

Measuring cognitive load in the presence of educational video: Towards a multimodal methodology

Jan-Louis Kruger

Macquarie University, Australia; North-West University, South Africa

Stephen Doherty

The University of New South Wales, Australia

The use of video has become well established in education, from traditional courses to blended and online courses. It has grown both in its diversity of applications as well as its content. Such educational video however is not fully accessible to all students, particularly those who require additional visual support or students studying in a foreign language. Subtitles (also known as captions) represent a unique solution to these language and accessibility barriers, however, the impact of subtitles on cognitive load in such a rich and complex multimodal environment has yet to be determined. Cognitive load is a complex construct and its measurement by means of single indirect and unidimensional methods is a severe methodological limitation. Building upon previous work from several disciplines, this paper moves to establish a multimodal methodology for the measurement of cognitive load in the presence of educational video. We show how this methodology, with refinement, can allow us to determine the effectiveness of subtitles as a learning support in educational contexts. This methodology will also make it possible to analyse the impact of other multimedia learning technology on cognitive load.

Introduction

The use of video in education is hardly a new concept and can be traced back as far as the Second World War in training situations (Chandler & Cypher, 1948). While there is now hardly any educational context where it is not used, it is particularly in blended and online education that this mode has become ubiquitous. In the field of instructional design, multimedia learning forms an important focus, and many authors discuss the effective use of video to enhance learning as well as the impact of video on learning itself (cf. Guo, Kim, & Rubin, 2014; Kay, 2012; Schmidt et al., 2014).

Our interest in educational video is located primarily in the complex interaction between different channels of incoming information identified in the cognitive theory of multimedia learning (Mayer, 2009; Mayer & Moreno, 2003), namely the visual and the auditory processing channels. This theory is premised on the seminal works on learning through multiple channels: Paivio (1990) in his work on mental representations, Baddeley (1997) in his working memory model, and Engelkamp (1998) in his multimodal theory. Here, we are specifically interested in the role of language in both visual and auditory form, in other words, the nexus where words can either maximise the capacity of working memory by coding information in both channels (auditory narration and visual text), or overburden working memory by combining on-screen text with animation or pictures. Our ultimate goal is to determine to what extent subtitling, as a mostly redundant transcript of spoken words presented at the bottom of the screen, can be used to support and enhance learning in an educational context.

The cornerstone of all discussions on multimedia learning is the management of cognitive load. While the concept of cognitive load is well established and its role in multimedia learning has received considerable attention (cf. Mayer, Heiser, & Lonn, 2001; Mayer & Moreno, 1998, 2003; Mayer, Moreno, Boire, & Vagge, 1999; Moreno & Mayer, 2000, 2002; Mousavi, Low, & Sweller, 1995; Paas, Tuovinen, Tabbers, & Van Gerven, 2003), the actual measurement of cognitive load in multimedia environments has proven to be rather complicated (cf. Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Paas et al. (2003) provide an overview of a number of uni- and multidimensional methods that are used to measure cognitive load, but also point out that “researchers have measured the total cognitive load and have not been able to use one of the measurement techniques to differentiate between these three cognitive load components” (2003; p. 67), namely: intrinsic, extraneous, and germane cognitive load.

In moving to address this observed limitation, this paper first presents a critical overview of the dominant measurements of cognitive load ranging from subjective rating of cognitive effort (paper-based), to psychophysiological measurements (e.g., pupillometrics, eye movement, and heart rate), task and performance measures (primary and dual-task measurements, e.g., reaction time and error rate) and direct measures of brain activity (e.g., electroencephalography). Building upon this, our discussion will then focus on the benefits and drawbacks of these measures in the context of multimedia learning and educational technology involving the use of video with and without subtitles. Finally, we present a multidimensional methodology that can be used to measure cognitive load arising from educational video using a triangulation of the above measures. Current limitations and future refinements will then be discussed.

Cognitive load in multimedia learning

Cognitive load theory (CLT) (Paas et al., 2003; Paas, Van Merriënboer, & Adam, 1994; Plass, Moreno, & Brünken, 2010; Sweller, Ayres, & Kalyuga, 2011) is based on the notion of a limited working memory and processing capacity. According to Paas et al. (2003; p. 63), “central to CLT is the notion that working memory architecture and its limitations should be a major consideration when designing instruction”. Paas et al. (2003; p. 64) also identify the fact that cognitive load “can be defined as a multidimensional construct representing the load that performing a particular task imposes on the learner’s cognitive system”, that it has “a causal dimension reflecting the interaction between task and learner characteristics, and an assessment dimension reflecting the measurable concepts of mental load, mental effort, and performance”.

The three components of cognitive load according to CLT are: intrinsic load (inherent to the subject), extraneous load (those aspects of the learning experience that impose cognitive effort and impede learning), and germane load (the level of cognitive activity that is required for learning to take place) (cf. Mayer & Moreno, 2003). The main goal of instructional design in multimedia learning is to reduce the extraneous load in order to avoid cognitive overload and maximise the cognitive capacity to be assigned to germane load, thus ensuring the learner can reach the learning outcomes. In multimedia learning, this conception of cognitive load was formalised in the work of Mayer and Moreno (2003) (see also Mayer, 2009) in their cognitive theory of multimedia learning which posits two processing channels for the acquisition of information, namely the visual (pictorial) channel and the auditory (verbal) channel. According to this theory, the combination of the two channels in instructional design can be used to maximise the capacity of working memory, whereas using the same channel to present more than one source of information could lead to cognitive overload.

