Computers & Education 160 (2021) 104034

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Learning from text, video, or subtitles: A comparative analysis

Christian Tarchi [a,*], Sonia Zaccoletti [b], Lucia Mason [b]

[a] University of Florence, Italy
[b] University of Padua, Italy

ARTICLE INFO

Keywords:
Reading comprehension
Multimedia learning
Calibration
Instructional video
Instructional text

ABSTRACT

The present study investigated the influence of media (text, video, or subtitled video) on students' learning outcomes. Past studies have raised concerns about the effectiveness of learning from online videos over content-equivalent texts. Moreover, subtitled videos place additional demands on learning. Two-hundred and forty-seven undergraduate students were randomly assigned to a text, video, or subtitled-video condition, in a pretest, posttest, and delayed posttest design. The topic assigned was stem cells. Literal, inferential, and transfer questions were used to assess comprehension and learning outcomes. Results from the study confirmed the substantial equivalence of all conditions in immediate comprehension. Conversely, results confirmed the disadvantage of subtitled videos for deep learning outcomes.

1. Introduction

Nowadays, learners encounter many of the information sources they engage with online. Teachers increasingly assign projects that require students to study a text or a video retrieved online. While research on learning from offline and online texts has deep roots in educational psychology, recently the attention has shifted towards learning from instructional videos (Expósito et al., 2020; List, 2018; Shoufan, 2019). Indeed, since the popularization of video-sharing platforms (e.g., YouTube in 2005) videos have become a major component of leisure time and an integral component of educational environments (Merkt et al., 2011).

Results from past studies have raised concerns about the effectiveness of videos for learning purposes, especially if compared to content-equivalent texts (*e.g.,* van der Molen & Voort, 2000; Wannagat et al., 2017). The comparative analysis of learning from videos and texts calls into question the issue of calibration, that is, the contrast between students' predicted and actual performance (e.g., Alexander, 2013). Today's students perceive themselves as digital natives and may over-judge their competences in learning from digital sources (List, 2018). Although students may express a clear preference for digital texts over printed texts, they generally perform better in comprehension questions when reading in the print condition (Singer & Alexander, 2017). The situation is complicated by the fact that most of the available instructional videos are produced in English, which requires for non-native English speakers to process information in a foreign language and/or rely on subtitles in their first language (i.e., same-language subtitles, SLS).

Notwithstanding the spread of videos as learning medium, the way students process and learn from online videos (with and without subtitles) has been underinvestigated in comparison to more traditional mediums, like text (List, 2018). The present study aimed to contribute to our understanding of learning in different modalities by comparing undergraduates' immediate and delayed

* Corresponding author. Address: c/o Dept. of Education, Languages, Intercultures, Literatures and Psychology (University of Florence), via San Salvi, 12 - 50125 Florence, Italy.
E-mail address: christian.tarchi@unifi.it (C. Tarchi).

https://doi.org/10.1016/j.compedu.2020.104034
Received 3 March 2020; Received in revised form 15 September 2020; Accepted 25 September 2020
Available online 30 September 2020

comprehension of a scientific topic in three different conditions: learning from text on screen, learning from video, and learning from subtitled video.

1.1. Comprehending and learning in different modalities

Research on comprehending and learning from text has a long tradition in educational psychology. Most of these studies are based on Kintsch's foundational model (1998). This model identifies three types (or levels) of memory representations of the text: surface level, textbase level, and situation model. The surface model is a representation of the words included in the text, created through decoding processes. The textbase level is a representation of a network of concepts and propositions corresponding to the meaning of the text. The situation model is a coherent representation of the events described by the whole text, with the integration of prior knowledge to resolve coherence gaps. The situation model is assembled through different types of inference, including knowledge-based inferences, but most importantly is distinguished from the textbase model by its global coherence. Any critical or reflective analysis on the text should be grounded on this situation model.

The most prominent theoretical framework for research on learning not only from text but also from images is Mayer's Cognitive Theory of Multimedia Learning (CTML; Mayer, 2002), which is particularly relevant for the processes involved in learning from videos. Stemming from Paivio's dual-coding theory (1991), the CTML states that learners process multimedia by coordinating dual channels for visual/pictorial and auditory/verbal processing. However, each channel has limited resources to dedicate to processing. Learners should select relevant words from the presented text, select relevant images from the presented visuals, and organize them in coherent verbal/pictorial representations. Finally, learners should integrate the verbal and pictorial representations with each other, and with their prior knowledge. Videos with oral texts (e.g., narrating voice) or with on-screen text (e.g., subtitled videos) involve the processing channels differently. Narrated videos are processed in the auditory-verbal channel, whereas on-screen-texted videos are processed in the visual-verbal channel. Images instead are processed in the visual-pictorial channel (Mayer, 2002).

Subtitled videos differ also from static texts in terms of processes involved. Same-language subtitles (SLS) can be described as a fleeting text on a dynamic background, which places demands on learning that are different from those entailed by static texts on stable backgrounds. Moreover, similarly to what happens when learning from videos, the reader has to manage cognitive resources across different sources of information and adjust the reading pace to the pace with which subtitles appear on the screen. From the one hand, there is an extensive evidence that SLS have the potential to improve students' learning processes (Matthew, 2020), in line with the dual coding theory, which suggests that a combination of visual and verbal information improves information processing (Cuevas & Dawson, 2018; Paivio, 1991). On the other hand, the redundancy effect may hinder learning when students are exposed to information in multiple channels because of the limited capacity of our processing system (Rop et al., 2018). Of notice, subtitled videos differ from closed captioning. Whereas several instructional videos are simply based on translating spoken texts into running text on the low part of the video (see also YouTube automatic subtitling function), closed captioning also translates sound effects, relevant musical cues, and other relevant audio information, besides speech acts. Interestingly, past studies have shown that closed captioning has been effective in supporting English language learners' comprehension processes (*e.g.*, Shahsavandi, 2017). Moreover, on-screen labels (i.e., 'condensed' forms of written text on animations) have found to improve retention accuracy (Rop et al., 2018). Closed captioning seems also to reduce the redundancy effect (Ozdemir et al., 2016).

