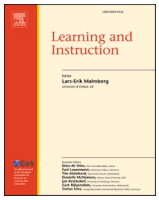
Learning and Instruction 90 (2024) 101863

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc



Learning from academic video with subtitles: When foreign language proficiency matters[1]

Maria Pannatier *, Mireille Bétrancourt

TECFA, Faculty of Psychology and Educational Sciences, University of Geneva, Villa Battelle (Bâtiment A) 7, Route de Drize, 1227, Carouge, Switzerland

ARTICLE INFO

Keywords:
Instructional video
Subtitling
Foreign language comprehension
Redundancy effect
Boundary conditions

ABSTRACT

Background: Previous multimedia research suggests that learning from an academic video in a foreign language may represent a boundary condition for the redundancy principle, such that subtitles assist learning, especially for low-proficiency learners.
Aim: The effects of the subtitle language and the learners' foreign language proficiency level plus any interaction between the two on learning from a subtitled video were examined.
Sample and methods: In an online study, 131 francophone students allocated to three English proficiency levels studied a video lecture under three conditions: subtitles in English (same as audio), subtitles in French or no subtitles (control). They were then asked to provide subjective ratings (cognitive load and interest) and perform comprehension tasks.
Results: Neither a main effect of the condition nor interactions were found on any measure. However, there was a main effect of language proficiency on inference and transfer outcomes as well as on extraneous cognitive load.
Conclusion: The findings did not provide evidence for any effect of subtitles in the same or native language, but confirm the leading role of foreign language proficiency in content learning from video in a foreign language.

1. Introduction

Academic expository video lectures delivered in a foreign language (FL) have become tremendously popular in online higher education as well as for supplementary learning material in formal and informal settings (Belt & Lowenthal, 2021). However, when delivered without any instructional support, they may pose comprehension-related accessibility challenges for students learning in a FL (Roussel et al., 2017). One way to support foreign-language comprehension is to provide concurrent on-screen texts of the audio narration. These on-screen texts are called "same-language subtitles" (SLS) when presented in the original language of the video and "native-language subtitles" (NLS) when translated into the learner's native language (NL). However, multimedia research has failed to reach a consensus regarding the effects of subtitling on learning from videos.

On the one hand, the existing literature on FL learning has identified advantages of SLS for listening comprehension and vocabulary learning (for reviews, see Montero Perez et al., 2013, g = 0.87–0.99; Vanderplank, 2016). On the other hand, multimedia research has found NLS to be detrimental to learning in a NL, e.g., based on the redundancy principle (Kalyuga & Sweller, 2021) and split-attention principle (Ayres & Sweller, 2021), whereas SLS have been shown to aid learning in a FL (Lee & Mayer, 2018). This suggests that learning from video in a FL could be a boundary condition for the modality and redundancy principles. Boundary conditions for multimedia design principles are defined as learning situations that exhibit exceptions to the existing principles and the conditions under which these exceptions apply (Mayer et al., 2020). Moreover, research on whether these effects are dependent on the subtitle language or the learner's FL proficiency is scarce. This study aimed to contribute to the further investigation of the multimedia "subtitles effect" identified by Lee and Mayer (2018) as a boundary condition for the multimedia principles of redundancy and split attention in the context of distance or blended learning in tertiary education. An experimental study with French-speaking adult students with different levels of FL proficiency was therefore conducted to compare their immediate comprehension of a scientific topic at three levels

* Corresponding author.
E-mail addresses: maria.pannatier@etu.unige.ch (M. Pannatier), Mireille.Betrancourt@unige.ch (M. Bétrancourt).

[1] The study received ethics approval from the local ethics committee at the Faculty of Psychology and Educational Sciences of the University of Geneva (decision number: PSE.20200303.03). Informed consent was obtained from each participant before the selection procedure. The data were managed and stored according to the ethical guidelines of the University of Geneva.

https://doi.org/10.1016/j.learninstruc.2023.101863
Received 13 April 2023; Received in revised form 21 November 2023; Accepted 24 November 2023
Available online 9 December 2023

(retention, recall and inferences, transfer) under three different conditions: learning from video with SLS, learning from video with NLS and learning from video without any subtitles.

1.1. Theoretical background

The foremost theoretical framework for research on learning from text and pictures is Mayer's Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2021), based on dual coding theory (Paivio, 2006) and cognitive load theory (Sweller, 2011), which is particularly relevant for the processes involved in learning from videos (Tarchi et al., 2021). The CTML states that learners process multimedia information by coordinating dual channels for visual/pictorial and auditory/verbal processing. In an effective audio-visual production, the auditory and visual channels complement each other - i.e., they are mutually supportive and not conflicting. However, each channel has limited processing resources. Before learning can take place, the learner must decide which elements to attend to and then integrate the components into a coherent representation. Narrated videos are processed in the auditory-verbal channel, whereas on-screen-texted videos are processed in the visual-verbal channel. Images, on the other hand, are processed in the visual-pictorial channel (Mayer, 2002).

In addition, non-native speakers who do not understand the language of the presentation and those learners less proficient in a FL are expected to understand even less. Those who struggle with the language are likely to miss important information along with clues that tell them what to focus on. Learning in a FL may thus pose additional cognitive constraints.

Moreover, on-screen-texted videos differ from narrated videos in that they are processed in the visual-verbal and visual-pictorial channels (if the audio is missing; otherwise, they involve the auditory-verbal channel too) (Tarchi & Mason, 2022). Tarchi et al. (2021) describe subtitles as a fleeting text on a dynamic background, which places learning demands different from those entailed by static texts on stable backgrounds and is more likely to overload learners' cognitive capacity during a video lecture study.

It is well-established that cognitive load is a central consideration in the design of multimedia instruction (Mayer & Moreno, 2003). The information presented to learners should be designed in such a way as to eliminate any avoidable load on working memory (Paas et al., 2003). Moreover, interest theory assumes that the learner's interest drives learning (Hidi & Harackiewicz, 2000). It is therefore essential to measure the impact of on-screen texts on several levels of learning – retention and transfer (Mayer, 2014) – as well as cognitive load and situational interest to determine the relevance of subtitles in multimodal educational design to learning outcomes and subjective experience.

1.1.1. Learning a foreign language from subtitled videos

Much of the prior research exploring the use of on-screen texts in video-based treatments was conducted in the context of FL learning and, in most cases, proved the beneficial effects of subtitles except for low-proficiency learners (Montero Perez et al., 2018; Schmidt, 2001; Vanderplank, 2016). SLS were reported to promote the understanding of video content by providing a simple means of controlling the verbal element of audio-visual material without substantial teacher intervention and offering crucial support for learners in informal and independent settings (Vanderplank, 2016). According to the noticing hypothesis (Schmidt, 2001), when oral narration in a FL is supported by concurrent SLS, learners can allocate attentional resources to unknown words to create the initial form-meaning links, an essential process for FL word acquisition (Montero Perez et al., 2018).

However, research on the effects of NLS remains inconsistent for the FL learning domain, with studies reporting mixed results depending on the task type, FL proficiency level, background knowledge or video type (Mitterer & McQueen, 2009; Vanderplank, 2016). As a result, the extent to which NLS might help learners with low English proficiency is not well understood and requires investigation (Markham et al., 2001; van der Zee et al., 2017).

1.1.2. Learning from subtitled videos in a native language

Multimedia learning research is another area that investigates native language video subtitles. The findings of multimedia studies have provided evidence that NLS displayed concurrently with graphics create a verbal redundancy (Kalyuga & Sweller, 2021; Mayer, 2021). According to cognitive load theory, providing subtitles concurrently with a spoken text is extraneous to learning and may cause working memory overload, thereby reducing the mental resources for essential cognitive processing and, in turn, hampering learning (redundancy effect) (Kalyuga & Sweller, 2021).

