch, W., & Beisch, N. (2020). Ergebnisse der ARD/ZDF-Onlinestudie 2020: Erneut starke Zuwächse bei Onlinevideo [Results of the ARD/ZDF online study 2020: Again large growth in online video]. *Media Perspektiven*, 9/2020, 482–500. https://www.ard-werbung.de/fileadmin/user_upload/media-perspektiven/pdf/2020/0920_Koch_Beisch_Korr_30-11-20.pdf

Kuhlthau, C. C., Heinström, J., & Todd, R. J. (2008). The 'information search process' revisited: Is the model still useful. *Information Research*, *13*(4), 13–14.

Lewandowski, D., & Kammerer, Y. (2021). Factors influencing viewing behaviour on search engine results pages: A review of eye-tracking research. *Behaviour & Information Technology*, *40*(14), 1485–1515. https://doi.org/10.1080/0144929X.2020.1761450

List, A. (2018). Strategies for comprehending and integrating texts and videos. *Learning and Instruction*, *57*, 34–46. https://doi.org/10.1016/j.learninstruc.2018.01.008

List, A., & Ballenger, E. E. (2019). Comprehension across mediums: The case of text and video. *Journal of Computing in Higher Education*, *31*(3), 514–535. https://doi.org/10.1007/s12528-018-09204-9

Liu, J., Cole, M. J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., Zhang, J., & Zhang, X. (2010). Search behaviors in different task types. In J. Hunter, C. Lagoze, L. Giles, & Y.-F. Li (Eds.), *Proceedings of the ACM International Conference on Digital Libraries* (pp. 69–78). ACM. https://doi.org/10.1145/1816123.1816134.

Marenzi, I., & Zerr, S. (2012). Multiliteracies and active learning in CLIL—the development of LearnWeb2.0. *IEEE Transactions on Learning Technologies*, *5*(4), 336–348. https://doi.org/10.1109/TLT.2012.14

Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607–633. https://doi.org/10.1007/s11251-008-9089-y

Mason, L., Scrimin, S., Tornatora, M. C., Suitner, C., & Moè, A. (2018). Internet source evaluation: The role of implicit associations and psychophysiological self-regulation. *Computers & Education*, *119*, 59–75. https://doi.org/10.1016/j.compedu.2017.12.009

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. Learning with print: The role of interactive features. *Learning & Instruction*, *21*(6), 687–704. https://doi.org/10.1016/j.learninstruc.2011.03.004.

Otto, C., Springstein, M., Anand, A., & Ewerth, R. (2019). Understanding, categorizing and predicting semantic image-text relations. In A. El Saddik, A. Del Bimbo, & Z. Zhang (Eds.), *Proceedings of the 2019 International Conference on Multimedia Retrieval* (pp. 168–176). ACM. https://doi.org/10.1145/3323873.3325049

Otto, C., Yu, R., Pardi, G., von Hoyer, J., Rokicki, M., Hoppe, A., Holtz, P., Kammerer, Y., Dietze, S., & Ewerth, R. (2021). Predicting knowledge gain during Web search based on multimedia resource consumption. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Proceedings of the 22nd International Conference on Artificial Intelligence in Education* (pp. 318–330). Springer. https://doi.org/10.1007/978-3-030-78292-4_26

Pardi, G., von Hoyer, J., Holtz, P., & Kammerer, Y. (2020). The role of cognitive abilities and time spent on texts and videos in a multimodal searching as learning task. In H. O'Brien, & L. Freund (Eds.), *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (pp. 378–382). ACM. https://doi.org/10.1145/3343413.3378001

Porter, M. F. (2001). Snowball: A language for stemming algorithms. http://snowball.tartarus.org/texts/introduction.html

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21. https://doi.org/10.3758/PBR.16.1.1

Rouet, J., & Britt, A. (2014). Multimedia learning from multiple documents. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 813–841). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.039

Roy, N., Moraes, F., & Hauff, C. (2020). Exploring users' learning gains within search sessions. In H. O'Brien, & L. Freund (Eds.), *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (pp. 432–436). ACM. https://doi.org/10.1145/3343413.3378012

Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, *36*(6), 1038–1051. https://doi.org/10.1111/jcal.12458

Salmerón, L., Sampietro, A., & Delgado, P. (2020). Using internet videos to learn about controversies: Evaluation and integration of multiple and multimodal documents by primary school students. *Computers & Education*, *148*, Article 103796. https://doi.org/10.1016/j.compedu.2019.103796.

