Interviewer presence can affect learning with interview videos

Fang Zhao
Faculty of Psychology, University of Hagen, Hagen, Germany

Abstract
Interviews with scholars and experts are becoming more and more popular as e-learning materials. Yet how an interview video should be edited is mostly based on personal preference rather than on rigorous scientific research. Thus this study tested whether showing the interviewer in educational interview videos can affect the learning outcome. Two interview learning materials on two topics (eye tracking and text–picture integration) were conducted by the author and edited in two versions. One version was with the interviewer and the other version was without the interviewer, the latter’s image and voice being edited out. Psychology students (N = 180) watched either the video with or the one without the interviewer and answered the corresponding questions. Results in an online experiment yielded a better learning outcome in the video without the interviewer than in the video with the interviewer. It is probable that the absence of the interviewer can protect participants from extraneous processing and a split-attention effect. The without-interviewer video, segmented by displaying interview questions in keywords on slides, seemed to assist participants in managing the essential processing. The absence of the interviewer may avoid the confusion of multiple instructors, which fosters the generative processing. This study provides practical and pedagogical implications and suggests that removing the image and voice of the interviewer is likely to promote learning.

Keywords
Interview video, multimedia learning, interviewer presence, online learning, cognitive load

Introduction
Watching interview videos is ubiquitous in our society. During the Covid-19 pandemic, scholars in virology are often interviewed to teach people how to avoid being infected.
In the educational context, instructors generally use one of two approaches for presenting interview videos to students. One approach is to present the interviewee together with the interviewer. The other approach is to present only the interviewee. Unfortunately, the decision of whether to present the interviewer is probably made based on personal preference or intuitions rather than on rigorous scientific research. Research is thus needed to determine the extent to which presenting the interviewer or not in interview videos differentially impacts learning.

The present study postulates that the absence of the interviewer in educational interview videos can prevent learners from experiencing extraneous processing and a split-attention effect. Participants viewed two short interview lessons conducted by the author on the topic of eye-tracking and text–picture integration. They viewed the videos either with or without the interviewer and answered the comprehension test. The goal of this research is to address whether interviewer presence can affect the learning outcome in educational interview videos. The results are aimed at providing practical implications for editing interview videos as instructional materials, which has rarely been studied before.

**Interview videos as a subset of instructional videos**

Educational videos are videos whose primary purpose is to educate, and they contain different subcategories. They are widely used in online computer-assisted communication platforms, such as massive open online courses (MOOCs). The need for an effective design of educational videos is currently of high interest, as many universities and schools have to produce educational videos during the Covid-19 pandemic. Educational videos range from instructional videos to documentaries or fiction films (Wijnker et al., 2019). Instructional videos are videos used to introduce new skills, knowledge or behaviour to be learned. They include knowledge clips (such as an explanation of Einstein’s relativity theory), how-to videos (e.g. a demonstration of how to analyse a t-Test by using the software R), and live demonstration (e.g. chemical reactions of hydrogen and oxygen).

Instructional videos can be categorized into classic classroom, weatherman, demonstration, learning glass, pen tablet, talking-head, slides on/off, and interview in terms of the presenting styles (Choe et al., 2019). The classic classroom video involves the instructor standing near a monitor displaying lecture slides and walking between the chalkboard and the monitor. ‘Weatherman’ refers to the instructor standing in front of a monitor. ‘Demonstration’ means the instructor illustrates scientific concepts through orchestrated experiments without lecture slides. ‘Learning glass’ means the instructor uses an LED-illuminated low iron glass as a whiteboard without lecture slides. With the pen tablet the instructor directs the presentation process with an interactive pen tablet by pointing to and drawing on the slides. The talking-head style video has the instructor’s head in a corner on the screen looking directly at the camera and the lecture slides are displayed in full-screen format. Slides on/off style refers to either the instructor or the lecture slides being in full-screen format.

Compared to the above-mentioned styles, interview videos have special set-ups. In an interview, the instructor sits in front of the camera in conversation with an interviewer. Interview videos lack detailed images, animations or lecture slides and do not present the same amount of didactic materials. They are thus inappropriate instructional videos for an entire course or complex course materials (Choe et al., 2019). Nevertheless, interview videos have high learning values due to the interactive quality of the style. They are suitable
as supplemental learning materials in a course, such as a response to frequently asked questions and misconceptions. As interview videos are seldom the focus of research, we reviewed the literature in the field of instructional videos. They are relevant to the current study, as interview videos are a subset of instructional videos and both aim to deliver knowledge.

**Multimedia learning**

Multimedia learning refers to learning from texts and pictures (Mayer, 2014a; Schnotz and Bannert, 2003), which is the case when learning from interview videos. While watching the interviews, students hear the auditory explanation of a research and see the images of the interviewer. Given that working memory only holds limited information for a short period (Atkinson and Shiffrin, 1968; Baddeley, 2000), instructors should design materials in order to meet the demands on cognitive capacity (Mayer, 2014a). The first demand is *extraneous processing* (see also extraneous cognitive load in cognitive load theory in Sweller et al., 2011), which refers to information that is irrelevant or unrelated to the instructional goal. This can occur by adding unrelated information to the learning materials. The second demand is to manage *essential processing*, which refers to processing the necessary and important information. It includes selecting the most relevant information in sensory memory and organizing it into coherent verbal and visual representations. The last cognitive demand is to foster *generative processing*, which is to exert effort to make sense of the material. This involves organizing and integrating the information with prior knowledge.