In a standard learning situation in which learners study in their NL, adding SLS requires them to split their visual attention between the subtitles and the graphic information. This increases the extraneous load if the different sources of information need to be mentally integrated (split-attention effect) (Ayres & Sweller, 2021). Learners may also miss something in the video, in which information is transient (Tversky et al., 2002), if they spend too much time reading the printed words (Mayer et al., 2014).

Early studies demonstrated the validity of the split-attention effect (Kalyuga et al., 1999; Mayer et al., 2001; Mayer & Moreno, 1998) and showed that students learnt better with material providing co-referring text-picture information close together (integrated) rather than separated on the page. Similarly, learners acquire knowledge more effectively from multimedia instruction containing visual materials and narration than from that containing visual materials, narration and written text (redundancy effect) or visual materials and written text (modality effect).

The existence of a redundancy effect has also been supported by more recent research yielding a negative effect of NLS on learner performance with no differences in learner satisfaction (Ritzhaupt et al., 2015; $\eta^2p = .07$). Similar results have shown SLS to be detrimental to deep-learning outcomes (Tarchi et al., 2021; $\eta^2p = .04$). The existence of a split-attention effect has been evidenced by an eye-tracking study, though has not been found to have any significant influence on learning performance, except for visual memory (Schmidt-Weigand et al., 2010. $\eta^2p = .05$).

1.1.3. Subtitles in learning from foreign language video as boundary condition

Building on evidence that the redundancy effect disappears with longer, more complex texts or for learners with reduced memory capacity, Lee and Mayer (2018) examined the idea that this principle might not apply when learning in a FL because the texts are more challenging to listen to. In line with research on FL learning, they assumed that seeing printed text may be especially important to non-native speakers as a way of minimising extraneous processing and managing essential processing (Lee & Mayer, 2018). In their study, they found that Korean-speaking university students performed better when learning from English (FL) video with narration and printed text (SLS) than from video with narration only, albeit with a small effect size ($d = 0.33$). In addition, the participants reported less difficulty ($d = 1.03$) and less effort ($d = 0.46$) for lessons containing video, narration and text. Lin et al. (2016) also found a reverse split-attention effect for Chinese-speaking undergraduates learning brain anatomy in English. Students learning from a narrated animation with SLS performed better ($\eta^2p = .06$) and expressed lower cognitive load ($\eta^2p = .34$) than their counterparts without SLS. Learning in FL thus seems to be a boundary condition for the redundancy principle, later called the "subtitle effect" (Kalyuga & Sweller, 2021; Mayer et al., 2020).

1.1.4. Same- or native-language subtitles for learning in a foreign language

Overall, research supports the hypothesis that learning in a foreign

language is a boundary condition for the split-attention and redundancy principles. Contrary to their expectations, many authors have failed to provide evidence for a negative effect of SLS when learning from video in a FL, compared to both the no-subtitle condition (Matthew, 2020; Mayer et al., 2014; Tarchi & Mason, 2022; van der Zee et al., 2017; Ozdemir et al., 2016) and NLS (Kruger et al., 2014; Kruger & Steyn, 2014). Nor were any differences between the subtitled and no-subtitle conditions ascertained by the same set of studies for perceived cognitive load or difficulty (Kruger et al., 2014; Kruger & Steyn, 2014; Matthew, 2020; van der Zee et al., 2017), enjoyment and motivation (Mayer et al., 2014; Ozdemir et al., 2016), or perceived value (Tisdell & Loch, 2017) (all differences close to zero).

A recent study by van der Zee et al. (2017) investigating the effects of SLS on learning from educational videos in a FL reported no main effect of SLS on learning (neither beneficial nor detrimental), nor any interactions. They did, however, find a main effect of video visual-textual information complexity, i.e., the lower complexity led to better short-term performance and a lower mental effort rating ($d = 0.62$). FL proficiency also had a main effect, i.e., students with a higher English language proficiency scored substantially higher than students with a lower proficiency, without any interaction ($d = 0.55$). Furthermore, videos with SLS had already been shown to have an impact on performance (Lin et al., 2016, $\eta^2p = .29$) and have since been found to boost deep learning only in highly skilled learners (Tarchi & Mason, 2022; $\eta^2 p = .01 - .05$).

Kruger et al. (2014) also examined the influence of SLS, this time in comparison with NLS and no-subtitle conditions, on immediate comprehension and delayed retention as part of an eye-tracking study. They reported no main effects of subtitling in either language on learning outcomes. However, the participants allocated less attention to the NLS (20%) and, when watching a video with a lecturer and subtitles, non-native speakers spent 43% of the time looking at the SLS (Kruger et al., 2014).

Recent research on "content-and-language integrated learning" claims that delivering content in a FL, especially without explicit instructional support, can be detrimental to both content and language learning. Roussel et al. (2017) conducted a study in which French students had to read a text under one of the three conditions: (1) in French (control condition), (2) in a FL, German or English (without any instructional support), and (3) both in FL (German or English) and with partial translation into French, the mother tongue (that included limited instruction in FL). The results showed that the condition with native language aids provided optimum instructional support for learners with low-intermediate FL proficiency ($\eta^2p = .165 - .273$).

1.2. The present study

The present study aimed to further investigate the potential effects of subtitles in the native language (NLS) or the same language as the video (SLS) on content learning from academic videos in FL. The target population was students in higher education or tertiary-level adults in continuing education, for whom the use of video-based learning resources has become commonplace (Fyfield et al., 2022). The study examined not only vocabulary acquisition or gist comprehension, as in learning a FL, but also retention and knowledge transfer, as in multimedia learning in a NL. The reviewed literature exhibited mixed or even contradictory findings, depending on several factors that were incorporated into this study. First, individual learners' characteristics were considered, particularly their FL proficiency level, since it was reported to be a critical factor for FL learning (Montero Perez et al., 2013; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019). FL proficiency has rarely been assessed and taken into consideration by recent research on multimedia learning in a FL (e.g., Lin et al., 2016; van der Zee et al., 2017; Tarchi and Mason, 2022).

Second, according to cognitive load theory, the design of instruction should limit non-essential cognitive processing (extraneous cognitive load) to free up mental resources for learning. This is especially critical for learning from video since the greater temporal constraint in processing a dynamic presentation could result in a higher extraneous cognitive load than for static graphics processing (Ayres & Paas, 2007). In line with multimedia learning design principles (Fiorella, 2021), adding on-screen texts increases visual processing and could lead to extraneous cognitive load if the auditory information makes it redundant (redundancy effect) or if its integration with other visual information is required (split-attention effect). Cognitive load is therefore measured alongside learning performance (Kruger et al., 2014).

The third issue to be considered for multimedia learning design is motivation. Motivational factors mediate learning by increasing or decreasing cognitive engagement (Moreno & Mayer, 2007). According to Hidi (2006), interest is a critical dimension of motivation in the context of a learning situation. A distinction is typically drawn between individual interest, i.e., schematically, the stable personal interest in a given topic, and situational interest, which is triggered by the learning situation. Situational interest is further composed of two categories, activated and maintained (Hidi, 2006). Since situational interest increases attention and effort, thus enhancing academic success (Harackiewicz et al., 2016), it was also taken into consideration in the present research.

Based on the reviewed literature, the following research questions guided the present study:

RQ1. Can providing subtitles for a video lecture delivered in a FL promote conceptual topic learning? More specifically, does the video format (SLS, NLS or no subtitles) influence learning performances (RQ1a), self-reported cognitive load (RQ1b) and situational interest (RQ1c)?