Salomon, G. (1984). 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*(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71–78). ACM. https://doi.org/10.1145/355017.355028

Schmidt, S., Zlatkin-Troitschanskaia, O., Roeper, J., Klose, V., Weber, M., Bültmann, A. K., & Brückner, S. (2020). Undergraduate students' critical online reasoning—Process mining analysis. *Frontiers in Psychology*, *11*, Article 3301. https://doi.org/10.3389/fpsyg.2020.576273

Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. Computers in Human Behavior. *27*(1), 22–28. https://doi.org/10.1016/j.chb.2010.05.007

Smith, A., Toor, S., & van Kessel, P. (2018). *Many turn to YouTube for* childreńs content, *news, how-to lessons.* Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/11/PI_2018.11.07_youtube_FINAL.pdf

Taibi, D., Fulantelli, G., Marenzi, I., Nejdl, W., Rogers, R., & Ijaz, A. (2017). SaR-WEB: A semantic web tool to support search as learning practices and cross-language results on the web. In M. Chang, N. Chen, R. Huang, Knshuk, D. G. Sampson, & R. Visiu (Eds.), *Proceedings of the 2017 International Conference on Advanced Learning Technologies* (pp. 522–524). IEEE. https://doi.org/10.1109/ICALT.2017.51

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education*, *160*, Article104034. https://doi.org/10.1016/j.compedu.2020.104034..

Tibau, M., Siqueira, S. W., Nunes, B. P., Bortoluzzi, M., Marenzi, I., & Kemkes, P. (2018). Investigating users' decision-making process while searching online and their shortcuts towards understanding. In G. Hancke, M. Spaniol, K. Osathanunkul, S. Unankard, & R. Klamma (Eds.), *Proceedings of the 2018 International Conference on Web-Based Learning* (pp. 54–64). Springer. https://doi.org/10.1007/978-3-319-96565-9_6

Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, *42*(1), 7–18. https://doi.org/10.1177/0165551515615833

von Hoyer, J., Pardi, G., Kammerer, Y., & Holtz, P. (2019). Metacognitive judgments in searching as learning (SAL) tasks: Insights on (Mis-) calibration, multimedia usage, and confidence. In R. Ewerth, S. Dietze, A. Hoppe, & R. Yu (Eds.), *Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information* (pp. 3–10). ACM. https://doi.org/10.1145/3347451.3356730

Whitelock-Wainwright, A., Laan, N., Wen, D., & Gašević, D. (2020). Exploring student information problem solving behaviour using fine-grained concept map and search tool data. *Computers & Education*, *145*, Article 103731. https://doi.org/10.1016/j.compedu.2019.103731.

Wopereis, I. G., & van Merriënboer, J. J. (2011). Evaluating text-based information on the World wide Web. *Learning and Instruction*, *21*(2), 232–237. https://doi.org/10.1016/j.learninstruc.2010.02.003

Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P., & Dietze, S. (2018). Predicting user knowledge gain in informational search sessions. In K. Collins-Thompson & Q. Mei (Eds.), *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 75–84). ACM. https://doi.org/10.1145/3209978.3210027.

Zlatkin-Troitschanskaia, O., Hartig, J., Goldhammer, F., & Krstev, J. (2021). Students' online information use and learning progress in higher education—A critical literature review. *Studies in Higher Education*, *46*(10), 1996–2021. https://doi.org/10.1080/03075079.2021.1953336

Appendix

Table A1. Overview of the 10 most visited information resources (only resource URLs considered, no SERPs or service Google pages included).