**Reducing extraneous processing**

Several principles are identified as being applied to reduce extraneous processing, especially the coherence principle, which refers to removing the irrelevant and unrelated information (cf. Mayer, 2014a). A previous study on learning multimedia materials on lightning formation and braking systems showed that the retention and transfer tests were better performed when no music was added to the materials compared to the condition where distracting or background music was added (Moreno and Mayer, 2000). In addition, decorative pictures can lead to extraneous processing. Lenzner et al. (2013) observed no improved learning outcome in a decorative photo of the forest compared to a simulation image to explain how shadows are formed. The subjects’ motivation and mood, however, were enhanced. In Sanchez and Wiley (2006), learners performed significantly better on materials with conceptual images in a topic on the Ice age than on materials with seductive details, such as flowers or mountains. All these studies have suggested a detrimental effect of irrelevant information on the learning outcome. In the current study, the coherence principle can be applied by removing the image and voice of the interviewer, which might diminish the extraneous processing.

The split-attention effect also provides an example of poor instructional design, which triggers extraneous processing. The split-attention effect refers to learning being more effective when the different information sources (e.g. texts and pictures) in the learning materials are integrated rather than spatially separated (Chandler and Sweller, 1992). Various studies have provided evidence that learners outperformed in an integrated condition of texts and pictures compared to a separated condition (e.g. Florax and Ploetzner, 2010; Johnson and Mayer, 2012; Mayer and Moreno, 1998). The underlying mechanism of the split-attention
effect may be due to the high cognitive load triggered by the spatial separation of the information (Paas and Sweller, 2014). Learners thus have to select and process the relevant information for learning.

When the interviewer is visible in an interview, there are two agents (i.e. interviewer and interviewee) on the screen, who communicate with facial and body expressions. Students might shift their attention back and forth from the interviewer to the interviewee. In this condition, the cognitive load can be increased and learners cannot easily integrate resources of information (Bodemer et al., 2004; Chandler and Sweller, 1992; Tarmizi and Sweller, 1988). Eye-tracking studies have documented fewer integrative eye movements when texts and graphs are spatially separated compared to integrated materials (Ballard et al., 1995; Johnson and Mayer, 2012). More specifically, Pouw et al. (2019) varied the spatial distance in pictorial stimuli, text–picture stimuli and authentic learning materials and asked participants to integrate them while concurrently performing a spatial-working memory secondary task. They found that separating pictorial stimuli (e.g. picture of a triangle in green vs picture of a triangle in red) at a large distance degraded the performance of the secondary task but did not affect the integration speed of the separated pictorial stimuli. This indicates the cognitive overload in the visuospatial sketchpad (cf. Baddeley, 1992), when processing spatially separated pictorial stimuli. When participants integrate textual and pictorial stimuli (e.g. a picture of four yellow triangles vs text reading ‘four yellow triangles’), the integration of textual and pictorial stimuli is slowed down but distance does not have a detrimental effect on the performance of the secondary task. This thus suggests no cognitive overload in the visuospatial sketchpad. When participants integrate complex learning materials with texts and pictures (e.g. transmissions and communication in the human nervous system), spatially separated information leads to less integration but enlarging the distance did not worsen the learning outcome. They therefore claimed that learning outcomes do not degrade simply due to the large distance between the information. It is likely that learners in split-attention situations change their learning strategies from a perceptual intensive strategy (i.e. leading to high accuracy and a greater number of saccades) to a memory-intensive strategy (i.e. more prone to errors and fewer saccades) (Ballard et al., 1995; Gray and Fu, 2004). Students can therefore have more difficulty in integrating the information from the interviewer and the interviewee and are prone to make more errors. Hence, the split-attention effect might take place in the with-interviewer condition, which might lead to a poor learning outcome.

**Managing essential processing**

According to the cognitive theory of multimedia learning (Mayer, 2014a; Mayer and Pilegard, 2014), the segmenting principle can assist learners to manage essential processing. This principle refers to learning being more effective with segmented animations than with continuous animations. The use of segmentation can especially benefit students with lower working memory capacity (Lusk et al., 2009). Segmented videos can also enhance the student engagement. Guo et al. (2014) analysed 6.9 million video sessions and suggested that the student engagement drops significantly after watching 6 minutes of the video. The videos are therefore recommended to be segmented into lengths of less than 6 minutes. Accordingly, the segments separated by slides in without-interviewer videos might help learners to concentrate on the content and might orientate them toward the topics involved.
Moreover, presenting words in spoken and written formats can ease the burden of the cognitive processing in one channel. Various cognitive theories (e.g. cognitive theory of multimedia learning by Mayer, 2014a; integrative model of text–picture integration by Schnotz and Bannert, 2003) assume that the integrative processing of texts and pictures takes place in a verbal channel and a pictorial channel. When all the information is given in a spoken form, it can cause a ‘traffic jam’ along the route from the ears to the sensory memory. According to the cognitive models (cf. Mayer, 2014a; Schnotz and Bannert, 2003), verbal processing starts with selectively perceiving texts through the ears in the verbal channel. The constructed phonological loops are formed as the propositional representation by making reference of the objects and events being described in the text to the relevant world knowledge. When the information is displayed in a spoken form as well as a written form – for instance slides with written interview questions – the congestion in the sensory memory through the ears can be diminished, leading to ‘smooth traffic’ in the cognitive processing. The spoken words are perceived in the same way as above. The written words are perceived actively through eyes as images and they are decoded into phonological loops in the pictorial channel, which are organized into the mental model by integrating them with the prior knowledge. During the process, the spoken and written words interact with each other in two channels in terms of mental model construction and mental model adaptation (Schnotz and Wagner, 2018; Zhao et al., 2014, 2020). In the without-interviewer condition, the interviewee expressed their views in the spoken form and the introduction and the questions proposed by the interviewer were in written form. The burden of selectively processing the texts in the spoken form through the ears can thus be alleviated by perceiving the texts in the written form through the eyes. It is therefore conceivable that when a certain amount of information is reached, presenting it in multimodal sensory systems (i.e. eyes or ears) can lead to a more effective learning outcome than presenting information in one mode (cf. Zhao and Mahrt, 2018).

