

## **Feasibility and use of psychophysiological responses based on cognitive load theory**

**Jeonghyun Kim & Il-Hyun Jo**

Ewha Womans University, South Korea

The purpose of this study was to identify the feasibility of using the pupil dilation and heart rate variability (HRV), to diagnose learners' changing cognitive load in a video learning setting. The pupil dilation and HRV of 23 participants were recorded during their learning progress, and the learning materials, including differences in task complexity, were presented to the groups of distinct prior knowledge. The results showed that the high prior knowledge group had lower pupil size than the low prior knowledge group, but the difference was not statistically significant. However, the ratio of low frequency to high frequency power (LF/HF ratio) was higher in the high prior knowledge group than in the low prior knowledge group throughout the entire learning process, and the two groups showed different arousal patterns. Overall, learners' changing cognitive load could be measured and visualised based on psychophysiological responses, the types of cognitive load could be assumed by using the two indices in parallel, and this study raised practical issues to the use of pupil size as an indicator of cognitive load in video learning.

### **Introduction**

Learning environments are being increasingly digitalised, distributed, and data-driven, and the scope and complexity of the types of learning are being transformed. While these changes are having positive effects on learning, supporting a learner within these new learning environments remains a challenge. Support for learners in the progress of learning is typically given in the form of instructional intervention, and such intervention is made based on instructional design. In most instructional design models, developing an instructional system starts with the analyses of learning needs and learning context requirements, and optimal strategies or tools are identified based on these analyses (Persico & Pozzi, 2015). However in the modern learning environment, this approach, taken by most instructional design models, is limited. First, it depends on a preliminary single-step measurement of the learners' level of understanding or their traits. Taking as an example the analysis, design, development, implementation, and evaluation (ADDIE) model, the most widely used model that represents the general form of the instructional design process (Persico & Pozzi, 2015), the information for intervening in the learning process is entirely based on the measurements of the analysis phase, and thus determining how to evaluate, supplement, or modify the implementation phase, where actual learning occurs, becomes an issue. Second, because the measurement generally relies on self-reporting, there is the problem of objectivity. Third, despite the importance of securing data to reflect learners' engagement in the learning process (Siemens, 2013) and the improvement of technologies that reflect the real-time context of learning, this approach of instructional intervention based on general instructional design, does not keep up with these trends.

Growing interest has been shown in learning analytics, and various attempts are being made to integrate educational big data to understand learner engagement, progression, and achievement at the point of actual learning. Learning analytics features the provision of timely interventions driven from real-time data, and when these interventions are made under the understanding of the states of learner, their effect can be maximised. In this context, psychophysiological measures are recognised as having potential, in that they provide an objective, unobtrusive, and immediately trackable way to measure learners' cognitive and affective states (Dirican & Göktürk, 2011). If a learner's cognitive load can be measured in real time based on psychophysiological responses, this can be a reasonable criterion for providing instructional interventions.

Psychophysiological measures are generally recognised to be applicable in the lab setting, and the empirical studies to measure learners' changing cognitive load by psychophysiological responses are still in the early stage, although their high reliability is asserted. Most related studies have focused on sophisticated measurement, or the tasks used in those studies were too simple or fragmented to be considered as a general learning situation (e.g., Bourisly, 2015; Brünken, Plass, & Leutner, 2003; Hansen, Johnsen, & Thayer, 2003; Hess & Polt, 1964; Zheng, & Cook, 2012). These insufficient examinations of cognitive measurement in

the real learning situation and the absence of a framework to align the collected data with the interventions are making cognitive load theory (CLT) less applicable to the actual learning field.

CLT has been considered as an influential framework for explaining learners' cognitive states and for conducting instructional design. The ground of this theory is that when it deals with novel information, working memory is limited in capacity and duration. Due to the nature of the working memory, the main concern of CLT is to minimise unnecessary working memory load and make those scarce resources available to focus on the core of learning. Because the total cognitive load is determined by the sum of intrinsic cognitive load, caused by the intrinsic nature of the task, and extraneous cognitive load, which depends on how knowledge is acquired, instructional design has been applied to eliminate extraneous cognitive load, and to optimise intrinsic cognitive load (Sweller, Ayres, & Kalyuga, 2011). In CLT, the working memory resources allocated by the intrinsic nature of the task are referred to as *germane resources*, while the resources devoted to deal with extraneous cognitive load are referred to as *extraneous resources*.

Levels of intrinsic and extraneous cognitive load are determined by element interactivity (Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). Students experience high cognitive load when the elements to be processed interact in a complex manner or when large numbers of elements must be processed simultaneously. If the interacting elements are due to extraneous cognitive load, removing the extraneous cognitive load factors can be the solution. However, if the interacting elements are due to intrinsic cognitive load, they are unalterable, because intrinsic cognitive load depends on the task complexity or the knowledge level of learners. In addition, task complexity, determined by the number of elements in a task and the interaction between the elements (Sweller & Chandler, 1994), depends not only on the learning material itself, but also on the interaction between the learning material and a learner's prior knowledge. Elements with multiple interactions for a learner with low prior knowledge could be handled as a single element by another learner with high prior knowledge (Sweller, Ayres, & Kalyuga, 2011). For this reason, Schnotz and Kürschner (2007) suggested that task complexity or instructional help should be provided under consideration of the levels of expertise.

For an adaptive intervention depending on a learner's cognitive load, it is essential to detect cognitive load differences depending on the learner's prior knowledge and task complexity. If the differences in psychophysiological responses among learners with different levels of prior knowledge or among the learning segments with different levels of task complexity are clearly observed, and if we can diagnose these differences based on CLT, providing adaptive interventions depending on learners' changing cognitive load can be realised in the progress of learning.

The purpose of this study was to identify the feasibility of using psychophysiological responses to diagnose learners' changing cognitive load in real learning settings, by utilising pupil size as an indicator of learners' cognitive load (Beatty & Lucero-Wagoner, 2000; Kahneman & Beatty, 1966) and heart rate variability (HRV) as an indicator of learner arousal (Malik, 1996; Riener, Ferscha, & Aly, 2009). Although pupil size only indicates the total cognitive load, and the different types of load could not be distinguished using pupil size alone, the parallel measurement of HRV may supplement the interpretation of types of learner's cognitive load. As we aimed to compare the cognitive load differences among the learners with different levels of prior knowledge and among the learning segments with different levels of task complexity, the extraneous cognitive load factors due to incorrect instructional design were as far as possible excluded, and learning materials for all participants were constant. The specific research questions are outlined below.

1. In the online learning process, can the cognitive load state of learners with different prior knowledge levels be compared based on psychophysiological responses?
2. In the online learning process, can psychophysiological responses be used to detect different task complexities?