RQ2. To what extent does FL proficiency influence learning performances (RQ2a), self-reported cognitive load (RQ2b) and situational interest (RQ2c)?

RQ3. Is there an interaction between FL proficiency level and subtitle format regarding learning performances (RQ3a), self-reported cognitive load (RQ3b) and situational interest (RQ3c)?

1.2.1. Hypotheses

Regarding RQ1, given the conflicting results in the existing literature, no confirmatory hypothesis concerning learning outcomes was developed (RQ1a). According to multimedia learning literature, providing on-screen text will increase visual processing and cognitive load (RQ1b), but not necessarily extraneous cognitive load, since learning in a foreign language seems to be a boundary condition for the redundancy and split-attention effect (Kalyuga & Sweller, 2021; Mayer et al., 2020). The effect on situational interest (RQ1c) is likewise exploratory.

Regarding **RQ2**, consistent with previous research (Lin et al., 2016; Mitterer & McQueen, 2009; Muñoz, 2017; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019; Tarchi & Mason, 2022), the authors expected a correlation between higher FL proficiency and higher learning gains (RQ2a). Overall, the authors expected the participants with elementary-level language proficiency to perform worse than the participants with intermediate-level proficiency and the advanced-level participants to score better than the intermediate-level participants. In line with the FL acquisition hypothesis, the higher the level of FL proficiency, the less cognitively demanding the instruction will be perceived to be by the participants since FL proficiency will help them cope with language difficulties (RQ2b). Similarly, the authors expected the activated and maintained interest shown by the participants under all experimental conditions to be greater, the higher their level of FL proficiency (RQ2c).

Regarding RQ3, the authors assumed an interaction between the experimental condition (subtitle format) and the level of FL proficiency

on all measures. First, native language input requires far fewer cognitive resources to operate in working memory (Sweller et al., 1998). NLS would therefore require less cognitive processing than SLS, which are in an unfamiliar language for low-FL-proficiency learners. The NLS allow learners to concentrate more on the other visual content as they can be processed more automatically in peripheral vision (Kruger et al., 2014). Given that low-FL-proficiency learners cannot easily understand the audio narration without an aid, the NLS condition should be more instructionally effective for them than the other two conditions.

According to prior eye-tracking research, learners with an intermediate level of FL proficiency can take advantage of subtitles in a FL (SLS) by reading them (65% of the time) while paying attention to other visual elements necessary for better integration of the presented video content (Winke et al., 2013). The authors thus expected intermediate-level participants to perform better in the SLS condition than in the other two. Finally, the advanced-level participants are likely to benefit more from the "no-subtitle" condition than from the other two. As their level of FL proficiency is close to that of native learners, subtitles could provoke a redundancy effect, such as that observed in learning in a NL, thereby causing increased extraneous cognitive load and an inferior learning performance (e.g., Kalyuga & Sweller, 2021). To sum up, the advanced levels will experience less difficulty in a "no-subtitle" condition, the intermediate levels in a "SLS condition" and the elementary levels in a "NLS condition". Accordingly, the advanced-level participants will express more situational interest in a "no-subtitle" condition, the intermediate levels in a SLS condition and the elementary levels in a NLS condition.

2. Methods

2.1. Participants and design

120 participants with various levels of English proficiency were recruited online via Prolific[2] and then redirected to Qualtrics[3] for further experimental procedures. Participants were considered eligible for the study if they had a student status and had French as their mother tongue or were bilingual in French and any language other than English. There was no pre-selection regarding prior knowledge of the video topic. They were paid 10 GBP online via the Prolific platform. The data were collected over two weeks in October 2020.

Our sample size was justified by the analysis performed in G*power 3.1 on Mac OS (Faul et al., 2007), based on $\alpha = 0.05$, $1 - \beta = 0.85$, and an estimated medium effect size ($f = 0.35$) based on the results of prior research (Lee and Mayer, 2018; Tarchi and Mason, 2022).

12 participants were rejected as they did not watch the video due to a technical problem caused by enabled browser security features. Prolific then recruited more participants until 120 responses were validated. 3 participants were rejected for being bilingual in English. The selection procedure ended with ***131 participants*** (age $M = 22.3$; $SD = 3.58$; women = 63, unwilling to state gender = 1; with 51% bilingual other than in English).

The general design of the experiment was a between-group factorial design. After evaluation of the participants' English proficiency, each of these was assigned to one of the three FL proficiency levels (elementary, intermediate and advanced, see 2.3.2). They were then randomly assigned to one of the three experimental conditions resulting from the format of the video lecture: SLS video, NLS video and no subtitles. This procedure ensured that participants with each level of proficiency were equally distributed across the three conditions.

2.2. Learning materials

To guarantee ecological validity, a genuine YouTube video with a typical recorded-lecture format for online or blended learning was used as a base material. The video is an open-source recorded lecture delivered by Professor Richard Mayer at Harvard University on 5 May 2, 014.[4] The lecture presents Mayer's research-based principles for multimedia learning and the video extract chosen presents the CTML. The language of the video was American English, which was foreign to all participants. The structure of the original video lecture (with the professor's image and the slides present) was retained. The video was segmented in iMovie (Apple Inc.) to present the information in meaningful instructional sequences (theoretical background, schematic presentation of the theory itself and demonstration of a modality principle as a practical example). The final video was 5 min long.

Though the presence of the lecturer's image in the multimedia materials is the subject of debate in existing literature (e.g., Sondermann & Merkt, 2023), the authors decided to retain it – first, to preserve the authentic format, second, because learners report that lectures are easier to follow and learn from when the instructor is present (Wilson et al., 2018), and, third, because multimedia research has shown the presence of a person in the video to promote learning and self-efficacy beliefs through social modelling (van Gog et al., 2014) and social interaction (Clark & Mayer, 2016).

The original video with no subtitles was supplemented by two new subtitled versions: one with SLS in English (750 words) and the other with NLS in French (779 words). The Flesch Reading Ease score was 58.7 according to the English text scale and 59.6 according to the French text scale, i.e., the subtitles were fairly difficult to read for English and standard/average for French.[5]

In both formats, the subtitles were created with iMovie (Apple Inc.) and are shown in the bottom part of the screen where they do not overlap with any other content (Fig. 1). The optimum subtitle font, size, colour, position on the screen, etc. were selected in line with the BBC basic guidelines for subtitles.[6] To generate the subtitle text, the original audio was transcribed word for word for the SLS format and, for the NLS format, translated into French by one of the authors, then proofread by a native French-speaking subject expert.

The materials were uploaded in mp4 format to a cloud-base storage service and then made available to the participants through a shared link. The materials were presented in 960*540 pixels resolution. All interactive features were disabled. The videos started and ended automatically through JavaScript functions in Qualtrics.

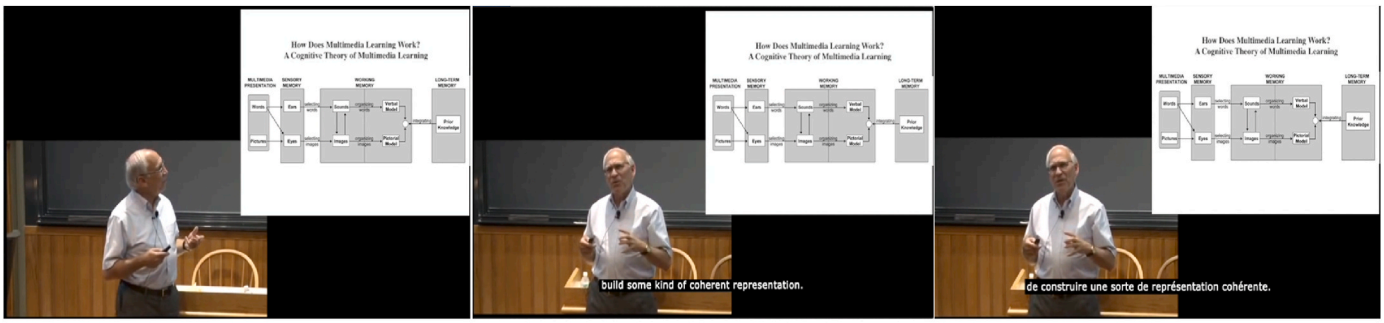
2.3. Measurements and scoring

2.3.1. Demographics, prior knowledge and subtitle experience

The participants' demographic data were collected via a background pre-questionnaire. Each participant was asked questions regarding age, gender, native language, subject and level of study in their native language (French).

