Learning and Instruction 61 (2019) 23–34

Contents lists available at ScienceDirect

Learning and Instruction

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

Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect

Guido Makransky[a,*], Thomas S. Terkildsen[a], Richard E. Mayer[b]

[a] Department of Psychology, University of Copenhagen, Copenhagen, Denmark
[b] Psychological and Brain Sciences, University of California Santa Barbara, CA, USA

ARTICLE INFO

Keywords:
Spatial contiguity
EEG
Eye tracking
Cognitive load
Multimedia learning

ABSTRACT

The main objective of this study was to investigate the potential of combining subjective and objective measures of learning process to uncover the mechanisms underlying the spatial contiguity effect in multimedia learning. The subjective measures of learning process were self-reported cognitive load ratings and the objective measures were eye-tracking and EEG measures. Learning outcome was measured by scores on retention and transfer posttests. A sample of 78 university students participated in a between-subjects design in which a multimedia slideshow lesson on how lightning storms develop was presented either with printed text as a caption at the bottom of each illustration (separated presentation) or with printed text placed next to the corresponding part of each illustration (integrated presentation). Regarding spatial contiguity, the integrated group spent significantly more time looking at the text ($d = 0.64$), but significantly less time looking at irrelevant illustrations ($d = 1.10$), and reported a significantly lower level of extraneous load ($d = 0.57$), compared to the separated group. As expected, they also scored significantly higher on the transfer test ($d = 0.49$). Students who performed best on posttests reported a lower level of extraneous load ($d = 0.56$). Furthermore, EEG based alpha band activity was predictive of intrinsic cognitive load but not predictive of extraneous cognitive load, and EEG based theta activity was not predictive of intrinsic or extraneous load. The results suggest that subjective and objective measures of cognitive load can provide different information to test the theoretical mechanisms involved in multimedia learning.

1. Introduction

1.1. Objective and rationale

According to cognitive theories of learning such as Cognitive Load Theory (CLT; Sweller, Ayres, & Kalyuga, 2011) and the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2014), effective instructional methods prime appropriate cognitive processing during learning which leads to superior learning outcomes. As an example of appropriate cognitive processing, intrinsic cognitive load (or essential processing) is cognitive processing during learning needed to represent the material and is a result of the complexity of the material for the learner. In contrast, ineffective methods of instruction prime inappropriate cognitive processing during learning, which leads to inferior learning outcomes. As an example of inappropriate cognitive processing, extraneous cognitive load (or extraneous processing) is cognitive processing during learning that does not serve the instructional goal and is caused by the way the material is presented.

Advances have been made in describing effective instructional methods (such as placing printed words next to the corresponding part of the graphic rather than as a caption) and in creating transfer posttests to measure learning outcomes (such as using open-ended questions that ask the learner to use the material in a new situation). In contrast, advances in cognitive theories of learning have been challenged by the need for useful measures of cognitive processing during learning. Scholars note the need for useful measures of cognitive processing during learning for further development of cognitive load theory and its applications (Paas, 1992; Plass, Moreno, & Brünken, 2010; Sweller, 1988; Sweller et al., 2011). This challenge has resulted in calls for investigation of potential objective measures including brain-activity measures, such as electroencephalography (EEG), and eye-tracking (e.g., deJong, 2010; Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014; National Research Council, 2011).

In the present study, we attempt to heed this call by examining how a well-known instructional design method influences subjective and objective measures of cognitive processing during learning from a

* Corresponding author. University of Copenhage, Oester Farimagsgade 2A, DK-1353, Copenhagen K, Denmark.
E-mail address: gm@psy.ku.dk (G. Makransky).

https://doi.org/10.1016/j.learninstruc.2018.12.001
Received 28 May 2018; Received in revised form 3 December 2018; Accepted 9 December 2018
Available online 09 January 2019

multimedia lesson. The instructional method is the spatial contiguity principle, in which we compare integrated presentation of words and graphics (i.e., printed words are placed next to the corresponding part of the graphic) and separated presentation of words and graphics (i.e., printed words are placed at the bottom of the screen as a caption far from the graphic). Separated presentation is expected to create more extraneous cognitive load (or extraneous processing) during learning than integrated presentation, that is cognitive processing that does not serve the instructional goal caused by poorly designed presentation of the material (Mayer & Fiorella, 2014; Sweller et al., 2011a). Intrinsic cognitive load depends on the number of elements that must be processed simultaneously in working memory and the expertise of the learner (van Merrienboer & Sweller, 2005). Therefore, we manipulate intrinsic cognitive load by dividing our sample into high- and low-performing learners and investigate which cognitive process measures are diagnostic of successful learning. The learning outcome is measured by an open-ended retention posttest and an open-ended transfer test. The retention test is designed to measure how much students remember, and the transfer test is designed to measure a student's understanding of the presented material (Mayer, 2014). The main new element in this study is to include a collection of learning process measures. These include both subjective measures–self-report ratings of cognitive processing during learning on a nine-point scale for three items–and objective measures–EEG measures of alpha and theta band power and eye-tracking measures of dwell times on various areas of the screen.

1.2. Measuring cognitive load in learning

Cognitive load refers to the load imposed on working memory by the cognitive processes that learning materials evoke (Sweller et al., 2011). Although different indirect measures of cognitive load were originally used to provide evidence for CLT including error rates (e.g., Ayres & Sweller, 1990) and learning times (e.g., Chandler & Sweller, 1991 or 1992), subjective measures based on rating scales became and remain the most common method for measuring cognitive load in education (Sweller et al., 2011). Other measures of cognitive load include efficiency measures (e.g., Paas & van Merriënboer, 1993), secondary tasks based on dual-task methodology (e.g., Brünken, Steinbacher, Plass, & Leutner, 2002), and physiological measures (e.g., Antonenko, Paas, Grabner, & Van Gog, 2010). In this paper we focus on subjective measures due to their long-standing influence in learning science and on objective measures as they represent a promising new direction in the continuous measurement of cognitive load (Sweller et al., 2011). Concerning objective measures, we employ EEG measures and eye-tracking measures to determine whether they add value to our understanding of multimedia learning processes.

Subjective Measures of Cognitive Load. A commonly used subjective measure of cognitive load in educational science is a one item measure developed by Paas (1992). The item is on a 9-point Likert scale ranging from *very, very low mental effort* (1) to *very, very high mental effort* (9) where learners have to rate their mental effort while learning. Paas, Tuovienen, Tabbers, and van Gerven (2003) have defined mental effort as "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). Although this scale is widely used, it is limited because it only provides an overall concept of cognitive load that does not reflect its multidimensional character (Ayres, 2006). Based on the assumption that intrinsic and extraneous cognitive load add to total cognitive load, researchers have argued that it is possible to assess intrinsic cognitive load by keeping extraneous cognitive load constant (e.g., Ayres, 2006), or to assess extraneous cognitive load by using different instructional methods to teach the same material (e.g., DeLeeuw & Mayer, 2008). In an attempt to measure these different types of cognitive load directly, there has been a common tendency to align the wording of the subjective items to assess different types of load (Sweller et al., 2011). For example, Cierniak, Scheiter, and Gerjets (2009, p. 318) use wordings such as, "How difficult was the learning content for you?" to assess intrinsic cognitive load, and, "How difficult was it for you to learn with the material?" to assess extraneous cognitive load. Although these types of measures are commonly used (e.g., Corbalan, Kester, & van Merriënboer, 2008; Gerjets et al., 2006, 2009), they have also been criticized on the grounds that learners may be incapable of making the required distinctions. For example, students may not be able to distinguish between whether they are experiencing difficulty due to the complexity of the material or inadequate instructional design (e.g., Kirschner, Ayres, & Chandler, 2011).

Using EEG to Measure of Cognitive Load. Electroencephalography (EEG) is a common neuroimaging technique that measures the gross electrical activity of the brain generated by millions of neurons firing at the same time, which produces a large enough electrical potential that it is measurable along the scalp (Breedlove & Watson, 2013; Pinel, 2011). EEG may have potential value for measuring cognitive load during learning due to having several potential advantages over subjective measures (Antonenko & Keil, 2017). These advantages include the potential to obtain objective measures of cognitive load, rather than relying on learners' subjective ratings. Furthermore, physiological measures are sensitive to variations over time and can be collected while learning is taking place, rather than relying on measuring cognitive load after the learning is complete (Van Gog, Rikers, & Ayres, 2008). EEG has a high temporal resolution enabling it to measure changes in cognitive load on the millisecond scale (Antonenko et al., 2010). Another advantage is that the method makes it possible to reflect various temporal types of load, such as instantaneous, peak, average, accumulated, as well as overall load (Antonenko & Niederhauser, 2010; Xie & Salvendy, 2000). In spite of these advantages, there are still many open questions and issues regarding the use of EEG for cognitive load measurement.