## Psychophysiological response for measuring cognitive load

To measure cognitive load, self-reporting questionnaires and physiological measures are usually performed (Plass, Moreno, & Brünken, 2010). Various self-reporting questionnaires are the most commonly used methods to measure cognitive load (de Jong, 2010; Sweller, 2017), and previous studies reported that these methods sensitively distinguish even small cognitive load differences (Paas, van Merriënboer, & Adam,

1994; Paas, 1992). However, the self-reporting questionnaire is the subjective and retrospective rating of perceived cognitive load, and differences in ratings between learners can come from factors of competency, motivation, or ability to concentrate of each participant (Martin, 2017). Above all, this method does not reflect the change of cognitive load with time (de Jong, 2010). On the other hand, because the human cognitive function physiologically affects the body (Brünken et al., 2003), a learner's cognitive or psychological state can be measured by the lapse of time through physiological responses. This method can easily be used to calculate or visualise a learner's cognitive load trend (Paas, Tuovinen, Tabbers, & van Gerven, 2003), and the distortion of the result by a learner's own judgment can be minimised.

Among the physiological measures, a promising method for measuring cognitive load is the measurement of pupil size (e.g., Goldinger & Papesch, 2012; Klingner, Kumar, & Hanrahan, 2008; Marshall, 2007). Activation of the cerebral cortex is largely influenced by the locus coeruleus (LC), which regulates arousal and autonomic function through the secretion of norepinephrine (Aston-Jones & Cohen, 2005). Norepinephrine plays a role in the working memory and attention functions of the prefrontal cortex, and is known to improve the integration of long-term memory (Ramos & Arnsten, 2007). Since the secretion of norepinephrine from the LC correlates with pupil dilation, pupil dilation has been researched as the summative index of brain activity resulting from cognitive load (Landgraf, van der Meer, & Krueger, 2010), and has been used as a representative indicator for cognitive load.

During a mentally demanding task, the pupil size has been found to dilate up to 0.5 mm (Beatty & Lucero-Wagoner, 2000). However, although previous studies have consistently reported the reliability of the pupil size as an indicator of cognitive load, only a few empirical studies have measured learners' pupil size as an index of changing cognitive load in actual learning situations. Additionally, this index is limited when interpreting the measured results, since the types of cognitive load cannot be distinguished by the pupil size alone.

HRV refers to the amount of heart rate fluctuation in the length of intervals between heart beats. It is considered as a representative index of autonomic nervous system (ANS) activity. The two branches of the ANS, the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), control the beat-to-beat differences of the heart (McDuff, Gontarek, & Picard, 2014), and these phenomena in the cardiovascular system provide information about the psychophysiological state experienced by humans. While the sympathetic nervous system represents the arousal aspect of the human nervous system, the parasympathetic nervous system has the reverse effect to that of the sympathetic nervous system, of relaxing and soothing. Therefore, many psychological investigations have used HRV to interpret the effects of stress, emotion, and work on the autonomic nervous system (Malik, 1996).

Among the HRV parameters, the LF/HF ratio is known to represent a sympathetic modulation. It is well known that physiological activation (arousal) is modulated by the sympathetic system, while deactivation functions are performed by the parasympathetic system (de Zambotti, Covassin, De Min Tona, Sarlo, & Stegagno, 2011). Usually, the low frequency (LF) component is interpreted as promoting sympathetic activity, while the high frequency (HF) component has a parasympathetic influence on the heart. Under a stressful situation, LF is increased, HF is decreased, and the LF/HF ratio is increased (Berntson & Cacioppo, 2004). Thus, the LF/HF ratio represents an evaluation of the balance of the autonomic nervous system (Malik, 1996), the increase of which implies a predominance of sympathetic over parasympathetic activity. In this context, we used the LF/HF ratio as an arousal index to interpret cognitive load more objectively.

## Related works

Although few studies have been performed that measure and interpret the pupil size or HRV in an actual learning situation, studies on the pupil size and HRV have been consistently performed with reference to task performance. One of the representative studies using pupil size, as a cognitive load index, was performed by Hess and Polt (1964), in which the answers to mathematical multiplication problems were requested with different levels of difficulty. In their study, the subjects' pupil sizes gradually increased in the order of problem difficulty, the pupil size reached a peak before responding with an answer, and restored to their original size after responding. Several researchers have shown a similar result. Bourisly (2015) reported that the pupil sizes of participants were larger in complex addition problems than in simple problems. Heitz, Schrock, Payne, and Engle (2008) found that the subjects' pupil sizes increased with the

level of task difficulty, and working memory capacity and levels of incentive condition affected pupil size. Recarte, Pérez, Conchillo, and Nunes (2008) reported that pupil size sensitively changed according to the level of mental load involved in a task, by comparing the NASA task load index, pupil dilation, and blink rate in a dual task condition. In addition, Iqbal, Zheng, and Bailey (2004) found that pupil size differed according to the difference in mental load through an experiment using time on task, self-reporting, and pupil size as cognitive load indexes. van Gerven, Paas, van Merriënboer, and Schmidt (2004) identified that mean dilation increases with memory load although the sensitivity of mean dilation decreases with age. Zheng and Cook (2012) measured learners' changing cognitive load during complex problem-solving tasks with the three indicators (i.e., peak load, average load, and accumulated load), and showed that as learners' cognitive load increases, pupil dilation increased.

Although pupil size has been proved to be a useful indicator to measure the cognitive load of learners, previous research has some limits, since most studies were conducted using several short tasks of different levels of difficulty at regular intervals. In real learning progress, learning contents are usually provided hierarchically and sequentially, and individual learning contents are not segmented. To interpret the pupil size as a cognitive index in learning, it is necessary to confirm whether pupil size can successfully measure a learner's cognitive load in an actual learning situation.

Previous research confirmed that learners' sympathetic nervous systems were more active with reference to task performance. Filaire, Portier, Massart, Ramat, and Teixeira (2010) conducted an experiment on the mental stress of 52 teachers that occurs both before and after a lecture to 200 students. The result showed that the lecture induced a significant decrease in HF component, and an increase in the LH/HF ratio. The authors explained that the result showed a tilt of autonomic balance towards sympathetic activation. Chung and Kim (2008) analysed the stress of learners during an oral test through HRV. In their study, before the test, the sympathetic nervous system became activated, and the LF/HF ratio was high; however, after the test, the sympathetic nervous system became stabilised, and the LF/HF ratio was low. Other studies show similar effects of autonomic balance toward sympathetic activation, although the LF/HF ratio was not used as a research index. Hansen et al. (2003) verified the reactivity to cognitive tasks through HRV and HR, and the results showed that the root mean square of successive differences, which reflects the parasympathetic nerve activity, was lower in baseline and task performance, than in recovery. In addition, Kennedy and Scholey (2000) identified that cognitively demanding tasks are associated with increased heart rates in participants, by examining the interaction between glucose administration, cognitive performance, and heart rate with three tasks of different levels of mental demand.

