Received: 7 February 2024 | Accepted: 25 November 2024

DOI: 10.1111/bjep.12729

ARTICLE

the british psychological society
promoting excellence in psychology

Re-examining cognitive load measures in real-world learning: Evidence from both subjective and neurophysiological data

Xiaobo Liu | Yu Zhang

Institute of Education, Tsinghua University, Beijing, China

Correspondence
Yu Zhang, Institute of Education, Tsinghua University, Beijing, China.
Email: zhangyu2011@tsinghua.edu.cn

Funding information
National Natural Science Foundation of China; the Beijing Educational Science Foundation of the Fourteenth 5-year Planning

Abstract

Background: Cognitive load theory is widely used in educational research and instructional design, which relies heavily on conceptual constructs and measurement instruments of cognitive load. Due to its implicit nature, cognitive load is usually measured by other related instruments, such as commonly-used self-report scales of mental effort or task difficulty. However, these concepts are different in nature, as they emphasize distinct perspectives on cognitive processing. In addition, real-world learning is more complex than simplified experimental conditions. Simply assuming that these variables will change in a monotonic way with workload may be misleading.

Aims: This study aims to examine whether these measures are consistent with each other, and to discover the neurophysiological basis underlying the potential discrepancy.

Sample: The study collected data in both a real-world (Study 1, 22 high school students in 13 math classes) and a laboratory setting (Study 2, 30 students in 6 lab-based math tasks).

Methods: In addition to self-report measures, the study also collected multimodal neurophysiological data, such as electroencephalography (EEG), electrodermal activity (EDA), and photoplethysmography (PPG).

Results: The results show that although the difficulty level can be perceived with difficulty ratings, it does not lead to the corresponding level of mental effort. Only within an appropriate level of load, can we observe a positive correlation between self-report difficulty and mental effort. Neurophysiological evidence also supports the conceptual discrepancies and group differences, indicating distinct neurophysiological mechanisms underlying these 'similar' constructs.

Conclusions: These findings also emphasize the need for combining these concepts to better evaluate students' cognitive load.

KEYWORDS

cognitive load, cognitive load measurement, neurophysiological representation, real-world learning

INTRODUCTION

Working memory plays a central role in human learning, but due to its limited capacity, it is important to minimize unnecessary loads through appropriate instructional design. Addressing this, Sweller (1988) proposed cognitive load theory, defining cognitive load as 'the load imposed on an individual's working memory by a particular (learning) task' (van Gog & Paas, 2012). Cognitive load is categorized into intrinsic, extraneous, and germane components, focusing on the intrinsic complexity of information, how information is extraneously presented, and the load for schema construction and automation, respectively (Sweller, 2010). The total load is comprised of intrinsic and extraneous components, while germane cognitive load redistributes working memory resources from extraneous load to intrinsic load (Sweller et al., 2019). The conceptualization of cognitive load has greatly facilitated the improvement of instructional design (de Jong, 2010; Paas et al., 2003). Since the theory's inception, several instructional effects have been proposed, aimed at reducing extraneous cognitive load and ultimately enhancing learning outcomes (Sweller et al., 2019).

Measuring cognitive load in learning environments has therefore become a critical issue for theoretical advancement and instructional practice (Paas et al., 2016). A variety of measurement methods have been proposed from different disciplines. Cognitive load can be evoked by tasks with various levels of working memory demands, and can indirectly be reflected by task performance, response time, or error rates (Sweller et al., 2011). Similarly, cognitive load can also be assessed by observing the performance of a dual task in laboratory settings (Brünken et al., 2002). In real-world scenarios, directly computing the intrinsic cognitive load of tasks is more complex. Consequently, subjective ratings have become widely used. Among them, the Task Load Index (NASA-TLX) (Hart & Staveland, 1988) has been extensively employed to evaluate general workload in various kinds of tasks. When assessing cognitive load in educational settings, single-item scales have been frequently used. Two commonly used concepts are mental effort (Paas, 1992; Paas & van Merriënboer, 1994) and task difficulty (Ayres, 2006; Marcus et al., 1996). Both have been shown to sensitively measure the change in cognitive load imposed by instructional manipulations (Sweller et al., 2011). Recent studies have explored measuring differentiated components of intrinsic, extraneous, and germane cognitive load (Klepsch et al., 2017; Leppink et al., 2013; Orru & Longo, 2019), providing a deeper explanation of instructional effects (Klepsch & Seufert, 2020).

However, despite the convenience of subjective ratings, there have been considerable discrepancies in the conceptualization of cognitive load and its measurements. Working memory can operate unintentionally and outside of conscious awareness, making it challenging for learners to directly report its usage (Hassin et al., 2009). Therefore, cognitive load is typically measured by concepts associated with it, but not working memory per se. The assessment factors of cognitive load involve three distinct dimensions of mental load, mental effort, and performance (Paas & van Merriënboer, 1994). However, many cognitive load measures focus on either one of these dimensions or simply add them up, potentially oversimplifying the complexity of human learning. As summarized in a systematic review (Mutlu-Bayraktar et al., 2019), a proportion of researches measured cognitive load using only mental effort or general workload. Although previous studies have identified potential inconsistencies among different assessment factors (van Gog & Paas, 2008), these inconsistencies have often been neglected in empirical studies (Kirschner et al., 2011), and the causal mechanism underlying the discrepancy is still under investigation.

Moreover, the monotonic relationship between cognitive load and the abovementioned instruments cannot be easily guaranteed in real-world learning, although this might be the case in laboratory settings. Unlike when performing a simplified task in a well-controlled experiment, which usually involves no new knowledge acquisition, students in real-world learning need to engage in complicated tasks and knowledge construction processes (Bada & Olusegun, 2015; von Glasersfeld, 1987). Under such conditions, the cognitive load may be regulated dynamically to adapt to the changing demands of tasks (Seufert, 2020; Sweller et al., 2019). This instantaneous change may not be reflected in post-hoc scales, which mainly measure average or accumulated cognitive load (Sarailoo et al., 2022). Moreover, if learners experience overload, particularly during high-demand tasks, they may not invest the necessary working memory resources in learning, which violates the assumption of a monotonic relationship between mental load and mental effort (Paas et al., 2004, 2005). These inconsistencies may undermine the validity of the aforementioned cognitive load measures. To address these issues, new experiments are necessary to investigate the complex relationship between cognitive load and its existing measures.