Video presents a compelling case as a form of multimedia in that it has the potential to present various sources of information simultaneously and consecutively that may tax both channels to the point where cognitive overload occurs and learning becomes less effective or even non-existent. Video typically presents different types of information in the visual channel: moving or static images such as illustrations, animations, and speakers or talking heads, as well as text in the form of captions on illustrations or identifying speakers, or even subtitles providing a transcript of spoken words. In the auditory channel it can present speech in the form of dialogue and narration, music, and sound effects. What makes the processing of video a complex phenomenon, is the fact that it introduces an element of continuous selection. In other words, learners have to self-select the relevant elements of the auditory and the visual information, organise it, activate prior knowledge, and integrate this knowledge with the relevant auditory and visual information in order to arrive at a coherent and unified representation (Kalyuga, 2012; Mayer, 2009).

Kalyuga (2012) provides a comprehensive review of cognitive load factors in the context of spoken words with direct relevance to the use of video in education. He discusses the use of both spoken and written words with pictures that are unintelligible on their own. We identify examples of this in the use of slides in a class or video where large amounts of text are placed on a slide and also largely read out. Although this is often considered to make use of the modality effect (simultaneous use of auditory and visual channels to increase cognitive capacity), and to accommodate learner preferences for either reading or listening, Kalyuga (2012) points out that “available evidence indicates that learning could be inhibited by the presentation of the same verbal information in both modalities” (p. 151), also known as the verbal redundancy effect. In particular, working memory could be overloaded when learners have to process pictorial information and visual text simultaneously (visual channel of working memory), while the load also increases in the auditory channel when “visual words are recoded into auditory form at some stage of cognitive processing” (Kalyuga, 2012; p. 151).

This verbal redundancy effect has been demonstrated empirically in a number of studies (Kalyuga, 2012). However, the verbal redundancy effect may be moderated by a number of factors. First, the interdependence of pictorial and verbal information means that in the absence of dependence, redundancy between modes should be avoided. Second, in terms of complexity of information, redundancy in low-complexity conditions does not result in overload. Finally, the length of the textual material could moderate the redundancy effect. For example, in long explanations, written-only formats are more effective than written and spoken formats, although written and spoken formats are more effective than spoken-only formats, which goes against the redundancy effect. Indeed, Kalyuga (2012) contends that the partitioning of text into small segments with time breaks between them would mean that a narration with concurrent text may not cause overload and may even improve learning.

The study Kalyuga refers to here, in particular, is that of Moreno and Mayer (2002). A study that uses the dual-processing theory of working memory and shows that redundant verbal explanations (e.g., a verbal explanation with its on-screen transcription) benefit learning provided that no other competing information is presented visually. When words are presented both visually and aurally, additional processing capacity is made available because the visual working memory and the auditory working memory work independently. An important proviso here is that the redundant information should be synchronised, that is, be available simultaneously. One situation in which redundancy between spoken and written text may be beneficial, according to Kalyuga (2012), is “for learners for whom the language of instruction is a second language” (p. 152).

Diao and Sweller (2007) found that written presentation concurrent with verbatim spoken presentation indeed resulted in the redundancy effect when participants in their first year of university performed worse on translation scores, subjective mental load ratings, and free recall performance in the combined condition than in the written only condition. However, this study did not investigate fluctuations in load, and presented only continuous text.

From the above findings and their referents, it is evident that presenting words in both spoken and written form introduces a verbal redundancy effect although this may be mediated if the text is segmented and synchronised with the verbal information. This particular moderating factor is extremely relevant in the case of subtitling, where the same text is available in both written and spoken modalities, but presented visually in short segments (typically no more than two lines of semantically segmented text at a time). This also explains the benefits that have been found with the use of subtitles in educational contexts.

The effect of individual characteristics (working memory capacity, language proficiency, prior knowledge, motivation, fatigue, familiarity with subtitles, etc.) on performance is critical in that the redundancy of subtitles is dependent on the individual’s dynamic needs at any given point in time, for example, reading the textual form of an unknown word that has just been heard in the dialogue. Individuals who wish or need to have additional visual support at such times arguably do not have a redundancy effect. This relates to the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003), where instructional design should be adapted to the learner as their domain-specific knowledge of the tasks increases. In the case of subtitles, however, the rate of change in the presentation of on-screen stimuli, both spatially and temporally, is much greater than in the static, text-based stimuli typically found in CLT tasks. In order to account for such high temporal resolution during tasks, online measures with a high degree of temporal precision at millisecond level are required.