HRV has been interpreted as a reliable indicator of stress, alertness, or arousal, according to the research context. However, taking into consideration the study of Poh, Swenson, and Picard (2010), which reported that sympathetic nerve activity during the class was low, unlike the situations where students themselves study or carry out homework, HRV can be used to compare the arousal level of learners in video learning. Paas and van Merriënboer (1994) insisted that HRV was invalid as a mental load index. Galy, Cariou, and Mélan (2012) explained that an index controlled by the autonomous nervous system can be used as a good indicator to represent the available cognitive resources at the time of task execution. In this context, we interpret this to mean that a learner, showing a higher sympathetic nerve activity in the learning, is in a higher arousal state, and the LF/HF ratio was used as the HRV indicator to measure learner arousal.

Based on the results of the previous studies, this study aimed to confirm whether differences of pupil size and HRV are observed depending on the level of learners' prior knowledge and task complexity in video learning. The analyses were conducted not only between two groups with different levels of prior knowledge, but also within each group. After that, this study intended to interpret the measurement results with CLT, based on the relationship between the cognitive load and the arousal level.

## Methods

### Participants

The data of 54 undergraduate students recruited in Seoul, Korea were collected. In the process of recruitment, the students whose major was mathematics were excluded, to control excessive gaps in the level of prior knowledge. Additionally, the participants were limited to second-year college students,

because some learning contents used as stimuli were included in the high school curriculum, and thus the time of exposure to the learning contents needed to be controlled. During the data preprocessing, the data of 16 students who did not participate faithfully in the experiment (e.g., showing extreme drowsiness) or who showed outlier data were additionally excluded. Among the 38 students filtered, a total of 23 subjects (high prior knowledge group - 10 students; low prior knowledge group - 13 students) were selected as the final subjects, in consideration of the tie and the median. Females made up 63.3% of the subjects and males 34.7 %.

## Measurement

### *Learning stimuli and task complexity*

Stimuli included two videos developed for the experiment, of about 10 minutes each, on *proposition* in mathematics. The first video was about universal and existential propositions in the high school curriculum, and provided prerequisite learning for the second video. The content covered the concepts, examples, and denials of universal and existential propositions. The second video, about a statement with quantifiers, was the content of the course for the college mathematics major, and was new information for the participants. Although the level and scope of the material were adjusted to be appropriate for the participants, the second video included fundamentally higher task complexity than the first video. The task of proving or denying a statement with quantifiers, which combines a universal and an existential proposition, was more complicated, because the included propositions, before and after, interacted with each other. In addition, the meaning of the assertion completely changed depending on the order of presentation of the quantifiers, and when proving or negating a statement with quantifiers, the students needed to be particularly careful about the order of steps.

As mentioned above, the first video represented the prior phase of the second video. Therefore, we confirmed the differences of cognitive load between the first (universal and existential proposition) and the second (a statement with quantifiers) video sections. The next comparison was between more detailed segments. We verified the differences between the more detailed segments that show clear differences in task complexity. More specifically, the second video included two types of a statement with quantifiers;  $(\forall, \exists)$  and  $(\exists, \forall)$  ( $\forall$  is universal quantifier, and  $\exists$  is existential quantifier). These two propositions,  $(\forall, \exists)$  and  $(\exists, \forall)$ , were designed to be equal in terms of the order of concept-example-negation. Because the negation component had relatively higher task complexity than the other components, the comparison was made between the negation component and the others (concept and example) for each type of proposition.

### *Cognitive load and arousal*

The pupil size was used to measure cognitive load. The researchers used the baseline corrected diameter calculated by subtracting the baseline pupil diameter from the averaged pupil diameter during the measurement of interest (Beatty & Lucero-Wagoner, 2000; Goldinger & Papesh, 2012). Tobii Pro X2-30 eye tracker unit sampling at 30 Hz was used for the measurement.

To precisely measure the cognitive pupil dilation and to interpret the cognitive load related to the learners' prior knowledge or task complexity, those instructional design factors that are known to cause extraneous cognitive load effects or experiment environment factors that might influence the subjects' pupil size were excluded from the design stage of the experiment. The screen of stimuli only included pictures and text with synced narration, and the screen design was consistent across all the screens. To minimise the influence of uncontrolled light and minimise eye fatigue caused by long-term watching the screen, the background color was maintained as the default color of the Kindle Reader, the default font color was maintained as black, and the highlighting text was maintained as low chroma colors. For HRV, a portable wrist-band type photoplethysmographic (PPG) measuring device was used. The LF/HF ratio, the index of sympathetic nerve activity, was calculated as the variable representing arousal.

### *Prior knowledge and achievement*

The level of prior knowledge was determined by the pre-test, and achievement was determined by the score gap by subtracting the pre-test score from the post-test score. The pre-test and the post-test were equivalent, and were about proving the truth of the given propositions.

## Procedure

Upon arrival at the lab, the participant completed a consent form, and received instructions for the experiment in accordance with the Institutional Review Board’s requirements. Then, a PPG experimenter was attached to each participant, and each participant listened to classical music for 3 minutes to relax. Immediately prior to watching the stimuli, the baseline pupil size was collected by having the users fixate on an “X” sign on the center of the screen. Then, the tasks were administered in the following order: taking the pre-test, watching the stimuli, and taking the post-test (Figure 1). Each task had a limited completion time.

Before the data analysis, the outliers of the dataset were preprocessed, and the preprocessed data were converted to the mean per second to meet the time standard between the two devices. Due to non-normal data distributions, the Wilcoxon rank sum test, the Wilcoxon sign rank test, and the Friedman test were used to confirm the differences. The statistical programs R and SPSS were used for the analysis.

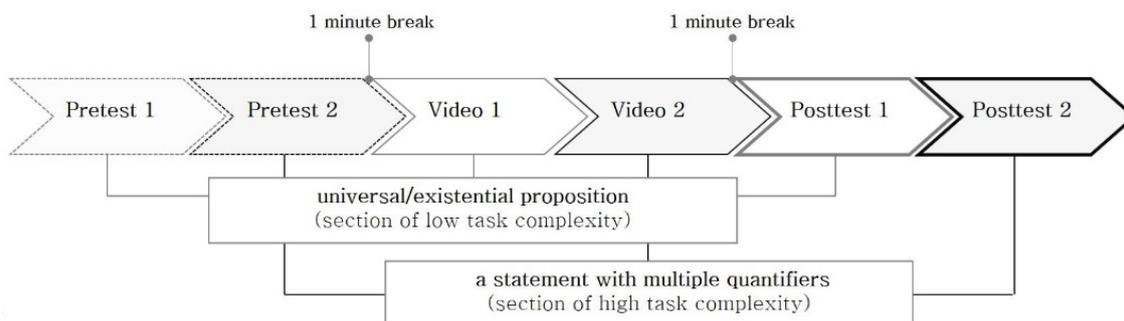


Figure 1. Steps of the experiment

## Results

### Descriptive statistics

We examined the differences in pupil size and HRV between the high and low prior knowledge groups. Table 1 shows the low prior knowledge group (henceforth referred to as the low group) showed a larger pupil size through the entire learning process, but the *SD* was higher in the high prior knowledge group (henceforth referred to as the high group) in all sections.