Preliminary studies have shown that oscillatory brain activity, as measured with EEG, vary predictably in response to changing levels of cognitive stimuli (Anderson & Bratman, 2008; Klimesch, 1999). A correlation between working memory load and power in distinct EEG spectral frequency bands has been observed across numerous empirical studies (Brouwer et al., 2012). Several studies have pointed to alpha (8–12 hz) and theta (4–8hz) bands in particular as being significant indicators of working-memory workload (Antonenko & Keil, 2017; Berka et al., 2007). Alpha is generally associated with attentional processes (Frey, Mühl, Lotte, & Hachet, 2014). There is widespread consensus among scholars that when demands of attentional processing increase, alpha spectral power is expected to decrease, particularly in parietal and occipital electrode sites (Klimesch, 1999; Mühl, Heylen, & Nijholt, 2015; Puma, Matton, Paubel, Raufaste, & El-Yagoubi, 2018). Alpha band power has been empirically demonstrated to decrease with increased task difficulty and increased memory load (Fairclough & Venables, 2006; Puma et al., 2018). A decrease of alpha power has also been linked to increase in cognitive arousal, resource allocation, and workload (Brouwer et al., 2012). Studies also have shown that suppression of alpha oscillations is strongly associated with semantic memory processing in particular, such as searching, accessing, and retrieval of information from long-term memory (Klimesch, 1996; Klimesch, 1999).

Theta frequency especially in frontal areas has been linked to working memory capacity across several studies (Puma et al., 2018). In these studies, increasing levels of spectral power in the theta band was proposed to reflect increasing memory load (Mühl et al., 2015). The positive relation between theta and mental effort has also been explicated in reviews by Klimesch (1996, 1997, 1999).

The combination of desynchronization of alpha and synchronization of theta has frequently been used as an index of cognitive workload (Berka et al., 2007; Matthews, Reinerman-Jones, Barber, & Abich, 2014; Rabbi, Zony, Leon, & Fazel-Rezai, 2012; Puma, 2018:; Gerjets

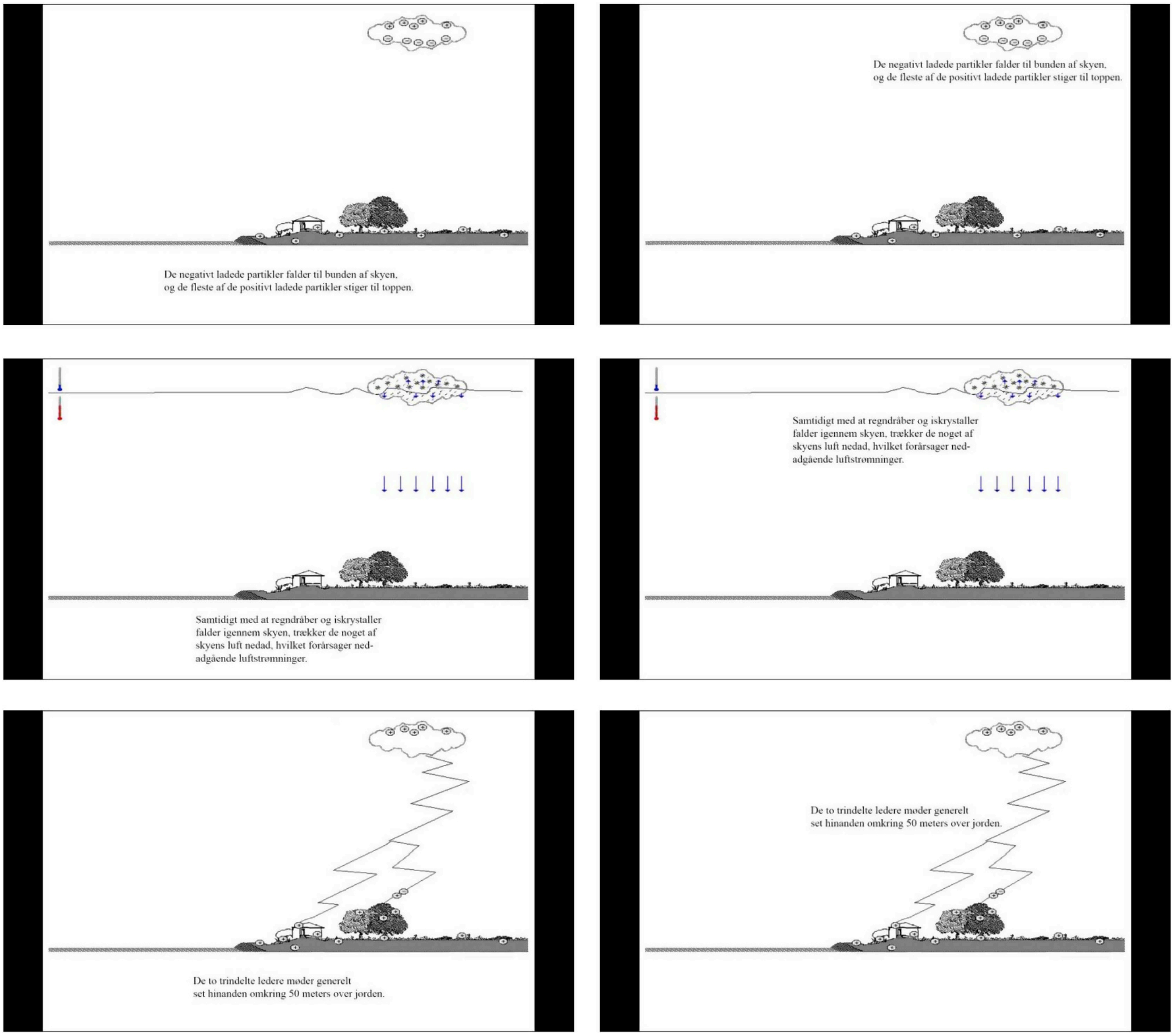

Fig. 1. Screenshots of the multimedia lightning lesson with the non-spatial contiguity (NSC) lesson on the left and the spatial contiguity lesson (SC) on the right.

et al., 2014; Antonenko et al., 2010; Givens et al., 1988; Sauseng et al., 2005). For instance, Gevins and Smith (2003) showed that theta and alpha frequency band effects correlated with task difficulty levels in simulated flight tasks and n-back tests. Mobile EEG methodologies seem particularly promising for learning science, because they offer a non-invasive and relatively unobtrusive means of measuring cognitive load with a high level of ecological validity (Gerjets et al., 2014).

There are several research examples of how EEG can be used to provide valuable information about the learning process within multimedia research. Antonenko and Niederhauser (2010) used subjective and EEG measures to investigate learning with hypertexts. They concluded that alpha, beta, and theta measures from EEG were significantly lower when hypertext was used. By using an EEG measure that was sensitive to instantaneous cognitive load, they were able to conclude that hypertext lead to lower cognitive load. In another study, Makransky, Terkildsen, and Mayer (2017) used EEG to measure cognitive load in learning with an immersive vs. a desktop version of a virtual reality (VR) simulation. They found that although the immersive VR version of the simulation lead to higher self-reported presence, it also led to lower learning outcomes. Based on the EEG measure of cognitive load they concluded that the immersive VR version created cognitive overload in students late in the learning process. These studies are examples of how the temporal resolution of EEG can provide important information about the learning process.

Although there seems to be potential for using EEG as a measure of cognitive processing during learning, more focused research evidence is needed. Several researchers have recommended combining EEG measures of cognitive load with outcome measures such as tests of retention and transfer and self-reported cognitive load (e.g Antonenko & Niederhauser, 2010; Lee, 2014), as well as the need to combine EEG process measures with other process measures such as eye tracking (Mills et al., 2017). This is the approach used in this study.

Using Eye Tracking to Gain a Better Understanding of Cognitive Processing During Learning. Eye tracking data within multimedia learning has been used to investigate how different design interventions such as spoken vs. written text and developing expertise affect processing of multimedia material. Eye-tracking can provide a process measure of cognitive load by calculating the amount of time spent gazing at a specific point in a multimedia lesson (Sweller et al., 2011; van Gog & Scheiter, 2010). Since there is quite a lot of emerging research on the use of eye tracking in multimedia learning (Holmqvist et al., 2011), we focus specifically on the research related to the use of eye tracking as a means of understanding the underlying mechanisms of the spatial contiguity principle, which is relevant for the present study.

Table 1
Which Measures of Learning Process are predicted to be Sensitive to Extraneous Cognitive Load and Diagnostic of Successful Learning Outcomes?.