Subtitles as verbal redundancy in educational videos

Subtitling is a mode of audiovisual translation that presents a transcript of the dialogue in a video at the bottom of the screen, typically in no more than two lines. Subtitles can be in the same language as the video or in a different language. Same-language subtitles (SLS) or intralingual subtitles, are in some contexts called captions, particularly when referring to subtitling for those who are deaf and hard of hearing. We will use SLS to refer to same-language subtitles that are not created primarily for a deaf or hard-of-hearing audience, although that group would also benefit from SLS. The main difference would be that SLS does not provide a description of sounds, or identify speakers as in the case of captions. Translation subtitles or interlingual subtitles, refer to subtitles that are in a different language from that in the video, mostly for viewers who do not understand the language used in the video. In both forms, the text is synchronised with

the speech so that the words that are spoken are on screen at the same time as the utterance is made. Although subtitles are primarily used to make video accessible to those who are deaf and hard of hearing and to viewers who do not understand the language of the video, the use of subtitles in language learning and education has increased over the years as part of a wider development in language and translation technologies (Doherty, 2016; Gernsbacher, 2015).

In terms of reading and listening comprehension, a number of discontinuous studies over decades have shown the possible benefits of subtitling. Bird and Williams (2002), for example, demonstrated the positive impact of subtitles on word recognition and comprehension. They also point out that subtitles allow the integration of reading and listening skills, something that is invaluable for students studying in a second language. The work of Vanderplank (2013) in particular points to the benefits of subtitling in a context where students are confronted with challenging auditory input. He found that SLS increase comprehension particularly because they neutralise accents and dialects (a problem, for instance, in educational contexts such as higher education where many lecturers and speakers in supplementary videos have accents that may be unfamiliar to students), and also foregrounds unfamiliar phrases and words, such as new concepts and terminology as well as everyday expressions. Similarly, Markham (1999; p. 326) found that “university-level ESL [English Second Language] students clearly derive substantial listening (specifically word recognition) benefits from viewing second-language captioned video material”. These findings, however, were limited and the studies did not look at longer-term benefits or academic performance, nor did they go beyond unidimensional paper-based, post-task, word-recognition tests.

Understanding the contribution of different sources of information to cognitive load will make it possible not only to determine and manage cognitive load more efficiently in instructional design involving video, but will also make it possible to test the complexities of verbal redundancy effects with much more precision. In this line of reasoning, Kalyuga (2012) posits that if continuous text is partitioned into logically complete and easily managed sequential segments with time breaks between them (as subtitles are constructed), a narration with concurrent visual text may not only eliminate negative effects of verbal redundancy, but actually improve learning. He notes that this could be particularly effective for learners for whom the language of instruction is a second language, and this written backup for brief spoken explanations is helpful. On the other hand, presenting lengthy textual descriptions simultaneously and continuously in both modalities may significantly increase cognitive load. The negative consequences of such increases are particularly relevant to learning a second or foreign language (Kalyuga, 2012).

In order to determine whether subtitles added to educational video result in the redundancy effect for either a first or second language audience, it is important to be able to measure the impact of text partitioned with minimal time breaks. This could also be considered a case of “lengthy textual descriptions” presented “simultaneously and continuously in both modalities” (Kalyuga, 2012; p. 155). Very few of the current measurements of cognitive load provide this level of nuance, particularly by combining global impact on cognitive load with online fluctuations in load in the context of continuous video.

Further issues of “weak research design” (Vanderplank, 2013; p. 6) and reliance on pseudo-experimental approaches have limited the application of subtitling in educational contexts. An additional limitation is that these studies also did not advance beyond offline methods of post-task questionnaires and simple word-recognition tests, as above, nor did they profile or attempt to account for individual differences in the participants in their research design or statistical analyses. Added to this, other barriers to the widespread adoption of subtitles have been identified in the lack of knowledge about subtitling itself, the limited amount of generalisable empirical evidence as to its efficacy, as we also argue here, and a disconnect between the above fields of research.

Managing cognitive load in multimedia learning

In order to design educational video that will result in optimal opportunity for learning, the management of cognitive load is of critical importance. In a comprehensive discussion of some of the most important causes of cognitive overload in multimedia learning, and indeed the solutions to these problems, Mayer and Moreno (2003) identified five problems that lead to overload along with their solutions presented (Table 1).

Table 1

Overload scenarios and solutions (adapted from Mayer & Moreno, 2003; p. 46)

Overload scenario	Example	Solutions
One channel is overloaded with essential processing demands	Presenting an animation and concurrent explanatory text	Off-loading (replace text with narration)
Both channels are overloaded with essential processing demands in working memory	Narrated animation with high intrinsic complexity due to rich content and fast pace	Segmenting (time between shorter segments) Pretraining (provide pretraining in names and qualities of components)
One or both channels overloaded by essential and incidental processing (extraneous material)	Narrated animation with music and/or inserted video clips with examples	Weeding (eliminate extraneous material) Signalling (produce cues for processing material)
One or both channels overloaded by essential and incidental processing (confusing presentation)	Images at the top of the screen and on-screen text at the bottom	Aligning (place text near corresponding part of image) Eliminating redundancy (avoid presenting same information in spoken and written format simultaneously)
One or both channels overloaded by essential processing and representational holding	Narration, followed by an animation	Synchronising (present narration and animation simultaneously) Individualising (give learners skills to hold mental representations)

Helpful as these instructions are in terms of the design of multimedia texts, the reality is that most educational videos routinely contribute to some or all of the overload scenarios. For example, in many educational contexts subtitles are used to provide support for students learning in a second language: (1) a video recording of a class in English; (2) a video clip with narration in English with English same-language subtitles as a transcription of the words of the lecturer; (3) or the narration translated from another language; or (4) the subtitles, or words of the lecturer, translated into the language of the student. In most educational videos at least scenarios 1, 3, and 4 emerge, yet without the subtitles the audience may not be able to make optimal use of the video, and listening to a class in a second language may result in higher extraneous load than having the redundant subtitles.