1.2. Comparative studies on learning across modalities

Research has demonstrated the greater efficacy of videos than texts in disseminating scientific concepts to the lay public (Walthouwer et al., 2015). The higher effectiveness of videos over texts has been shown for several learning-related aspects. As suggested by Mayer's CTML (2002), the acquisition of information through videos is easier than texts as multimedia environments reduce cognitive effort needed to process information. Individuals focus more attention on videos than texts (Alley et al., 2014). Moreover, visual/pictorial imagery generally produces more affective responses as compared to texts (Clark & Paivio, 1991). The study carried out by Yadav et al. (2011) demonstrated that a narrated video and a video + text (the video was juxtaposed to the left of the text narrative, thus it was static and not dynamic as it happens with subtitles) were more powerful in terms of affective responses than a text. Specifically, participants claimed to be more engaged during video watching as it elicited more emotions than texts. Participants found the video as a more realistic and powerful medium in comparison to the text, which was perceived "like reading a newspaper". However, no differences between mediums emerged in terms of cognitive processing. Therefore, participants recalled information equally across modalities. However, it must be noticed that the learning material was based on personal narratives about a scientific topic, rather than on expository texts (Yadav et al., 2011).

Conversely, a few studies seem to suggest that the preference of learners for videos over texts does not transfer to a better comprehension when learning from a video (e.g., Caspi et al., 2005). There is also some evidence supporting the idea that the use of video to communicate science information does not guarantee its comprehension (Mayer et al., 2005; Wilson & Wolf, 2009). Moreover, Merkt et al. (2011) noticed that a common feature of the research designs of these studies was that videos were broadcasted, not making it possible for the participants to control the flow of information. Thus, they suggested that a possible reason why participants underperform in the video condition is that the processing of videos can be controlled by recipients to a much lesser extent compared to texts. When reading, participants could re-read sentences or paragraphs, skip parts of the text, and adjust their reading pace, whereas none of these actions was possible in the video condition. Merkt et al. (2011) tested this hypothesis and confirmed that the effectiveness of interactive videos (i.e., in which viewers could self-regulate the processing of content) was at least comparable to that of reading on paper. The self-pacing hypothesis was also confirmed by another study with a sample of adult learning about the circulatory system:

System-paced study time was perceived as more effortful than self-paced study time (Rop et al., 2018).

So far, we have reviewed research comparing comprehending and learning from video versus text on screen and on paper. The state of the art of research on subtitles is more fragmented. The increase of the availability and frequency of the use of online education videos warrants more research on the topic, especially for non-native English speakers accessing English-produced material. Indeed, the Web Content Accessibility Guidelines 2.0 prescribe subtitles for any media with audio included to ensure a high level of accessibility (World Wide Web Consortium, 2008). Most of the studies on subtitles have focus on exploring the beneficial effects of subtitling on language learning, vocabulary acquisition, and listening comprehension, and found positive effects (e.g., Montero Perez et al., 2013). In contrast, very few studies have investigated the effects of subtitles on content learning (Zee et al., 2017), most of which have focused on L2 subtitles (Liu et al., 2018). Thus, our understanding of the influence of same-language subtitles on content learning as compared to the influence of text or narrated video is extremely limited.

A few indications derive from Mayer's Cognitive Theory of Multimedia Learning (CTML, 2002). In particular, according to the modality principle, people learn better from narrated video rather than from printed text (Mayer & Pilegard, 2014). In Mayer and Pilegard's review (2014), narrated videos were associated with higher learning performances than printed-text videos (median effect size with $d = 0.76$). Ginns (2005) conducted a meta-analysis on 43 independent effects and found that narrated videos lead to higher learning performances than printed texts. The theoretical explanation may be that learners experience an overload in their visual channel when processing graphics and printed words, creating an unwanted cognitive load, especially when the lesson is not self-paced (such as it happens in videos with subtitles). Instead, in narrated videos, words are processed in the auditory channel, freeing resources in the visual channel.

The modality principle, however, may be constrained by certain boundary conditions, for instance when words are unfamiliar or the material is complex for the learner (Schüler et al., 2013). Unfortunately, most of the research available was conducted on L2 subtitles, with a few exceptions. Moreno and Mayer (1999) compared learning performances across three conditions, viewing on-screen text presented near the animation or viewing on-screen text presented far from the animation, or concurrently listening to a narration. Besides confirming the modality principle (higher performances for the narrated video compared to both the conditions with printed text), the study also confirmed the spatial-contiguity principle, according to which students presented with on-screen text close to animations recall more and solve problems better than those presented with on-screen text far from animations. This last result is particularly relevant for the present study, as subtitles are generally included in the bottom section of a video, thus far from the animations. Kalyuga et al. (1999) confirmed the modality effect in learning about science using static documents combining graphics with printed words or narration. It must be noticed, however, that some of these studies (e.g., Kalyuga et al., 1999) presented multimedia material that resembled more a PowerPoint slide than a captioned video. Indeed, videos in which text-heavy slides are narrated may interfere with the reading of subtitles more than narrated videos with animations would. This consideration contributes to question the validity of the modality principle to explain the results, as the visual material required a higher level of split attention than subtitled videos typically would do.

1.3. The current study

To extend current research the aim of the present study was to investigate the effect of modality on students' learning performances, by comparing a narrated video, a subtitled video, and a text-on-screen condition. Of notice, students' performance depends on how comprehension is measured, thus it is important to employ a multi-dimensional assessment (Andreassen & Bråten, 2010). Comprehension and learning from a text are interdependent but not overlapping concepts. One may be able to construct an integrated model of a text immediately after reading it, but may not be able to apply what he or she has just learned to different situations (i.e., transfer), or may not be able to recall it over time. Immediate retention tests with literal and inferential questions can help us to infer how well the material is comprehended, whereas transfer and delayed comprehension questions can help us to tell how well the material is learned. Specifically, the following three research questions (RQ) guided the study:

RQ1: Does the modality of the reading/learning material presentation affect students' literal comprehension, inferential comprehension, and transfer (learning) immediately after watching/reading the material (immediate comprehension and learning)?

RQ2: Does the modality of the reading/learning material presentation also affect students' literal comprehension, inferential comprehension, and transfer (learning) six weeks after (delayed comprehension and learning)?

RQ3: Does the modality of the reading/learning material presentation also affect students' calibration?