In the background pre-questionnaire, participants were also asked to rate their prior knowledge about the topic of the video using a ten-point Likert scale: *please rate your knowledge of the Cognitive Theory of Multimedia Learning developed by an American researcher R. Mayer and his colleagues on a scale from 0 to 10 where "0" indicates no knowledge and "10" very good knowledge*). Subjective evaluation was chosen since the goal was to check whether participants had any knowledge of a specific topic (Mayer's theory of multimedia learning). Moreover, subjective evaluation of prior knowledge had been used previously in multimedia

[2] https://www.prolific.co.
[3] https://www.qualtrics.com.
[4] https://www.youtube.com/watch?v=AJ3wSf-ccXo.
[5] https://www.readabilityformulas.com/free-readability-formula-tests.php.
[6] https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/#Editing-text.



a. No subtitles b. English subtitles (SLS) c. French subtitles (NLS)

Fig. 1. Screenshots of video lectures for the three experimental conditions.

research to eliminate a confounding influence of expertise (e.g., Mayer et al., 2014). Participants who scored 5 or higher (e.g., who self-reported a moderate amount of prior knowledge) were excluded from the study. This led to the automatic elimination of nine participants by Qualtrics and the total sample ($n = 131$) reported their prior knowledge as low ($M = 0.74$; $SD = 1.72$).

As previous experience with watching subtitled videos can positively influence the learning outcomes (Vanderplank, 2016), the participants were asked to rate their personal experience of watching subtitled videos on a ten-point Likert scale *(where "0" indicates no experience with watching subtitled videos and "10" regular experience with watching subtitled videos)*. The level of experience with watching subtitled videos of the total sample ($n = 131$) was above average ($M = 7.55$; $SD = 2.74$).

2.3.2. Foreign language proficiency

The participants were asked to take a short (25-question) English proficiency test.[7] This placement test developed by Cambridge Assessment was successfully used in previous research (Roussel et al., 2017) and rates the participant's English proficiency according to CFRL[8] scales by the number of correct responses: elementary (0–13); – intermediate (14–21); – advanced (22–25). The test was integrated into the online experimental procedure, with immediate feedback on the level displayed on the screen to allow even distribution of the participants with various levels of English proficiency across the experimental conditions.

2.3.3. Learning performance

A post-test in French, the native language of the participants, was developed to assess the level of comprehension of the content presented in the video. In line with standard practice in multimedia research (e.g., Mayer, 2021), a test combining the assessment of retention (remembering facts and definitions from the video) and comprehension (recall, inference and transfer) was developed. The test comprised three parts, each designed to measure a dependent variable (DV).

Part 1. Labelling task: The participants were asked to label the diagram of the CTML theory as shown in the slides presented in the video by dragging and dropping its elements into the correct sections. To assess the task, the original diagram was divided into 14 elements that included both the boxes and the processes linking them. A point was awarded for each correct element (max 14). The task had an acceptable level of reliability with *Cronbach's alpha* = 0.638 given the different formats of assessment.

Part 2. Questionnaire: eight multiple-choice questions and two short-answer questions assessed retention and inferences. For example, the wording of the retention question was as follows: "According to R. Mayer, which of the cognitive processes describes the moment when the comprehension of multimedia information takes place? Your answer in one word: …." An example of an inference question was: "Which cognitive process is involved when a split-attention problem occurs?" with three choices: selection, organisation or both. Each correct answer was allocated one point (max 10).

Part 3. One open transfer question (maximum 10 sentences) assessed the application of the principles explained in the video to a design situation. The question was: "What can be done to improve this multimedia instruction for better learning? This instruction is presented on the computer screen and you can use an animation or sound. Use the arguments from R. Mayer's video to support your ideas".

The response was evaluated using a four-point scale, each point relating to a criterion expressed in the main argument of the video lecture: 1) position of the written text on the screen; 2) amount of text written and spoken; 3) replacing the text by voice-over narration, audio; 4) explaining why, argumentation. All transfer answers were assessed by a first author and then cross-rated by a second author (two times 25% of the answers) until the inter-rater reliability measured by Cohen's Kappa reached the value of 0.87 (almost perfect agreement). It should be noted that 11.5% of the participants chose to answer the transfer question in English instead of French.

2.3.4. Cognitive load and situational interest

Cognitive load was evaluated through a self-reported mental effort questionnaire. The psychometric questionnaire used was the one tested, enlarged and recommended by Leppink et al. (2014) in their Study 2. The same eight items were used for this study, the first four measuring intrinsic load and the last four measuring extraneous cognitive load on a ten-point Likert scale. According to Leppink et al. (2014) germane cognitive load can be reconceptualised so as to refer to the actual working memory resources devoted to dealing with intrinsic cognitive load (Sweller, 2011). The questionnaire was adopted for the present research and translated into French. Internal consistency was excellent, with *Cronbach's alpha* = 0.868 for intrinsic cognitive load and *Cronbach's alpha* = 0.872 for extraneous cognitive load (Appendix A).

Activated situational interest and maintained situational interest were measured with the same ten-point Likert scale used by Desiron (2020). The questionnaire was presented to the participants in their native French language. Items 1–4 were intended to record activated situational interest (max 10 for each item), e.g., "This video on multimedia learning is so interesting that it easily captures my attention", while items 5-8 addressed maintained situational interest, e.g., "During this video viewing I learnt something about multimedia learning that I will use later" (Appendix B).

Both questionnaires had an excellent level of internal consistency with *Cronbach's alpha* = 0.927 for activated and *Cronbach's alpha* = 0.891 for maintained situational interest.

[7] https://www.cambridgeenglish.org/fr/test-your-english/general-english/.

[8] CFRL – Common European Framework of Language References https://rm.coe.int/16802fc1bf.

2.4. Procedure

Fig. 2 summarises the main steps of the experiment. After recruitment in Prolific, the participants accessed Qualtrics, where they were guided through the online experimental procedure. First, their native language and student status were verified by the system for inconsistent screening responses. Participants with native languages other than French were excluded from the study. All participants started by accepting a consent form describing the general procedure and conditions of the experiment, and what was expected of them. It confirmed that all personal data would be confidential and that they could stop at any time.

The study started with a short pre-questionnaire requesting basic demographic information followed by two questions for the participants' self-evaluation of their knowledge of the topic of the video lecture, i.e., the CTML by R. Mayer, and their experience of watching subtitled video. Having received the language test results on the screen, each participant was redirected to a learning phase with one of the three format conditions. After the learning phase, each participant was redirected to the two cognitive load and interest questionnaires, and finally to the three-part post-test. After the responses were registered by the software, a final debriefing message was displayed on the screen with a link to the correct responses to the comprehension test.

The language of the experimental flow was French. Access to the study flow was provided through an anonymous link, with no IP data collected, and no second attempt was allowed. All video interactivity features were disabled, note-taking was not allowed, the video started and ran automatically, and only a single viewing was possible for each participant. The experiment took about 30–45 min to complete.