More importantly, measuring the impact of these scenarios on cognitive load is extremely complicated, particularly due to the fact that overall load or total load does not tell us much about impact of the sources of information individually or in combination, as these sources may change constantly due to the dynamic nature of video. Current measurements of cognitive load do not possess this level of precision (Paas et al., 2003) and typically rely on either subjective or objective measures, not both (cf. Brünken, Steinbacher, Plass, & Leutner, 2002). For these reasons, the measuring of cognitive load in the presence of video acquires a particular complexity that calls for a more robust multimodal suite of measurements that will allow measurement of instantaneous load over the course of the video, cumulative load, and total load. In the subsequent section we will detail these different measures of cognitive load and discuss the benefits and limitations of each in the context of educational video before moving to present their combination as a unified multimodal measurement.

Measuring cognitive load

Self-rating of cognitive effort

One of the most widely-used measurements of cognitive load involves the post-hoc measurement of self-reported mental effort, which measures perceptions of mental effort caused by a task. This measure is typically paper-based and provides an indication of overall cognitive load. It can be multidimensional, measuring groups of associated variables such as mental effort, fatigue and frustration, or simply unidimensional measuring only perceived mental effort. According to Paas et al. (2003), mental effort is “the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load” (p. 64) and it has been found sensitive to small differences in cognitive load, valid, reliable, and nonintrusive. The most widely-used subjective measures of cognitive load are adapted versions of the NASA-Task Load (NASA-TL) Index (cf. Hart & Staveland, 1988) or a 9-point mental effort scale asking the participant to rate the amount of mental effort a task required from “*very very low mental effort*” to “*very very high mental effort*” (Paas et al., 2003).

Recent studies by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013, 2014) present attempts to distinguish between intrinsic, extraneous, and germane load using self-rating items. Their findings provide “some support for the assumption that intrinsic and extraneous cognitive load can be differentiated using a psychometric instrument” (2014; p. 40). They also find support for the reconceptualisation of germane cognitive load “as referring to the actual working memory resources devoted to dealing with intrinsic cognitive load” (2014; p. 40), as discussed by Kalyuga (2011), Sweller (2010), and Sweller, Ayres, and Kalyuga (2011). The items designed by Leppink and colleagues in these two studies provide an essential psychometric instrument for measuring accumulated cognitive load, as discussed comprehensively by the authors. As such, the multimodal measurement of cognitive load should include such an instrument, as it can be triangulated against more online psychophysiological measurements.

Psychophysiological measurements

Psychophysiological measures address the subjectivity of self-rating measures and operate on the assumption that changes in physiological variables reflect changes in cognitive functioning (cf. Paas et al., 2003). Examples of psychophysiological measurements include heart rate, eye movements, and brain activity. According to Paas et al. (2003), however, heart rate has been shown to be intrusive, invalid, and insensitive to subtle fluctuations in cognitive load, we therefore do not expand upon it here.

Eye movements

Eye movements offer a range of measures – some of which remain contested – that have been operationalised in studies of cognitive load. Although Paas et al., (2003; p. 66) consider cognitive pupillary response to be “a highly sensitive instrument for tracking fluctuating levels of cognitive load”, its use outside of highly controlled, single-concept experimental paradigms has been consistently questioned. Klingner, Kumar, and Hanrahan (2008), for example, argue that task-evoked pupillary response “does not occur reliably for any one single episode of mental effort” and therefore “several pupillary responses must be combined” (p. 70). This time-aggregated style yields a coarse measure of cognitive load as it averages to a single size, which is itself widely varied due to individual pupil sizes and thresholds of constriction and dilation. In complete contrast however the field of cognitive science typically aligns and averages data over several identical trials, offering a unique solution to increase reliability, but with limited usage due to its vastly simplified experimental contexts (cf. Klingner et al., 2008). Furthermore, pupil size is sensitive to lighting, time of the day, stimulants, and other environmental factors that are beyond the experimental control of all but a few tightly controlled experiment designs. Due to its very nature, video is characterised by continuous changes in luminosity that also directly impact upon pupil size.

Gagl, Hawelka, and Hutzler (2011) also point out that although cognitive effort is indeed reflected in pupil dilation, pupil size itself, which is central to accurate measurement of its dilation, could be susceptible to changes in gaze direction. Similarly, Brisson et al. (2013) report systematic errors of pupil size estimation using three different eye trackers, and conclude by calling the use of task-evoked pupillometry into

question. Although both of these studies offer recommendations for correcting for this error for individual participants, the nature of continuous reading, let alone reading in the context of video, means that this measure is not sufficiently reliable in this domain (see also Antonenko, Paas, & Grabner, 2010).