Based on prior studies, for both RQ1 and RQ2 we hypothesized that the static and fixed-background characteristics of the text condition would bring an advantage to reading comprehension as compared to the video and subtitles modalities (van der Molen & Voort, 2000). Following the modality principle, we also hypothesized the worst reading performance for the subtitles condition over the other two conditions because of the high processing demands of integrating dynamic images with a dynamic text (Mayer, 2002).

For RQ3, we expected a higher calibration error for the video condition due to young adults' self-perception as digital natives who are able to understand information presented in non-traditional mediums or modalities (Singer & Alexander, 2017).

To ensure equivalence among modalities, we took into consideration the following: i) the content was exactly the same across the three conditions; ii) participants could re-read the text, and pause and/or re-watch the video as many times as they wanted, and linger on parts of the text, or pause the video; iii) the audio track was silenced in the subtitles condition to avoid competition in cognitive resources to be assigned to verbal channel processing, Mayer, 2002); and iv) we included screenshots from the video in the text. In this way, in all three conditions participants were given control over the material and had to combine visual and verbal information. It must be noticed that while the text condition was naturally self-paced, the other two conditions had a certain pace that had to be over-ruled by the viewers.

Moreover, research hypotheses were tested after controlling for potentially confounding effects due to participants' habits (i.e., frequency of learning from online texts and videos), perceived competence (*i.e.*, self-efficacy), topic knowledge, and topic interest. These variables may concur in influencing participants' calibration processes (Alexander, 2013). Indeed, considering ourselves as someone who have a large learning experience in a specific medium, feeling a high degree of confidence in a learning situation, and perceiving ourselves as knowledgeable and interested in a topic are all factors that concur in inducing a person to perceive the learning task as easy and simple, with possibly negative consequences on learning outcomes. As Dunning, Johnson, Ehrlinger, & Kruger (2003) suggested, people use their preconceived beliefs about their skills to estimate how well they will perform a task.

2. Method

2.1. Participants

Participants were undergraduate students from a University located in central Italy. Students were enrolled in an Educational Psychology course included in an undergraduate program in Education, and voluntarily participated in the study. Out of the 281 students who initially expressed interest in participating, 18 students dropped out before the beginning of the study, and 16 began the study but did not complete the learning tasks. Thus, the final sample included 247 participants (231 females, $M_{age} = 20.96$, $SD_{age} = 4.30$). The high number of female participants is representative of the population of students enrolled in a program in Education in Italy. All participants were Italian and spoke Italian as their primary language. The sample was relatively homogeneous (i.e., middle class) regarding socioeconomic status, as assessed through parents' occupation. The study followed all the indications of the Declaration of Helsinki (World Medical Association, 2013) and the guidelines of the University Ethics Committee.

2.2. Procedure

Participants were randomly assigned to a modality condition: text-on-screen (n = 82), video with sound (n = 84), and video with subtitles without sound (n = 81). First, we assessed performances related to control variables. Second, students were assigned the learning material of each condition. Third, immediately after reading/watching the material, students were asked to answer to reading comprehension and learning questions. Finally, six weeks after, students were asked to answer again to the same reading comprehension and learning questions.

2.3. Materials and measures

2.3.1. Learning material

Students were assigned a learning material about the topic of stem cells. The learning material was developed by an educational publisher and was targeted for upper secondary school students. In its original version, the learning material was a video, available at <u>https://www.youtube.com/watch?v=hDw_q_Zm87Y&t=4s. </u>This material introduced the definition of stem cells, their use in medicine, their importance, issues debated, and legislation in Europe. The video was created as an animated slideshow with a narrative voice embedded (see Fig. 1). Of notice, the video had a background music that was kept in the video condition but removed in the other two conditions for technical constraints. Although some studies has shown that music may be beneficial for learning, this may apply to complex material, such as lectures (Dosseville et al., 2012), or anxiety-inducing contexts, such as foreign language learning (Dolean, 2016), mostly by inducing positive emotions in the students. Given the characteristics of the learning material (i.e., short videos or texts) and the task (low-anxiety task) we hypothesized that the background music would not play a confounding effect.

The video was 5 min and 56 s long. In the video condition, students were provided with the original instructional video. In the subtitles condition, we removed the audio-track from the video and included subtitles reproducing the exact content of the original audio-track, in sync with the corresponding slide. The amount of text in each slide was similar to closed-captioned videos (1–2 lines of text, from a minimum of 4 to a maximum of 14 words per caption). In the printed text condition, students were assigned the text transcribed from the original audio track (speech only). Thus, all learners were exposed to the same amount of text. The word count

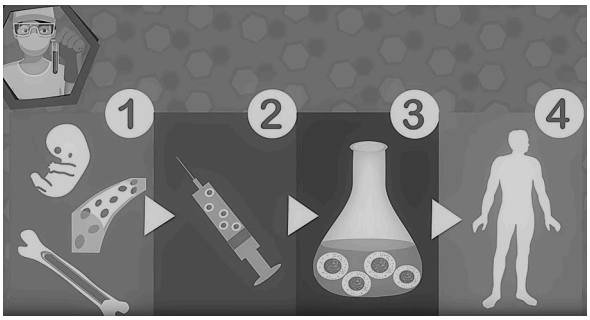

Fig. 1. Snapshot from the instructional video (the original is colored).

Table 1
Transfer Questions in Immediate and Delayed Comprehension and Models of Correct and Complete Answers (information from learning material is underlined).