Recent studies have demonstrated the potential of neurophysiological data for assessing cognitive load (Ayres et al., 2021) and for exploring the neural mechanisms underlying cognitive load (Medeiros et al., 2021). Working memory functions in multiple brain regions, including the prefrontal cortex (Cohen et al., 1997; Curtis & D'Esposito, 2003). As an indicator of working memory activities, cognitive load can be evaluated with brain activity measures including electroencephalography (EEG) (Chikhi et al., 2022; Zhou et al., 2021) and functional near-infrared spectroscopy (fNIRS) (Yuksel et al., 2016). Additionally, under a high cognitive load, mental stress induced by stimuli will affect sympathetic activity (Callister et al., 1992), activating a series of physiological responses including electrodermal and cardiovascular activities to manage the increased cognitive demands. Therefore, physiological measures of these activities, specifically electrodermal activity (EDA) and photoplethysmography (PPG), are employed as methods of measuring cognitive load (Lyu et al., 2015; Qi et al., 2020; Setz et al., 2009; Solhjoo et al., 2019).

As summarized in the review by Ayres et al. (2021), most neurophysiological studies rely on laboratory-based experiments. However, learning is situated in naturalistic environments (Barsalou, 2008; Immordino-Yang & Gotlieb, 2017), which are far more complex than laboratory settings. Bridging the gap between laboratory-based studies and real-world learning has thus become a critical issue (Miller, 2016; Stafford-Brizard et al., 2017). The advancement of portable biosensors enables educational researchers to analyse cognitive and emotional states in real-world learning environments (Dikker et al., 2017; Xu & Zhong, 2018; Zhang et al., 2018), providing empirical evidence for relevant theories. Cognitive load theory is also grounded in real-world learning (Murphy et al., 2016), emphasizing the need for naturalistic data collection. While a few studies have attempted to analyse the neural correlates of cognitive load under complex learning tasks, these features were less sensitive than those in traditional lab-based tasks (Castro-Meneses et al., 2020; Larmuseau et al., 2019), highlighting the need for additional empirical evidence.

Therefore, we have posed the following research question for re-examination: Are mental load and mental effort, as two subdimensions of cognitive load, different constructs in real-world learning? Two hypotheses are formulated regarding the research question: (i) The relationship between mental load and mental effort does not follow a monotonic pattern, and (ii) these measures are associated with different neural correlates. If the data are in line with these hypotheses, they will provide clear evidence that mental load and mental effort are indeed distinct constructs. The present study attempted to address this question by conducting both real-world research and laboratory-based research in the context of high school mathematics. During both studies, subjective measures of cognitive load and neurophysiological data were simultaneously collected. This study re-examines the conceptual differences in cognitive load measures in real-world learning, emphasizing the need for a multidimensional assessment construct. The findings also demonstrate the importance of using multimodal learning analytics to measure students' cognitive load more accurately in real-world learning environments.

MATERIALS AND METHODS

Framework

To examine the discrepancies among concepts, two experiments were conducted, as illustrated in Figure 1. Study 1 involved collecting data in a real-world classroom without interfering with teaching. The experience sampling method (ESM) was employed to record self-report measures of cognitive load during each class session, which has previously been validated in classroom learning settings (Dykstra & Paul, 2018; Xie et al., 2019). In Study 2, participants were presented with three types of mathematics tasks at two difficulty levels. In both studies, neurophysiological data, including EEG, EDA, and PPG data, were simultaneously collected. Relevant features were subsequently extracted to examine the discrepancies between mental load and mental effort.

In both studies, cognitive load was assessed through three concepts: task difficulty, perceived difficulty, and mental effort. The first two concepts pertain to the mental load dimension. As shown in Figure 1c, the re-examination detailed the analysis of the relationship between measures. First, the association between task difficulty and perceived difficulty was analysed to verify the measurement of mental load (Analysis [1]). Then, the association between mental load and mental effort was analysed to examine the consistency of the concepts (Analyses [2] and [3]). Finally, the association between cognitive load concepts and neurophysiological features was analysed (Analysis [4]), further examining the underlying neural basis.