**Fostering generative processing**

According to the cognitive–affective theory of learning with media (Moreno, 2006), motivation can promote learning by fostering generative processing. The personalization principle is aimed at motivating learners to make sense of the materials by adding social cues to the instruction (Clark and Mayer, 2011). Social cues are, for instance, using polite words, humour, personal experience, and conversational words. Personalized materials enable learners to perceive others or feel connected to others and increase learning satisfaction, course completion and learning achievement (Mayer, 2014b; Oyarzun et al., 2018; Richardson et al., 2017; Swan and Shih, 2005). When learning from video lectures, learning is more effective when showing only the instructor’s hand rather than the instructor’s entire body (Fiorella and Mayer, 2016, Experiment 4). It is probable that viewing the instructor’s body can be distracting, as the movements of a human hand is enough to trigger the social cue.

Two agents are involved in the communication in an interview and this can be considered as a teaching condition with multiple instructors. Previous studies have suggested some advantages and disadvantages of multiple instructors who work as a team in a class (Anderson and Speck, 1998; Carpenter et al., 2007). From the learner’s perspective, the benefits are rich in knowledge, perspectives, teaching and assessment styles, as teaching with multiple instructors can bridge valuable information to the class, such as expert thinking, communication and discussion (Jones and Harris, 2012). Multiple instructors are
appropriate when the learning goals are to improve scholarly skills and attitudes and to gain interdisciplinary knowledge. In contrast, the disadvantages are confusion and adjustment caused by the variety of perspectives and expertise. It is thus recommended that the multiple instructors should be well organized so as to avoid confusion and poor communication among the instructors themselves and between the instructors and students. When the interviewer is visible or the voice is heard, the interviewer and the interviewee are like a collaborative teaching team, who interact with each other with different perspectives and experiences. The conversation loop is closed and the different perspectives on a topic might cause confusion. Consequently, students might feel poorly connected and might be confused about the various perspectives. In contrast, when the interviewer is invisible, the conversation loop is open because only the interviewee can be seen and heard. Students might feel more involved in the communication and confusion of different perspectives can be avoided.

**Research question and hypothesis**

Interviews are a subset of instructional videos, which focus on the research and the personal perspective of the interviewee (for an example, see Cope et al., 2005). In the current study, the author interviewed two famous psychologists regarding their opinions on the research and theories in specific domains. Although there are many video lectures from MOOC producers and digital media offices at universities, not enough consideration has been given to the pedagogical affordances of video (Thomson et al., 2014). Educational interview videos have rarely been focused on in previous studies. The following research question and hypothesis are proposed to provide scientific evidence for designing interview videos.

- **Research question:** Can the interviewer’s presence affect the learning outcome in educational interview videos?
- **Hypothesis:** It is assumed that the learning outcome is better in interview videos without the interviewer than with the interviewer.

**Method**

The current study expands on a pilot study (for more details about the pilot study see Appendix).

**Participants**

An a priori power analysis using G*Power 3.1 (Faul et al., 2009) compared the difference of learning outcomes between two topics in two video styles using repeated-measures ANOVA within-between interaction. Results indicated that a sample size of 64 would allow the detection of an effect size of $\eta_p^2 = 0.05$ at $\alpha = 0.05$ with a statistical power $(1 - \beta) = 0.95$. One hundred and eighty undergraduate psychology students (130 females, 48 males and 2 others) from semesters in a range from 1 to 7 completed the study at the University of Hagen (a state-run distance teaching university in Germany). The age of the participants was not recorded due to a technical error. Participants were offered participation credits for compensation. We deleted the trials (the learning performance) that contained less than two correct answers (13.3%) per topic. In total, there were 360 trials (two topics $\times$ 180 participants) and 200 were used.
Materials

There were three phases of video production: planning, filming and editing. The interviewer planned the interview and recorded the interviews. The videos were edited by a professional editor in the video production.

Two interviews. Two famous male psychologists (Dr Kenneth Holmqvist from Sweden and Dr Wolfgang Schnotz from Germany), who have over 30 years’ research experience, were interviewed separately on two topics: eye tracking and text–picture integration. The interviewer is a female research assistant from China. The interview on eye tracking with Dr Holmqvist was conducted on 4th December, 2018 from 10:00 to 14:00 at the University of Regensburg, Regensburg in Germany. In the video, Dr Holmqvist not only answered questions but also showed and explained the functions of five eye trackers. The interview on text–picture integration with Dr Schnotz was performed on 18th February, 2019 from 20:00 to 21:30 at the University Koblenz-Landau, Landau in der Pfalz in Germany. In the video, Dr Schnotz mainly explained his theories and his previous research. The interviews were in English. The videos had not been shown to the students before the experiment.

Excerpts of the eye-tracking video

Interviewer: What is eye tracking?

Dr Holmqvist: Eye tracking is the measurement of the movement of the eyes. We can use it for two things: either to just study how the eyes move, which is good if you want to study neurology, for instance. So, just to know how people move them. But most people are interested in attention. Presumably, it indicates attention. Then you can use it for anything that is visual.

Interviewer: How do you see eye tracking as a research method?

Dr Holmqvist: errr . . . Eye tracking is mostly – or eye tracking could be – a measuring process. It allows us to measure continuously the direction of the eyes. What people are looking at millisecond by millisecond – literally.