Fixation duration could be a very useful measure to determine cognitive load, since there seem to be “functional links between what is fixated and cognitive processing of that item – the longer the fixation, the ‘deeper’ the processing” (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & van de Weijer 2011; p. 382). Therefore, longer mean fixation durations within the same activity could reflect increased cognitive processing and cognitive load (Kruger et al., 2015), a finding consistently reported in other studies (e.g., Doherty & O’Brien, 2014; Doherty, O’Brien, & Carl, 2010). In this regard, however, it is important to consider that activities like reading tend to elicit more, shorter fixations than scene perception, which elicits fewer, longer fixations (cf. Holmqvist, et al., 2011). With these considerations accounted for, fixation duration remains well suited to measuring instantaneous load over the course of a video.

Other eye-movement derived measures that have been shown to offer more reliable online measures of attention allocation and cognitive load include blink rate, blink duration, and blink latency. Goldstein, Bauer, and Stern (1992) use these measures in comparing the reading of two characters to that of six characters, which makes it difficult to extrapolate to more ecologically valid contexts like video. As in the case of fixation duration, these measures provide online measures of instantaneous load. In particular, there seems to be an inverse relationship between cognitive load and blink rate with blink rate decreasing as the cognitive workload increases (Bagley & Manelis, 1979, Brookings, Wilson, & Swain, 1996; Chen, Epps, Ruiz, & Chen, 2011). Blink latency has also been found to increase with higher cognitive load (Chen et al., 2011).

Fixation count and fixation duration are both linked to the eye-mind hypothesis (Just & Carpenter, 1980), namely that the mind attends to where the eye fixates. According to Irwin (2004), there are a number of factors that complicate this hypothesis, mainly centred around the fact that cognitive processing is not limited to the periods during which the eyes are still (fixations), but could also occur while they eyes are moving (saccades). Furthermore, “the size of the functional field of view depends upon the nature of the task, the number of items in the visual field, and whether other cognitive demands are placed on the subject” (Irwin, 2004; p. 107). In the context of educational video, this means that the size of the functional field of view could be quite variable, depending on the nature of the scene, and the number of competing sources of information.

Nevertheless, eye movements give us the only way to approximate the location of visual attention allocation, and combining it with other measures such as electroencephalography means that we can gain a vastly better understanding of the sources of instantaneous cognitive load during learning. We therefore consider fixation counts and fixation duration within specific areas of the screen to be important measures of attention allocation. However, the constant movement of fixated objects, time limitations of displayed text elements, and adapted reading strategies of the viewers are factors that all have to be taken into account, especially in multimodal and dynamic environments (e.g., Kruger & Steyn, 2014). Even though eye tracking offers a reliable source of information on what a student looks at and how, it allows limited insight into cognitive processing.

Electroencephalography

Electroencephalography (EEG) is a neuroimaging technique used in many disciplines as an electrophysiological measure of electrical activity of the brain. Antonenko et al. (2010) provide a discussion of EEG as offering “new and promising approaches to educational psychology research” particularly as it can serve as an “online, continuous measure of cognitive load detecting subtle fluctuations in instantaneous load, which can help explain effects of instructional interventions when measures of overall cognitive load fail to reflect such differences in cognitive processing” (2010; p. 425).

Gerlic and Jausovec (1999) already applied this method to video and measured changes in the power of neural oscillations with EEG to examine cognitive processes in multimedia learning. They found that when text was compared to picture (text, sound and image) and video (text sound and video), the activity in the occipital lobe (associated with vision and imagery) and temporal lobe (associated with auditory input) increased in the video and picture presentations, whereas the activity in the frontal lobes (associated with cognitive control and working memory) increased in the presence of text. Although no difference was found

between the picture and video presentations, alpha power in the frontal locations and left central location was significantly lower for text when compared to picture and video (indicating higher mental effort). However, for the occipital and temporal locations, alpha was lower for video and picture than for text. The study is important as it was one of the first studies to use EEG in the context of video, although it had some limitations in terms of replicability (due to limited information on the nature and comparability of the content of the three texts), as well as the fact that the alpha power for the best 30 seconds out of a 1 one-minute video was selected by visual inspection, and then a spectral power average was calculated. This method of data analysis lacks sufficient detail to allow for extrapolation from the findings.

Although most neurocognitive EEG research focused on event-related potential (ERP) indices, these reflect brain responses to specific events and they are calculated by “averaging the continuous EEG signal over many trials so that the oscillatory background activity, considered as noise, is cancelled out” (Antonenko et al., 2010; p. 429). Functional networks, however, focus on the dynamics of brain oscillations, and according to Antonenko et al. (2010, p. 429), two oscillatory components of continuous EEG are sensitive to task difficulty manipulations, namely alpha and theta. Alpha consists of oscillations in the 8–13 waves per second (Hz) range. “When the eyes are opened, a suppression (or desynchronisation) of alpha activity occurs indicating alert attention ... The general consensus is that the localization of recording sites is determined by where these brain wave rhythms are most prominent – parietal areas for alpha” (Antonenko et al., 2010; p. 430).

Antonenko and Niederhauser (2010) measured fluctuations in alpha, beta, and theta rhythms on two locations, the prefrontal cortex (F7) and the parietal lobe (P3) to investigate the impact of leads in hypertext on cognitive load. They used the event-related desynchronisation percentage (ERD%) for these three rhythms as online measures of brain activity, with increased cognitive load being associated with higher brain wave desynchronisation for alpha and beta rhythms, and higher brainwave synchronisation for the theta rhythm. Their results indicated lower cognitive load in the condition with leads where the EEG measures are concerned (focusing on the 20 seconds during which leads were accessed), but no difference between the leads and no-leads conditions based on self-rated effort. Also, the time on task for the leads condition was longer. Finally, the leads resulted in improved performance. Their explanation for this discrepancy between EEG and self-rated measures is that the EEG measured instantaneous load, and the self-ratings measured overall load. The positive impact on performance found with the condition with leads seems to suggest that they resulted in more cognitive capacity being made available.