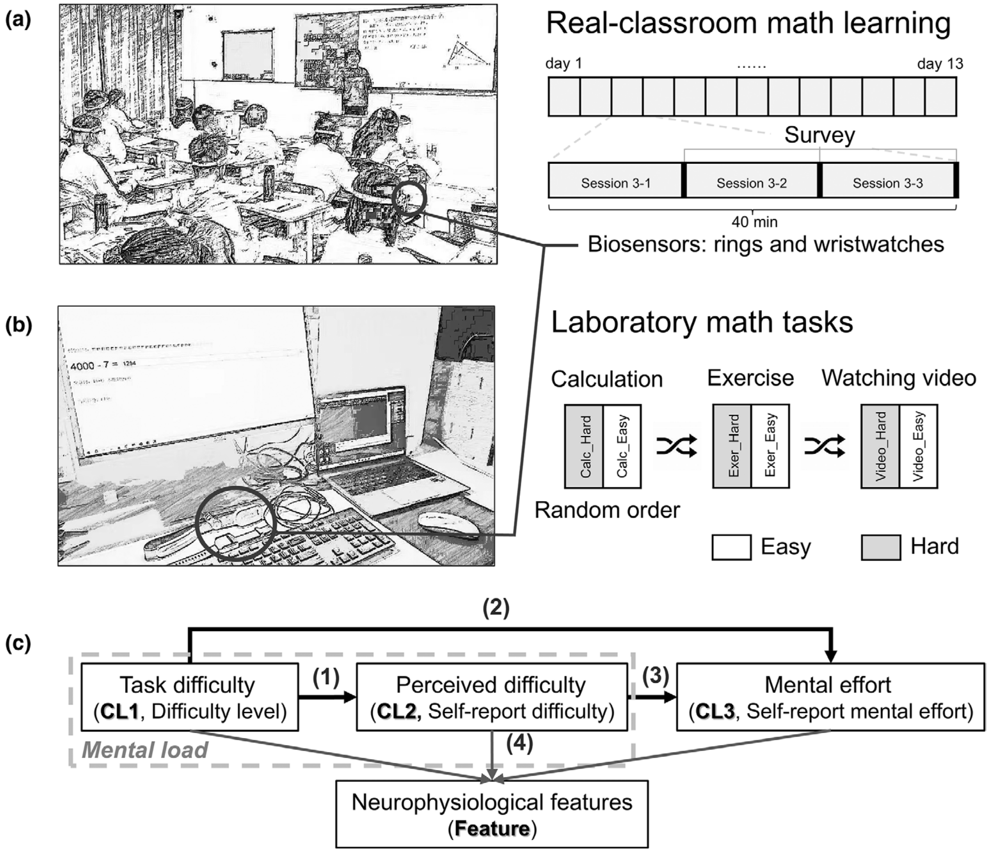

FIGURE 1 Schematic framework of the studies. (a) Study 1, which took place in a real-classroom setting. (b) Study 2, which involved learning tasks conducted in a laboratory setting. (c) Analytical framework, which examines the relationship between mental load and mental effort, as well as their neural correlates.

Study 1: Real-classroom learning

Participants

All participants were recruited from a G10 class in a typical high school in Beijing. In the class, 22 out of 25 students participated in the study (9 girls and 13 boys aged 15–16 years; all of them were right-handed). The study complied with Chinese laws and the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of the Department of Psychology, Tsinghua University. All participants and their legal guardians read and signed the informed consent form. Data collection took place in March 2023 and involved 13 math lessons that focused on planar vector and trigonometric functions.

Instruments

In each lesson (40 min), the teacher was requested to pause twice, dividing the lesson into three sessions of approximately 10–15 min each. The ESM survey was administered during the break after each session. In the survey, students were invited to report their cognitive load for the previous 10 min. This approach ensured alignment of the duration of each session. Cognitive load was measured on the ESM scale using the single-item scales of mental effort (Paas, 1992) and difficulty (Ayres, 2006). Students were invited to report their degree of mental effort and perceived difficulty during the preceding 10 min via a 5-point Likert scale (Camp et al., 2001; Leppink et al., 2013). The scale details are shown in Table S1.

To evaluate the difficulty level of real-classroom learning, which aligns with the construct of task difficulty in laboratory settings, the difficulty level for each session was determined by averaging the self-report difficulty scores from all the students (Centra, 2003).

Neurophysiological data collection

A head belt and a wristwatch were used to collect neurophysiological data. The head belt records EEG signals with dry electrodes at Fp1 and Fp2 over the forehead at a sampling rate of 250 Hz (Brainno, SOSO H&C, South Korea). The device has been previously utilized in educational contexts to explore differences across various disciplines (Chen et al., 2023). The wristwatch collects EDA signals at a sampling rate of 40 Hz and collects PPG signals at 20 Hz (Psychorus, China). This device has been utilized in prior studies that focused on real-classroom scenarios (Liu et al., 2021; Zhang et al., 2018, 2021). In this study, signals from the 10-min interval before each survey were retained for further analysis.

Data analysis

In Study 1, mixed effect models were applied because the repeatedly measured data exhibited a high degree of intra-personal similarity, as evidenced by the intraclass correlation (ICC) reported in Table S2. The variance of the residuals is composed of two parts: within-person variance and between-person variance. Thus, the models are set up with clustered standard error:

$$CL2_{ij} = \beta_{10} + \beta_{11} CL1_{ij} + \mu_{1j} + \epsilon_{1ij} \tag{1}$$

$$CL3_{ij} = \beta_{20} + \beta_{21} CL1_{ij} + \mu_{2j} + \epsilon_{2ij} \tag{2}$$

$$CL3_{ij} = \beta_{30} + \beta_{31} CL2_{ij} + \mu_{3j} + \epsilon_{3ij} \tag{3}$$

where CL1, CL2, CL3 represent the difficulty level of the session, self-report difficulty, and self-report mental effort, respectively. The subscripts i and j represent session i and student j. The coefficient β_{n1} (i.e., β_{11}, or β_{21}, or β_{31}) represents the effect of the model, and β_{m0} (i.e., β_{10}, or β_{20}, or β_{30}), μ_j, and ϵ_{ij} respectively represent the intercept, the between-person residual, and the within-person residual.

Regarding the neurophysiological representation of cognitive load, different measurements of cognitive load may be represented by different kinds of neurophysiological activities, and the model is constructed as:

$$Feature_{ij} = \beta_{40} + \beta_{41} CL_{ij} + \mu_{4j} + \epsilon_{4ij} \tag{4}$$

where $Feature_{ij}$ represents the neurophysiological feature (e.g., the power of delta, theta, and alpha bands in EEG; see Section 2.4 Neurophysiological data processing for a complete list and detailed discussion) of student j in session i, and CL_{ij} represents the corresponding self-report cognitive load measures (CL1, or CL2, or CL3). The coefficient β_{41} represents the effect between self-report measures and neurophysiological features, and β_{40}, μ_{4j}, and ϵ_{4ij} are the intercept, the between-person residual, and the within-person residual, respectively. All variables were standardized before fitting the model (through the z-score).

Subgroup analysis

As suggested by relevant theories (Paas et al., 2004; Paas et al., 2005), the relationship between mental load and mental effort may be influenced by the gap between task demand and learner ability. Therefore, Study 1 categorized participants into high- or low-performing students and learning sessions into easy or hard sessions. This division resulted in a 2×2 subgroup analysis, creating varying task demands and learner abilities.

The subgroup division was conducted as follows. Before the data were collected, an exam was administered to assess the math ability of the students. Of the 22 participants in the class, eleven were placed in the top half, and the rest were placed in the bottom half. These two subgroups, referred to as high-performing and low-performing students, were defined for further analysis. Similarly, all sessions were categorized as easy or hard based on the estimated difficulty level, using the average score as a threshold. This categorization resulted in 21 easy sessions and 18 hard sessions.

Study 2: A supplemental laboratory-based experiment

Participants

In addition to the 22 participants in Study 1, 8 additional participants (6 girls and 2 boys) were recruited from the same student cohort. All newly added participants and their legal guardians read and signed an informed consent form.