Measure	Target cognitive processes during learning	Sensitive to extraneous load	Diagnostic of successful learning outcome
Self-report (rating)			
Mental effort	Experienced effort	NO	YES
Experienced difficulty	Experienced difficulty of learning	YES	YES
Eye-tracking (dwell times)			
Relevant illustrations	Manifested ease of learning (Essential and germane processing)	YES	YES
Irrelevant illustrations	Manifested ease of learning (Extraneous processing)	YES	YES
Text	Manifested ease of learning (Essential and germane processing)	YES	YES
Outside AOI	Manifested ease of learning (Extraneous processing)	YES	YES
EEG			
Alpha band activity	Attentional processing, semantic load, task difficulty	YES	YES
Theta band activity	Mental effort	NO	YES
Learning outcome			
Transfer test	Selecting, organizing, integrating	YES	na
Retention test	Selecting	NO	na

Schmidt-Weigand, Kohnert, and Glowalla (2010) used an animation on lightning formation in two experiments to uncover how learners split their visual attention between the animation and the printed words. Their conclusion was that learners' viewing behavior is largely guided by the text. Johnson and Mayer (2012) used eye tracking in three experiments and found that the integrated groups made significantly more eye-movements from the text to the corresponding part of the diagram than the separated groups. They were able to conclude that spatial contiguity encourages more attempts to integrate words and pictures and enables more successful integration of words and pictures during learning based on the eye-tracking data. Similarly, Bauhoff, Huff, and Schwan (2012) used eye tracking to investigate the spatial contiguity principle and found that gaze shifts decreased, and dwell times increased as the distance between two sources of information increased. The authors contended that as gaze shifts decreased, demands on working memory increased, because larger chunks of information had to be held in working memory between gazes. Dwell times are defined as the amount of time a student fixates or gazes within or outside defined AOIs (Holmqvist et al., 2011). In usability research, the average dwell time for a certain AOI is proposed to indicate the amount of cognitive and visual processing allocated to elements in this area (Poole & Ball, 2006; iMotions, 2018; Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002).

1.3. Spatial contiguity principle in multimedia learning

Consider a situation in which a learner views a multimedia lesson on how lightning storms develop consisting of a series of slides with captions at the bottom of each slide, as shown in left side of Fig. 1. In terms of cognitive processing during learning, this situation can place a heavy cognitive load on working memory, because learners have to split their attention between and mentally integrate the illustrations on the top of the screen and the corresponding text at the bottom of the screen. One attempt to alleviate this split attention situation is to move the text next to part of the illustration that it refers to, as shown in the right side of Fig. 1. This approach to spatially integrating words and illustrations is intended to reduce cognitive load by signaling to learners how to connect words with corresponding parts of the illustration. The right side of Fig. 1 reflects an implementation of the *spatial contiguity principle* based on the cognitive theory of multimedia learning (CTML; Mayer, 2009), which states that people learn better from multimedia lessons when printed words are placed next to the corresponding part of the illustration. The rationale behind the principle is that extraneous cognitive load is increased by the need to mentally integrate the multiple sources of information that are not in spatial congruence, and that too much extraneous cognitive load can overload the cognitive system, which detracts from learning the presented material.

Why does spatial contiguity aid learning? The theoretical mechanism underlying the spatial contiguity principle is that extraneous cognitive load is higher for separated presentations such as shown on the left side of Fig. 1 than for integrated presentations as shown on the right side of Fig. 1. The spatial contiguity principle is one of the most researched multimedia principles (Mayer & Fiorella, 2014). A meta-analysis of the spatial contiguity principle identified 37 studies and found a mean effect size of $d = 0.72$ favouring the integrated group over the separated group on measures of learning outcome (Ginns, 2006). More recently, a meta-analysis by Mayer and Fiorella (2014) found that students performed better on transfer posttests with integrated presentations rather than separated presentations in 22 out of 22 experiments, yielding a median effect size of $d = 1.10$.

Although a lot is known about the spatial contiguity principle, most previous studies have investigated the principle in experiments using learning outcome posttests and/or self-report measures of cognitive load administered after the lesson. Although these measures provide useful information, in the present study we wish to add objective process measures recorded during learning in order to provide a more direct test of the theoretical mechanism assumed to be involved in the spatial contiguity principle. In particular, we explore whether the proposed objective measures are sensitive to the expected changes in extraneous cognitive load caused by the instructional method manipulation.

1.4. Theoretical background and predictions

Which Learning Outcome Measures Are Sensitive to Extraneous Cognitive Load? According to the *spatial contiguity principle*, students learn better from multimedia lessons in which printed words are placed near to rather than far from corresponding graphic elements on the screen (Mayer & Fiorella, 2014). To test the principle, we compare the learning outcome posttest performance (particularly on a transfer posttest) of students who learned with an integrated versus separated version of the lesson containing the same words and graphics. As shown in Table 1, we expect the integrated group to perform better than the separated group on a transfer test which assesses the quality of the acquired schemas during the learning phase, but not necessarily on a retention test. This is an indirect way to assess cognitive processing during learning. Specifically, we posit that the integrated group can more efficiently use its limited processing capacity for selecting, organizing, and integrating, which leads to better transfer test performance. In contrast, the separated group uses more of its limited processing capacity for extraneous processing (i.e., inefficiently) leaving some remaining capacity for the cognitive process of selecting, but not for the additional deeper processing of organizing and integrating. According to this analysis, the transfer test is intended to be a gauge of extraneous processing, because increased extraneous processing saps processing capacity that cannot be used for essential processing needed to support transfer. Consistent with previous research, hypothesis one is that the integrated group will perform better on a transfer test than the separated group but not necessarily on a retention test.

Which Learning Process Measures Are Sensitive to Extraneous Load? The major new contribution in this study concerns the addition of direct measures of cognitive processing during learning. The underlying learning mechanism consists of appropriate cognitive processing during; learning, including attending to relevant rather than irrelevant material in the lesson, organising the relevant material into a coherent structure, and integrating corresponding verbal and pictorial information with each other and with relevant prior knowledge activated from long-term memory. In the present study, we expect the separated presentation method to cause more extraneous processing, such as scanning the screen for how to connect printed words with corresponding parts of the graphics (Mayer, 2009; Mayer & Fiorella, 2014).

As summarized in Table 1, this allows us to administer a collection of subjective and objective measures to determine their sensitivity to extraneous cognitive processing, that is, their ability to distinguish between the integrated and separated groups. Some measures are intended to gauge extraneous processing and some gauge overall cognitive effort. As can be seen in the second column of Table 1, eye-tracking measures of dwell times are intended to provide an objective indication of whether cognitive resources are being spent on processing relevant material or whether resources are being used unnecessarily to process extraneous material. Based on this, we argue that dwell times within and outside of defined areas of interest can provide insight into the students' learning process and be considered as an operationalised indicator of extraneous processing. In particular, dwell times on irrelevant illustrations and on outside areas indicate inefficient visual scanning (high extraneous processing) whereas dwell times on text and relevant illustrations indicate efficient scanning (low extraneous processing). Hypothesis 2 is that the integrated group will spend less dwell time on irrelevant material and spend more time on processing relevant material than the separated group.

As also can be seen in Table 1, a self-report item about the ease or difficulty of learning experienced by the learner can be seen as a subjective measure of extraneous processing. Hypothesis 3 is that the integrated group will produce a lower mean rating of how much difficulty they experienced than the separated group.

In contrast, the self-report measure of mental effort and the EEG measure of theta band activity is intended to provide an indication of the overall level of cognitive effort. We expect students in both groups to use their available cognitive capacity, with those in the integrated group to mainly engage in appropriate cognitive processing (i.e., selecting, organizing, and integrating) and those in the separated group to divert some of the processing to extraneous processing (and away from organizing and integrating). Thus, the groups could engage in equivalent levels of cognitive effort, as would be indicated by no difference on the self-report item that focuses on mental effort (hypothesis 4) and by no significant difference in theta band power (hypothesis 5). In line with hypothesis 3, it is expected that mean spectral power in the alpha band frequency will be lower in the separated condition, because the added extraneous processing will lead to higher levels of task difficulty, which desynchronization of alpha activity has been shown to reflect (hypothesis 6).

Which Learning Process Measures Are Diagnostic of Successful Learning? Another way to validate learning process measures is to compare subjective and objective measures for students who score high versus low on learning outcome score after viewing a multimedia lesson. In short, we wish to know which learning process scores are diagnostic of learning outcome scores. This can be seen as a manipulation of intrinsic cognitive load because it depends on the number of elements that must be processed simultaneously in working memory and the expertise of the learner (van Merrienboer & Sweller, 2005). Therefore, a large number of interacting elements for one person might be a single element for another more experienced person who has a schema that incorporates the elements.