Transfer questions	Scores	*Models of answers*
1. "Recently, it has been possible to produce in the laboratory certain staminal cells called iPS (induced pluripotent stem cells), starting from cells from the skin or other parts of the organism. In this way, cells are "rewound" from their process of differentiation. Which problems in the research on staminal cells does this discovery allow to solve?"	Correct answer (3 points)	*The discovery allows to bypass the ethical issues of using embryonic staminal cells, and it also avoids the rejection risk, as it provides a considerable quantity of staminal cells from the same patient we want to cure.*
	Partially correct answer (2 points)	*It allows the creation of stem cells capable of transforming into every type of cell in the body, so with this discovery it would be possible to create embryonic cells without taking them from the embryo but from adult subjects.*
	Partially incorrect answer (1 point)	*Induced pluripotent stem cells could be useful for research because they could replace cells that are destroyed due to certain diseases.*
	Incorrect answer (0 points)	*The discovery allows to solve skin-related diseases.*
2. Which typology of cells can provide us with important information on early development?	Correct answer (3 points)	*Embryonic stem cells provide invaluable information on early development. Embryonic cells are like the big bang for the universe, biologist can study these cells to get cues about how a single cell can transform into billions of cells, with different forms and functions.*
	Partially correct answer (2 points)	*In my opinion, the stem cells that can give us important information about early development are embryonic stem cells.*
	Partially incorrect answer (1 point)	*Cells that can give us important information about early development are those from the umbilical cord and bone marrow.*
	Incorrect answer (0 points)	*Early development is linked to the production of hormones that stimulate the function of the ovaries and testicles. The latter can lead to the formation of diseases, if there is already the presence of a person with the same problem in the family (rare genetic diseases), and above all it can cause, going on with growth, serious forms of juvenile discomfort (excessive height, excessive development of female/male organs/ apparatus and excessive bone development).*
3. In biology, cloning is a technique that allows to obtain an individual that is genetically identical to another one. This can be done in two ways. Either in the first phases of division of the embryo, by separating the cells that compose it, each of these can originate genetically identical embryos. Or in the same way as in the case of Dolly the sheep, with the technique of nuclear transfer: the genetic heritage is extracted from the nucleus of a cell of an adult organism and inserted within a cell, from which the nucleus had been removed. What is the relationship between cloning and staminal cells?	Correct answer (3 points)	*To clone adult tissues, it is necessary to start from an embryonic staminal cell. With cloning the differentiating process can start from zero and we can get in vitro totipotent cells, that can be produced in big quantities and are genetically identical to the donor individual. The advantage is that in future we can generate in vitro genetically identical tissues or organs for transplant*
	Partially correct answer (2 points)	*The cloning of healthy stem cells would allow the creation of new tissues to replace diseased ones without the risk of rejection by the body, as they have the same DNA as the original.*
	Partially incorrect answer (1 point)	*Cloning allows the creation of an individual genetically identical to another, while stem cells are intended to regenerate cells from diseased tissue.*
	Incorrect answer (0 points)	*In the cloning of Dolly the sheep, the nucleus was transferred into a stem cell.*
4. In the past, the umbilical cord was thrown away, but today it represents a useful resource of blood stamina cells. Which could be some potential applications of blood staminal cells?	Correct answer (3 points)	*Umbilical blood contains blood staminal cells, that can generate all the other blood cells, including the ones of the immune system. Transplant of blood staminal cells from the umbilical cord can be used to treat several types of blood disease. These transplants should cause less cases of incompatibility of the immune system.*
	Partially correct answer (2 points)	*Blood stem cells could be used to treat diseases of the blood or immune system.*
	Partially incorrect answer (1 point)	*They are useful in the medical field to treat diseases such as leukemia, or diseases of the immune system, but often there are some problems such as the incompatibility of these cells with the patient to be treated, so there are negative aspects such as the regeneration of the disease.*
	Incorrect answer (0 points)	*They can be used to treat genetic diseases.*

across conditions was 808. To maintain equivalences across conditions, in the text condition we also included 9 significant images from the video. The subtitled video and the text are available in the supplementary materials. The time spent watching/reading the learning material by participants was recorded.

2.3.2. Outcome variables

Outcome variables included an immediate comprehension task, a judgment of comprehension task, and a delayed comprehension task.

2.3.2.1. Immediate comprehension. Immediately after reading/watching the material, students were asked 25 multiple-choice questions with four alternatives (19 literal comprehension, $\alpha = 0.70$; and 6 inferential comprehension questions, $\alpha = 0.55$) and four open-ended transfer questions ($\alpha = 0.50$, interrater reliability = 98%). All questions were expressly devised by the authors for the purposes of the study.

Students had no access to the learning material when answering the questions. Literal comprehension questions tapped on information that could be retrieved in the text. An illustrative example is: *"Stem cells are: A. non-specialized primitive cells; B. specialized primitive cells; C. non-specialized differentiated cells; D. specialized differentiated cells."* In the section in which stem cells are defined, the text explicitly states that "Stem cells are non-specialized primitive cells".

Inferential questions tapped on logical consequences of information stated in the text. An illustrative example is: *"The stage in which the stem cells potency is higher is: A. zygote stage; B. embryonic stage; C. adult stage; D. None of these options, the potency level is constant across stages."* The text claims that cells in the zygotic stage are totipotent, that is cells can replicate themselves in all types of cells). Thus, the reader needs to infer that potency is defined as the ability to replicate to all types of cells and attribute this feature to the zygotic stage.

Finally, transfer questions tapped on related issues that are not directly discussed within the learning material, but that can be answered by using some information from the learning material to support a line of reasoning. Each question could be answered by reflecting on two pieces of information included in the text and going beyond the text content. A score from 0 to 3 was assigned to each answer: 0 points were assigned for no answers given or for answers completely based on personal opinion or commonsense; 1 point was assigned for answers based on non-relevant information from the learning material; 2 points were assigned for partially complete answers in which the reasoning was based on only one piece of information from the learning material; and 3 points were assigned to complete answers in which the reasoning was based on both relevant pieces of information from the learning material. See Table 1 for a list of the transfer questions and models of complete answers (pieces of information included in the learning material are underlined).

2.3.2.2. Judgment of comprehension. Students' judgment of comprehension was assessed by asking them how many questions they thought they had answered correctly in the reading comprehension test. The difference between judgment of comprehension and correct answers in the immediate comprehension test was calculated to determine the calibration error.

2.3.2.3. Delayed comprehension. Six weeks after the immediate comprehension questions, participants were asked the same literal ($\alpha = 0.72$), inferential ($\alpha = 0.60$), and transfer questions ($\alpha = 0.55$; interrater reliability = 98%) to determine their reading comprehension performances over time.

2.3.3. Control variables

A number of potentially confounding variables were also measured.

2.3.3.1. Topic knowledge. It was assessed using a 14-item multiple choice test ($\alpha = 0.55$).

2.3.3.2. Topic interest. Students were asked to rate their interest in the topic of stem cells on a 6-point Liker-type scale (one item, 1 = not interested at all, 6 = very interested).

2.3.3.3. Frequency of learning from online texts. Students were asked to self-report their frequency of learning from online texts and

Table 2
Descriptive statistics of control variables and results of preliminary analyses.