Tasks

There were three pairs of mathematical tasks, each consisting of two levels of difficulty. The calculation tasks, adopted from Qi et al. (2020), were validated to induce different levels of cognitive load for high school students. In addition, this study also involved two other kinds of tasks: exercise and video watching, in order to simulate real learning tasks. The exercise tasks were selected from a database of high-school math exams. The difficult problems involved more math concepts and steps, whereas the easy problems involved fewer concepts. With high element interactivity, the difficult problems were assumed to induce higher intrinsic cognitive load (Sweller, 2010). The videos, both produced by the same teacher blogger to ensure consistency in teaching style, were selected from an open-source video website. Both videos involved solving math problems, with the difficult video working through a complex problem and the easy video working through a simple one. All the materials used in the experiment were selected by the researchers according to the specific knowledge domain relevant to the sampled students, and were further double confirmed by their math teacher. The teacher also examined the difficulty of all problems and videos using a binary rating.

Procedures

Before the experiment, the participants were instructed to wear the same devices for neurophysiological data collection as in Study 1. After a short relaxation period, the participants sequentially completed three sets of tasks. Upon the completion of each task, the participants were asked to rate their mental effort and perceived task difficulty using the same scale applied in Study 1. The average duration of each task was 5 min and the duration of relaxation between the tasks was 4 min. The order of tasks within each pair was randomly assigned to each participant to control for order effects and potential biases.

Neurophysiological data processing

The EEG processing protocol was designed to automatically remove artefacts (Chen et al., 2023). The major sources of artefacts in the EEG signal include missing data, transient signals caused by losing contact, slow drifts, and ocular artefacts. To address these issues, missing data were first identified automatically. Subsequently, robust detrending (de Cheveigné & Arzounian, 2018) was applied, followed by bandpass filtering between 1 Hz and 40 Hz, and ocular artefact removal (Kanoga et al., 2019). These automatic processing techniques cannot guarantee complete removal of artefacts. Therefore, epochs with data points exceeding ±150 μV, which were regarded as containing artefacts, were excluded. In this study, a 5-s epoch length was considered suitable for analysis (Figure S2). Ultimately, an average of 49.6% ($SD = 24.3\%$) of epochs were retained for each trial, a rate that is considered acceptable in classroom settings. During feature extraction, the delta (1 Hz – 4 Hz), theta (4 Hz – 8 Hz), alpha (8 Hz – 13 Hz), beta (13 Hz – 30 Hz), and gamma (30 Hz – 45 Hz) band powers were estimated through fast Fourier transform. To avoid interruptions caused by extreme values, the median value of epochs within each trial was considered the feature of that corresponding trial.

After identifying and deleting missing epochs, the EDA data were downsampled to 10 Hz and then smoothed. Filtering methods were applied to remove background noise (Zhang et al., 2021). The EDA data were then decomposed into two components using the cvxEDA python toolbox (Greco et al., 2016): the tonic skin conductance level (SCL) and phasic skin conductance response (SCR). Both components were segmented into 10-s epochs, from which the mean value of the SCLs and the integration of the SCRs (iSCR) were extracted. To derive features for each trial, the means and standard deviations (std) of SCL and iSCR (Zhang et al., 2018) were computed, generating four features, namely, the SCL mean, SCL std., iSCR mean and iSCR std.

Heart rate (HR) and pulse rate variability (PRV) were estimated by sliding window analysis of the PPG signals, performed with software provided by Psychorus, China. The values were subsequently averaged over the 10-minute period.

In both Study 1 and Study 2, neurophysiological data were processed using the aforementioned pipeline in the same manner. The quality of multimodal signals is validated in Figure S1, demonstrating the feasibility of data collection and processing. To address missing data, mean imputation was employed in Study 1, replacing the missing data with the mean value for each student (Waljee et al., 2013). Given the small sample size in Study 2, missing values were excluded from the analysis. Details of the missing data can be found in Figure S3.

RESULTS

Discrepancies among concepts in study 1

In Study 1, the mean value of students' mental effort was 3.84 ($SD = 0.81$) and the mean value of perceived difficulty was 2.27 ($SD = 0.95$) ($n = 813$, for a detailed distribution, see Figure S4). The mean level of session difficulty was 2.27 ($SD = 0.25$) among 39 learning sessions.

Table 1 presents the relationships between the difficulty level of the sessions, self-report difficulty, and self-report mental effort. Columns (1)–(3) report the coefficients of models (1)–(3) respectively. A significant positive relationship between the difficulty level and self-report difficulty was found ($\beta = 0.242, SE = 0.021, p < .001$). However, the relationship between the difficulty level and self-report mental effort was not significant ($\beta = -0.012, SE = 0.022, p = .571$), suggesting that mental effort did not change with the difficulty level in real-classroom settings. The trend is also illustrated by the scatter plot in Figure 2 (for the correlation analysis results, see Table S3).

The relationship between self-report difficulty and self-report mental effort was significantly positive for the whole sample ($\beta = 0.118, SE = 0.034, p < .001$). However, when considering subgroups, a positive relationship was found to be significant only for low-performing students when learning easy sessions (LE: $\beta = 0.159, SE = 0.067, p = .019$), and for high-performing students when learning hard sessions (HH: $\beta = 0.310, SE = 0.071, p < .001$). Conversely, there was no significant relationship for high-performing students when learning easy sessions (HE: $\beta = -0.025, SE = 0.068, p = .718$), or for low-performing students when learning hard sessions (LH: $\beta = -0.118, SE = 0.077, p = .127$).

TABLE 1 Results of the behavioural data in Study 1.