As mentioned, an increase in theta band power has been associated with increased mental effort and increased working memory capacity. Thus, it is expected that a significant increase in theta will be observed for the high-performing students as we expect them to have invested more mental effort and therefore managed to process more information in working memory overall (hypothesis 7). Alpha band power is also expected to be diagnostic of successful learning as alpha band power has been empirically demonstrated to decrease with increased task difficulty and increased semantic memory load (Fairclough & Venables, 2006; Puma et al., 2018). Alpha should be lower for low-performing students, who are expected to have found the material more difficult than high-performing students. Therefore, hypothesis 8 is that alpha band power will be lower in the low-performing group than the high-performing group.

Similarly, measures of ease of learning process such as indicated by eye dwell times on relevant rather than irrelevant areas could be diagnostic because high-performing students have better executive control of their learning process (direction of attention etc.). Thus, high-performing students should exhibit higher percentages of time spent dwelling on relevant material than low-performing students (hypothesis 9).

We expect subjective ratings of mental effort and experienced difficulty of learning to be diagnostic, with the high-performing group reporting more mental effort (hypothesis 10) and experiencing lower levels of perceived task difficulty than the low-performing group (hypothesis 11).

2. Method

2.1. Participants and design

The sample consisted of 78 college students recruited from a European university. The mean age was 23.59 years (SD = 3.46) with a range from 19 to 45. There were 47 women and 31 men. The mean score on a prior knowledge survey was 3.35 (SD = 2.40) out of 11 which indicates low prior knowledge. In a between-subjects design, 41 students served in the spatial contiguity (SC) group, which received an integrated presentation, and 37 served in the no spatial contiguity (NSC) group, which received a separated presentation. All participants were required to be fluent in Danish and have normal vision without the use of eyeglasses to avoid potential problems with the eye tracking measures. The study protocol received IRB approval and was conducted according to the Helsinki Declaration in which every participant received both oral and written information about the study. The participants all gave written informed consent.

2.2. Procedure

Participants were randomly assigned to the SC or NSC group prior to arriving at the experimental location. Students were tested individually in a sound-proof learning lab at a European university. After giving the participant initial information and the informed consent form the experimenter fitted the participant with the ABM EEG system, and gave oral instructions on how to complete the following EEG benchmark.

The experimenter left the room each time after instructions were provided, so the participant was alone in the room when the experimental tasks were performed. The participant first completed the participant questionnaire. Then, the participant was presented with the multimedia lightning lesson corresponding to their assigned treatment group (SC or NSC). Immediately after the lesson, the participant was administered the retention test with a 4-min time limit and transfer test with 2 min allowed for each item. Finally, the participant completed the self-report survey. The average run time for each participant was about 45 min. Each participant was compensated for their time with a gift card valued at 100 Danish crowns (about 15 Euros) upon completion. The difference in sample sizes between the SC (n = 41) and NSC (n = 37) conditions occurred because approximately 11% of the respondents who were scheduled did not show up to the experiment.

2.3. Materials and apparatus

The materials used in the study included a spatial contiguity version of a multimedia lightning lesson (i.e., integrated presentation), no spatial contiguity version of a multimedia lighting lesson (i.e., separated presentation), participant questionnaire, retention test, transfer test, and self-report survey designed to measure different forms of cognitive load. All of the information was presented in Danish. Furthermore, all of the materials were administered on a high-end desktop computer and presented to the participants on an external 24-inch computer monitor (HP EliteDisplay E232) with a resolution of 1920 × 1080 and a screen luminance of 250 lumen. Screen and ambient luminance in the lab was kept stable for all participants across conditions. Students' eye movements were tracked by a Tobii eye-tracking system (Tobii X2-30) and students' brain activity was monitored using a 9-channel AMB EEG headset system (B-Alert X10).

Multimedia lightning lessons. The two multimedia lessons consisted of 16 PowerPoint slides describing how a lightning storm develops based on materials used by Mayer and Moreno (1998). Each slide contained an illustration and printed text. As shown in Fig. 1, for the NSC lesson, the words were presented as a caption at the bottom of the slide; whereas for the SC lesson, the words were presented next to the corresponding part of the illustration. The length of time that a slide was presented was determined based on previous literature that has used this lesson and a pilot test where learners were presented with each slide, and the time required to sufficiently process the information was recorded. For each lesson, the total presentation was two and a half minutes.

Participant questionnaire. The participant questionnaire solicited demographic information including age and gender and asked about prior knowledge of lightning. A prior knowledge rating was used instead of a pretest to avoid signaling important content before beginning the lesson (Mayer, 2014). The prior knowledge rating consisted of eight items where the first one asked participants to rate their knowledge of how lightning works with the categories: very high, somewhat high, medium, somewhat low, very low. The other seven items asked students to place a checkmark next to the things that they had done and included items such as: "I know what AC and DC is" and "I know what this symbol means" (followed by a symbol for cold front or warm front). The Cronbach's alpha reliability of the prior knowledge measure was 0.77.

Retention and transfer test. The retention and transfer tests were identical to the ones used in the original experiment by Mayer and Moreno (1998). The retention test contained the following instructions at the top of a blank Microsoft word sheet: "Please write down an explanation of how lightning works." This was followed by the instructions at the bottom of the sheet: "Please keep working until you are told to stop." Students were given 4 min to work on this part of the test. The transfer test consisted of the following four questions, each presented on a separate sheet: "What could you do to decrease the intensity of lightning?" "Suppose you see clouds in the sky, but no lightning. Why not?" "What does air temperature have to do with lightning?" and "What causes lightning?" Each of these questions was followed by the text: "Please keep working until you are told to stop." Participants were given 2 min to work on each transfer question. The Cronbach's alpha reliability of the knowledge test was 0.73, and was calculated by counting the number of major idea units in each of the 16 slides. The inter-rater reliability of the retention and transfer tests was assessed by comparing the scores from the two independent raters. An acceptable correlation was obtained between the two ratings for the retention test (r = 0.818), and the transfer test (r = 0.808). Therefore, the average of the scores from the two raters is used in the subsequent analyses.

Self-report survey. The self-report survey consisted of rating items intended to measure general cognitive load, and extraneous cognitive load. The item for general cognitive load asked students to rate their "perceived amount of mental effort" in the lesson they just finished on a 9-point rating scale consisting of "I invested: 1. very, very low mental effort/2. very low mental effort/3. low mental effort/4. rather low mental effort/5. neither low nor high mental effort/6. rather high mental effort/7. high mental effort/8. very high mental effort/9. very, very high mental effort" and was based on Paas (1992). Reported effort is seen as an index of general cognitive load (see Paas, Van Merrienboer, & Adam, 1994, p. 420) and this item is commonly used in the literature. One item was used with the intention of measuring extraneous load (from Cierniak et al., 2009). The item asked participants to: "Please choose the category (1, 2, 3, 4, 5, 6, 7, 8 or 9) that applies to you: To learn from the lesson was 1. very, very easy/2. very easy/3. easy/4. rather easy/5. neither easy nor difficult/6. rather difficult/7. difficult/8. very difficult/9. very, very difficult".

Measurements with EEG. The EEG data was collected using an Advanced Brain Monitoring (ABM) X-10, wireless 9-channel EEG system digitalized at 256 Hz. The X-10 records data in real-time from nine Ag/AgCl electrodes spread across the scalp in accordance with the International 10–20 electrode placement system (F3, Fz, F4, C3, Cz, C4, P3, POz, P4), along with a linked mastoid reference. Impedances levels were measured for all electrodes and were kept below ABM's recommendations of a 40 kOhm threshold (ABM, 2018).

In order to prepare the EEG signals for power spectral analysis, the signals were first bandpass filtered between 0.5 Hz and 100 Hz and then re-referenced to an average reference. A notch filter was then applied at 50 Hz to reduce the line noise that was visible in the spectra plot. After initial filtering, the time-series domain of the signals were plotted and visually inspected for excessive noise artifacts, which were then rejected. Following the manual visual artifact rejection, ICA was performed specifically to remove components containing eye blinks and horizontal eye movement artifacts. These processing steps were performed using the open-source EEGLAB toolbox (Delorme & Makeig, 2004).

The Neurospec toolbox (Neurospec 2.0, Neurospec.org) was employed to extract the frequency-domain estimates of the signal from the nine electrodes through a Discrete Fourier Transform (DFT) with a Hanning window length of 2000 ms specified as a segment length of 2^9 and sample rate of 256. This defines a spacing of the Fourier frequencies returned as 0.5 Hz (Bloomfield, 2000; Halliday et al., 1995; Nielsen, Conway, Halliday, Perreault, & Hultborn, 2005). The signals were normalized and log-transformed to reduce the potential unit variance and distribution skew and to compute the log spectral density estimates. Mean estimates for the alpha and theta bands were then computed (decibel power/hz).