No.	URL	Source (Translation)	Title (Translation)	# Participants visiting	*M (SD)* of viewing time per participant (in minutes)[a]	*M (SD)* of viewing time per participant (in %)[a]	Overall viewing time across all participants (in hours) [a]
1	https://www.youtube.com/watch?v=BGDVGWhknwk	YouTube channel: Physik – simpleclub (Physics – simpleclub)	Entstehung von Blitz und Donner (Formation of lightning and thunder)	79	5.67 (4.08)	22.09 (15.99)	7.47
2	https://www.weltderphysik.de/thema/hinter-den-dingen/gewitterentstehung-und-vorhersage/	Welt der Physik (World of Physics)	Wie entstehen Gewitter und warum geht die Vorhersage oft schief? (How do thunderstorms form and why do forecasts often go wrong?)	63	3.13 (2.58)	11.96 (9.75)	3.29
3	https://www.planet-schule.de/mm/die-erde/Barrierefrei/pages/Wie_entstehen_Gewitter.html	Planet Schule (Planet School)	Wie entstehen Gewitter? (How do thunderstorms form?)	54	1.97 (1.12)	7.71 (4.53)	1.78
4	https://www.weltderphysik.de/thema/hinter-den-dingen/gewitterblitze/	Welt der Physik (World of Physics)	Wie entstehen Gewitterblitze? (How do thunderstorm lightnings form?)	52	4.38 (3.85)	17.87 (14.94)	3.80
5	https://de.wikipedia.org/wiki/Gewitter	Wikipedia	Gewitter (Thunderstorm)	47	3.55 (2.53)	13.11 (10.10)	2.63
6	https://www.wissenschaft-im-dialog.de/projekte/wieso/artikel/beitrag/wie-entstehen-blitz-und-donner/	Wissenschaft im Dialog (Science in Dialog)	Wie entstehen Gewitter? (How do thunderstorms form)	43	1.55 (1.13)	6.58 (5.33)	1.12
7	https://www.youtube.com/watch?v=wO-NL8Bbu_c	YouTube channel: objektiv (objective)	Entstehung von Gewitter (Formation of thunderstorms)	33	2.98 (1.72)	10.53 (6.14)	1.64
8	https://de.wikipedia.org/wiki/Blitz	Wikipedia	Blitz (Lightning)	32	2.87 (3.37)	10.96 (12.61)	1.54
9	https://www.nela-forscht.de/2011/06/08/wie-entsteht-ein-gewitter	Nela forscht (Nela investigates)	Wie entsteht ein Gewitter? (How does a thunderstorm form?)	32	2.98 (3.68)	12.13 (17.23)	1.60
10	https://www.youtube.com/watch?v=KzB-X4AIuE8	YouTube channel: Ilovesisteract (I love Sister Act)	Wie entsteht Gewitter? (How does a thunderstorm form?)	28	1.82 (0.53)	8.36 (4.45)	0.86

[a]Only includes participants who accessed the information resource.

Table A2. Overview of the 5 most used (stemmed) words in German (English translation in parentheses) per concept group as well as the average number of found words per concept group (overall, in text, in video, in text & video).

No.	Concept group		1	2	3	4	5	*M* (*SD*) overall (*n* = 108)	*M* (*SD*) text (*n* = 107)	*M* (*SD*) video (*n* = 91)	*M* (*SD*) overlap text / video (*n* = 90)
1	Convection	#participants	44	38	27	26	18	4.25 (3.00)	3.87 (2.80)	3.29 (2.76)	2.97 (2.39)
		word	steigt (rises)	feucht (moist)	oben (above)	warm (warm)	wasserdampf (water steam)				
2	Condensation	#participants	31	22	20	20	19	4.86 (4.88)	4.35 (4.41)	3.75 (3.71)	3.07 (3.10)
		word	kondensiert (condenses)	wasserdampf (water steam)	energi (energy)	wassertröpfch (water droplets)	kondensation (condensation)				
3	Air circulation	#participants	9	4	2	2	2	0.50 (1.45)	0.47 (1.35)	0.26 (1.05)	0.22 (0.93)
		word	aufwind (upwind)	stark (strong)	eiskristall (ice crystal)	gewitterwolk (thundercloud)	luft (air)				
4	Cloud characteristics	#participants	10	9	7	7	7	1.83 (4.41)	1.70 (4.08)	0.85 (2.36)	0.66 (1.95)
		word	stratosphär (stratosphere)	horizontal (horizontal)	gewitterwolk (thundercloud)	grenz (border)	tropopause (tropopause)				
5	Icing phase	#participants	31	28	17	15	15	3.09 (3.42)	2.35 (2.70)	2.80 (3.02)	1.91 (2.14)
		word	gefri (freeze)	eiskristall (ice crystal)	wolk (cloud)	nullgradgrenz (zero degree limit)	wassertröpfch (water droplets)				
6	Thunderstorm electricity	#participants	86	85	75	56	40	11.18 (4.56)	9.32 (4.34)	9.56 (4.03)	6.48 (3.33)
		word	negativ (negative)	positiv (positive)	unt (below)	gelad (loaded)	wolk (cloud)				
7	Pre-discharge	#participants	11	9	8	7	7	1.58 (3.34)	1.46 (3.24)	0.79 (1.84)	0.63 (1.61)
		word	negative (negative)	blitzkanal (lightning channel)	positive (positive)	kanal (channel)	leitblitz (lead lightning)				
8	Main discharge	#participants	24	19	15	14	13	3.05 (2.52)	2.62 (2.47)	2.48 (2.02)	1.97 (1.91)
		word	spannung (voltage)	entlad (discharge)	kommt (comes)	blitz (lightning)	entlädt (discharge)				
9	Other aspects	#participants	38	26	21	21	18	5.23 (4.22)	4.06 (3.70)	4.23 (3.35)	2.82 (2.50)
		word	donn (thunder)	luft (air)	dehnt (stretches)	schnell (fast)	erde (ground)				