Interviewer: What research perspectives in research on multimedia learning could specifically profit from eye tracking?

Dr Holmqvist: Multimedia is by nature very visual. Right? So, it’s often about the combination of texts and graphs and . . . and images. And there are a lot of open questions, as I see it. Take, for instance, that . . . images and graphs seem to benefit low-ability students more than high-ability students. Why is that?

Excerpts of the text–picture integration video

Interviewer: What is the main difference between a text and a picture with respect to cognitive processing?
Dr Schnotz: That’s a good, complicated question. There are many differences between texts and pictures. There are also similarities. But the main differences are that texts consist of symbols and they have symbols that denote relationships in the natural language in the form of verbs and prepositions and things like that. You find nothing like that in pictures. They also have signs but they have no relationship signs. So one kind of representation function is the description of something and the other is a kind of model of something. Or I could possibly simply say that text describes the subject matter and the picture shows the subject matter. They are very different things. This also includes, for example, structural aspects. Texts are not linear things. They are read linearly, usually word by word, but, in fact, they have hierarchical structures. But still there is a constraint of this linear kind of processing – the linear structure of information. And this linear structure is highly constrained. We expect students to read word by word. But in a picture, the situation is different. There are lots of… there is a higher degree of freedom: how you can look at pictures, where you start and where you go then. So, this lower level of constraints…, this higher level of degrees of freedom, allows another kind of processing. So, the texts have simply a specific different function compared to pictures, namely texts have conceptual guidance and construction of knowledge. Whereas pictures are more an external support of the construction of the internal model of the subject matter. Thus, the internal mental model should be considered as a kind of internal analogue presentation. So there are differences in terms of the principle for presentation and in the kind of processing.

Interviewer: What is text–picture integration?

Dr Schnotz: Text–picture integration is just the conjoint processing of text and pictures for the purpose of constructing a coherent presentation of a subject matter. It’s whenever you receive information about some subject matter in a pictorial as well as in a verbal format. And this happens all the time. And people are not expected to construct different kinds of representations in verbal model and the model of the picture or picture-based model. I expected that they construct one model of the subject matter.

Interviewer: Our society now predominantly delivers information by using videos, gifs, stickers and pictures with funny words, etc. on social media, news agencies like the BBC… Do you see any risk of getting information mainly from pictures?

Dr Schnotz: No, I don’t see a risk. I mean, there is generally the danger of manipulating people’s thinking and manipulating their views. But this can be done very well with text. We only have to listen to political speeches. It can also be done with the help of pictures by selecting pictures and by manipulating them and so on. So this is a general risk. This is not a risk specific to pictures or a specific risk of texts. Generally speaking, I don’t think there will be a kind of hostile takeover of texts by pictures so that we have more and more pictures in our communication and less texts, as we do still write and we do communicate verbally: we exchange words not pictures. We talk about pictures but we communicate with words. And as I mentioned before, because language has this special or text, verbal text has its special capability to guide the cognitive processes of your communication partner in a much more constrained way than pictures. We will never give it up. So it’s just impossible to give up verbal communication. So, just because there is such a fundamental difference between depictive and descriptive representations, the one can never live without the other. So, this is why I’m not afraid of any takeover.
Two styles. The interview videos were edited in two styles: with and without the interviewer (see Figures 1 and 2). In the with-interviewer condition, the interviewer was visible. She introduced the interviewee by sitting near him. After the introduction, only the interviewee could be seen, but the voice of the interviewer could be heard. In the without-interviewer condition, the interviewer was invisible. The interviewer’s image and her voice were removed. All the questions were displayed in keywords on slides. The with-interviewer condition is slightly longer than the without-interviewer condition (see Table 1).

Figure 1. The screenshots of the with-interviewer videos: Prof. Dr Kenneth Holmqvist (panel a) was interviewed on eye tracking and em. Prof. Dr Wolfgang Schnottz (panel b) was interviewed on text–picture integration. The person on the right in both videos is the interviewer.

Figure 2. The screenshots of the without-interviewer videos. The image and the voice of the interviewer were removed. The introduction of the interviewees and all the questions were shown on slides with the keywords such as in panel a and panel c. Only the interviewees can be seen, as in panel b and panel d.
Tests. Participants answered 15 comprehension questions in English on each topic and 30 questions in total. The post-study tests were multiple-choice questions, which could contain more than one correct answer. One question for the eye-tracking video was: ‘What is eye tracking? A. To measure where people look; B. Measurements of the movement of the eyes; C. Usage of gaze as a proxy for attention; D. It is used by researchers who are more interested in how people move their eyes rather than in attention.’ The correct answer was ABC, and D was incorrect as most researchers are interested in attention (see the first excerpt of the eye-tracking video). One question for the text–picture integration video was ‘What are the main differences between a text and a picture with regard to the representation form? A. Texts contain relational symbols, like verbs, prepositions; B. Texts are descriptions; C. Pictures contain icons; D. Pictures are depictions.’ The correct answer was ABCD (see the first excerpt of the text–picture integration video). Participants only received one point, when they chose all the correct options. One could receive maximally 30 points and minimally 0 points.

Procedure

The study was conducted online with Unipark (Questback GmbH, 2019). Participants first agreed with the informed consent. Then the aim and the procedure of the study was introduced, and participants were told to understand the content as much as they could. After participants filled in their demographic data, the videos were shown, with the instruction that they should watch the video once and answer questions after watching the video. To counterbalance the sequence effect, participants were randomly assigned to four conditions: (1) eye tracking and text–picture integration with the interviewer; (2) both topics without the interviewer; (3) text–picture integration without the interviewer and eye tracking with the interviewer; (4) text–picture integration with the interviewer and eye tracking without the interviewer. After watching each video once, they received a comprehension test about the corresponding video. There was no time limit and they could not go back to the last screen. The experiment took, on average, around 41 min.