Control Variables	Text		Video		Subtitles		*F*	*p*
	M	SD	M	SD	M	SD		
Time on learning task	12.88	6.82	13.14	5.23	14.34	6.15	1.00	.37
Frequency reading online	3.12	1.27	3.23	1.20	3.14	1.05	.20	.82
Frequency video online	3.02	1.33	2.96	1.33	2.97	1.25	.07	.94
Topic interest	3.46	1.48	3.68	1.32	3.60	1.39	.60	.57
Reading self-efficacy	49.57	7.06	51.12	7.72	50.72	7.23	1.08	.34
Video self-efficacy	54.29	6.23	54.26	5.50	54.69	5.27	.16	.85
Topic knowledge	7.12	1.93	6.98	2.16	7.38	1.61	1.06	.35
Judgment of comprehension	14.14	3.73	14.43	3.23	14.01	3.38	.34	.72

videos on a 6-point Likert-type scale (one item, 1 = very rarely, 6 = very frequently).

2.3.3.4. Self-efficacy in learning from texts and videos. It was assessed using two self-report questionnaires (9 items for each scale on a 10-point Likert-type scale, adapted from Anmarkrud and Braten (2009); α = 0.64 for learning from texts and α = 0.60 for learning from videos).

3. Results

3.1. Descriptive statistics and correlations

Although reliability estimates for some measures were certainly rather modest, reliability estimates in the 0.50s can still be considered within the acceptable range for measures developed and used for research purposes (Nunnally, 1978). Reliability estimates in this study are also in line with prior studies on reading comprehension (e.g., Strømsø et al., 2010). Variables were all normally distributed. Descriptive statistics (means and standard deviations) for each condition are reported in Tables 2 and 3. Preliminarily, we conducted a series of ANOVAs on control variables to verify equivalence across conditions. No significant differences emerged as reported in Table 2.

Table 4 reports correlations among the examined variables. Frequency of reading online sources and watching online videos for learning purposes were correlated. These data along with the similarity in the means for these variables suggest that the frequency of use is similar across media. However, frequencies of reading/watching did not significantly correlate with any outcome variable, except frequency of reading online that was associated with immediate transfer. Topic interest and reading self-efficacy correlated with most comprehension measures in both immediate and delayed post-test tasks. Conversely, video self-efficacy was not significantly correlated with any of the outcome variables, except for delayed literal comprehension. A paired-sample comparison of means suggests that students perceived a higher self-efficacy when watching videos than when reading a text-on-screen [$t = -8.00$, $p < .001$, *95% CI* = −4.77; −2.88]. Topic knowledge at pre-test was significantly correlated with all the outcome measures. Calibration error significantly correlated with literal and inferential comprehension in both the immediate and the delayed post-test tasks. Expectedly, most of the outcome measures were correlated.

3.2. RQ1: modality effects on immediate comprehension and learning

A MANOVA was conducted to investigate the effects of modality (condition) on immediate comprehension. Preliminarily, all the assumptions were verified. One case was identified as multivariate outlier after assessing Mahalanobis distances among the participants. Condition was included as the independent variable and measures of immediate comprehension as dependent variables. The model was not statistically significant [Wilks' $\lambda = 0.98$, $F(6, 514) = 1.10$, $p = .36$, $\eta^2_p = .01$]. The substantial equivalence in immediate comprehension outcomes across modalities was confirmed with a series of Bayesian ANOVAs: literal comprehension ($BF_1 = 0.29$), inferential comprehension ($BF_1 = 0.01$), and transfer ($BF_1 = 0.004$). According to Jeffreys (1961) guidelines for interpreting Bayesian Factors, these scores can be considered as a strong (scores between 0.33 and 0.03) or very strong support (scores between 0.03 and 0.001) to the null model. The substantial equivalence across modalities indirectly provides evidence supporting the hypothesis that background music did not play a confounding effect with the material and task used in the present study.

3.3. RQ2: modality effects on delayed comprehension and learning

A MANCOVA was then conducted to investigate the effects of condition on delayed comprehension. Preliminarily, all the assumptions were verified. One case was identified as multivariate outlier after assessing Mahalanobis distances among the participants. Condition was included as the independent variable, scores at the immediate comprehension measures as covariate, and measures of delayed comprehension as dependent variables. The multivariate effect of condition was statistically significant [Wilks' $\lambda = 0.94$, $F(6, 468) = 2.61$, $p = .02$, $\eta^2_p = .03$]. The analysis of univariate tests showed that the main effect of condition was statistically significant on delayed transfer [$F(2, 241) = 3.68$, $p = .03$, $\eta^2_p = .03$], but it was not statistically significant on delayed literal comprehension [$F(2, 241) = 2.56$, $p = .08$, $\eta^2_p = .02$] and delayed inferential comprehension [$F(2, 241) = 1.24$, $p = .29$, $\eta^2_p = .01$] (see Table 5).

Table 3
Descriptive statistics of outcome variables.

Outcome Variables	Text		Video		Subtitles	
	M	SD	M	SD	M	SD
Immediate literal comprehension	14.86	2.74	14.12	3.15	14.22	3.21
Immediate inferential comprehension	4.86	1.11	4.87	1.22	4.74	1.23
Immediate transfer	4.44	2.22	4.66	2.21	4.57	2.54
Calibration error	5.50	3.52	4.63	4.33	4.69	4.05
Delayed literal comprehension	12.49	3.44	13.00	3.49	12.85	3.58
Delayed inferential comprehension	4.16	1.17	4.29	1.22	4.31	1.25
Delayed transfer	4.71	2.23	4.88	2.30	4.11	2.56

Table 4
Correlations among variables.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Frequency reading online	1	.55**	.08	.07	.22**	-.01	.10	-.07	-.01	.17**	-.13*	-.01	-.09	.12
2	Frequency video online		1	.07	-.08	.20**	-.13*	.05	-.05	-.04	.06	-.10	-.09	-.06	.08
3	Topic interest			1	.14*	.26**	.21**	.27**	.18**	.15*	.08	-.04	.22**	.22**	.11
4	Reading self-efficacy				1	.34**	.13*	.27**	.16*	.12	.23**	-.07	.17**	.22**	.19**
5	Video self-efficacy					1	.05	.18**	.08	.02	.09	-.09	.14*	.11	.06
6	Topic knowledge						1	.20**	.16*	.15*	.22**	.004	.20**	.22**	.24**
7	Judgment of comprehension							1	.28**	.12	.17**	-.59**	.26**	.21**	.20**
8	Immediate literal comprehension								1	.08	.17**	.56**	.62**	.26**	.09
9	Immediate inferential comprehension									1	.08	.25**	.14*	.38**	.08
10	Immediate transfer										1	.01	.14*	.12	.68**
11	Calibration error											1	.30**	.13*	-.08
12	Delayed literal comprehension												1	.43**	.16*
13	Delayed inferential comprehension													1	.19**
14	Delayed transfer														1

Note. *p < .05, **p < .01.