3. Results

3.1. Data analysis

Most of the quantitative data were automatically recorded by Qualtrics software and saved in several log files for each participant. These log files were imported to MS Excel and processed to extract the necessary data. All statistical analyses were performed using Jamovi.[9] The learning performance and questionnaire data were analysed using ANCOVAs. The Type 1 sum of squares was employed for the significant main effect of ANCOVA. The power and effect sizes are reported as partial ETA squared (η^2p) and Cohen's d interpreted with 0.2 as small, 0.5 as medium and 0.8 as large effect in case of a significant effect. While running ANCOVAs, the homogeneity of variance (Levene's test, $p > .01$) was checked. In case of a significant effect, the pairwise comparisons of the means were checked using the Tukey HSD.

3.2. Participants' characteristics

The descriptive results are reported in Table 1 and Tables C1 (Appendix C).

Preliminary analyses showed that the conditions did not significantly differ in terms of FL proficiency, both overall between the experimental groups (F (2,128) = 0.6, p = .54), and within each language level group: elementary (F (2,18) = 0.1, p = .09); intermediate (F (2,52) = 0.43, p = .65); advanced (F (3,51) = 0.48, p = .07). There were no statistically significant differences in age (F (2,128) = 0.43, p < .647), nor in the prior knowledge of the video topic, both overall between experimental groups (F (2,128) = 0.08, p = .92) and within each language level group: elementary (F (5,15) = 1.34, p = .03); intermediate (F (5,49) = 0.56, p = .72); advanced (F (3,51) = 0.37, p = .77). Nor did the experience with subtitled viewing differ across experimental groups (F (2,128) = 0.75, p = .476) or for each language level group: elementary (F (5,15) = 0.99, p = .45); intermediate (F (5,49) = 0.63, p = .68); advanced (F (3,51) = 0.19, p = .09).

A correlation analysis was carried out involving all control variables, both the performance and subjective variables (Table C2 in Appendix C). Prior knowledge was positively associated only with the score for the labelling task. Experience with subtitles was positively associated with the questionnaire and maintained situational interest scores. Prior knowledge and experience with subtitles were therefore included as covariates in subsequent variance analyses.

3.3. Learning performance

In line with factorial experimental design, ANCOVAs were conducted on the three types of learning scores as dependent variables, with condition and FL proficiency as independent variables, and prior knowledge and experience with subtitles as covariates (Table 2). The ANCOVA models were significant for the questionnaire and transfer scores, though not for the labelling task. There was no significant effect of condition on any of the performance measures. Prior knowledge was a significant covariate for the labelling task, and experience with subtitles for questionnaire performance.

Moreover, the analysis indicated a significant effect of FL proficiency for the transfer question and questionnaire, but not for the labelling task. Post-hoc comparisons revealed significant differences between the performance of the language groups. More specifically, for the transfer task, the advanced-level participants (M = 2.04, SD = 0.99) performed significantly better (p = .016, d = 0.542) than the intermediate-level participants (M = 1.49, SD = 0.94) and the elementary-level participants (M = 1.24, SD = 1.04, p = .019, d = 0.752). For the multiple-choice recall and inference task, there was a significant difference (p = .006) between the advanced-level group (M = 4.98, SD = 1.67; d = 0.8568) and the elementary-level group (M = 3.38, SD = 1.47) as well as between the intermediate-level (M = 4.78, SD = 1.57) and elementary-level (p = .015, d = 0.7575) groups. In both cases, the group with the lower language proficiency level performed significantly worse.

Finally, the analysis revealed no significant interaction between viewing format and FL proficiency on either of the performance measures (Fig. 3). Overall, these results suggest that learning performances in FL are not influenced by the presence of subtitles. Conversely, it is FL proficiency and, to a lesser extent, prior knowledge and experience that contribute to participants' performance.

3.4. Cognitive load and situational interest

As for the learning performances, ANCOVAs were conducted for the two types of cognitive load and situational interest (activated and maintained) as dependent variables, with the three viewing format conditions and the three FL proficiency levels as fixed factors (IV). Experience with subtitles and prior knowledge were entered as covariates (see Table 3).

Regarding cognitive load, there was no significant effect of viewing format on either score. Moreover, there was no significant effect of FL proficiency on intrinsic cognitive load, but a significant effect on extraneous cognitive load was observed. Post-hoc comparisons showed that the higher-proficiency-level participants in all formats evaluated their learning material (M = 1.52, SD = 1.56) as less cognitively demanding (p = .043, d = 0.470) than intermediate-level (M = 2.35, SD = 2.04) and elementary-level participants (M = 2.74, SD = 1.65) (p = .059, d = 0.631). There was no significant interaction between viewing format and FL proficiency level on either cognitive load score.

As for situational interest, there was no significant effect of conditions, FL proficiency or interaction on the activated or maintained situational interest scores.

[9] https://www.jamovi.org.

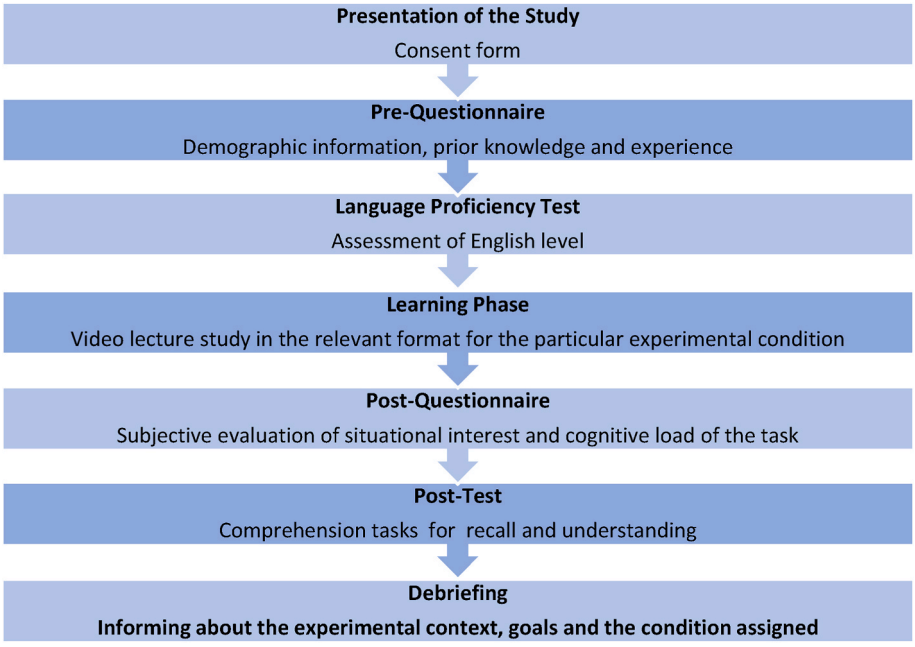


Fig. 2. General experimental procedure.

Table 1
Descriptives for the dependent variables.