In a slightly different context, Kruger, Soto-Sanfiel, Doherty, and Ibrahim (2016) use EEG to measure beta coherence between prefrontal and parietal regions as an indication of immersion in film with and without subtitles. They report that while this EEG measure correlates with other established, subjective measures of immersion, the multimodal measurement of psychological constructs such as immersion and cognitive load in multimedia contexts requires further validation. We therefore argue here that before measures such as EEG and eye movement behaviour can be utilised as robust measures of cognitive load in the context of video, they will have to be rigorously tested and validated in experimental studies.

Multimodal measurement of cognitive load in video and other multimedia contexts

In a comprehensive discussion of the potential relevance of cognitive neuroscience for the development and use of technology-enhanced education, Howard-Jones, Ott, van Leeuwen, T., and De Smedt (2015) stress that technology-enhanced education has to be conceptualised in terms of all levels (namely brain, mind, and behaviour – including social behaviour), and has to consider the interrelation of concepts across all levels. It will therefore remain important to measure cognitive load in terms of more than just psychophysiological and subjective measures, and with a sensitivity to individual differences and behaviour in social context.

In our methodology, we build on Antonenko and Niederhauser’s (2010) contention that cognitive load should be conceptualised as a dynamic process; it should be “assessed using a comprehensive analytical framework that integrates measures to target both the temporal dimensions of cognitive load like instantaneous load, peak load, average load, accumulated load, and overall load, and the contributions of structural load including intrinsic, extraneous, and germane load” (p. 148). In particular, in the context of educational video, the complexity of the learner’s engagement with different parts of the video involving

different combinations of static pictures, written words, spoken words, and moving images, means that a better understanding of the dynamic nature of cognitive load is required. This can only be achieved with a combination of several measurements of cognitive load.

Based on that study as well as the above review of literature on the measurement of cognitive load, we therefore propose the following as a multimodal methodology for measuring cognitive load in the presence of all modes of stimuli and combinations thereof, e.g. subtitled video. Table 2 provides an overview of the alignment between CLT constructs and the operationalised measures of the proposed methodology. Each component is then detailed further: psychometric, eye tracking, and electroencephalography.

Table 2
Alignment of CLT constructs with operationalised measures from the proposed multimodal methodology (adapted from Kruger, Doherty, Fox, & de Lissa, 2017)

CLT construct	Operationalised measures	Nature
Average load	Psychometric instrument Averaged mean fixation duration (across stimulus)	Offline and online
Overall load	Psychometric instrument Averaged mean fixation duration (across stimulus)	Offline and online
Instantaneous load	Mean fixation duration (at particular nodes) Average fixation count Blink rate Blink latency Alpha power	Online
Extraneous load	Psychometric instrument triangulated with averaged mean fixation duration, average fixation count, blink rate and blink latency and alpha power	Offline and online
Intrinsic load	Psychometric instrument triangulated with averaged mean fixation duration, average fixation count, blink rate and blink latency and alpha power	Offline and online
Germane load	Psychometric instrument triangulated with alpha power	Offline and online

Psychometric component

Here we propose the use of an instrument based on the items developed and validated by Leppink et al. (2013; 2014) to distinguish between intrinsic and extraneous load while also looking at possible correlations between online psychophysiological measures of cognitive load (eye tracking and EEG) and germane load. This may address the limitation identified by Leppink et al. (2014) in the accurate measurement of germane load. The instrument could be used for repeated measurements to study video segments or other multimedia texts.

Eye tracking component

In terms of eye movements, we propose the use of average fixation count (normalised per character), averaged mean fixation duration, blink rate, blink duration, and blink latency as online measures of cognitive load. Due to the exact nature of these measurements, it is essential to apply validated protocols

in setting up experiments and also to report the calibration and algorithms used for fixation detection. Caution should be used here however, not to compare these measures across activities that will necessarily elicit different eye behaviour (such as reading, visual search, and watching a stable point such as the face of a speaker). Particularly due to the fact that multimodal texts typically do combine a variety of activities, the research questions should be operationalised carefully to avoid such invalid comparisons.

However, the advantage of eye tracking is situated in the fact that it makes it possible to identify particular elements of a multimodal text and the behaviour of participants' eye movements and blinks in the presence of these elements in order to pinpoint the contribution of these elements to instantaneous cognitive load. Eye tracking synchronised with EEG measurements also makes it possible to interpret the EEG measurements in a contextualised manner.