Measurements with Eye tracking. Binocular eye tracking data was collected using a Tobii X2-30 recorded on the iMotions research software platform. The Tobii X2-30 is a screen-based (also called remote or desktop) eye tracker with a sampling frequency of 30 Hz and an operating distance between 40 and 90 cm. The continuously measured distance between the participants and the screen during the lesson was on average 57.93 cm with a standard deviation of 3.41. Before the study lesson started, each participant performed a nine-point gaze calibration designed by Tobii, where they had to follow a white dot with their gaze

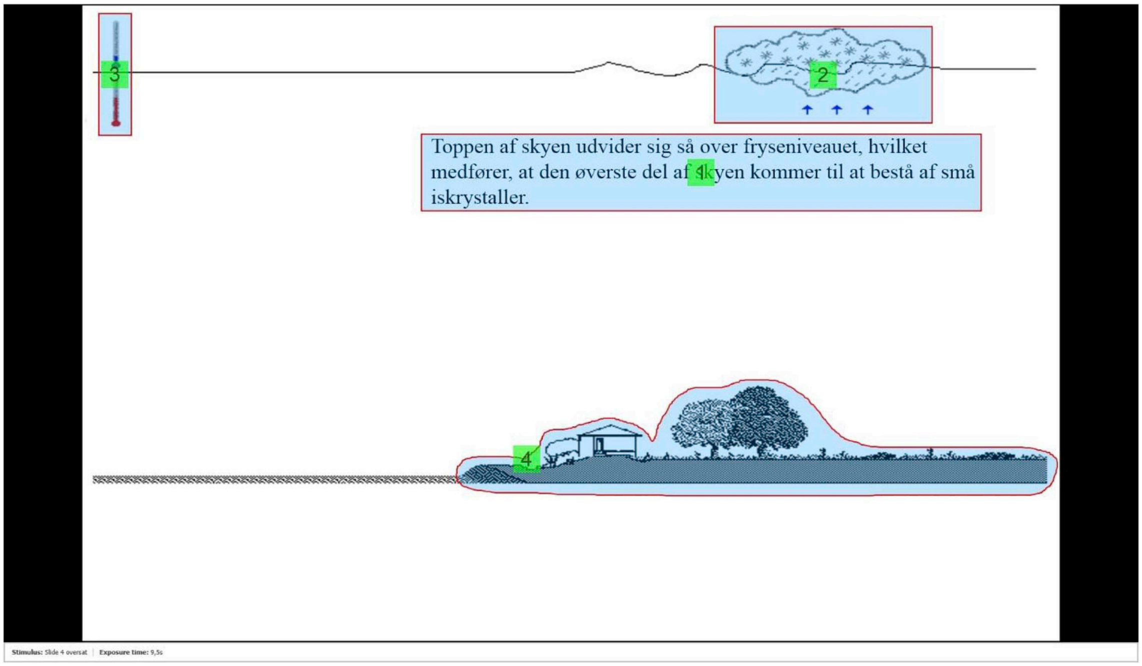

Fig. 2. An illustration of the four AOIs from slide 4 in the SC-condition.

around the screen. Students continued to the lesson when Tobii's calibration validation stated that a good or excellent calibration had been achieved based on their accuracy and precision thresholds. The threshold for classifying a good calibration requires a mean distance of measured gaze data from the target calibration point being $\mu(x,y) \leq 40$ pixels, whereas an excellent-calibration necessitates a mean distance of $\mu(x,y) \leq 20$ pixels. Fig. 2 shows an example of the four areas of interests (AOIs) that were defined and identified on each slide for both lessons: (1) text, (2) cloud and warm/cold front illustration, (3) thermometer illustration and (4) ground illustration. The three illustration AOIs were further divided into two categories, relevant (2 and 3) and non-relevant illustrations (4). The text used on the slides were written in Times New Roman with a font size of 26.

The key measures were: percentage of time spent on text; percentage of time-spent on relevant illustrations; percentage of time-spent on non-relevant illustrations and percentage of time spent outside of AOIs. The percentage of time spent measures were based on calculations of how long a participant's gaze and fixations were recorded inside of the AOIs compared to the total duration of the slides. Measures based on gaze-location and fixations were calculated using a duration dispersion-based algorithm, that continuously searches for gaze point inputs that fall within 1-degree radius of each other for a minimum duration of 100 ms. The fixation-centroid (x, y coordinates) is also recalculated continuously as more gaze points are added to the fixation. The gaze data and fixations were then compared to the location of the AOIs to produce the fixation- and gaze-based measures mentioned above. The percentage time spent measures are advocated to indicate the respective dwell time within or outside the defined AOIs. Fixation counts were not included as they are reflected in the dwell time-measures and preliminary analysis revealed that these measures resulted in equivalent conclusions. Eye tracking data quality, i.e. the amount of valid gaze data that was recorded for all slides across both condition averaged 95.44% of the maximum amount of possible gaze data (30 inputs a second). The gaze samples recorded that were deemed invalid are either not within the valid coordinates on the screen, result of blinking or participants looking away, or the participants temporarily moved outside of the operating viewing distance.

Data from the surveys, tests, the EEG, and the eye tacking data were collected using the iMotions research software platform, which permits synchronization of all of these measures and allows for accessible data analysis of these measures (see iMotions.com for further information regarding the platform). The data was exported to IBM SPSS version 23.0 for statistical analyses.

3. Results

3.1. Scoring

The ABM EEG did not work for three of the participants in the study due to various technical issues; so, EEG measures of cognitive load were only obtained for 75 participants. Data was available for all participants for the other variables in the study. Two raters who were blind to the treatment group of the participants independently scored the retention and transfer tests. A retention score was computed for each participant by counting the number of major idea units (out of 16 possible) that the participant produced on the retention test. Each of the idea units was linked to the main idea of one of the 16 slides in the multimedia lesson. The list of acceptable answers was based on the original article by Mayer and Moreno (1998). Answers on the transfer test were coded for correct answers, based on an acceptable list of answers generated by Mayer and Moreno (1998). The sample was divided at the median into high and low performers by combining the transfer and retention tests. The tests were combined by transforming the scores on each test to a z-score and summing them together.

3.2. Did the groups differ on basic characteristics?

A preliminary step is to determine whether the SC and NSC groups were equivalent on basic characteristics. Based on t-tests, the groups did not differ significantly on mean age, t (76) = 0.599, p = .551, or mean prior knowledge score, t (76) = 0.867, p = .389, and based on a chi-square test, X^2 (2, N = 78) = 0.744, p = .819, the groups did not differ significantly in the proportion of men and women. We conclude that the groups did not differ on basic characteristics that existed before the start of the experiment.

Table 2
Means, SD, P-values, and effect sizes for the tests, eye tracking data, and cognitive load measures by learning condition.

Variables	SC (n = 41)		NSC (n = 37)		p-value	d
Tests	Mean	SD	Mean	SD		
Transfer	5.17	3.32	3.62	3.05	**.036**	.49
Retention	13.07	5.97	12.14	6.07	.494	.15
Eye tracking						
% time text	72.60%	8.65	65.71%	12.76	**.006**	.64
% time relevant illustrations	10.01%	4.19	9.85%	5.01	.875	.03
% time non-relevant illustration	3.18%	1.78	5.42%	2.29	**.000**	1.10
% time outside AOIs	14.20%	6.03	19.02%	11.52	**.021**	.55
Self-report						
Experienced difficulty	4.88	1.73	5.89	1.81	**.014**	.57
Mental effort	5.32	1.65	5.41	1.48	.805	.06

Note. SC = spatial contiguity group; NSC = non-spatial contiguity group; AOI = area of interest; Bold font indicates significant effect at $p < .05$.

3.3. Which Learning Outcome Measures Are Sensitive to extraneous cognitive load?

Hypothesis 1 is that the SC group would perform better on the transfer test than the NSC group but not necessarily on the retention test. The top line of Table 2 shows the mean (and *SD*) for each group on the transfer test. Consistent with predictions, the SC group ($M = 5.17$, $SD = 3.32$) scored significantly higher than the NSC group ($M = 3.62$, $SD = 3.05$) on the transfer test, t (76) = 2.14, $p = .036$, $d = 0.49$. Line two in Table 2 shows that the superiority of the SC group ($M = 13.07$, $SD = 5.97$) over the NSC group ($M = 12.14$, $SD = 6.07$) did not reach statistical significance on the retention test, t (76) = 0.687, $p = .494$, $d = 0.15$. These results replicate previous findings supporting the spatial contiguity effect and are consistent with the idea that placing words next to corresponding parts of the illustration helps promote deeper learning of the material. We conclude that hypothesis 1 was supported.