Results

Learning outcomes

The $2 \times 2$ within-between mixed analysis of variance (ANOVA) was performed on the correct rates with topic (eye-tracking vs. text-picture integration) as within-subject factor and with video style (with vs. without interviewer) as between-subject factor. Most interestingly, there was a main effect of video style, $F(1,196) = 9.40, p = .002, \eta^2_p = .05$, and an interaction effect of video style and topic, $F(1, 196) = 4.97, p = .03, \eta^2_p = .03$, indicating a better learning outcome with the video style without the interviewer ($M = 35.2\%, SD = 13.3\%$) than with ($M = 30.2\%, SD = 11.2\%$) and the effect differs across topics (see Figure 3 and Table 2).
No effect of topic was shown, \( F(1,196) = 1.67, \ p = .20, \ \eta^2 = .01 \). Participants answered, on average, 31.2% correct answers (\( SD = 12.3\% \)) on eye tracking and 33.9% (\( SD = 12.6\% \)) on text–picture integration.

**Discussion**

This study examined whether the visibility of the interviewer in educational interview videos can affect the learning outcome. Confirming the hypothesis, the results in the current study suggest that students performed better in videos without the interviewer than in videos with the interviewer. The results are discussed in the framework of the cognitive theories in multimedia learning.

**Coherence principle and Split-Attention effect**

According to the cognitive theory of multimedia learning (Mayer, 2014a) and the cognitive load theory (Sweller et al., 2011), learning instructions should be designed to reduce extraneous processing, which is caused by irrelevant or unrelated information. Instructional design with the coherence principle is suggested to diminish extraneous processing. Removing the irrelevant music (Moreno and Mayer, 2000), or irrelevant pictures
(Lenzner et al., 2013; Sanchez and Wiley, 2006) can lead to a better learning outcome compared to conditions with distracting information. In the current study, the comprehension tests focused only on the contents elaborated by the interviewee. The role of the interviewer was to introduce the interviewee and to raise questions. The visibility of the interviewer is thus irrelevant for the comprehension test. The appearance of the interviewer can be considered as distracting information, which can lead to extraneous processing. When removing the image and voice of the interviewer, participants might be less distracted and might concentrate on the content instead of the unrelated information.

The with-interviewer video may enable learners to split their attention between two visual objects: the interviewee as well as the interviewer. The split-attention effect has not only been found in learning from static texts and pictures (e.g. Johnson and Mayer, 2012), but also in learning from instructional videos with slides and the instructor’s presence (e.g. Colliot and Jamet, 2018). According to Pouw et al. (2019), the split-attention effect is not simply due to the spatial distance of information, but also to the change of the learner’s strategy of information processing. It is likely that learners in the with-interviewer video change their perceptual intensive strategy (i.e. with more integrative eye movements and fewer errors) to a memory-intensive strategy (i.e. with less integrative eye movements but more errors) (Ballard et al., 1995; Gray and Fu, 2004).

In the current study, while watching the videos with the interviewer, two agents are on the screen. They each use different gestures, facial expressions and sometimes different hand gestures to express information. Their voices and pitches are different and they have different accents. As two people are presented in the depictive forms (cf. Schnotz and Bannert, 2003) and textual information is presented in spoken forms, participants might split their visual and auditory attention between the interviewer and the interviewee. They thus could have difficulty in integrating the information from two people and cannot concentrate on the content of the interview.

**Segmenting principle and multimodal processing**

The cognitive theories on multimedia learning have suggested using the segmenting principle as a design instruction to manage essential processing (Mayer and Pileggiard, 2014; Moreno, 2007). Videos are difficult to comprehend due to their quick temporal changes compared to static texts (Lowe, 2003). In the without-interviewer videos, the interviewer’s questions were segmented by the keywords on slides (see Figure 2). For instance, one question on the text–picture integration was ‘What is the main difference between a text and a picture with respect to cognitive processing?’ In the without-interviewer video, ‘Text vs Picture’ was displayed as a keyword on the slide. The entire video was segmented by slides into several small segments with subtopics. For example, the text–picture integration video had subtopics, such as text vs picture, text–picture integration, and risk of picture dominance (see the excerpts of the text–picture integration video). On the one hand, these segments can help participants to be more concentrated, as 6 min videos are more effective in learning than long videos (cf. Guo et al., 2014). On the other hand, the slides can guide participants through the videos in the form of keywords. It is recommended in the design of presentation that keywords rather than long sentences on a slide can make the presentation more effective, as slides with concise language can quickly guide the audience through the topic (Alley, 2003).
Furthermore, the sensory memory can be eased by perceiving information in multimodal channels rather than a single channel. According to the cognitive theory of multimedia learning (Mayer, 2014a) and integrated model of text–picture comprehension (Schnotz and Bannert, 2003), spoken words are selectively perceived through the ears in the verbal channel and written words are selectively perceived through the eyes in the pictorial channel. Therefore, learning can be more effective in the without-interview video when the videos are segmented by topics and the questions are presented in the written and spoken formats.

Multiple instructors

Furthermore, although previous studies suggested that team teaching can be beneficial in providing knowledge of the expertise (Anderson and Speck, 1998; Carpenter et al., 2007), multiple instructors can make students confused due to the variety of perspectives and expertise of the instructors (Jones and Harris, 2012). The with-interviewer video can be considered as a kind of team teaching. The instructors interact with each other and share different perspectives and experiences. The confusion and lack of involvement with students may affect the generative processing, when the interviewer is visible compared to when the interviewer is absent.