	English level	*n*	Video format						All groups (n = 131)	
			SLS* (n = 41)		NLS** (n = 46)		No subtitles (n = 44)			
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Labelling (Max = 14)	Elementary	21	4.71	2.50	5.00	2.76	5.88	3.91	5.24	3.06
	Intermediate	55	6.89	3.74	7.17	3.52	7.28	3.53	7.11	3.54
	Advanced	55	6.53	3.70	7.27	3.01	7.06	3.56	7.00	3.34
Questionnaire (Max = 10)	Elementary	21	3.43	1.51	3.67	1.21	3.13	1.73	3.36	1.47
	Intermediate	55	4.53	1.58	4.83	1.29	5.00	1.85	4.78	1.57
	Advanced	55	4.73	1.53	4.64	1.43	5.61	1.85	4.98	1.64
Transfer question (Max = 4)	Elementary	21	0.857	1.07	1.83	1.17	1.13	0.835	1.24	1.04
	Intermediate	55	1.32	0.885	1.44	0.784	1.72	1.13	1.49*	0.94
	Advanced	55	2.13	0.834	2.05	1.25	1.94	0.802	2.04	0.99
Intrinsic cognitive load (Max 10)	Elementary	21	4.0	2.22	4.88	2.92	4.43	2.17	4.40	2.32
	Intermediate	55	4.22	1.70	3.85	2.58	3.88	2.36	3.99	2.20
	Advanced	55	3.55	1.77	3.58	2.11	2.94	2.07	3.36	2.00
Extraneous cognitive load (Max 10)	Elementary	21	2.50	2.02	2.71	1.65	2.97	1.49	2.74	1.65
	Intermediate	55	2.18	2.11	2.78	2.27	2.08	1.74	2.35	2.04
	Advanced	55	1.38	1.74	1.61	1.58	1.53	1.45	1.52	1.56
Activated situational interest (Max 10)	Elementary	21	4.96	2.62	5.88	2.26	4.97	1.34	5.23	2.04
	Intermediate	55	6.05	2.34	5.39	2.14	5.93	2.47	5.80	2.30
	Advanced	55	6.22	2.30	6.11	2.40	6.67	2.13	6.32	2.26
Maintained situational interest (Max 10)	Elementary	21	5.50	2.62	5.50	2.59	5.88	1.41	5.64	2.04
	Intermediate	55	7.37	2.34	6.40	2.01	7.01	2.29	6.94	2.30
	Advanced	55	7.25	2.30	6.73	2.40	7.38	1.41	7.08	2.26

*SLS - same-language subtitles in English
**NLS - native-language subtitles in French

Table 2
Results of ANCOVAs for learning performance measures.

	Labelling Task			Questionnaire			Transfer Question		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Model	1.4153	0.181	–	2.45917	0.01*	–	1.876	0.055*	–
Format	0.3658	0.694	0.006	1.07757	0.344	0.018	0.688	0.505	0.011
FL proficiency	2.6094	0.078	0.042	5.75984	0.004**	0.114	5.957	0.003**	0.090
Format*FL proficiency	0.1273	0.972	0.004	0.82685	0.511	0.000	1.121	0.350	0.036
Prior knowledge	7.6452	0.007**	0.060	0.00278	0.958	0.000	0.398	0.529	0.003
Experience	0.0486	0.826	0.000	7.60665	0.007**	0.060	0.591	0.444	0.005

*p < .05, **p < .01, ***p < .001.

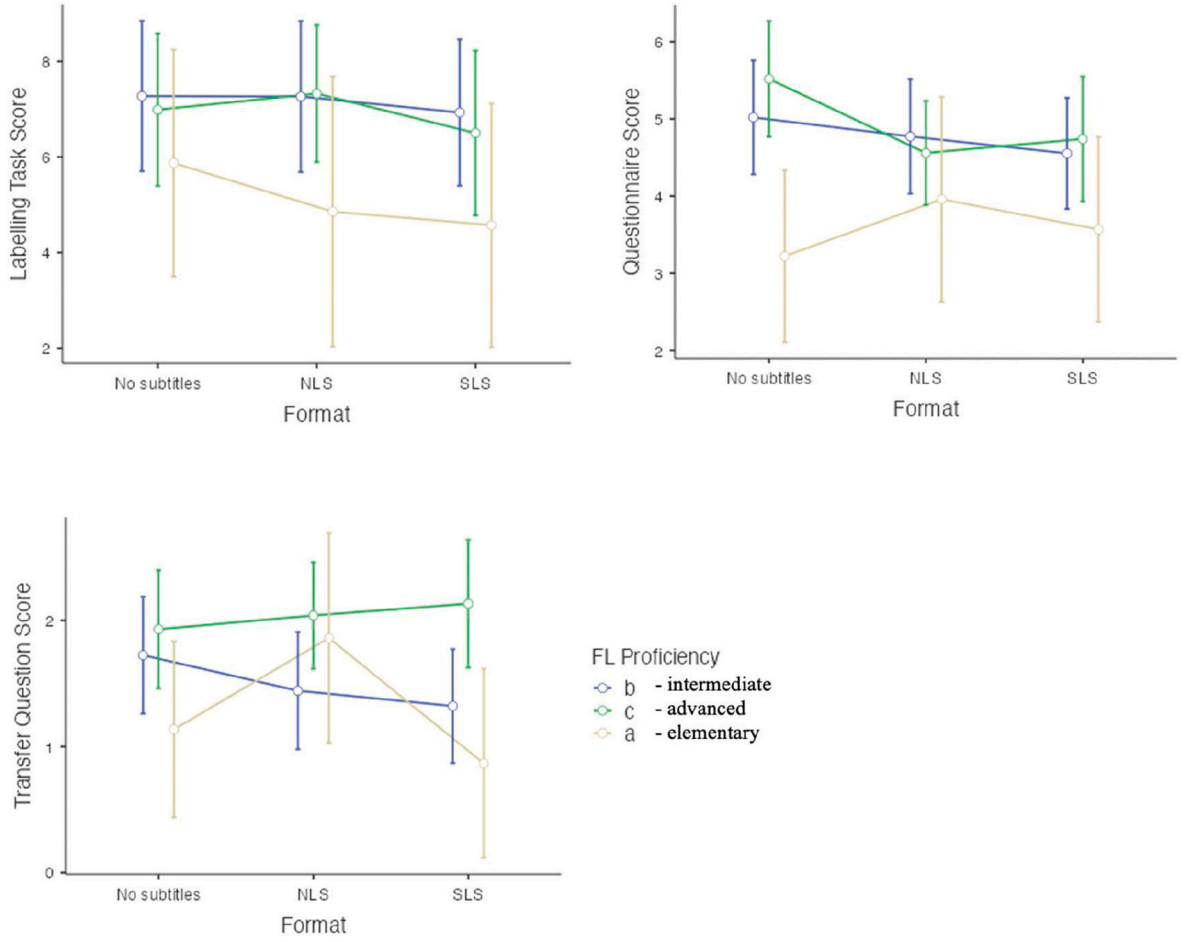


Fig. 3. Interaction Plots for the Performance Measures (not significant).
SLS - same-language subtitles in English
NLS - native-language subtitles in French

Table 3
Results of ANCOVAs for subjective measures.

	Cognitive load						Situational interest					
	Intrinsic			Extraneous			Activated			Maintained		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Model	0.722	0.702	–	1.474	0.157	–	0.659	0.760	–	1.403	0.187	–
Format	0.299	0.742	0.005	0.258	0.773	0.004	0.143	0.867	0.002	0.892	0.412	0.015
FL proficiency	2.039	0.135	0.033	4.392	0.014**	0.068	1.641	0.198	0.027	1.959	0.145	0.032
Format*FL proficiency	0.320	0.864	0.011	0.276	0.893	0.009	0.450	0.772	0.015	0.235	0.918	0.008
Prior Knowledge	0.891	0.347	0.007	2.697	0.103	0.022	0.578	0.448	0.005	5.00e-5	0.994	0.000
Experience	0.378	0.540	0.003	1.636	0.203	0.013	0.646	0.423	0.015	7.384	0.008**	0.058

*p < .05, **p < .01, ***p < .001.

4. Discussion

The study sought to expand current research on multimedia learning by comparing three different viewing format conditions for a video lecture delivered in a FL: video with SLS, video with NLS and no-aid video. Building on prior literature, it investigated the effect of subtitle language on learning as well as on cognitive load and situational interest, and the role played by FL proficiency level.