Table 5
Results of the MANCOVA: Univariate tests.

	Delayed literal comprehension			Delayed inferential comprehension			Delayed transfer		
	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p
Model	31.08	<.001	.40	12.37	<.001	.21	43.07	<.001	.48
Condition	2.56	.08	.02	1.24	.29	.01	3.68	.03	.03
Immediate literal comprehension	132.06	<.001	.36	15.95	<.001	.06	.60	.44	<.01
Immediate inferential comprehension	5.13	.02	.02	33.15	<.001	.12	.41	.52	<.01
Immediate transfer	.25	.62	.001	.79	.38	.003	201.51	<.001	.46

Table 6
Results of pairwise comparisons.

		Literal			Inferential			Transfer		
(I) condition	(J) condition	(I-J)	*p*	*95%CI*	*(I-J)*	*p*	*95%CI*	*(I-J)*	*p*	*95%CI*
Text	Video	.86	.05	−1.71; −.01	-.16	.36	-.50; .18	.11	.69	-.44; .66
Text	Subtitles	-.84	.05	−1.69; .02	-.27	.12	-.61; .07	.70	.01	.15; 1.26
Video	Subtitles	.03	.95	-.81; .86	-.11	.51	-.44; .22	.59	.03	.05; 1.14

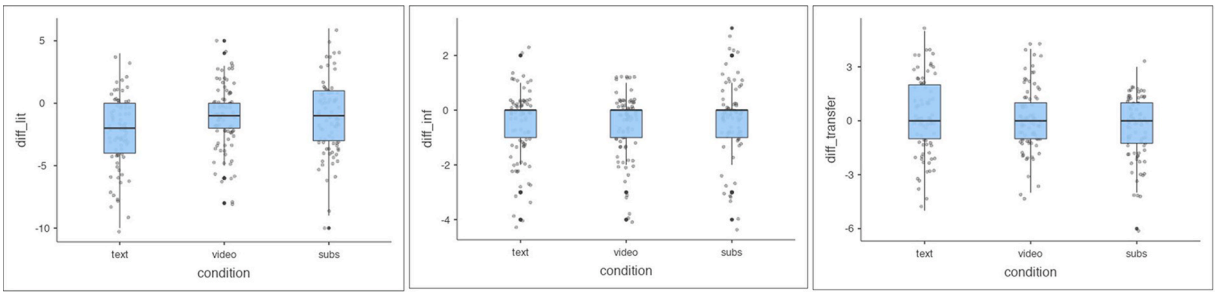

Fig. 2. Graphical representations of data distribution across conditions of the differences in performance between the immediate and delayed comprehension.

The analysis of post-hoc comparisons showed that for the delayed transfer measure, participants in the text condition significantly outperformed those in the subtitles condition (see Table 6 and Fig. 2).

3.4. RQ3: modality effects on calibration performance

An ANOVA was conducted to investigate the effects of condition on calibration error. Preliminarily, all the assumptions were verified. The model was not statistically significant [$F(2, 267) = 1.41$, $p = .25$, $\eta^2_p = .01$]. Thus, condition was not significantly associated with differences in calibration error. The substantial equivalence in the accuracy of the fit between predicted and actual performance was confirmed with a Bayesian ANOVA, which provided a strong support to the null model ($BF_1 = 0.15$).

4. Discussion

The study sought to extend current research on multimedia learning by comparing three conditions differing for modality: narrated video versus subtitled video versus text with graphics. Past studies suggest a higher effectiveness for video on learning-related processes (e.g., reduction of cognitive effort, Mayer, 2002; increased attention, Alley et al., 2014; Koehler et al., 2005; more affective responses, Clark & Paivio, 1991). Conversely, calibration error may be higher when learning from videos than when learning from texts due to young adults' self-perception as digital natives (Singer & Alexander, 2017). Finally, research on learning outcomes seems to support the hypothesis of an advantage for the text condition (Mayer et al., 2005; van der Molen & Voort, 2000; Wilson & Wolf, 2009). Based on prior literature, we hypothesized that: i) the text condition would bring an advantage to reading comprehension as compared to the video and subtitles conditions, and that ii) the subtitles condition would be characterized by a lower performance as compared to the other two conditions. We also hypothesized a higher calibration error for the video condition as compared to the text condition. Results of the study only partially confirmed our hypotheses.

Performance in immediate comprehension outcomes was equivalent across modalities, disconfirming the first hypothesis. This result may depend on the type of learning material used in the present study. Indeed, efforts were made in order to ensure equivalence across modalities for potentially confounding effects due to differences in text, degree of control, competition between auditory and visual channels in decoding speech, and competition between graphic and verbal components of the material. In this way, in all three

conditions participants were given control over the material and had to combine visual and verbal information. The average time spent on the learning material was not statistically different across conditions. This result represents an indirect evidence supporting the equivalence of the learning material across conditions. Differences in each of these components may lead to different results. The same text was used in the three modalities, whereas in previous research often subtitles are translated from the audio version (thus audio is in L2 and subtitles in L1), or subtitles are often phrased with less words than in the narrated version. Finally, the choice of wording and the structure of the discourse may differ from narrated videos to expository texts. In both video and subtitles conditions control was given to the viewers: Participants could pause and play, and watch the video as many times as they wanted, similarly to what happens when reading a text. In this way, we avoided what happened in some previous studies when the learners do not have control over learning videos (if videos are streamlined for instance) which can cause a difference in performances due to non-equivalence across conditions (see Merkt et al., 2011). The audio track was silenced in the subtitles condition, whereas in previous investigations often subtitles were presented in combination with the narration, forcing learners to split their attention (see redundancy effect, e.g., Kalyuga et al., 1999). Finally, we included screenshots from the video in the text, but videos may be more effective in reproducing scientific phenomena through animations, an aspect that has no equivalent in printed texts.