Table 1  
Descriptive statistics for pupil size

Section	Group	<i>M</i>	<i>SD</i>	Minimum	Maximum
Videos 1 & 2	High	0.1514	0.2495	-0.5986	0.5998
	Low	0.1798	0.2262	-0.5991	0.5999
Video 1	High	0.1760	0.2462	-0.5954	0.5990
	Low	0.1898	0.2227	-0.5980	0.5995
Video 2	High	0.1316	0.2504	-0.5986	0.5998
	Low	0.1715	0.2287	-0.5991	0.5999

Table 2 shows that the high group showed a higher LF/HF ratio for the entire learning. The low group showed a larger mean drop in the LF/HF ratio between video 1 and video 2 than the high group. Because HF values were higher during each section in the low group, these results could be interpreted as the degree of sympathetic nerve activity.

Table 2  
Descriptive statistics for HRV

Parameter	Section	Group	M	SD	Minimum	Maximum
LF/HF ratio	Videos 1 & 2	High	2.6411	1.0799	0.7787	6.5828
		Low	1.8347	0.9887	0.4123	7.0487
	Video 1	High	2.6534	1.1781	0.8227	6.3117
		Low	2.0246	1.1941	0.4591	7.0487
	Video 2	High	2.6314	0.9960	0.7787	6.5828
		Low	1.6995	0.7840	0.4123	5.4807
HF	Videos 1 & 2	High	413.6247	260.0102	56.9760	1460.8150
		Low	685.8376	523.7985	122.1498	2958.4910
	Video 1	High	368.4173	249.7171	56.9760	1460.8150
		Low	622.2160	451.7973	130.3435	2514.4780
	Video 2	High	449.1144	262.4367	76.4768	1096.0470
		Low	731.1159	565.2061	122.1498	2958.4910

**Differences between the high and low groups**

*Change in pupil size*

From the graphs shown in Figures 2 and 3, we compared the trends of psychophysiological responses between the high and low groups during video 1 (with low task complexity) and video 2 (with high task complexity), and the Wilcoxon rank sum test was then performed to verify the difference between the groups. The left side in Figure 2, based on the dotted line in the middle, shows the change during video 1, and the right side shows the change during video 2. At the end of each section, the end of the duration time of video 1 or video 2 is displayed, and only the data of the learners who carried out further learning is involved. A smooth function was used in the center to show the pupil size changes more intuitively. Figure 2 shows no obvious difference in pupil size between the high and low groups. However, the overall average height was slightly lower in the high group, and this difference was more apparent in video 2. The low group showed a relatively small overall pupil variability during the video sections, especially in the video 2 section. Wilcoxon rank sum test showed no statistically significant difference in the pupil size between the two groups during video 1 ( $p = 0.8315$ ) and video 2 ( $p = 0.5629$ ). The components after video 1 and video 2 duration time could not be interpreted as the mid-value of all learners, and thus no further interpretation was performed.

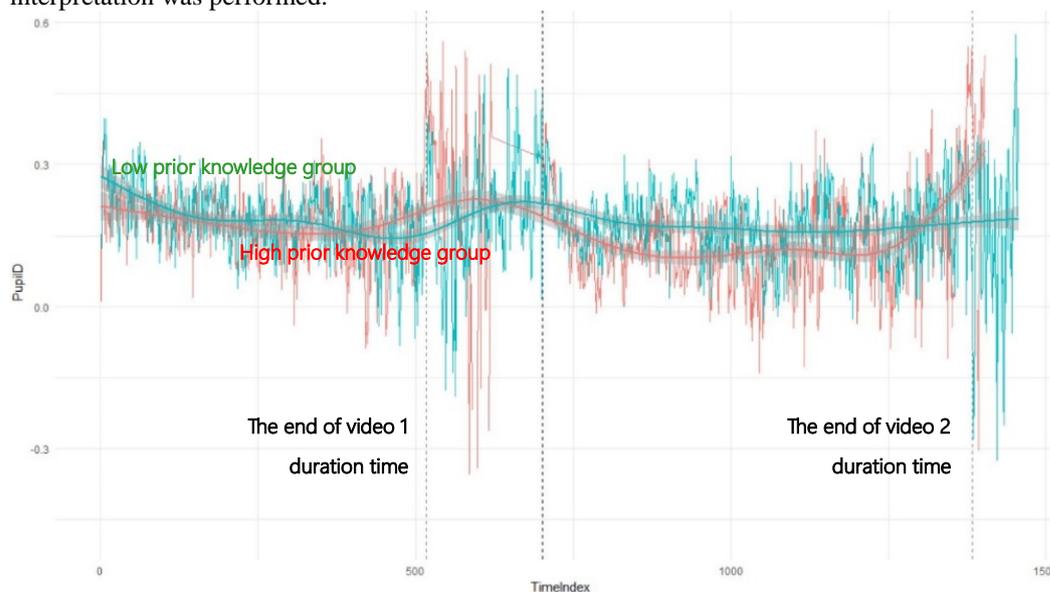


Figure 2. Pupil size change

### Change in HRV

Figure 3 shows the mean change in LF/HF ratio during the overall learning process. The high group maintained a higher LF/HF ratio than the low group during the whole time. During video 1, the LF/HF ratio of the high group increased at the beginning, and then declined, while the low group maintained a moderate increase. During video 2, the two groups showed differences in the magnitude of the mean, but they showed a similar appearance in the graph form. For the same reason as that for the pupil size, the graph after video 1 and video 2 duration time was not interpreted.

The Wilcoxon rank sum test showed no statistically significant difference in HRV during video 1. However, during video 2, a significant difference was shown between the two groups,  $W = 171$ ,  $p = 0.0008$ . The LF/HF ratio median of the high group was 2.4715, and that of the low group was 1.6840.