Subgroup	Analysis (1)	Analysis (2)	Analysis (3)
All (n = 813)	0.242***	−0.012	0.118***
	(0.021)	(0.022)	(0.034)
LE (n = 217)	0.205***	0.018	0.159*
	(0.036)	(0.038)	(0.067)
LH (n = 179)	0.176***	−0.047	−0.118
	(0.039)	(0.043)	(0.077)
HE (n = 225)	0.136***	−0.086*	−0.025
	(0.040)	(0.042)	(0.068)
HH (n = 192)	0.125*	0.040	0.310***
	(0.050)	(0.052)	(0.071)

Note: *$p < .05$, ** $p < .01$, *** $p < .001$. Standard errors in parentheses. The mixed effects models of Analyses (1)–(3) are:

1. $CL2_{ij} = \beta_{10} + \beta_{11} CL1_{ij} + \mu_{1j} + \epsilon_{1ij}$
2. $CL3_{ij} = \beta_{20} + \beta_{21} CL1_{ij} + \mu_{2j} + \epsilon_{2ij}$
3. $CL3_{ij} = \beta_{30} + \beta_{31} CL2_{ij} + \mu_{3j} + \epsilon_{3ij}$

where CL1, CL2, CL3 are respectively difficulty level, self-report difficulty, and self-report mental effort. In these models, the coefficients β_{11}, β_{21}, and β_{31} were estimated, and their values are provided in Table 1.

Abbreviations: HE, high-performing students learning easy sessions; HH, high-performing students learning hard sessions; LE, low-performing students learning easy sessions; LH, low-performing students learning hard sessions.

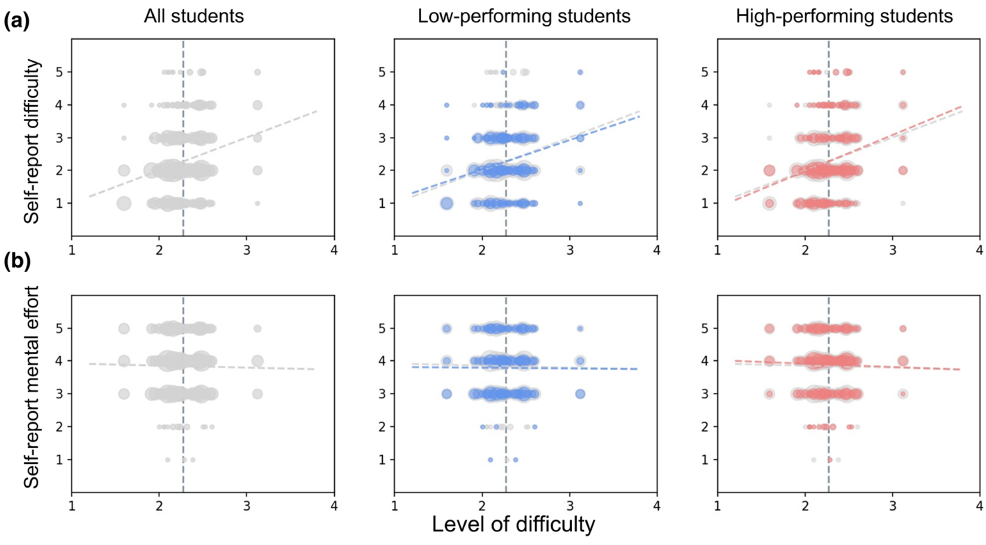

FIGURE 2 Scatter plot of the behavioural data in Study 1. (a) Difficulty level and self-report difficulty (Analysis [1]). (b) Difficulty level and self-report mental effort (Analysis [2]). The size of the bubble corresponds to the count of repeated sample points. The vertical dashed line represents the threshold separating easy and hard sessions. The coloured dashed line represents the trend, estimated through the ordinary least squares method.

Results from laboratory math tasks

In Study 2, the mean value of students' mental effort was 4.57 ($SD = 0.65$), and the mean value of perceived difficulty was 2.38 ($SD = 1.05$) ($n = 175$). Figure 3 illustrates the relationship between the cognitive load measures, echoing findings from Study 1. Different levels of task difficulty were significantly associated with perceived difficulty, as demonstrated by pairwise t-tests: Calculation ($t = 5.22, df = 29, p < .001$), Exercise ($t = 8.27, df = 29, p < .001$), and Watching video ($t = 4.00, df = 26, p < .001$). However, these levels were not significantly associated with self-report mental effort: Calculation ($t = -1.14, df = 29, p = .264$), Exercise ($t = -0.701, df = 29, p = .489$), or Watching video ($t = -0.328, df = 26, p = .746$), which was consistent with the findings in Study 1. Similarly, there was no significant correlation between self-report mental effort and perceived difficulty in any of the three tasks: Calculation ($r = -.069, p = .599, n = 60$), Exercise ($r = -.136, p = .300, n = 60$), or Watching Video ($r = -.142, p = .303, n = 55$).

Neurophysiological representation in study 1

The neurophysiological analysis from Study 1 showed that the difficulty level and self-report mental effort correlated with different neurophysiological features. As detailed in Table 2, for the whole sample, significant associations were found between the sessions' difficulty level and several features including EEG theta ($\beta = -0.069, SE = 0.025, p = .006$), alpha ($\beta = -0.069, SE = 0.024, p = .004$), beta ($\beta = -0.043, SE = 0.015, p = .003$), SCL mean ($\beta = 0.075, SE = 0.030, p = .013$), and HR ($\beta = -0.143, SE = 0.025, p < .001$). Self-report mental effort was significantly associated with quite different features, or showing different directions, including EEG delta ($\beta = 0.119, SE = 0.042, p = .004$), theta ($\beta = 0.119, SE = 0.040, p = .003$), alpha ($\beta = 0.092, SE = 0.038, p = .015$), and SCL mean ($\beta = -0.117, SE = 0.046, p = .011$).

In addition to the general correlations observed in the whole sample, the neurophysiological representations of the cognitive load measurements differed across the various subgroups (see Figure 4 and Table S4 and S5). Specifically, the relationship between difficulty level and SCL mean was significant only in LE