Image vs voice of the interviewer

In the current study, the presence of the interviewer involves the image as well as the voice of the interviewer. The image of the interviewer was present at the beginning and end of the videos and the voice of the interviewer was heard throughout the video in the with-interviewer video. Previous studies show that even though learners prefer to see the image of the instructor, adding the image of the instructor leads to higher cognitive load (Homer et al., 2008) and learners only focused on the image of the instructor less than half of the learning time (Colliot and Jamet, 2018). It is thus questionable whether other settings with different degrees of interviewer presence moderate the learning outcome. The effect of the degree of visibility of the interviewer on the learning outcome should be further examined, such as by showing either the image or the voice of the interviewer. Future studies should also compare the conditions where the interviewer’s image is continuously present or only shown at the beginning or end of the interviews. On the one hand, constantly displaying the image of the instructor throughout the instructional video can enhance the social cues and supplement the narrative comprehension by lip-reading or non-verbal communication (Kizilcec et al., 2015). On the other, constantly presenting the image of the instructor may distract the learners’ attention, with a higher number of eye movements between visual sources, and asynchronous processing of the relevant information (Walker et al., 1994).

Motion and video scenery

The interaction effect of the video style and topic in the main experiment indicated that the effect of the presence of the interviewer is stronger in the eye-tracking video than in the text–picture integration video. The findings may be due to the different settings in the videos in terms of motion and background scenes. In the eye-tracking interview, the interviewee used relatively frequent movements to introduce the eye trackers at the end of the interview. He demonstrated how to use the eye trackers and described the advantages and
disadvantages of them, which might enhance the social cues. In addition, the eye-tracking interview had a background of an eye-tracking book and eye trackers (see Figure 1). The text–picture interview was recorded in front of a wall without content-relevant background. A recent study has revealed that learners tend to prefer the content-relevant scenery than only walls, although it neither hampers nor enhances the learning outcome (Merkt et al., 2019). The interaction effect might be due to video settings (e.g. motion, background), which can moderate the effect of the interviewer’s presence.

Limitations and future studies

There are several limitations to the current research. The effect of the gender as interviewer should be investigated, as studies have shown that men in a higher position may behave differently when a younger female who holds a lower position interviews them (e.g. Arendell, 1997; Pini, 2005). It might also be intriguing to test the validity of the effect on individual differences in working memory capacity, which refers to actively selecting relative information and connecting it to the related information in the long-term memory. Working memory capacity plays an essential role in learning, especially when distraction is present (Conway et al., 2001). Previous studies show that participants with lower working memory capacity were more vulnerable to irrelevant information than those with higher working memory capacity (Fenesi et al., 2016). In addition, the participation time in the online experiment was much less than it should be. It is possible that students have selectively paused or skipped some parts of the videos. A reasonable approach to tackle this issue would be to investigate the effect of selectively pausing or skipping the video on learning outcome (Guo et al., 2014). Moreover, one common concern in online learning is that it inhibits students in establishing a sense of belonging with the instructor, peer students and the institution (Haythornthwaite, 2002), as the sense of belonging seems to facilitate deeper learning (Mayer, 2014b). A recent study (Fiorella et al., 2019) has provided evidence for the positive learning effect of eye contact with the instructor by enhancing the social presence of the instructor. Therefore, future study should measure the effect of social presence in interview videos. Another possible area of future research would be to measure the moderation effect of the learner’s cognitive learning style on interviewer presence. Homer et al. (2008) showed that the cognitive load triggered by the image of the instructor can be moderated by the learner’s cognitive learning style. For example, when the instructor’s image is absent, learners with high visual preference have higher cognitive load, whereas learners with low visual preference have higher cognitive load when the instructor’s image is present. Lastly, in the interviews, neither the interviewer nor the interviewees are English native speakers. It might be interesting to examine whether the accent of the interviewer affects the learning outcome.

Conclusion

This empirical study gives the first glimpse of the effect of instructional design on educational interview videos. Tentative results have suggested that students might benefit from observing interview videos when the image and the voice of the interviewer are absent and when interview videos are segmented with keywords on slides. This study provides scientific evidence on how to edit educational interview videos. Further research is needed to investigate how the potential factors moderate the effect of interviewer presence on the learning outcome.
Author’s Note
Fang Zhao is also affiliated with the research cluster D²L² “Digitalization, Diversity and Lifelong Learning – Consequences for Higher Education”, FernUniversität in Hagen, Hagen, Germany.

Acknowledgement
I gratefully acknowledge the support of Kenneth Holmqvist and Wolfgang Schnitzl for participating in the interviews voluntarily. I would like to thank Robert Gaschler for his suggestions on experiment design, Alexander Reinshagen, Matthias Heine-Bohnes and Carina Kötter for organizing the recording devices, editing the videos and their support in performing the interviews.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author(s) received the financial support from the Female Postdoctoral Stipend at FernUniversität in Hagen in Germany for the publication of this article.