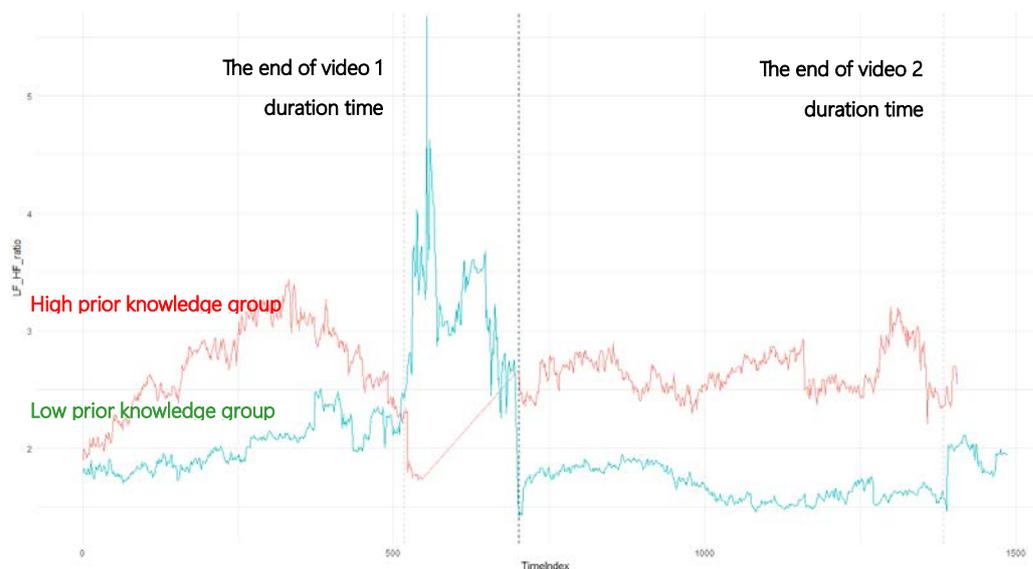


Figure 3. HRV change

### Difference within each group

To identify the differences in the physiological responses by task complexity within each group, the mean pupil size and HRV between the video 1 and video 2 were compared using the Wilcoxon signed rank test. But for each group, the difference between the two videos was not statistically significant. Next, we further divided video 2 into small learning sections showing distinct task complexities, and compared the differences in the psychophysiological responses among these divided sections. This was because by calculating the mean of the psychophysiological responses during the entire video, the task complexity differences among the small learning sections could be hidden. Video 2 was suitable for testing task complexity differences among the small learning sections, because two types of a statement with quantifiers (with the same format and task complexity) were clearly distinguished. We extracted the part of a statement with quantifiers ( $\forall, \exists$ ), organised in order of universal proposition - existential proposition, into three small learning sections: concept, example, and negation example. The same operation was then performed as that for the part of a statement with quantifiers ( $\exists, \forall$ ), organised in order of existential proposition - universal proposition. Finally, the psychophysiological responses among the small learning sections (concept, example, and negation example) within each statement were compared.

The results of the Friedman test showed a statistically significant difference in the LF/HF ratio, depending on the types of section. In the case of the statement with quantifiers ( $\forall, \exists$ ), the test statistic value was  $\chi^2(2) = 10.316$ , and the corresponding  $p$  value was 0.006. The statement with quantifiers ( $\exists, \forall$ ) showed that the test statistic value was  $\chi^2(2) = 8.629$ , and the corresponding  $p$  value was 0.013. Post hoc analysis of the Wilcoxon signed-rank tests was conducted by applying a Bonferroni correction, resulting in a significance level set at  $p < 0.017$ . In the case of the statement with quantifiers ( $\forall, \exists$ ), the medians of the LF/HF ratio for the sections of the concept, example, and negation example were 1.5307 (.7498 to 3.0299), 1.5719 (.0000 to 2.1647), and 2.7756 (2.0928 to 3.1726), respectively. No significant differences of the LF/HF

ratio were observed between the example section and the concept section ( $Z = -1.260, p = .208$ ), or between the negation example section and the concept section ( $Z = -2.191, p = .028$ ). However, a statistical difference of the LF/HF ratio was shown between the negation example section and the example section ( $Z = -2.701, p = 0.007$ ). These results were also similar in the statement with quantifiers ( $\exists, \forall$ ). The medians of the LF/HF ratio for the sections of the concept, example, and negation example were 1.5307 (0.7498 to 3.0299), 1.7956 (0.0000 to 2.1645), and 2.7456 (1.8913 to 3.4266), respectively. Again, statistically significant difference was only shown between the negation example section and the example section ( $Z = -2.547, p = 0.011$ ).

As the two groups showed different patterns in the LF/HF ratio, we verified whether these arousal patterns were related to the learning achievement. Figure 4 shows the boxplots for the pre-test 1 (left) and pre-test 2 (right) by each group (high and low), while Figure 5 shows the boxplots for the score gap between post-test 1 and pre-test 1 (left), and between post-test 2 and pre-test 2 (right).