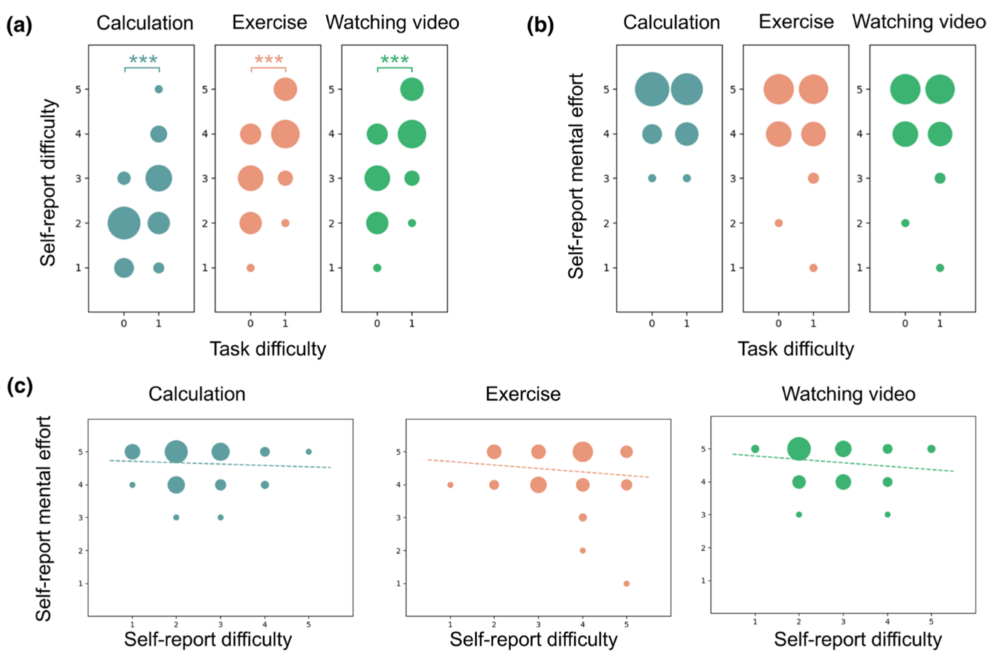

FIGURE 3 Scatter plot of the behavioural data in Study 2. (a) Task difficulty (0 for the easy task and 1 for the hard task) and self-report difficulty (Analysis [1]). (b) Task difficulty and self-report mental effort (Analysis [2]). (c) Self-report difficulty and self-report mental effort (Analysis [3]). The size of the bubble corresponds to the count of repeated sample points. $^{***}p < .001$.

($\beta = 0.175, SE = 0.057, p = .002$) and HE ($\beta = 0.201, SE = 0.059, p = .001$) subgroups. Regarding the association between difficulty level and HR, it was significant in LH ($\beta = -0.101, SE = 0.044, p = .022$), HE ($\beta = -0.120, SE = 0.057, p = .034$), and HH ($\beta = -0.128, SE = 0.061, p = .036$) subgroups, demonstrating the stability of the association. In the LH subgroup, self-report mental effort was significantly associated with SCL std. ($\beta = -0.251, SE = 0.096, p = .009$), iSCR mean ($\beta = -0.257, SE = 0.097, p = .008$), and iSCR std. ($\beta = -0.257, SE = 0.097, p = .008$), which was not significant in the whole sample.

Neurophysiological representation in study 2

In Study 2, significant correlations were observed between cognitive load measures and neurophysiological features, providing additional evidence for the neurophysiological meaning of these measures. These correlations are shown in Figure 5 and Table S6.

Mental load and mental effort also correspond to different neurophysiological features in Study 2, echoing the results of Study 1. However, some different findings were observed. HR was negatively associated with self-report mental effort in calculation tasks ($r = -.411, p = .004, n = 47$), especially in the hard calculation task ($r = -.670, p < .001, n = 23$), but the association was significantly positive in the easy video watching task ($r = .495, p = .037, n = 18$). This finding contrasts with Study 1, where HR was negatively associated with task difficulty. Additionally, significant and negative associations were found between self-report mental effort and SCL std., iSCR mean, and iSCR std. in both the hard exercise task (SCL std. ($r = -.770, p < .001, n = 23$); iSCR mean ($r = -.795, p < .001, n = 23$); iSCR std. ($r = -.789, p < .001, n = 23$)) and the hard video watching task (SCL std. ($r = -.619, p = .002, n = 23$); iSCR mean ($r = -.516, p = .012, n = 23$); iSCR std. ($r = -.581, p = .004, n = 23$)). These associations

TABLE 2 Results of the neurophysiological data analysis in Study 1.

	EEG features					EDA features				PPG features	
	Delta	**Theta**	**Alpha**	**Beta**	**Gamma**	**SCL mean**	**SCL std**	**iSCR mean**	**iSCR std**	**HR**	**PRV**
	All students (n = 813)										
Difficulty level	−0.048	−0.069**	−0.069**	−0.043**	−0.014	0.075*	−0.028	−0.053	−0.046	−0.143***	0.021
	(0.027)	(0.025)	(0.024)	(0.015)	(0.015)	(0.030)	(0.031)	(0.032)	(0.031)	(0.025)	(0.030)
Self-report difficulty	0.030	−0.008	−0.009	−0.016	−0.020	0.078	0.077	0.004	0.019	0.010	−0.079
	(0.041)	(0.039)	(0.037)	(0.022)	(0.023)	(0.045)	(0.045)	(0.046)	(0.046)	(0.039)	(0.044)
Self-report mental effort	0.119**	0.119**	0.092*	0.032	−0.007	−0.117 *	−0.055	−0.046	−0.062	0.031	0.005
	(0.042)	(0.040)	(0.038)	(0.023)	(0.024)	(0.046)	(0.047)	(0.047)	(0.047)	(0.040)	(0.046)

Note: The effects were estimated through the mixed effects model: $Feature_{ij} = \beta_{40} + \beta_{41} CL_{ij} + \mu_{4j} + \epsilon_{4ij}$ (Analysis [4]). In the model, the coefficient β_{41} was estimated, and the values are provided in Table 2. *$p < .05$, ** $p < .01$, *** $p < .001$. Standard errors in parentheses.