Hypothesis 2 is that the integrated group would spend less time on irrelevant parts of the lesson but spend more time on the relevant parts of the lesson compared to the separated group. A one-way multivariate analysis of variance (MANOVA) was conducted to test the hypothesis that there would be one or more mean differences between the SC and NSC students on the four eye tracking variables. A statistically significant MANOVA effect was obtained, Pillais' Trace = 0.282, $F_{(3, 74)} = 9.687$, $p < .001$. The multivariate effect size was 0.282, which implies that 28.2% of the variance in the canonically derived dependent variables was accounted for. A series of one-way ANOVA's on each of the four dependent variables was conducted as follow-up tests to the MANOVA. Significant one-way ANOVA's were obtained for time spent on text $F_{(1, 76)} = 7.930$, $p = .006$, time spent on irrelevant illustrations $F_{(1, 76)} = 23.279$, $p < .001$, and time spent outside AOIs $F_{(1, 76)} = 5.524$, $p = .021$, but the difference was not significant on time spent on relevant illustrations $F_{(1, 76)} = 0.025$, $p = .875$. The next four lines in Table 2 show that the SC group ($M = 72.60\%$, $SD = 8.65$) spent a significantly higher percentage of time looking at the text than the NSC group ($M = 65.71\%$, $SD = 12.76$); the NSC group ($M = 5.42\%$, $SD = 2.29$) spent a significantly higher percentage of time looking at the non-relevant parts of the illustration than the SC group ($M = 3.18$, $SD = 1.78$); the NSC group ($M = 19.02$, $SD = 11.52$) spent significantly more time looking outside of the AOIs compared to the SC group ($M = 14.20$, $SD = 6.03$), but the superiority of the SC group ($M = 10.01\%$, $SD = 4.19$) compared to the NSC group ($M = 9.85\%$, $SD = 5.01$) in percentage of time spent looking at the relevant parts of the illustration did not reach statistical significance. We conclude that hypothesis 2 concerning the use of eye tracking as an objective measure of understanding the differences between SC and NSC groups was partially supported.

Hypothesis 3 is that the integrated group would produce a lower mean rating of how much difficulty they experienced compared to the separated group. The next line in Table 2 shows the mean self-rating (and *SD*) on the subjective experienced difficulty (extraneous load) item. As predicted, the t-test indicated that the NSC group ($M = 5.89$, $SD = 1.81$) reported that they experienced significantly greater difficulty than the SC group ($M = 4.88$, $SD = 1.73$), t (76) = 2.527, $p = .014$, $d = 0.57$. We conclude that hypothesis 3 concerning a subjective measure of extraneous load was supported.

Hypothesis 4 is that the groups were not expected to differ on the self-report item that focuses on effort. The final line in Table 2 shows that the groups did not differ significantly on mental effort, t (76) = 0.248, $p = .805$, $d = 0.06$. We conclude that hypothesis 4 concerning self-reported mental effort was supported.

Hypothesis 5 is that there would not be significant differences in theta band activity between the SC and NSC groups. A one-way multivariate analysis of variance (MANOVA) was conducted to test the hypothesis that there would be one or more mean differences between the SC and NSC group on the 9 electrodes for the mean theta waves. The MANOVA was not significant, Pillais' Trace = 0.112, $F_{(9, 75)} = 0.911$, $p = .521$. Table 3 reports the mean (and SD) theta and alpha values for each group. We conclude that hypothesis 5 that there would be no difference on theta band activity between the groups was supported.

Hypothesis 6 is that there would be significant differences in alpha band activity between the SC and NSC groups. A one-way multivariate analysis of variance (MANOVA) was conducted to test this hypothesis. The MANOVA for alpha waves was not significant, Pillais' Trace = 0.134, $F_{(9, 75)} = 1.119$, p = .363. We conclude that hypothesis 5 concerning the use of alpha band activity as an objective measure of extraneous processing was not supported.

3.4. Which Learning Process Measures Are Diagnostic of Successful Learning?

Hypothesis 7 is that there would be significant differences in theta band activity between high and low performing students. A one-way

Table 3
Mean theta and alpha band values by learning condition.

Variables	SC (n = 40)		NSC (n = 35)		p-value	d
Theta	Mean	SD	Mean	SD		
POZ	−0.12	0.12	−0.10	0.11	0.358	0.012
FZ	−0.09	0.14	−0.12	0.12	0.348	0.012
CZ	−0.04	0.11	−0.05	0.09	0.449	0.008
C3	−0.26	0.11	−0.24	0.12	0.426	0.009
C4	−0.28	0.12	−0.26	0.11	0.485	0.007
F3	−0.31	0.13	−0.29	0.13	0.445	0.008
F4	−0.35	0.16	−0.33	0.16	0.505	0.006
P3	−0.13	0.09	−0.13	0.10	0.771	0.001
P4	−0.11	0.12	−0.10	0.11	0.682	0.002
Alpha						
POZ	−0.27	0.18	−0.28	0.14	0.826	0.001
FZ	−0.37	0.15	−0.43	0.14	0.069	0.045
CZ	−0.32	0.14	−0.36	0.13	0.210	0.021
C3	−0.42	0.19	−0.45	0.17	0.470	0.007
C4	−0.44	0.19	−0.44	0.18	0.884	0.000
F3	−0.55	0.21	−0.58	0.20	0.532	0.005
F4	−0.61	0.24	−0.62	0.23	0.835	0.001
P3	−0.26	0.14	−0.30	0.13	0.216	0.021
P4	−0.23	0.16	−0.27	0.15	0.306	0.014

Note. SC = spatial contiguity group; NSC = non-spatial contiguity group; Alpha (8–12 hz); Theta (4–8hz).

Table 4
Mean theta and alpha band values for the high- and low-performing students.

Variables	High (n = 37)		Low (n = 38)		p-value	d
Theta	Mean	SD	Mean	SD		
POZ	−0.10	0.09	−0.12	0.14	0.669	0.003
FZ	−0.10	0.11	−0.11	0.15	0.929	0.000
CZ	−0.04	0.10	−0.04	0.11	0.972	0.000
C3	−0.27	0.09	−0.24	0.13	0.310	0.014
C4	−0.29	0.10	−0.25	0.13	0.172	0.025
F3	−0.29	0.12	−0.30	0.15	0.698	0.002
F4	−0.32	0.14	−0.35	0.17	0.399	0.010
P3	−0.13	0.10	−0.13	0.10	0.772	0.001
P4	−0.11	0.11	−0.11	0.12	0.975	0.000
Alpha						
POZ	**−0.24**	**0.13**	**−0.32**	**0.18**	**0.031**	**0.062**
FZ	**−0.36**	**0.14**	**−0.43**	**0.15**	**0.044**	**0.054**
CZ	**−0.30**	**0.15**	**−0.38**	**0.11**	**0.008**	**0.093**
C3	−0.40	0.18	−0.47	0.17	0.100	0.037
C4	−0.40	0.17	−0.48	0.20	0.051	0.051
F3	**−0.51**	**0.19**	**−0.61**	**0.21**	**0.042**	**0.055**
F4	**−0.54**	**0.22**	**−0.68**	**0.24**	**0.008**	**0.091**
P3	−0.26	0.15	−0.29	0.12	0.271	0.017
P4	**−0.21**	**0.15**	**−0.29**	**0.15**	**0.018**	**0.074**

Note. Alpha (8–12 hz); Theta (4–8hz); Bold font indicates significant effect at $p < .05$.

Table 5
Means. SD. P-values. And effect sizes for the cognitive load measures for the high- and low-performing students.

Variables	Learning score					
	High (n = 39)		Low (n = 39)		p	d
Tests	Mean	SD	Mean	SD		
Eye tracking						
% time text	69.33	11.26	68.94	12.96	.763	.03
% time relevant illustrations	10.75	4.57	9.12	4.48	.114	.36
% time non-relevant illustration	3.88	1.88	4.60	2.67	.172	.32
% time outside AOIs	15.64	7.99	17.33	10.52	.425	.18
Self-report						
Experienced difficulty	**4.91**	**1.84**	**5.91**	**1.69**	**.015**	**.56**
Mental effort	5.40	1.47	5.31	1.69	.821	.06

Note. AOI = area of interest; Low = students who performed below the median on posttests; high = students who performed above the median on posttests. Bold font indicates significant effect at $p < .05$.

multivariate analysis of variance (MANOVA) was conducted to test this hypothesis. The MANOVA was not significant, Pillais' Trace = 0.150, $F_{(9,\ 75)} = 1.271$, $p = .270$. We conclude that hypothesis 7 concerning the use theta band activity as an objective measure of intrinsic cognitive load was not supported.