ORCID iD
Fang Zhao https://orcid.org/0000-0001-9363-2581

References
Alley M (2003) The Craft of Scientific Presentations. New York, NY: Springer.
Anderson RS and Speck BW (1998) “Oh what a difference a team makes”: Why team teaching makes a difference. Teaching and Teacher Education 14(7): 671–686.
Arendell T (1997) Reflections on the researcher–researched relationship: A woman interviewing men. Qualitative Sociology 20(3): 341–368.
Atkinson RC and Shiffrin RM (1968) Human memory: A proposed system and its control processes. In: Sternberg RJ, Spence KW and Spence JT (eds) The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 2). New York, NY: Academic Press, pp.89–195.
Baddeley AD (1992) Working memory. Science (New York, NY) 255(5044): 556–559.
Baddeley AD (2000) The episodic buffer: A new component of working memory? Trends in Cognitive Sciences 4(11): 417–423.
Ballard DH, Hayhoe MM and Pelz JB (1995) Memory representations in natural tasks. Journal of Cognitive Neuroscience 7(1): 66–80.
Bodemer D, Ploetzner R, Feuerlein I, et al. (2004) The active integration of information during learning with dynamic and interactive visualisations. Learning and Instruction 14(3): 325–341.
Carpenter DMI, Crawford L and Walden R (2007) Testing the efficacy of team teaching. Learning Environments Research 10(1): 53–65.
Chandler P and Sweller J (1992) The split-attention effect as a factor in the design of instruction. British Journal of Educational Psychology 62(2): 233–246.
Choe RC, Scuric Z, Eshkol E, et al. (2019) Student satisfaction and learning outcomes in asynchronous online lecture videos. CBE Life Sciences Education 18(4): 1–14.
Clark RC and Mayer RE (2011) E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. San Francisco, CA: Pfeiffer.
Colliot T and Jamet É (2018) Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research and Development* 66(6): 1415–1433.

Conway ARA, Cowan N and Bunting MF (2001) The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review* 8(2): 331–335.

Cope B, Kalantzis M and Lankshear C (2005) A contemporary project: An interview. *E-Learning and Digital Media* 2(2): 192–207.

Faul F, Erdfelder E, Buchner A, et al. (2009) Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41(4): 1149–1160.

Fenesi B, Kramer E and Kim JA (2016) Split-attention and coherence principles in multimedia instruction can rescue performance for learners with lower working memory capacity. *Applied Cognitive Psychology* 30(5): 691–699.

Fiorella L and Mayer RE (2016) Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology* 108(4): 528–546.

Fiorella L, Stull AT, Kuhlmann S, et al. (2019) Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology* 111(7): 1162–1171.

Florax M and Ploetzner R (2010) What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction* 20(3): 216–224.

Gray WD and Fu W-T (2004) Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science* 28(3): 359–382.

Guo PJ, Kim J and Rubin R (2014) How video production affects student engagement: An empirical study of MOOC videos. In: *The First ACM conference on learning at scale*, Atlanta, GA, USA, 4–5 March, 2014, pp. 41–50. New York, NY: Association for Computing Machinery.

Haythornthwaite C (2002) Building social networks via computer networks: Creating and sustaining distributed learning communities. In: Renninger KA and Shumar W (eds) *Building Virtual Communities: Learning and Change in Cyberspace*. Cambridge: Cambridge University Press, pp.169–190.

Homer BD, Plass JL and Blake L (2008) The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior* 24(3): 786–797.

Johnson CI and Mayer RE (2012) An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology. Applied* 18(2): 178–191.

Jones F and Harris S (2012) Benefits drawbacks using multiple instructors to teach single courses. *College Teaching* 60(4): 132–139.

Kizilcec RF, Bailenson J and Gomez CJ (2015) The instructor’s face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107(3): 724–739.

Lenzner A, Schnotz W and Müller A (2013) The role of decorative pictures in learning. *Instructional Science* 41(5): 811–831.

Lowe RK (2003) Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction* 13(2): 157–176.

Lusk DL, Evans AD, Jeffrey TR, et al. (2009) Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology* 40(4): 636–651.

Mayer RE (2014a) Cognitive theory of multimedia learning. In: Mayer RE (ed.) *The Cambridge Handbook of Multimedia Learning*. 2nd ed. Cambridge, NY: Cambridge University Press, pp.43–71.

Mayer RE (2014b) Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In: Mayer RE (ed.) *The Cambridge Handbook of Multimedia Learning*. Cambridge, NY: Cambridge University Press, pp.345–368.

Mayer RE and Moreno R (1998) A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology* 90(2): 312–320.
Mayer RE and Pilegard C (2014) Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In Mayer RE (ed.) The Cambridge Handbook of Multimedia Learning. Cambridge, NY: Cambridge University Press, pp. 316–344.

Merkt M, Lux S, Hoogerheide V, et al. (2019) A change of scenery: Does the setting of an instructional video affect learning? Journal of Educational Psychology 112(6): 1273–1283.

Moreno R (2006) Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. Journal of Computer Assisted Learning 22(3): 149–158.

Moreno R (2007) Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. Applied Cognitive Psychology 21(6): 765–781.

Moreno R and Mayer RE (2000) A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. Journal of Educational Psychology 92(1): 117–125.

Oyarzun B, Stefaniak J, Bol L, et al. (2018) Effects of learner-to-learner interactions on social presence, achievement and satisfaction. Journal of Computing in Higher Education 30(1): 154–175.

Paas F and Sweller J (2014) Implications of cognitive load theory for multimedia learning. In: Mayer RE (ed.) The Cambridge Handbook of Multimedia Learning. Cambridge, NY: Cambridge University Press, pp.27–42.

Pini B (2005) Interviewing men: Gender and the collection and interpretation of qualitative data. Journal of Sociology 41(2): 201–216.

Pouw W, Rop G, de Koning B, et al. (2019) The cognitive basis for the split-attention effect. Journal of Experimental Psychology: General 148(11): 2058–2075.

Questback GmbH Q (EFS Winter (2019) [Computer software]. Oslo: Questback GmbH.

Richardson JC, Maeda Y, Lv J, et al. (2017) Social presence in relation to students’ satisfaction and learning in the online environment: A meta-analysis. Computers in Human Behavior 71: 402–417.

Sanchez CA and Wiley J (2006) An examination of the seductive details effect in terms of working memory capacity. Memory & Cognition 34(2): 344–355.