4.1. The effect of providing subtitles in native or foreign language

Since the existing literature provided inconsistent findings regarding the effect of subtitles when learning in a FL, this study aimed to examine whether learning with a subtitled video in a FL was a boundary condition for the redundancy and split-attention effects – in which case the two effects would not appear in this situation. According to the data analysis, neither of the subtitle conditions aided (or hampered) learning in terms of recall, inferences and transfer. These results are consistent with previous research (Kruger et al., 2014; Mayer et al., 2014; Tarchi & Mason, 2022; van der Zee et al., 2017). It should be noted that, on average, comprehension for the three tasks was generally satisfactory, thus ruling out any ceiling or floor effect (42% on the transfer question, 48% on the labelling task and 46% on the questionnaire). In other words, learners could make sense of the content conveyed in a video in a FL with or without subtitles.

However, our findings contradict two lines of research reviewed in

the introduction, one reporting significantly positive effects of both types of subtitles for FL learning and the other showing significantly negative effects of subtitles for multimedia learning in NL. This suggests that learning from video in a FL could be a boundary condition for the redundancy and split-attention multimedia principles (Mayer et al., 2020). At the same time, no evidence was found of reverse effects whereby subtitles benefited learning, as described in Lee and Mayer (2018). One potential interpretation of the absence of any effect could be the constant presence of the instructor in the video materials, with this acting as a seductive detail and distracting the learner's attention from the relevant visual and verbal content (Wilson et al., 2018). The presence of the speaker, as an additional focal point, could further split attention between the textual and graphic/visual information. In addition, social agency theory states that seeing the instructor is a social cue that can foster a sense of social relationship and improve learning (Mayer, 2014), but can also be distracting and a source of extraneous load (Sondermann & Merkt, 2023). Eye-tracking research shows that a substantial amount of time is allocated to the speaker's face (van Wermeskerken et al., 2018), especially in FL (Kruger et al., 2014), which could vastly reduce the attention paid to subtitles.

Nor did the viewing format groups differ in terms of cognitive load or interest. However, it should be noted that the participants reported low average intrinsic (38%) and extraneous (21%) cognitive load ratings. Our research found no significant difference in extraneous cognitive load, which can be explained by the fact that the learning phase was too short (5 min) to significantly affect either type of cognitive load, as suggested by Leppink et al. (2014). Moreover, the presence of subtitles could also increase the feeling of understanding and thus reduce the level of perceived difficulty. The contradiction between perceived understanding and measured outcomes has been addressed in the existing literature (e.g., Tarchi et al., 2021), which argues for more sophisticated measures of cognitive load in future research.

4.2. The effect of FL proficiency

The main effect of FL language proficiency was confirmed as predicted, i.e., that the highest level of FL proficiency enhances learning outcomes. FL proficiency influenced two learning performance outcomes, the transfer question and the questionnaire assessing retention and inference. It was only marginally significant for the labelling task ($p = .078$). We also found a main effect of FL proficiency on extraneous cognitive load, with advanced-level learners expressing a significantly lower cognitive load than elementary learners, which confirmed our prediction.

The findings support the hypothesis regarding the leading role of FL proficiency in learning content from video instruction in FL, where the more proficient participants will normally perform better ('the rich get richer' effect) and with less perceived difficulty. This is in line with the findings of previous studies that have demonstrated the effect of FL proficiency when learning content from subtitled videos (Lin et al., 2016; van der Zee et al., 2017; Tarchi and Mason, 2022). This result suggests that providing subtitles in either language does not compensate for low FL proficiency when learning from video in FL.

4.3. The interaction between FL proficiency and the subtitle condition

The authors predicted an interaction between the subtitle condition and FL proficiency, with low-proficiency learners benefitting more from NLS, intermediate learners from SLS and advanced learners from the no-subtitle format. No significant interaction was observed between the viewing format and FL proficiency levels for any of the learning outcomes or subjective measures. In the latter case, the participants reported rather moderate scores above the mid-range (between 49 and 74%) on the interest questionnaires. The reported cognitive load tended to be low (25–30%), though was slightly higher (38%) for elementary-level participants in a NLS video condition. This was probably due to a slight redundancy effect of the NLS or the greater attention paid by the participants to the NLS. This should be addressed in a follow-up eye-tracking study.

4.4. Limitations and future directions

First, as mentioned in the discussion, it is possible that the presence of the speaker's face diverts attention from the subtitles. This should be investigated in future research with eye-tracking methodology involving variation of the subtitle type and presence of the speaker's face.

In terms of language proficiency, our research was limited by the number of elementary-level participants ($n = 22$) compared to those with intermediate- and advanced-level proficiency ($n = 55$ for each). This limitation may be due to using a Prolific crowdsourcing platform for participant recruitment. Crowdsourcing platforms have become a popular tool in cognitive science research, especially during the covid pandemics. They provide a larger sample size and allow shorter and easier experiment management than with lab experiments (Stewart et al., 2017). However, as the audience is highly experienced in experiments and languages (especially English), there are few participants with low proficiency levels.

Moreover, the FL proficiency level was evaluated through a short general placement test. As listening comprehension is a critical competence for video viewing and most results indicating the positive effects of SLS are obtained in the FL learning research domain, more precise measurements and results are required here through further research building on Tarchi and Mason, 2022. Furthermore, as the study's focus was on content learning in a FL and not on FL learning, the comprehension test was presented to the participants in their mother tongue, French, while the video was delivered in a FL. Conducting the comprehension test in a FL could thus influence the results and can be examined in future research. Similarly, prior knowledge was self-assessed using a Likert scale because the material dealt with a very specific topic and participants would know if they had encountered it before. However, a pre-test and post-test design would be a more reliable means of assessing learning gains and controlling for guessing strategies.

The languages of the video subtitles in this research were limited to English and French, which are relatively close languages (same alphabet and similar structures). Previous research reports positive results of SLS only for Chinese and Korean native speakers (Lin et al., 2016; Lee and Mayer, 2018) and not for other languages. Yet, Winke et al. (2013) found differences in the processing of subtitles in various non-native languages, e.g., different amounts of time spent on Arabic, Russian, Spanish and Chinese captions by English-speaking learners. Future research should therefore address the issue of reading subtitles in different non-native languages.

This study measured only immediate non-native comprehension. As previous research shows that SLS are beneficial for short-term performance, that NLS could result in better long-term performance (Kruger et al., 2014) and that SLS promote deeper learning (Tarchi et al., 2021), a delayed comprehension measure could be introduced to further investigate how the presence/absence of subtitles affects schema formation from the point of view of instructional design.

An important variable to consider in learning research is the complexity of the content. In the present study, the complexity of the content was assessed only through indicators of text complexity (average in French and fairly difficult in English, according to the Flesch Reading Ease score). As the main effect of complexity has been evidenced in prior research involving subtitles (van der Zee et al., 2017), it would be important for further studies to compare the effects of providing subtitles for low and high content complexity on learning performance and its possible interaction with FL proficiency.

Nor did the present study take into consideration how non-native participants engage with videos and interact with multimedia content. For instance, the participants were not allowed to re-watch or pause videos, take notes, look up unknown words, subsequently use the

subtitle texts or employ any other strategy that might help to offset the lack of FL proficiency. In particular, prior research has shown that providing control over the pace of the video supports learning (Merkt et al., 2011) and also that control can alleviate the split-attention effect when using on-screen text (Tabbers et al., 2004). The way in which interactive features and individual learning strategies influence learning performance could thus be a further subject of investigation.

Finally, no data were collected in this study on how the subtitles were read and how frequently they were processed by the participants. Accordingly, more research is needed on if and how students read and process subtitles in either language. Online methods such as eye-tracking could be used to provide more triangulation of outcome variables, including objective measures of cognitive load and attention distribution (Kruger & Steyn, 2014).