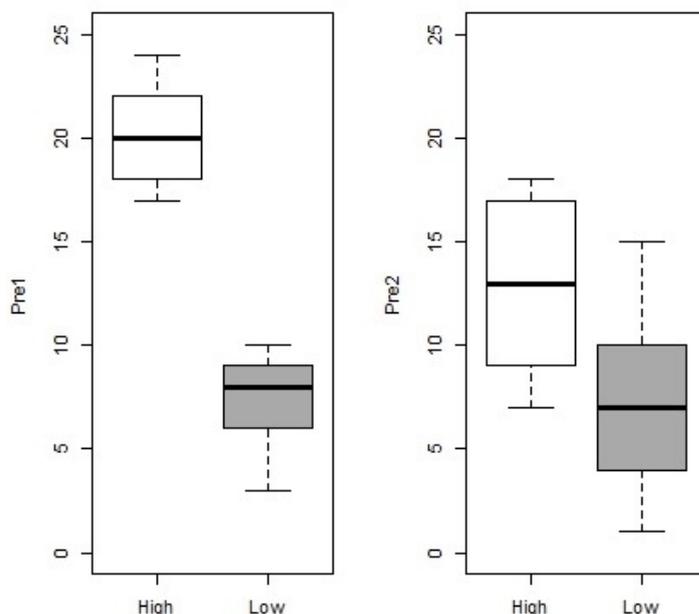


Figure 4. Comparison of pre-test

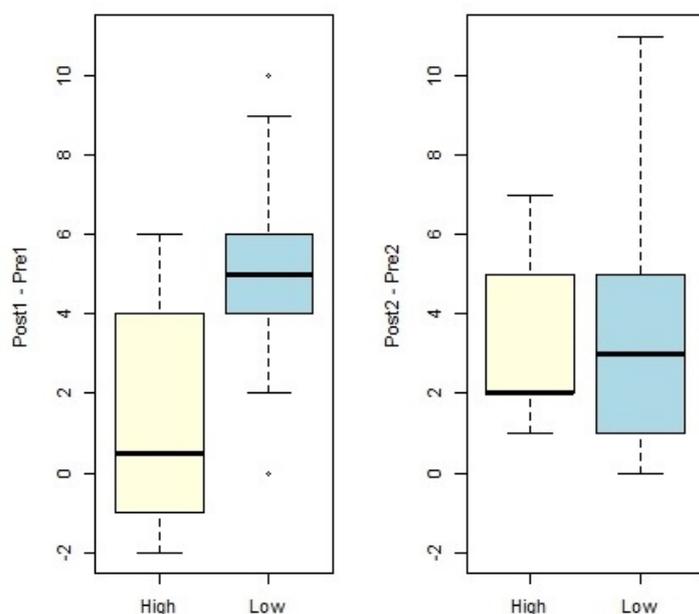


Figure 5. Score gap between post-test and pre-test

In this study, participants were given the same learning materials after taking pre-tests 1 and 2, Moreover, post-tests 1 and 2 were equivalent tests to pre-tests 1 and 2, respectively. Figure 4 shows that there were clear score differences between the high and low group in both pre-test 1 and pre-test 2, and these differences were statistically significant ( $p < 0.5$ ). The score gap between post-test 1 and pre-test 1, which was about video 1 where significant arousal difference was not observed, was statistically significant between the two groups ( $p = 0.0113$ ). However, in the case of the score gap between post-test 2 and pre-test 2, which was about video 2 where the level of arousal in the high group was significantly higher, there was no statistically significant difference between the two groups ( $p = 0.8252$ ). Although the low group showed more difference ( $M = 3.923$ ,  $SD = 1.9119$ ) between the post-test 2 and the pre-test 2 than the high group ( $M = 3.1$ ,  $SD = 3.3799$ ), the score gap (i.e., improvement) was quite small, given that the pre-test score of the high group was originally high.

## Discussion

### Difference in psychophysiological responses depending on prior knowledge

In this study, the psychophysiological response showed different tendencies, depending on the learners' level of prior knowledge. From a descriptive viewpoint, the high group showed a smaller pupil size than that of the low group, regardless of the task complexity, although the gap between the two groups was not statistically significant. On the other hand, the HRV showed different process flows depending on the learners' level of prior knowledge. During video 1, HRV of the high group increased at the early stage, and decreased after the middle stage, while during video 2, this group maintained high HRV. However, regardless of the task complexity, the low group maintained low HRV throughout the entire learning process.

According to Sweller et al. (2011), the total cognitive load is the sum of extraneous and intrinsic cognitive loads that are imposed by the nature and structure of the learning materials. On the other hand, germane cognitive load, which refers to germane resources, reflects the working memory capacity devoted to process the information that is relevant to learning. In other words, intrinsic and extraneous cognitive load are what inhere in a learning material, while the germane cognitive load is in a distinct ontological category (i.e., cognitive process). Given these points, the sum of actually allocated germane resources and extraneous resources can vary from the estimated total cognitive load by the task. Mental effort is separately described as the actual cognitive load allocated to deal with the demands imposed by the task (Paas et al., 2003).

In this study, extraneous cognitive load factors due to poor instructional design were excluded from the design stage, and the two groups studied the same learning material. Therefore, extraneous cognitive load due to instructional design was kept constant for all participants, and the given task complexity was basically the same. Under these conditions, if each group allocated their working memory resources corresponding to the sum of extraneous and intrinsic cognitive load estimated by the task, as well as their prior knowledge, the actually allocated cognitive load of the low group must be larger than the other. This is because the low group has lower-level schemas, and thus this group should handle more element interactivity. However, in this study, there was no statistically significant difference in the allocated cognitive load, measured by pupil size, between the two groups and this result raises two possibilities: The low group did not devote the required working memory resources as expected, or the high group allocated working memory resources above the expected requirement of the task.

However, considering the observed HRV values, it seems to be because the low group did not devote sufficient germane resources. The high group maintained higher arousal throughout the entire learning process, while the low group showed lower arousal, even in the sections with high task complexity. Although we could not conclude whether the level of arousal shown by the high group was absolutely or relatively high, the facts that the difference in arousal between the two groups was statistically significant, and that the low group, which should devote more amount of germane resources than the other to understand the given learning materials, maintained lower arousal, makes it difficult to interpret the allocated cognitive load of the low group as germane resources. Additionally, the fact that compared to the high group, the low group, which showed much lower scores in the pre-test, showed no statistically significant difference in the score gap between post-test 2 and pre-test 2, which was about video 2 where the level of arousal in the low group was significantly lower, also supports this assumption.

Another noteworthy point is the arousal aspect of the high group. In video 1 (already learned and low task complexity), the high group showed a falling tendency in arousal after mid-term. However, this group maintained high levels of arousal from the beginning to the end of video 2, which was never learned, and featured high task complexity. This contrasts with the low group, which maintained a lower level of arousal throughout the entire learning process. The higher arousal of the high group, in the high task complexity section than in the low task complexity section, can be interpreted as due to the learner characteristics of this group, such as the ability to efficiently allocate cognitive resources, motivation, and self-regulated learning ability. According to Debie and van de Leemput (2014), when learners are intrinsically motivated, they can devote more cognitive resources to learning, and their performance improved. Although the exact psychological construct could not be affirmed by this experiment alone, the result indicates well that there is interaction between the allocated cognitive load and learner characteristics. CLT explains that learners' intrinsic cognitive load depends on their prior knowledge shows that cognitive load is dependent on learner characteristics (Seufert, Jänen, & Brünken, 2007), and there has been criticism that CLT only involves the factors related to learning, despite the degree of a learner's engagement in learning depending on emotional factors as well, such as motivation (Plass et al., 2010). More research should follow on the influence of learner characteristics on cognitive load.

### **Within-group psychophysiological responses depending on task complexity**

When we compared the difference in psychophysiological responses between video 1 and video 2, no significant difference was observed between the groups. However, when we performed separate tests on the different combinations of the smaller sections, the high group showed significantly higher HRV in the negation example, which had a higher task complexity than the example section, which had a lower task complexity, in both types of statements with quantifiers. The fact that the difference in the learners' HRV observed when the level of the task complexity was finely divided indicates that the learners' physiological status sensitively changes, depending on the task complexity and the scope of the analysis. Similarly, Iqbal et al. (2004) insisted that a task generally does not require a sustained cognitive load throughout the entire task, and the required cognitive load could vary, depending on the scope of the analysis. Seen from these empirical verifications, to provide personalised intervention in the process of actual video learning, it is necessary to develop instructional intervention that is detailed to the smallest learning unit, combined with exquisite analysis of the task complexity involved in a course. When these sophisticated analyses and designs are supported, the cognitive load intervention based on psychophysiological responses can be truly effective, corresponding to the sensitively changing cognitive load of learners.