Schnotz W and Bannert M (2003) Construction and interference in learning from multiple representations. Learning and Instruction 13(2): 141–156.

Schnotz W and Wagner I (2018) Construction and elaboration of mental models through strategic conjoint processing of text and pictures. Journal of Educational Psychology 110(6): 850–863.

Swan K and Shih LF (2005) On the nature and development of social presence in online course discussions. Online Learning 9(3): 115–136.

Sweller J, Ayres PL and Kalyuga S (2011) Cognitive Load Theory. New York: Springer.

Tarmizi RA and Sweller J (1988) Guidance during mathematical problem solving. Journal of Educational Psychology 80(4): 424–436.

Thomson A, Bridgestock R and Willems C (2014) “Teachers flipping out” beyond the online lecture: Maximising the educational video potential of video. Journal of Learning Design 7(3): 67–78.

Walker JH, Sproull L and Subramani R (1994) Using a human face in an interface. In: SIGCHI conference on human factors in computing systems pp. 85–91. New York, NY: Association for Computing Machinery.

Wijnker W, Bakker A, van Gog T, et al. (2019) Educational videos from a film theory perspective: Relating teacher aims to video characteristics. British Journal of Educational Technology 50(6): 3175–3197.

Zhao F and Mahrt N (2018) Influences of comics expertise and comics types in comics reading. International Journal of Innovation and Research in Educational Sciences 5(2): 2349–5219.

Zhao F, Schnotz W, Wagner I, et al. (2014) Eye-tracking indicators of reading approaches in text–picture comprehension. Frontline Learning Research 2(5): 46–66.

Zhao F, Schnotz W, Wagner I, et al. (2020) Texts and pictures serve different functions in conjoint mental model construction and adaptation. Memory & Cognition 48(1): 69–82.
Author Biography

Fang Zhao is a Junior Research Group Leader at the research cluster D²L² “Digitalization, Diversity and Lifelong Learning – Consequences for Higher Education” of the FernUniversität in Hagen, Germany. Her research interests include comprehension of texts and pictures, learning from videos, interactive learning with simulation and sequence learning under multitasking in experimental and educational psychology.

Interviewer Presence Can Affect Learning with Interview Videos

Appendix

Pilot Study

Method

Participants. Seven psychology students participated in the pilot study and one was excluded due to the low number of correct answers (3.3%). Thus, data from six participants were reported (M_age = 28.7 years, SD = 11.3, 5 females and 1 male). The standard deviation for age was higher than that in many laboratory studies in cognitive psychology as students at our university (a state-run distance teaching university) are older and more heterogeneous in age than students at other universities. All participants had normal or corrected-to-normal vision acuity and hearing ability. The experiment took around 1 hour and 20 min. Participants were offered participation credit for compensation.

Procedure. The study was conducted individually in a lab environment. Participants first signed the written informed consent and were introduced to the aims of the study and encouraged to understand the content as well as they could. The videos were played on a 24-inch screen with either a headphone or a loudspeaker. To counterbalance the sequence effect, participants were randomly assigned to four conditions: (1) eye tracking and text–picture integration with the interviewer; (2) both topics without the interviewer; (3) text–picture integration without the interviewer and eye tracking with the interviewer; (4) text–picture integration with the interviewer and eye tracking without the interviewer. After watching each video once, they received a paper–pencil comprehension test about the corresponding video. There was no time limit for the comprehension test.

Results

Learning outcomes. A 2 × 2 within-between mixed analysis of variance (ANOVA) was performed on the correct answers with topic (eye tracking vs text–picture integration) as within-subject factor and with video style (with vs without interviewer) as between-subject factor. Most interestingly, the effect of video style reached significance, F(1, 8) = 7.47, p = .03, η²_p = .48, and the interaction of topic × video style did not, F(1, 8) = 1.50, p = .26, η²_p = .16. This indicated that participants outperformed in the without-interviewer condition (M = 41.3%, SD = 12.8%) compared to the with-interviewer condition (M = 27.6%, SD = 10.5%), regardless of the topic (see Figure 1). There was no significant effect of topic,
Confirming the hypothesis, the pilot study suggested that, regardless of topics, participants had a better learning outcome in videos without the interviewer compared to in videos with the interviewer. As the information provided by the interviewer is irrelevant for the comprehension test, the appearance of the interviewer can be considered as distractive information, which can lead to extraneous processing (see also Lenzner et al., 2013; Moreno & Mayer, 2000; Sanchez & Willey, 2006). As two agents were on the screen, participants might have had difficulty in integrating the information, leading to the split-attention effect (cf. Chandler & Sweller, 1992). The visibility of the interviewer might not be as effective as the segmented videos with slides, as it was easier for participants to concentrate on the subtopics in multimodal formats.

Table 1 Overview of the correct rates on two topics in two styles

| Exp              | Topics/Video styles          | With interviewer | Without interviewer | Average |
|------------------|------------------------------|------------------|---------------------|---------|
| Pilot study      | Eye tracking                 | 31.1% (11.5%)    | 53.3% (9.4%)        | 37.8% (15.6%) |
|                  | Text–picture integration     | 24.4% (10.1%)    | 33.3% (6.7%)        | 28.9% (9.1%) |
|                  | Average                      | 27.6% (10.5%)    | 41.3% (12.8%)       | 33.3% (13.0%) |

$F(1, 8) = 4.70, p = .06, \eta_p^2 = .37$, indicating no difference in learning outcome on the topic eye tracking ($M = 37.8\%, SD = 15.6\%$) and the topic text–picture integration ($M = 28.9\%, SD = 9.1\%$).