4.5. Implications and conclusion

This experiment has a number of implications for the design of an academic video. In line with content learning in FL research and unlike FL learning research that studies the use of films and entertainment videos, our study considered academic videos. The inferential statistics of the present study indicate that there is neither benefit nor harm in enabling subtitles for these kinds of videos, regardless of the language and level of FL proficiency. In other words, neither SLS nor NLS impaired learning performance and no significant effect on cognitive load was found in any video format group. This research therefore supports the hypothesis that learning content in a FL is a boundary condition for the redundancy and split-attention principles or subtitle effect (Mayer et al., 2020). In this respect, the subtitles cannot be considered harmful and, if designers have the option of enabling subtitles in whatever language, they should probably do it – even though the research provides no definitive answer regarding the choice of language (native or same language as narration). Some learners may prefer subtitles and others not, and subtitles might simply become a user-controlled preference. Moreover, subtitles can be an effective strategy to improve content accessibility, for example for learners with hearing disorders. Conversely, they can be much more challenging for people with reading disorders. The design of subtitles to improve accessibility in case of video in a FL is beyond the scope of the present study, though is certain to be a new avenue worth investigating (e.g., Lervåg, 2020).

Our results also imply that the design of videos with subtitles for FL academic comprehension should be closely geared to the audience's level of FL proficiency. The significant medium effect of FL proficiency and the post-hoc analyses showed advanced learners to perform better than intermediate ($d = 0.5$) and elementary learners ($d = 0.75$) on the transfer questions, which assess the deepest level of learning outcomes. Even for recall and inference tasks, intermediate ($d = 0.76$) and advanced ($d = 0.86$) learners performed significantly better with larger effect sizes than elementary learners. This means that learning content from instructional videos in FL can also be successfully offered to intermediate learners.

CRediT authorship contribution statement

Maria Pannatier: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Project administration, Investigation, Formal analysis, Data curation. **Mireille Bétrancourt:** Writing – review & editing, Conceptualization, Methodology, Project administration, Supervision.

Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2023.101863.

References

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*(6), 695–700. https://doi.org/10.1002/acp.1343

Ayres, P., & Sweller, J. (2021). The split-attention principle in multimedia learning. In L. Fiorella, & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 199–211). Cambridge University Press. https://doi.org/10.1017/9781108894333.020.

Belt, E. S., & Lowenthal, P. R. (2021). Video use in online and blended courses: A qualitative synthesis. *Distance Education, 42*(3), 410–440. https://doi.org/10.1080/01587919.2021.1954882

Clark & Mayer. (2016). Applying the personalization and embodiment principles: Use conversational style, polite wording, human voice, and virtual coaches. In *E-learning and the science of instruction* (pp. 179–200). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119239086.ch9.

Desiron, J. (2020). *Designing multimedia documents for struggling readers: Effects of text cohesion with static or animated depictions*. University of Geneva. https://doi.org/10.13097/archive-ouverte/unige:142063

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiorella, L. (2021). Multimedia learning with instructional video. In R. E. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 487–497). Cambridge University Press. https://doi.org/10.1017/9781108894333.050.

Fyfield, M., Henderson, M., & Phillips, M. (2022). Improving instructional video design: A systematic review. *Australasian Journal of Educational Technology, 38*(3), 155–183.

van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323–327. https://doi.org/10.1016/j.compedu.2013.12.004

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 220–227. https://doi.org/10.1177/2372732216655542

Hidi, S. (2006). Interest: A unique motivational variable. *Review of Educational Research, 1*(2), 69–82. https://doi.org/10.1016/j.edurev.2006.09.001

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 10*(2), 151–179.

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

Kalyuga, S., & Sweller, J. (2021). The redundancy principle in multimedia learning. In R. E. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 212–220). Cambridge University Press. https://doi.org/10.1017/9781108894333.021.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5). https://doi.org/10.16910/jemr.7.5.4. Article 5.

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. https://doi.org/10.1002/rrq.59

Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology, 32*(5), 648–654. https://doi.org/10.1002/acp.3436

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Lervåg, I. K. (2020). Role of subtitles in L2 acquisition and comprehension: Evidence from hearing-impaired learners [Master thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2781868.

Lin, J. J. H., Lee, Y. H., Wang, D. Y., & Lin, S. S. J. (2016). Reading subtitles and taking enotes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Educational Technology & Society, 19*(4), 47–58.

Markham, P. L., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. Target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439–445. https://doi.org/10.1111/j.1944-9720.2001.tb02083.x

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*(0). 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. Academic Press.

Mayer, R. E. (2014). In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning/* (2nd ed..). New York, N.Y.: Cambridge University Press.

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008.

Mayer, R. E., Fiorella, L., & Stull, A. (2020). *Five ways to increase the effectiveness of instructional video*. Educational Technology Research and Development. https://doi.org/10.1007/s11423-020-09749-6

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective: Multimedia learning in a second language. *Applied Cognitive Psychology, 28*(5), 653–660. https://doi.org/10.1002/acp.3050

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 9. https://doi.org/10.1037/0022-0663.90.2.312

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

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*(6), 687–704. https://doi.org/10.1016/j.learninstruc.2011.03.004

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One, 4*(11), Article e7785. https://doi.org/10.1371/journal.pone.0007785

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review, 19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System, 67*, 77–86. https://doi.org/10.1016/j.system.2017.04.015

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. *Educational Technology & Society, 19*(4), 1–10.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical approach*. L. Erlbaum Associates.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *Language Learning Journal, 47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Ritzhaupt, A. D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior, 45*, 222–227. https://doi.org/10.1016/j.chb.2014.12.020

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction, 52*, 69–79. https://doi.org/10.1016/j.learninstruc.2017.04.007

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.003.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100–110. https://doi.org/10.1016/j.learninstruc.2009.02.011

Sondermann, C., & Merkt, M. (2023). Like it or learn from it: Effects of talking heads in educational videos. *Computers & Education, 193*, Article 104675. https://doi.org/10.1016/j.compedu.2022.104675

Stewart, N., Chadler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences, 20*(10), 736–748. https://10.1016/j.tics.2017.06.007.

Suárez, M. del M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *Language Learning Journal, 47*(4), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Sweller, J. (2011). Chapter two - cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

Sweller, J., Van Merrinboer, & Pass. (1998). Cognitive architecture and instructional design. *Educational Psycology Review, 10*(3), 46.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71–81.

Tarchi, C., & Mason, L. (2022). Learning across media in a second language. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-022-00652-7

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education, 160*, Article 104034. https://doi.org/10.1016/j.compedu.2020.104034

Tisdell, C., & Loch, B. (2017). How useful are closed captions for learning mathematics via online video? *International Journal of Mathematical Education in Science & Technology, 48*(2), 229–243. https://doi.org/10.1080/0020739X.2016.1238518

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-50045-8

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education, 122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal, 97*(1), 254–275. https://doi.org/10.1111/j.1540-4781.2013.01432.x

van der Zee, T., 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: Theories, Methods, and Applications, 29*(1), 18–30. https://doi.org/10.1027/1864-1105/a000208

Maria Pannatier obtained her PhD in Education at TECFA, the Faculty of Psychology and Education, Geneva University, Switzerland. Her doctorate research focused on multimedia learning from online video lectures. She is also interested in instructional design for open, distance and e-learning, as well as technology-enhanced language teaching and learning.

Mireille Bétrancourt is a full Professor in Information Technology and Learning Processes at the Faculty of Psychology and Education, Geneva University, Switzerland. She is the head of TECFA, a research and teaching unit in the field of learning and teaching technology in academic education, adult training, and distance learning.