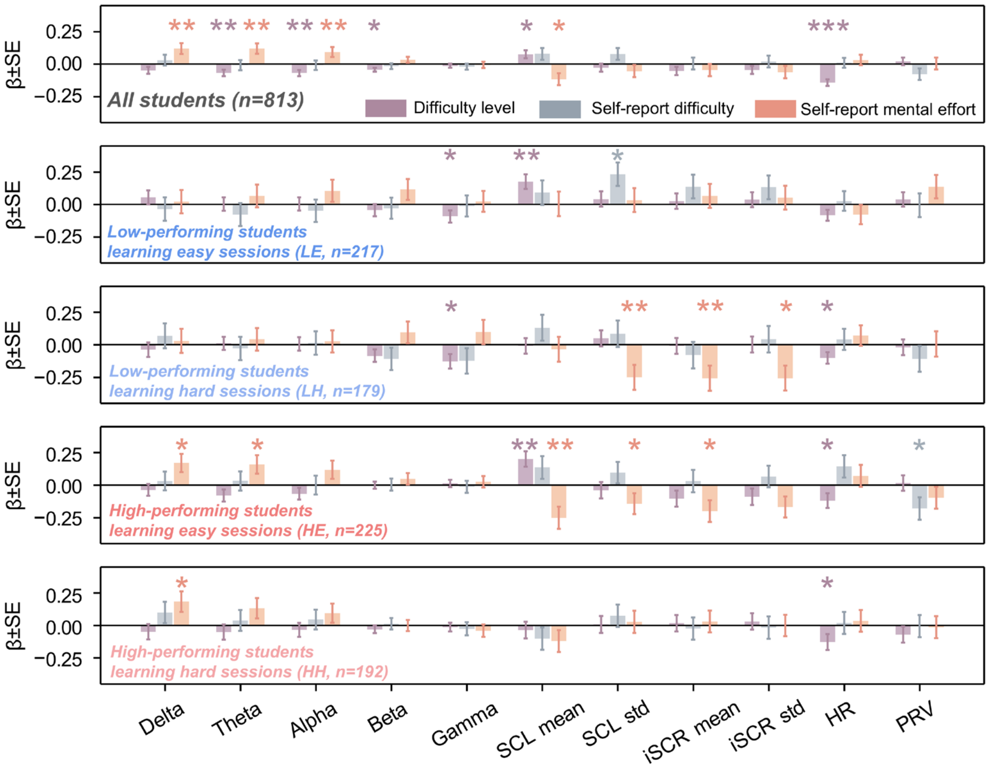

FIGURE 4 Results of the neurophysiological data in Study 1 by subgroups. The effects, β_{41}, were estimated through model (4). * $p < .05$, ** $p < .01$, *** $p < .001$.

were consistent with the LH subgroup in Study 1, suggesting a similar response when students face challenging tasks.

DISCUSSION

Cognitive load is usually estimated with the predetermined difficulty level of tasks, or measured with self-report items of difficulty or mental effort. Our findings, however, indicate that the association between these variables does not follow a monotonic pattern, aligning with the findings of previous theoretical and empirical studies (Minkley et al., 2021; Sweller et al., 2019; van Gog & Paas, 2008). In both real-world and laboratory settings, students' mental effort did not correspond to the difficulty level of the learning tasks. Moreover, these measures were accompanied by different neurophysiological representations, further demonstrating their discrepancy.

Both mental load and mental effort are associated with the activity of working memory, but in different ways. Mental load indicates the expected working memory capacity demand while mental effort indicates the actual working memory capacity allocated to meet the task demand (Paas & van Merriënboer, 1994). This distinction can be understood as the difference between task demand and actual cognitive investment. In addition, the inconsistent relationship can also be explained with the interplay between cognitive load and self-regulation (Seufert, 2020; Wang & Lajoie, 2023). Specifically, in the effort monitoring and regulation (EMR) framework (de Bruin et al., 2020), mental effort can be regulated based on task experience or predetermined goals (data- or goal-driven). This regulatory process can lead to varying dynamic patterns of mental effort, different from those of mental load.