Hypothesis 8 is that there would be significant differences in alpha band activity between high and low performing students. A one-way multivariate analysis of variance (MANOVA) was conducted to test this hypothesis. A statistically significant MANOVA effect was obtained for alpha band activity, Pillais' Trace = 0.241, $F_{(9,\ 75)} = 2.294$, p = .026. The multivariate effect size was 0.241, which implies that 24.1% of the variance in the canonically derived dependent variables was accounted for. As expected, the alpha power band activity was higher for the high performing students compared to the low performing students indicating higher intrinsic cognitive load (see Table 4). A series of one-way ANOVA's on each of the nine dependent variables was conducted as follow-up tests to the MANOVA. Significant differences were obtained for six electrodes including POZ $F_{(1,\ 73)} = 4.842$, $p = .031$; FZ $F_{(1,\ 73)} = 4.203$, $p = .044$; CZ $F_{(1,\ 73)} = 7.452$; $p = .008$; F3 $F_{(1,\ 73)} = 4.265$, $p = .042$; F4 $F_{(1,\ 73)} = 7.347$, $p = .008$; P4 $F_{(1,\ 73)} = 5.835$, $p = .018$. We conclude that the hypothesis 8 concerning the use of alpha band activity as an objective measure of intrinsic cognitive load was partially supported.

Hypothesis 9 is that high performing students would spend less time on irrelevant parts of the lesson but spend more time on the relevant parts of the lesson compared to the separated group. The means and standard deviations for the different groups are presented in Table 5. A one-way multivariate analysis of variance (MANOVA) was conducted to test the hypothesis that there would be one or more mean differences between the SC and NSC students on the four eye tracking variables. The MANOVA was not significant, Pillais' Trace = 0.077, $F_{(3,\ 74)} = 2.070$, $p = .111$. We conclude that the hypothesis 9 concerning the use of eye tracking as a process measure to understand the differences between high and low-performing students was not supported.

Hypotheses 10 and 11 are that the higher-performing group would produce lower ratings of experienced difficulty and mental effort than the lower-performing group. The two bottom lines of Table 5 indicate that the higher performing group ($M = 4.91$, $SD = 1.84$) reported significantly lower experienced difficulty (extraneous load) compared to the lower-performing group ($M = 5.91$, $SD = 1.69$), t (76) = 2.498, $p = .015$, $d = 0.56$. This finding supports hypothesis 10 indicating that experienced difficulty is sensitive to intrinsic cognitive load. However, the difference between the two groups on mental effort was not significant t (76) = 0.248, $p = .805$, $d = 0.06$. We conclude that hypothesis 11 was not supported.

4. Discussion

This study is the first to our knowledge to jointly use EEG, eye tracking, and self-report measures of learning process to investigate the mechanisms underlying the spatial contiguity effect. In particular, this is the first study to investigate the degree to which each of these subjective and objective measures would be sensitive to this cognitive load manipulations across groups. The results of the study suggest that the use of different process measures can provide a better understanding of the fundamental mechanisms underlying multimedia learning by illuminating different aspects of cognitive processing during learning. More specifically, we found that self-report as well as eye-tracking measures were diagnostic of extraneous cognitive load imposed on the lesson based on the special contiguity effect. Alternatively, self-report ratings as well as EEG alpha band activity were diagnostic of intrinsic cognitive load which was manipulated based on differences between high- and low-performance learners.

While many of the findings in this study were consistent with previous research, there were also several important discrepancies. The finding that the SC group scored significantly higher than the NSC group on transfer tests, but not on retention tests is consistent with previous research such as Johnson and Mayer (2012). This reaffirms that transfer tests can provide an indirect measure of extraneous processing during learning. The higher transfer score is assumed to be caused by differences in extraneous processing between the SC and NSC groups, in which learners in the SC group could use their limited capacity for making sense of the material rather than spending resources on extraneous processing.

While the use of spectral power in alpha and theta frequencies ranges is well-substantiated as an index of cognitive load in general, this study found that significant changes in alpha band power was only induced when intrinsic processing load was manipulated and not when extraneous processing load was manipulated. This could indicate that alpha is not necessarily associated with all types of task difficulty as significant differences in perceived task difficulty was reported for both the manipulations (NSC-SC condition and the high-low-learner group), but alpha power was specifically related to changes in difficulty with intrinsic semantic processing tasks (high-low learner group).

Increases in theta band activity have been established to reflect mental effort across several studies (Klimesch, 1996; Klimesch, 1997; Klimesch, 1999). It was therefore hypothesized that no differences would be observed in theta band activity between NSC and SC as the students were expected to invest equal amounts of mental efforts in the task. This hypothesis was supported. The self-reported ratings for mental effort supported the underlying rational for this hypothesis by demonstrating that there were no differences in perceived mental effort invested across these conditions. When comparing the high and low learning groups, it was hypothesised that a significant increase in theta activity in the high-performance group would be observed. This is because we expected that their higher learning outcomes could be attributed to the investment of more mental effort. However, this hypothesis was not supported. A possible explanation for this finding could be that the self-report ratings of perceived mental effort between high and low learners did not differ as we had expected and as such if theta is a measure of mental effort then no differences should be expected.

Eye-tracking technology allows for objective measures and visualization of cognitive processing during learning. In accordance with both Johnson and Mayer (2012) and Schmidt-Weigand et al. (2010), learning was largely text-directed in that much more time was directed at the text compared to the illustrations for both groups. However, unlike these two studies the SC group in this study spent significantly more time on the text compared to the NSC group, yet there was no difference in proportion of time spent looking at the relevant part of the illustration. Conversely, the NSC group spent significantly more time on non-relevant parts of the illustration and they spent significantly more time outside of the AOIs compared to the SC group. Therefore, eye tracking data provides a clear picture that the spatial contiguity principle causes students to spend more time looking at the relevant parts of a multimedia lesson while the lack of spatial contiguity results in students spending time and effort focusing on irrelevant information. This pattern eye-tracking data is consistent with the idea that NSC group engaged in more extraneous processing that the SC group. The eye tracking measures were surprisingly not related to differences between high- and low-performance students. This indicates that the differences between the groups did not come from different viewing patterns but rather from different cognitive processes. Those processes have previously been linked to the observed changes in alpha activity like attention allocation, task difficulty and semantic memory load (searching, accessing and retrieval of useful information from long-term memory).

The only significant difference between the two groups on the subjective cognitive load measures was on the self-reported extraneous cognitive load (experienced difficulty). This is consistent with the idea that a reduction in extraneous cognitive load might be the primary marker of the effectiveness of spatial contiguity. The group of students who learned most also report that it was easier to learn from the lesson, however there were no differences on the self-report measure of mental effort indicating that students in both groups put in equal effort in the lesson.

4.1. Theoretical implications

The findings in the study demonstrate that objective measures of cognitive processing during learning have something useful to contribute to understanding the mechanisms underlying instructional features supposed to affect extraneous cognitive load and intrinsic cognitive load. This is consistent with the notion that cognitive load is a multifaceted construct as proposed by Sweller et al. (2011) and Mayer (2009), rather than a unidimensional construct, reflecting an overall amount of cognitive resources allocated to a task. There are various debates about this conceptualization of cognitive load and questions regarding whether the different types of cognitive load can be distinguished and measured (De Jong, 2010). From a theoretical perspective the use of process measures such as eye tracking and EEG provide a framework to investigate the process by which intrinsic and extraneous cognitive load impact demands on working memory. The results in this study suggest that the working memory demand that is induced from the difficulty of the material (intrinsic cognitive load) may be different from that induced from poor instructional design (extraneous cognitive load).

4.2. Practical implications

Being able to measure cognitive load continuously would provide instructional designers with valuable information that could allow them to design learning material optimally for an individual or a group of students. An ultimate instructional goal would be an instantaneous assessment of cognitive load leading to an immediate online adaptation of instructional material in cases where learners are getting overwhelmed by the difficulty; or bored because the material is too easy compared to their current working memory capacity (Gerjets et al., 2014). This would allow all students to work at their zone of proximal development (Csikszentmihalyi, 1990). Valid measurement techniques that continuously, unobtrusively, and accurately assess cognitive load are necessary for reaching this goal. However, the results in this study provide initial evidence that EEG and eye tracking have potential to obtain this goal. More specifically, this study provides tentative practical insights that EEG and eye tracking measures of learning processes might be sensitive to different types of cognitive load. Specifically, alpha band activity demonstrated sensitivity towards intrinsic load and eye tracking measures of AOI dwell times showed differences when extraneous processing was manipulated.

4.3. Limitations and future directions

There are several important limitations and future directions that need to be considered given that the use of process measures such as EEG and eye tracking in multimedia research is relatively novel. A challenge to this field of study is that cognitive load is hypothesised to be a multifaceted construct including intrinsic and extraneous load (Sweller et al., 2011). Different forms of cognitive load contribute differently to learning and are, therefore, highly relevant to distinguish (De Jong, 2010). At this stage, most of the research in this field has focused on the feasibility of measuring cognitive load with EEG and not to distinguish between intrinsic and extraneous load which is necessary to have a real impact on the science of learning. The research that has identified alpha and theta band activity as potential measure of cognitive load has varied in terms of the interpretations related to the specific cognitive processes that these frequency bands are sensitive to. Therefore, more research using experiments that isolate different forms of cognitive load and investigate whether process measures are sensitive to them is needed.