Electroencephalography (EEG) component

Based on the literature review, we would recommend using alpha power in investigating the impact of different components of multimodal texts on cognitive load. Data from EEG recordings have to be processed offline and ocular artefact correction has to be performed through independent component analysis to isolate and remove systematic distortions caused by blinks and saccades. The EEG data then has to be transformed into alpha power through band-pass filtering to retain the alpha band between 8–12 Hz. Artefact rejection of extreme values has to be performed by replacing outlying values (e.g. exceeding six standard deviations) with mean values for each electrode. It is also important to calculate the mean alpha band power for particular intervals in reference to a baseline. As in the case of eye tracking it is essential to apply validated protocols in terms of setup, equipment, and reporting. The main advantage of EEG lies in the fact that it has the potential to provide insights concerning instantaneous load in the presence of particular components of multimodal texts in isolation or in combination. This could bring us closer to identifying the contribution of such components and their combination to intrinsic and extraneous load, particularly when triangulated with psychometric data.

Statistical analysis

Finally, we argue that refinements to current practices of statistical analysis in CLT can be made in order to account for the time-course nature of online measures. Growth curve modelling (for an overview, see Curren, Obeidat, & Losardo, 2010) offers several advantages in analysing data from experimental research. In its basic form, the statistical model comprises fixed and random effects to estimate individual and group performance over time (e.g., within a single experiment or a series of experiments). Fixed effects are values that are derived from the population, while random effects are the random probability distribution around the fixed effect. In an experiment that measures instantaneous load for example, the fixed effect would represent the mean alpha power in the sample, and the random effect would represent the individual variance of participants around this group mean. As the proposed measurement of cognitive load and its subtypes combines both online and offline measures, this approach is ideal as it allows for the analysis of group-level effects while accounting for individual differences in repeated measure designs. Growth curve modelling has become standard in cognitive science research but has yet to gain traction in CLT.

Furthermore, growth models can account for missing data, a common limitation of online measures, and repeated measures designs. Growth models can also include predictors upon which fixed and random effects can be conditioned. Such advantages are arguably ideal for the multimodal measurement of cognitive load and also address the shortcomings of traditional factorial designs and analysis of variance between groups reviewed in the literature above which are not as statistically robust for time course data or as suitable to account for individual differences in performance over time.

Conclusion

The measurement of the different components of cognitive load remains an important area in instructional design. The multimodal methodology proposed in this article is intended to provide a framework for the measurement of cognitive load in the presence of educational video and other multimedia environments. This will be instrumental in gaining a more comprehensive understanding of the dynamic nature of cognitive load in such texts. The methodology holds advantages for evidence-based refinements to the measurement of CLT constructs, in particular the measurement of different types of cognitive load at

specific points (instantaneous load) or over a task or set of tasks (average load and overall load). It could also provide instructional designers with robust data on the impact of redundancy (such as verbal redundancy when the words of a teacher are subtitled) or other qualities of multimodal texts on cognitive load.

In order to refine such a methodology, it is essential to first validate the multimodal measurement of cognitive load in the context of video. Our next steps will therefore be to validate the proposed multimodal methodology in both simple and complex tasks in uni-, bi-, and multimodal contexts in order to then apply it to educational subtitling. Once this has been achieved, it will be possible to investigate the way in which viewers process video and other multimodal texts. It will also be possible, for example, to study the way in which students process educational videos with and without subtitles, and with different kinds of subtitles (e.g., verbatim, reduced or keyword subtitles), in order to establish to what extent subtitling can be used to benefit learning, such as for students studying in a second language or students who require additional visual support, for example, those who are deaf, hard of hearing, or cognitively impaired. For these students, it has to be determined whether cognitive load, visual attention, and learning scores are correlated in both laboratory-based research as well as real-life educational contexts over short-term and long-term durations. All of these areas of application, and indeed many others not detailed in this paper, stand to benefit from the accuracy and precision offered by the multidimensional measurement of cognitive load in multimodal contexts. Such measurement can then realise benefits in terms of managing cognitive load while viewing videos to improvement educational and student outcomes and even the adaptation of educational content to the learner's cognitive load in real time.

References

- Antonenko, P. D., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior*, 26(2), 140–150. <http://dx.doi.org/10.1016/j.chb.2009.10.014>
- Antonenko, P. D., Paas, F., & Grabner, R. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22, 425–438. <http://dx.doi.org/10.1007/s10648-010-9130-y>
- Baddeley, A. D. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.
- Bagley, J., & Manelis, L. (1979). Effect of awareness on an indicator of cognitive load. *Perceptual and Motor Skills*, 49(2), 591–594. <http://dx.doi.org/10.2466/pms.1979.49.2.591>
- Bird, S. A., & Williams, J. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, 23(4), 509–533. [http://dx.doi.org/10.1017.S0142716402004022](http://dx.doi.org/10.1017/S0142716402004022)
- Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eye trackers. *Behavior Research Methods*, 45(4), 1322–1331. <http://dx.doi.org/10.3758/s13428-013-0327-0>
- Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42(3), 361–377. [http://dx.doi.org/10.1016/0301-0511\(95\)05167-8](http://dx.doi.org/10.1016/0301-0511(95)05167-8)
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, 49(2), 109–119. <http://dx.doi.org/10.1023/B:TRUC.0000021812.96911.c5>
- Chandler, A. C., & Cypher, I. F. (1948). *Audio-visual techniques for enrichment of the curriculum*. New York, NY: Noble and Noble.
- Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye activity as a measure of human mental effort in HCI. *Proceedings of the 16th ACM International Conference on Intelligent User Interfaces, Palo Alto, CA*, 315–318. <http://dx.doi.org/10.1145/1943403.1943454>
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, 11(2), 121–136. <http://dx.doi.org/10.1080/15248371003699969>
- Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction*, 17(1), 78–88. <http://dx.doi.org/10.1016/j.learninstruc.2006.11.007>
- Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication*, 10, 947–969.