The comparison of performances in the delayed tasks across modalities presented a different scenario and partially confirmed the research hypothesis. In the long-term, the text condition allowed participants to apply what they learned to different scenarios. Video conditions may have focused learners' attention on the surface details rather than inducing a deep elaboration as it also emerged in other investigations. For instance, in Koehler et al.'s study (2005) video versions of informative stories were rated as more engaging than the text versions, but no main effect of modality was found for recall of information. Higher attention or engagement does not necessarily imply depth of elaboration.

Results of the present study confirmed the hypothesized detrimental effect on learning of subtitles. The subtitle condition was characterized by a decrease of performance in transfer over time. This finding suggests that the modality principle can be extended to the comparison between texts and videos with printed text. The automatically-paced video may overload the learners' visual channel, because they have to split their attention between the dynamic text and the graphics (Ginns, 2005).

What concerns calibration differences, students perceived higher self-efficacy in learning from a video rather than learning from text, however this difference did not transfer to their judgment of comprehension, or to their calibration performance. Thus, the results of the study confirmed a general tendency towards overconfidence when learning from videos (Singer & Alexander, 2017), but this overconfidence was reduced when students reported on their comprehension after learning a specific material.

4.1. Educational implications

Although more research is needed to have deeper and more solid knowledge of advantages and disadvantages of learning from digital texts and various types of video, some implications for educational practice can be drawn from the current findings. The first implication regards the "affordances" of the various modalities that can be used to convey complex disciplinary material. The choice of the modality to use should be based on the specific purpose of the learning content (Mayer, 2014). If the purpose is maintained deep learning as reflected in the ability to apply the learned concepts, digital texts appear to be superior to other modalities. Videos, which are more preferred than texts by today's students, are also positively associated with immediate learning tasks.

The second educational implication concerns the use of videos with subtitles as they seem inferior to the other types of video for transfer of learning over time. Yet, educational materials available on the Web often include subtitles as they are originally produced in English, so they need to be accompanied by texts in other languages to be useable by students in countries with a different native language. Students must be made aware by teachers and instructors that they need to put extra effort when they try to learn from dynamic graphics and text because the simultaneous processing of the two types of representation may overload their visual channel.

The third educational implication refers to the issue of students' calibration, which is crucial for the impact that it may have on subsequent decision to restudy the learning material or not. This study confirms an overconfidence in learning from videos but also suggests that this overconfidence may not impair actual learning performances. Thus, this result seems to support the introduction of videos as a learning material in classroom practices. It must be noticed, however, that in this study several precautions were taken in order to maintain equivalence across conditions, and differences between the video and the text condition were minimal. The substantial equivalence in calibration may depend on the fact that both sources require an integration of verbal and visual material. Educators should carefully analyze the learning sources in terms of demands for the cognitive system to control for calibration performances.

Overall, classroom practitioners may use either instructional narrated videos or instructional texts to support deep learning processes on a relatively small amount of content, whereas subtitled videos may overload students' cognitive processing.

4.2. Limitations

When interpreting the findings of the current study, some limitations should be taken into account. Firstly, it must be noticed that, while the text condition is naturally self-paced, the other two conditions had an automatic pace that had to be over-ruled by the viewers. Learners' behavioral interactions with videos may play a moderating role (Merkt et al., 2011), thus future studies should also collect and compare process data while students are learning from text versus video versus subtitled video. Secondly, to ensure equivalence across conditions, we included screenshots from the video in the text, thus we did not collect data about performance when reading texts without graphics, that is, when readers do not have to split their visual attention between words and graphics. Thirdly, subtitles were included at the bottom of the video, as it is usually done in videos in foreign language. However, according to

the spatial-contiguity principle, when on-screen text is presented close to animations, students show higher recall and problem-solving performance (Moreno & Mayer, 1999). Thus, the detrimental effect of subtitled videos may be reduced by varying the position and the quantity of on-screen text. On the one hand, text-heavy videos can be detrimental to students' learning performances (Kalyuga et al., 1999). On the other hand, carefully designed subtitles or closed captions may be effective in supporting learners' performances (see Shahsavandi, 2017).

Fourth, due to technical reasons, in this study the video condition had a background music, whereas the subtitles and text conditions did not have it. Past studies have suggested that background music may have an influence on students' learning performances. For instance, Dosseville et al. (2012) tested the effect of classic music on the learning performances of undergraduate students. According to their results, the students in the experimental group, who watched a 1-h video-lecture with classic music in the background, had a higher learning performance than the students in the control group, who watched the same video-lecture without music. In the present study the video was much shorter, thus results from the study of Dosseville et al. (2012) cannot be directly applied. Moreover, the substantial equivalence of performance across modalities in the immediate comprehension assessment seems to support, at least indirectly, the neutrality of the background music. Nevertheless, future investigation should replicate the research design of this study by either eliminating music from all the conditions, or by adding it to all the conditions.

Finally, following the aptitude-treatment approach, each condition could favour a specific group of students. For instance, the learning styles hypothesis suggests that people have a certain style through which they prefer to process information (e.g., auditory learners may prefer narrated videos). A past study conducted by Cuevas and Dawson (2018) assessed university students' learning styles and prompted them to process material via either imagery or verbal means. According to the results, the condition (visual versus verbal prompt), but not the learning style, had a significant main effect on retention, and no interaction effect was found. Nevertheless, future studies should examine the possible moderating effect of learning styles on learning complex material (texts and videos rather than simple statements) in different conditions, and by taking into account the amplitude of learning style preference rather than classifying students by highest scores.

4.3. Conclusions

Despite the aforementioned limitations, the present study contributes to our understanding of multimedia learning. The pervasiveness of digital media in the lives of young people inside and out of the classroom calls into question the issue of equivalence of the many options through which people can learn new material. Exposure to instructional videos and online texts in both formal and informal education will continue to increase. Results from the study confirm the substantial equivalence of all conditions in immediate comprehension and the substantial equivalence of learning from text and learning from a video (in L1) in delayed comprehension. Conversely, results confirm the disadvantage of subtitled videos for deep learning outcomes (Mayer, 2002), which suggests introducing subtitled material (likely videos in L2 subtitled in L1) in a more conscious and careful manner, drawing from research conducted on EFL learners (Shahsavandi, 2017).