### **Additional discussion on cognitive load analysis in learning process**

In this study, the pupil size was not statistically different among the sections showing different task complexities. This result was not in agreement with the conclusions of most previous studies, in which it was reported that the pupil size sensitively captures even subtle changes in cognitive load (Beatty & Lucero-Wagoner, 2000; Bourisly, 2015; Kahneman, 2011). It should be noted that most previous studies have been conducted with short or segmented tasks that involved distinct levels of cognitive load. For example, many studies used arithmetic tasks (e.g., Bourisly, 2015; Hess & Polt, 1964; Szulewski, Fernando, Baylis, & Howes, 2014) or reading tasks (e.g., Hyönä et al., 1995; Just & Carpenter, 1993) separated by different task difficulty; and dual tasks (e.g., Marshall, 2007; Recarte et al., 2008) were also used frequently. In contrast to the results of most previous studies, Schultheis and Jameson (2004) cast doubt over the sensitivity of pupil size. In their research, while reading speed, survey, and event-related potential (ERP) showed consistent results according to task difficulty, the pupil size result was not consistent. The authors explained this result in association with the strictly controlled settings of the previous studies, and they insisted pupil size could not be a useful index of cognitive load for adaptive hypermedia systems that involve continuous reading. These conclusions are also consistent with Zheng and Cook's findings (2012). In their study, although the mean pupil diameter decreased with reduced cognitive load, the researchers questioned the sensitivity of the pupillary response in complex learning tasks, based on the small magnitude of the effects in pupillometric analyses.

In a general video learning situation, tasks with different levels of intrinsic cognitive load are presented hierarchically and sequentially, and learning is never free from the order effect. The preceding learning experience effects the subsequent learning experience. Moreover, the differences that need to be found

through cognitive load measurement may be small. In case of this study, the learning materials were about proving propositions. Although there were relative differences in task complexity, the given material was basically a series of contents which correspond to learning with understanding of high element interaction information. That is, in a general learning situation, the differences in task complexity to be compared may not a stark difference, and in that sense, the approach of this study concurrently using the psychophysiological responses which reflect different psychological constructs of learners, can be useful to interpret the degree or type of cognitive load. Although we yet lack access to measure the absolute degree of cognitive load into a numerical value, and to exquisitely separate the types of cognitive load, by accumulating empirical verifications manipulating the level of task complexity, learners' prior knowledge, learner characteristics and cognitive load in various learning settings, we will be able to identify more rigorous methodologies.

In addition, it should be noted that both the high and low groups showed lower pupil size in video 2, where the task complexity was high. This suggests that the influence of fatigue should be considered in measuring cognitive load in an online learning environment. The pupil size, after the eyes have adapted to constant illuminance, is known to decrease, due to fatigue or drowsiness (Goldwater, 1972; Tryon, 1975). In a general video learning situation, learning should last quite a long time seamlessly, and thus it seems difficult to exclude the effects of fatigue. Because the activity of fatigue is an obstacle to compare and interpret cognitive load that changes over time, future works will have to include the methods to control these unnecessary factors in the measurement process.

## **Conclusion and future steps**

The results of this study are worthwhile, and show that first, the learners' allocated cognitive load during the video learning was quantified into numerical values by psychophysiological responses. Moreover, based on these values, the learner's changing cognitive load status during the learning process was visualised and interpreted. If learners' changing cognitive load could not be quantified in real time and interpreted in conjunction with events of instruction at each learning point, the educational interventions in accordance with the learners' cognitive load are simply bound to stay in plausible discourse. Although such an approach is still in its early stage, the accumulation of this kind of research will surely work as the driving force for the measurement of cognitive load itself, as well as for practical applications to provide adaptive interventions.

Second, we verified the feasibility of the approach, measuring the pupil size and HRV in parallel, to diagnose learners' changing cognitive load. If learners exhibit low cognitive load at the section of high task complexity or vice versa, this can provide direct evidence that instructional interventions should be given. Moreover, while monitoring learners' cognitive load, the level of arousal can be utilised as the information for diagnosing the cognitive load types, and providing proper intervention. Although only two psychophysiological responses were verified in this study, by examining more psychophysiological responses representing various aspects of learning state, we can move toward more sophisticated diagnosis about the level and type of cognitive load.

Third, the practical issues of fatigue and the order effect, for using the pupil size as a cognitive load indicator in video learning were reconfirmed. Although previous studies have demonstrated the effectiveness of psychophysiological responses, there are still plenty of challenges to be overcome in the practical verification process. To examine the unclear points of pupil size measurement in video learning, and then to apply this index in an actual learning situation, repetitive verifications in various learning environments should be followed.

The result showed the need for follow-up studies. First, the influence of diverse learner characteristics on cognitive load should be actively verified. During a learning program, students do not always use all their available cognitive resources in learning (van Merriënboer & Sweller, 2005), and diverse learner characteristics may act in this process. However, the CLT field has been passive in coping with this argument, although several empirical studies have shown that cognitive load levels of learners can vary according to learner characteristics (e.g., Galy et al., 2012; Paas, Camp, & Rikers, 2001; Plass, Chun, Mayer,

& Leutner, 1998; Wallen, Plass, & Brünken, 2005). The only individual difference officially acknowledged in CLT is the prior knowledge level, and this presents confusion to the diagnosis of cognitive load in the real learning process.

Second, further research to examine diverse psychophysiological responses under the various cognitive load conditions should be followed. The absence of reliable measures to distinguish the different types of cognitive load is a main obstruction to threaten the explanatory and predictive power of the theory (Plass et al., 2010). As confirmed in this study, incorporating psychophysiological responses, which help interpret cognitive load in various aspects, can be a practical approach. In addition, considering the features of the video-learning environment in which various stimuli are continuously presented, combining stimulated recall interviews with psychophysiological responses can also provide concrete evidence to interpret the different types of cognitive load based on the exact timing of cognitive load occurrence.

Finally, statistical analysis techniques suitable for psychophysiological data should be suggested. Statistical analysis methods, generally used in social sciences, are based on a mid-value, such as mean, and these approaches lead to considerable waste of psychophysiological data, featuring vast amounts of raw data. In addition, these kinds of statistical analyses have limits on reflecting the change of learners' learning state over time. Therefore, by exploring more rigorous statistical measures, like time series analysis, which can include a massive change in the psychophysiological response, learners' changing cognitive load can be identified more clearly.

## Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5B6036244).