- Doherty, S., & O'Brien, S. (2014). Assessing the usability of raw machine translated output: A user-centred study using eye tracking. *International Journal of Human Computer Interaction*, 30(1), 40–51. <http://dx.doi.org/10.1080/10447318.2013.802199>
- Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine Translation*, 24(1), 1–13. <http://dx.doi.org/10.1007/s10590-010-9070-9>
- Engelkamp, J. (1998). *Memory for actions*. Hove: Psychology Press.
- Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behavior Research Methods*, 43(4), 1171–1181. <http://dx.doi.org/10.3758/s13428-011-0109-5>
- Gerlic, I., & Jausovec, N. (1999). Multimedia: Differences in cognitive processes observed with EEG. *Educational Technology Research and Development*, 47(3), 5–14. <http://dx.doi.org/10.1007/BF02299630>
- Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 195–202. <http://dx.doi.org/10.1177/2372732215602130>
- Goldstein, R., Bauer, L. O., & Stern, J. A. (1992). Effect of task difficulty and interstimulus interval on blink parameters. *International Journal of Psychophysiology*, 13(2), 111–117. [http://dx.doi.org/10.1016/0167-8760\(92\)90050-L](http://dx.doi.org/10.1016/0167-8760(92)90050-L)
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the first ACM conference on Learning@ scale conference, Atlanta, GA*, 41–50. <http://dx.doi.org/10.1145/2556325.2566239>
- Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, Netherlands: Elsevier.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Howard-Jones, P., Ott, M., van Leeuwen, T., & De Smedt, B. (2015). The potential relevance of cognitive neuroscience for the development and use of technology-enhanced learning. *Learning, Media and Technology*, 40(2), 131–151. <http://dx.doi.org/10.1080/17439884.2014.919321>
- Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 105–134). New York, NY: Psychology Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <http://dx.doi.org/10.1037/0033-295X.87.4.329>
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, 23(1), 1–19. <http://dx.doi.org/10.1037/0022-0663.93.3579>
- Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, 7(2), 145–159. <http://dx.doi.org/10.1016/j.edurev.2011.12.002>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4
- Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820–831. <http://dx.doi.org/10.1016/j.chb.2012.01.011>
- Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. *Proceedings of the 2008 ACM Symposium on eye tracking research and applications, Savannah, GA*, 69–72. <http://dx.doi.org/10.1145/1344471.1344489>
- Kruger, J. L., Doherty, S., Fox, W., & de Lissa, P. (2017). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign language viewers. In I. Lacruz, & R. Jääskeläinen (Eds.), *New Directions in Cognitive and Empirical Translation Process Research*. London: John Benjamins.
- Kruger, J. L., Soto-Sanfiel, M. T., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive audiovisual translology: Subtitles and embodied cognition. In Ricardo Muñoz (Ed.), *Reembedding Translation Process Research*. (pp. 171–194). London: John Benjamins Publishing Company.
- Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105–120. <http://dx.doi.org/10.1002/rq.59>
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. <http://dx.doi.org/10.3758/s13428-013-0334-1>

- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & 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. <http://dx.doi.org/10.1016/j.learninstruc.2013.12.001>
- Markham, P. L. (1999). Captioned video-tapes and second language listening word recognition. *Foreign Language Annals, 32*(3), 321–328. <http://dx.doi.org/10.1111/j.1944-9720.1999.tb01344.x>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.
- 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), 187–198. <http://dx.doi.org/10.1037/0022-0663.93.1.187>
- 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), 312–320. <http://dx.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. http://dx.doi.org/10.1207/S15326985EP3801_6
- Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638–643. <http://dx.doi.org/10.1037/0022-0663.91.4.638>
- Moreno, R., & Mayer, R. E. (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. <http://dx.doi.org/10.1037/0022-0663.92.1.117>
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156–163. <http://dx.doi.org/10.1037/0022-0663.94.1.156>
- Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334. <http://dx.doi.org/10.1037/0022-0663.87.2.319>
- Paas, F., G., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. http://dx.doi.org/10.1207/S15326985EP3801_8
- Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419–430. <http://dx.doi.org/10.2466/pms.1994.79.1.419>
- Paivio, A. (1990). *Mental representations*. Oxford: Oxford University Press.
- Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. New York, NY: Cambridge University Press.
- Schmidt, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., ... Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers and Education, 72*, 271–291. <http://dx.doi.org/10.1016/j.compedu.2013.11.002>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. <http://dx.doi.org/10.1007/s10648-010-9128-5>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.
- Vanderplank, R. (2013). ‘Effects of’ and ‘effects with’ captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching, 49*(2), 235–250. <http://dx.doi.org/10.1017/S0261444813000207>

Corresponding author: Jan-Louis Kruger, janlouis.kruger@mq.edu.au

Australasian Journal of Educational Technology © 2016.

Please cite as: Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19–31. <http://dx.doi.org/10.14742/ajet.3084>