Credit author statement

Christian Tarchi: Conceptualization; Methodology; Validation; Formal analysis; Investigation; Writing - Original Draft; Writing - Review & Editing; Project administration. **Sonia Zaccoletti:** Formal analysis; Writing - Review & Editing; Visualization. **Lucia Mason:** Conceptualization; Methodology; Writing - Review & Editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2020.104034.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction, 24*, 1–3. https://doi.org/10.1016/J.LEARNINSTRUC.2012.10.003

Alley, S., Jennings, C., Persaud, N., Plotnikoff, R. C., Horsley, M., & Vandelanotte, C. (2014). Do personally tailored videos in a web-based physical activity intervention lead to higher attention and recall? An eye-tracking study. *Frontiers in Public Health, 2*. https://doi.org/10.3389/fpubh.2014.00013

Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading, 33*, 263–283. https://doi.org/10.1111/j.1467-9817.2009.01413.x

Anmarkrud, O., & Braten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences, 19*, 252–256. https://doi.org/10.1016/j.lindif.2008.09.002

Caspi, A., Gorsky, P., & Privman, M. (2005). Viewing comprehension: Students' learning preferences and strategies when studying from video. *Instructional Science, 33*, 31–47. https://doi.org/10.1007/s11251-004-2576-x

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*, 149–210. https://doi.org/10.1007/BF01320076

Cuevas, J., & Dawson, B. L. (2018). A test of two alternative cognitive processing models: Learning styles and dual coding. *Theory and Research in Education, 16*, 40–64. https://doi.org/10.1177/1477878517731450

Dolean, D. D. (2016). The effects of teaching songs during foreign language classes on students' foreign language anxiety. *Language Teaching Research, 20*, 638–653. https://doi.org/10.1177/1362168815606151

Dosseville, F., Laborde, S., & Scelles, N. (2012). Music during lectures: Will students learn better? *Learning and Individual Differences, 22*, 258–262. https://doi.org/10.1016/j.lindif.2011.10.004

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87. https://doi.org/10.1111/1467-8721.01235

Expósito, A., Sánchez-Rivas, J., Gómez-Calero, M. P., & Pablo-Romero, M. P. (2020). Examining the use of instructional video clips for teaching macroeconomics. *Computers and Education, 144*, 103709. https://doi.org/10.1016/j.compedu.2019.103709

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331. https://doi.org/10.1016/j.learninstruc.2005.07.001

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371. https://doi.org/10.1002/(SICI)1099-0720 (199908)13:4<351::AID-ACP589>3.0.CO;2-6.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Koehler, M. J., Yadav, A., Phillips, M., & Cavazos-Kottke, S. (2005). What is video good for? Examining how media and story genre interact. *Journal of Educational Multimedia and Hypermedia, 14*, 249–272.

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

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers & Education, 127*, 190–200. https://doi.org/10.1016/j.compedu.2018.08.025

Matthew, G. (2020). The effect of adding same-language subtitles to recorded lectures for non-native, English speakers in e-learning environments. *Research in Learning Technology, 28*. https://doi.org/10.25304/rlt.v28.2340

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*, 256–265. https://doi.org/10.1037/1076-898X.11.4.256

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Signaling, pre-training, and modality. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). Cambridge University Press.

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

van der Molen, J. W., & Voort, T. H. A. (2000). The impact of television, print, and audio on children's recall of the news. *Human Communication Research, 26*, 3–26. https://doi.org/10.1111/j.1468-2958.2000.tb00747.x

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*, 720–739. https://doi.org/10.1016/j.system.2013.07.013

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368. https://doi.org/10.1037/0022-0663.91.2.358

Nunnally, J. C. (1978). *Psychometric theory*. McGraw Hill.

Ozdemir, M., Izmirli, S., & Sahin-Izmirli, O. (2016). The effects of captioning videos on academic achievement and motivation: Reconsideration of redundancy principle in instructional videos. *Journal of Educational Technology & Society, 19*, 1–10.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*, 255–287. https://doi.org/10.1037/h0084295

Rop, G., Schüler, A., Verkoeijen, P. P. J. L., Scheiter, K., & van Gog, T. (2018). The effect of layout and pacing on learning from diagrams with unnecessary text. *Applied Cognitive Psychology, 32*, 610–621. https://doi.org/10.1002/acp.3445

Schüler, A., Scheiter, K., & Gerjets, P. (2013). Is spoken text always better? Investigating the modality and redundancy effect with longer text presentation. *Computers in Human Behavior, 29*, 1590–1601. https://doi.org/10.1016/j.chb.2013.01.047

Shahsavandi, G. H. (2017). Effects of closed-caption programs on EFL learners' listening comprehension and vocabulary learning. *Journal of Studies in Learning and Teaching English, 6*(1), 47–70. http://jslte.iaushiraz.ac.ir/article_532362.html.

Shoufan, A. (2019). Estimating the cognitive value of YouTube's educational videos: A learning analytics approach. *Computers in Human Behavior, 92*, 450–458. https://doi.org/10.1016/j.chb.2018.03.036

Singer, L. M., & Alexander, P. A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*, 155–172. https://doi.org/10.1080/00220973.2016.1143794

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*, 192–204. https://doi.org/10.1016/j.learninstruc.2009.02.001

Walthouwer, L. M. J., Oenema, A., Lechner, L., & De Vries, H. (2015). Comparing a video and text version of a web-based computer-tailored intervention for obesity prevention: A randomized controlled trial. *Journal of Medical Internet Research, 17*. https://doi.org/10.2196/jmir.4083

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing, 18*, 491–504. https://doi.org/10.1007/s10339-017-0820-y

Wilson, E. A. H., & Wolf, M. S. (2009). Working memory and the design of health materials: A cognitive factors perspective. *Patient Education and Counseling, 74*, 318–322. https://doi.org/10.1016/j.pec.2008.11.005

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 27*, 2191–2194. https://doi.org/10.1001/jama.2013.281053

World Wide Web Consortium. (2008). *Web content accessibility guidelines (WCAG) 2.0*. https://www.w3.org/WAI/GL/WCAG20/.

Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education, 23*, 15–37. https://doi.org/10.1007/s12528-011-9042-y

Zee, T. Van Der, Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from online education videos. *Journal of Media Psychology, 29*, 18–30. https://doi.org/10.1027/1864-1105/a000208