## References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403-450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Beatty, J., & Lucer-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge, MA: Cambridge University Press.
- Berntson, G.G., & Cacioppo, J.T. (2004). Heart rate variability: stress and psychiatric conditions. In M. Malik & J. Camm (Eds.), *Dynamic electrocardiography* (pp.57-64). New York, NY: Futura.
- Bourisly, A. K. (2015). Pupil response diameter is modulated as a function of cognitive load during mental addition: A psychophysiological protocol and study. *Journal of Advanced Neuroscience Research*, 2(1), 1-6. <https://doi.org/10.15379/2409-3564.2015.02.01.1>
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53-61. [https://doi.org/10.1207/S15326985EP3801\\_7](https://doi.org/10.1207/S15326985EP3801_7)
- Chung, S., & Kim, J. (2008). The correlation between HRV and questionnaires before and after oral test. *The Korean Journal of Stress Research*, 16(4), 363-367.
- Dehue, N., & van De Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, 5(1099), 1-12. <https://doi.org/10.3389/fpsyg.2014.01099>
- 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>
- de Zambotti, M., Covassin, N., De Min Tona, G., Sarlo, M., & Stegagno, L. (2011). Sleep onset and cardiovascular activity in primary insomnia. *Journal of Sleep Research*, 20(2), 318-325. <https://doi.org/10.1111/j.1365-2869.2010.00871.x>
- Dirican, A. C., & Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3, 1361-1367. <https://doi.org/10.1016/j.procs.2011.01.016>

- Filaire, E., Portier, H., Massart, A., Ramat, L., & Teixeira, A. (2010). Effect of lecturing to 200 students on heart rate variability and alpha-amylase activity. *European Journal of Applied Physiology*, 108(5), 1035-1043. <https://doi.org/10.1007/s00421-009-1310-4>
- Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83(3), 269-275.
- Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, 21(2), 90-95. <https://doi.org/10.1177/0963721412436811>
- Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*, 77(5), 340-355. <https://doi.org/10.1037/h0032456>
- Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, 48(3), 263-274. [https://doi.org/10.1016/S0167-8760\(03\)00073-4](https://doi.org/10.1016/S0167-8760(03)00073-4)
- Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, 45(1), 119-129.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611), 1190-1192. <https://doi.org/10.1126/science.143.3611.1190>
- Hyönä, J., Tømmola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48(3), 598-612. <https://doi.org/10.1080/14640749508401407>
- Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the ACM conference on Human Factors in Computing Systems* (pp.1477-1480). <https://doi.org/10.1145/985921.986094>
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(2), 310-339. <https://doi.org/10.1037/h0078820>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583-1585. <https://doi.org/10.1126/science.154.3756.1583>
- Kennedy, D. O., & Scholey, A. B. (2000). Glucose administration, heart rate and cognitive performance: effects of increasing mental effort. *Psychopharmacology*, 149(1), 63-71. <https://doi.org/10.1007/s002139900335>
- Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (pp. 69-72). Savannah, GA. <https://doi.org/10.1145/1344471.1344489>
- Landgraf, S., van der Meer, E., & Krueger, F. (2010). Cognitive resource allocation for neural activity underlying mathematical cognition: A multi-method study. *ZDM*, 42(6), 579-590. <https://doi.org/10.1007/s11858-010-0264-7>
- Malik, M. (1996). Heart rate variability. *Annals of Noninvasive Electrocardiology*, 1(2), 151-181. <https://doi.org/10.1111/j.1542-474X.1996.tb00275.x>
- Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine*, 78(5), B165-B175.
- Martin, J. (2017). A critical analysis of the theoretical construction and empirical measurement of cognitive load. In Zheng, R. Z. (Ed.), *Cognitive load measurement and application* (pp. 29-44). New York, NY: Routledge.
- McDuff, D., Gontarek, S., & Picard, R. (2014, August). *Remote measurement of cognitive stress via heart rate variability*. Paper presented at the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL. <https://doi.org/10.1109/EMBC.2014.6944243>
- 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., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology*, 93(1), 181-186. <https://doi.org/10.1037/0022-0663.93.1.181>
- 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. [https://doi.org/10.1207/S15326985EP3801\\_8](https://doi.org/10.1207/S15326985EP3801_8)

- Paas, F., & van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122-133.
- Paas, F., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1), 419-430. <https://doi.org/10.2466/pms.1994.79.1.419>
- Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2), 230-248. <https://doi.org/10.1111/bjet.12207>
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology*, 90(1), 25-35.
- Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744>
- Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5), 1243-1252. <https://doi.org/10.1109/tbme.2009.2038487>
- Ramos, B. P., & Arnsten, A. F. T. (2007). Adrenergic pharmacology and cognition: Focus on the prefrontal cortex. *Pharmacology & Therapeutics*, 113(3), 523-536. <https://doi.org/10.1016/j.pharmthera.2006.11.006>
- Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, 11(2), 374-385. <https://doi.org/10.1017/S1138741600004406>
- Riener, A., Ferscha, A., & Aly, M. (2009, September). *Heart on the road: HRV analysis for monitoring a driver's affective state*. Paper presented at the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Essen. <https://doi.org/10.1145/1620509.1620529>
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469-508. <https://doi.org/10.1007/s10648-007-9053-4>
- Schultheis, H., & Jameson, A. (2004, August). *Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods*. Paper presented at the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Eindhoven. [https://doi.org/10.1007/978-3-540-27780-4\\_26](https://doi.org/10.1007/978-3-540-27780-4_26)
- Seufert, T., Jänen, I., & Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Computers in Human Behavior*, 23(3), 1055-1071. <https://doi.org/10.1016/j.chb.2006.10.002>
- Siemens. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, 55, 37-76. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Sweller, J. (2017). The role of independent measures of load in cognitive load theory. In R. Z. Zheng (Ed.), *Cognitive load measurement and application* (pp. 3-8). New York, NY: Routledge.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition & Instruction*, 12(3), 185-233. [https://doi.org/10.1207/s1532690xci1203\\_1](https://doi.org/10.1207/s1532690xci1203_1)
- Szulewski, A., Fernando, S. M., Baylis, J., & Howes, D. (2014). Increasing pupil size is associated with increasing cognitive processing demands: A pilot study using a mobile eye-tracking device. *Open Journal of Emergency Medicine*, 2, 8-11. <https://doi.org/10.4236/ojem.2014.21002>
- Tryon, W. W. (1975). Pupillometry: A survey of sources of variation. *Psychophysiology*, 12(1), 90-93. <https://doi.org/10.1111/j.1469-8986.1975.tb03068.x>
- van Gerven, P. W., Paas, F., van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41(2), 167-174. <https://doi.org/10.1111/j.1469-8986.2003.00148.x>
- van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147-177. <https://doi.org/10.1007/s10648-005-3951-0>
- Wallen, E., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology, Research & Development*, 53(3), 59-72. <https://doi.org/10.1007/BF02504798>

Zheng, R., & Cook, A. (2012). Solving complex problems: A convergent approach to cognitive load measurement. *British Journal of Educational Technology*, 43(2), 233-246.  
<https://doi.org/10.1111/j.1467-8535.2010.01169.x>

---

**Corresponding author:** Il-Hyun Jo, [ijo@ewha.ac.kr](mailto:ijo@ewha.ac.kr)

**Please cite as:** Kim, J., & Jo, I. (2019). Exploring the feasibility and use of psychophysiological responses based on cognitive load theory. *Australasian Journal of Educational Technology*, 35(3), 150-165.  
<https://doi.org/10.14742/ajet.4163>