A limitation to using realistic learning tasks such as the ones used in this study is that they may introduce potential confounds such as different perceptual or motor requirements which can impact the metrics from EEG. Consequently, future research should combine experiments from realistic learning tasks with tasks from working memory research to obtain more comprehensive evaluations of the validity of the process measures that are used. Furthermore, this study and previous research suggests that eye tracking and EEG might be valuable tools for assessing students' cognitive processing during learning. However, more research is needed that investigates the generalisability across multiple learning interactions such as reading, lectures, or using a virtual learning simulation in a domain-independent fashion.

In this study we used eye tracking as a process measure of learning, and a limitation in this study was the use of a fairly slow eye tracker (30 Hz). Future research should investigate the use of other metrics that have been related to cognitive load including saccadic distance (e.g., Inamdar & Pomplun, 2003; Kane, Bleckley, Conway, & Engle, 2001; Litchfield & Ball, 2011; Phillips & Edelman, 2008; Theeuwes, Belopolsky, & Olivers, 2009; Godijn & Theeuwes, 2011) as this was not pursued in this study due to reliability concerns regarding saccade extraction given the low sampling rate of our eye tracker and the low duration and high velocity of saccades.

A future potential limitation in this study was the use of single item self-report measures. These were used because they are the most commonly used measures of cognitive load in multimedia learning. Future research should investigate if the results from this study generalize when psychometrically validated scales (e.g., Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013) are used to measure different types of cognitive load. In general, future research should use subjective and objective process measures of cognitive processing while learning to investigate the mechanisms underlying other instructional design principles.

Acknowledgments

This research was funded by Innovation Fund Denmark. We would like to thank Magnus Omdahl for assisting in the data collection process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2018.12.001.

References

ABM. (2018). Retrieved September 24, 2018, from http://support.advancedbrainmonitoring.com/2015/12/22/what-should-my-impedances-be-what-is-good-impedance/.

Anderson, C. W., & Bratman, J. A. (2008). June). Translating thoughts into actions by finding patterns in brainwave. *Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems* (pp. 1–6). New Haven, CT: Yale University.

Antonenko, P. D., & Keil, A. (2017). *Assessing working memory dynamics with electroencephalography. Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*.

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. https://doi.org/10.1016/j.chb.2009.10.014.

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*(4), 425–438. https://doi.org/10.1007/s10648-010-9130-y.

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*(3), 287–298. https://doi.org/10.1002/acp.1245.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *American Journal of Psychology, 103*, 167–193.

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology, 26*(6), 863–871. https://doi.org/10.1002/acp.2887.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space & Environmental Medicine, 78*(5), 231–244.

Bloomfield, P. (2000). *Fourier analysis of time series: An introduction.* Weinheim: Wiley-Interscience.

Breedlove, S. M., & Watson, N. V. (2013). *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience* (7th ed.). Sinauer Associates, Inc.

Brouwer, A., Hogervorst, M. A., Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering, 9*(4), https://doi.org/10.1088/1741-2560/9/4/045008 045008.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 109–119.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233–246.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*, 733–756.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal performance*.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223–234. https://doi.org/10.1037/0022-0663.100.1.223.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology, 71*(1), 100–110. https://doi.org/10.1016/j.biopsycho.2005.03.007.

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2014). Review of the use of electroencephalography as an evaluation method for human-computer interaction. *Proceedings of the International Conference on Physiological Computing Systems* (pp. 214–223). .

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*, 104–121.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*, 360–367.

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience, 8*(385)https://doi.org/10.3389/fnins.2014.00385.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*, 113–131.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40*(1), 79–91.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001.

Godijn, R., & Theeuwes, J. (2011). Overt is no better than covert when rehearsing visuo-spatial information in working memory. *Memory & Cognition, 40*(1), 52–61. https://doi.org/10.3758/s13421-011-0132-x.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA 02*. https://doi.org/10.1145/507079.507082.

Halliday, D., Rosenberg, J., Amjad, A., Breeze, P., Conway, B., & Farmer, S. (1995). A framework for the analysis of mixed time series/point process data—Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Progress in Biophysics and Molecular Biology, 64*(2–3), 237–278. https://doi.org/10.1016/s0079-6107(96)00009-0.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking.* Oxford, UK: Oxford University Press.

iMotions (2018). Eye tracking: The Complete Pocket Guide. Retrieved from https://imotions.com/guides/.

Inamdar, S., & Pomplun, M. (2003). *Comparative Search Reveals the Tradeoff between Eye Movements and Working Memory Use in Visual Tasks*.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178–191. https://doi.org/10.1037/a0026923.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169–183. https://doi.org/10.1037/0096-3445.130.2.169.

Kirschner, P., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory: the good, bad and the ugly. *Computers in Human Behavior, 27*, 99–105.

Klimesch, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology, 24*, 61–100.

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology, 26*(1–3), 319–340. https://doi.org/10.1016/s0167-8760(97)00773-3.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3.

Lee, H. (2014). Measuring cognitive load with electroencephalography and self-report: focus on the effect of English-medium learning for Korean students. *Educational Psychology, 34*(7), 838–848. https://doi.org/10.1080/01443410.2013.860217.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072.

Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *The Quarterly Journal of Experimental Psychology, 64*, 649–656.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2017). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and*

Instruction. https://doi.org/10.1016/j.learninstruc.2017.12.007.
Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2014). The Psychometrics of Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(1), 125–143. https://doi.org/10.1177/0018720814539505.
Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 43–71). (2nd ed.). New York: Cambridge University Press.
Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 336–381). (2nd ed.). Cambridge: Cambridge University Press.
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. https://doi.org/10.1037/0022-0663.90.2.312.
Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., & D'Mello, S. K. (2017). Put your thinking cap on: Detecting cognitive load using EEG during learning. *Published in: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, Canada, March 13-17, 2017* (pp. 80–89). . https://doi.org/10.1145/3027385.3027431.
Mühl, C., Heylen, D., & Nijholt, A. (2015). *Affective Brain-Computer Interfaces. The Oxford Handbook of Affective Computing.* https://doi.org/10.1093/oxfordhb/9780199942237.013.024.
National Research Council (2011). *Learning science through computer games and simulations. committee on science learning: Computer games, simulations, and education.* Washington, DC: National Academies Press.
Nielsen, J. B., Conway, B. A., Halliday, D. M., Perreault, M., & Hultborn, H. (2005). Organization of common synaptic drive to motoneurones during fictive locomotion in the spinal cat. *The Journal of Physiology, 569*(1), 291–304. https://doi.org/10.1113/jphysiol.2005.091744.
Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429.
Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.
Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737–743.
Paas, F., Van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual & Motor Skills, 79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419.
Phillips, M. H., & Edelman, J. A. (2008). The dependence of visual scanning performance on saccade, fixation, and perceptual metrics. *Vision Research, 48*(7), 926–936. https://doi.org/10.1016/j.visres.2007.12.020.
Pinel, J. P. J. (2011). *Biopsychology* (8th ed.). Boston, MA: Pearson Education, Inc.
Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive load theory*. Cambridge, UK: Cambridge University Press.
Poole, A., & Ball, L. J. (2006). *Eye Tracking in HCI and Usability Research.* Encyclopedia of Human Computer Interaction211–219. https://doi.org/10.4018/978-1-59140-562-7.ch034.
Puma, S., Matton, N., Paubel, P. V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology, 123*, 111–120.
Rabbi, A. F., Zony, A., Leon, P. D., & Fazel-Rezai, R. (2012). Mental workload and task engagement evaluation based on changes in electroencephalogram. *Biomedical Engineering Letters, 2*(3), 139–146. https://doi.org/10.1007/s13534-012-0065-8.
Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., ... Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience, 22*(11), 2917–2926.
Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226–237. https://doi.org/10.1002/acp.1554.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.
Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132*(2), 106–114. https://doi.org/10.1016/j.actpsy.2009.01.005.
Van Gog, T., Rikers, R. M. J. P., & Ayres, P. (2008). Data collection and analysis: Assessment of complex performance. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.). *Handbook of research on educational communications and technology* (pp. 783–789). (3 ed.). London/England: Routledge.
Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99. https://doi.org/10.1016/j.learninstruc.2009.02.009.
Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177.
Xie, B., & Salvendy, G. (2000). Review and reappraisal of modeling and predicting mental workload in single- and multi-task environments. *Work & Stress, 14*, 74–99.