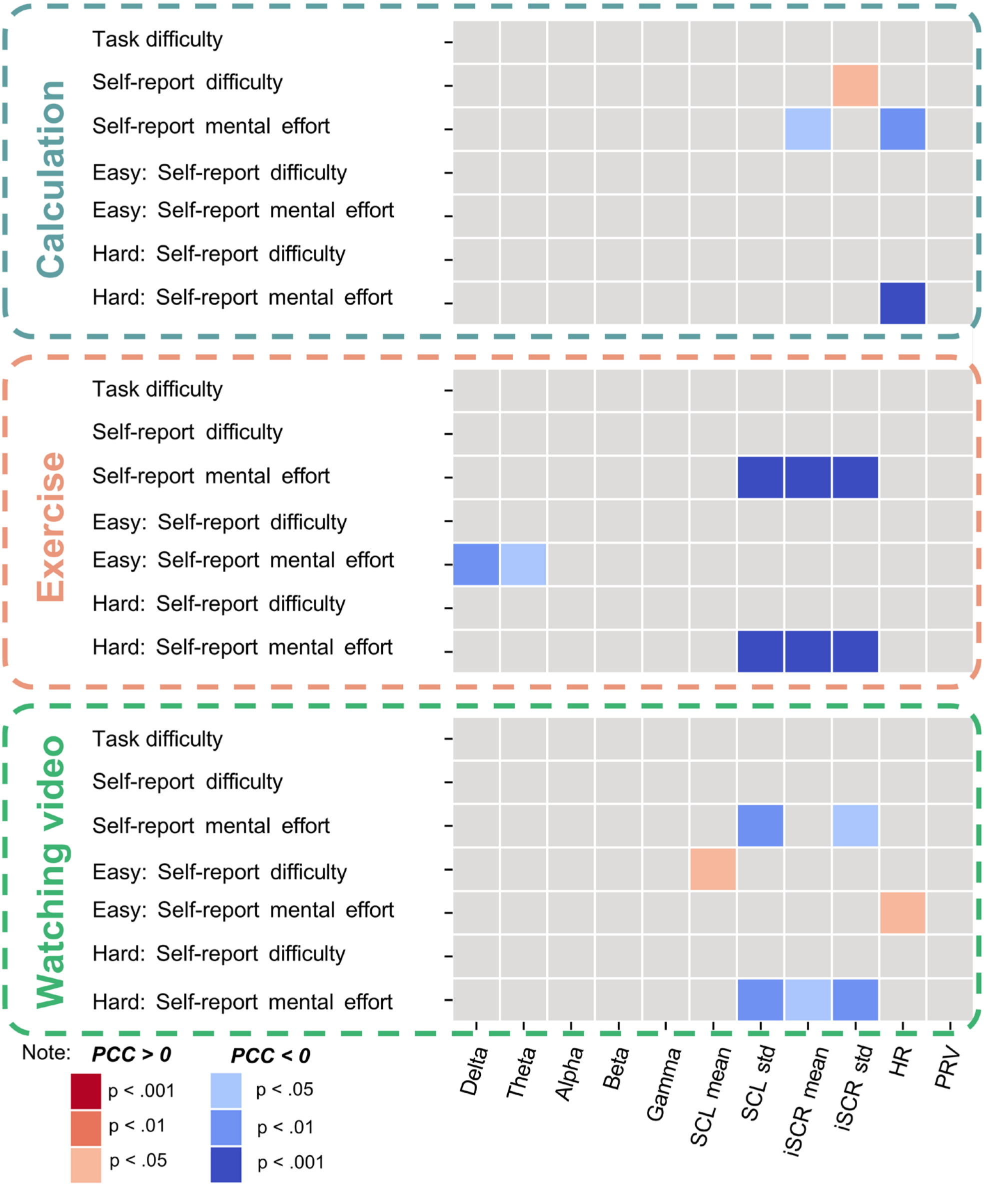

FIGURE 5 Results of the neurophysiological data analysis in Study 2. PCC: Pearson's correlation coefficient.

Neurophysiological evidence also supports this conceptual discrepancy. Increased task difficulty was associated with EEG alpha oscillations, which may indicate a 'wakefulness' state of the brain (Chikhi et al., 2022; Klimesch et al., 2007). It was also associated with increased EDA activity (SCL mean), which is an indicator of high stress and arousal (Setz et al., 2009). The increased task difficulty was also associated with decreased HR. HR is influenced by both the sympathetic and parasympathetic nervous systems (Hughson et al., 1994). Although sympathetic activation typically results in an increased HR, the parasympathetic system may regulate this response under mental demands, as evidenced by a similar study involving clinical reasoning tasks (Solhjoo et al., 2019). In contrast, mental effort exhibited a significant positive association with EEG theta, which is regarded as an indicator of cognitive control during working memory tasks (Cavanagh & Frank, 2014; Jensen & Tesche, 2002; Riddle et al., 2020).

Notably, for several features (such as theta, alpha, and SCL mean), the coefficients for different measures displayed opposite trends, further underscoring this inconsistency. Therefore, when students perceive difficulty, a range of subjective feelings induced by the task may co-occur (Erinosho, 2013; Kraft et al., 2005). When individuals invest mental effort, their related cognitive processing activities may be activated, which are not necessarily induced by the perceived difficulty.

Another noteworthy finding is that no physiological indicators were significantly associated with self-report difficulty in Study 1. A possible explanation for the result is the complex factors influencing self-report difficulty. The difficulty level, as averaged across all students' reports, provides a relatively objective reflection of the cognitive processing demands of the task. Mental effort directly reflects the mental processing expenditure of learners. However, compared with the other two measures, self-report difficulty is influenced by the interaction between tasks and learners, particularly the interaction between prior knowledge and task demand (Amadieu et al., 2009). Therefore, the neurophysiological features extracted in this study are not sufficient to sensitively reflect these complex factors.

To achieve a monotonic relationship between the measures, the cognitive load level should be neither too low nor too high (Paas et al., 2005). As revealed in the subgroup analysis, when low-performing students were learning easy sessions (LE group) and when high-performing students were learning hard sessions (HH group), mental effort was significantly correlated with perceived difficulty. In these two groups, task demand is more appropriate for learners' ability, which is favourable for investing in corresponding effort (van Gog & Paas, 2008).

In contrast with the appropriate level of cognitive load, cognitive overload in the learning environment can be disruptive for practitioners. In the present study, low-performing students in the hard sessions (LH group) may have experienced higher cognitive load due to a deficiency in prior knowledge and schema automation (Amadieu et al., 2009; Endres et al., 2023). This overwhelming cognitive load made it difficult for them to invest sufficient effort. Therefore, their mental effort did not significantly correspond to their perceived difficulty (compared with that of the LE and HH groups). Additionally, several theories can explain this phenomenon from the motivational perspective. For example, according to achievement motivation theory (Capa et al., 2008; Weiner, 1985), students are likely to avoid effort investment when perceiving challenging tasks when they are failure-avoidant. Similar models can also offer explanations, including self-efficacy (Bandura, 1989), compensatory control (Hockey, 1997) and the expected value of control (EVC) (Shenhav et al., 2021).

Neurophysiological analysis provides another perspective to explain this phenomenon. In the LH group, their self-report mental effort was found to be negatively associated with EDA features (SCL std., iSCR mean, and iSCR std). The observed negative effect of the variation in SCL over 10-s epochs (SCL std) suggested that certain classroom events may have resulted in higher activation levels in these students, potentially disrupting the stability of students' signals and ultimately leading to a decrease in mental effort. Similarly, the negative effects observed in the SCR features, both the mean value of the skin conductance response (iSCR mean) and its variation over epochs (iSCR std), indicate that the students' responses to certain events could also reflect this disruption. These findings suggest that during challenging tasks, mental effort may predominantly serve as regulating rather than maintaining a high arousal (Christopoulos et al., 2019). It is also worth noting that, in the difficult exercise and video watching tasks in Study 2, these three EDA features also showed negative associations with self-report mental effort. With a high average self-report difficulty (4.13 and 2.92), these two tasks may have triggered an overload state for the whole sample. Our results suggest that overload, as indicated by high EDA features, may disrupt students' learning with stress and arousal, leading to a disparity between perceived difficulty and invested mental effort.

Based on the subjective and neurophysiological analyses, both mental load and mental effort are indispensable factors of cognitive load assessments. Neurophysiological measures can help in distinguishing between these dimensions. In this study, some features (e.g., EEG alpha band power, EDA SCL mean, HR) were associated with the mental load dimension, while some other features (e.g., EEG theta band power) were linked to the mental effort dimension. This finding advances the work of Ayres et al. (2021) by taking a step further in identifying specific features that are sensitive to either mental load or mental effort. Additionally, by modelling these features, neurophysiological data are promising to provide separate representations for the mental load and mental effort dimensions.

It has also become promising to distinguish three components of cognitive load with a combined measure of mental load and mental effort. Intrinsic difficulty, extraneous factors, and schema construction processes may exhibit different patterns of mental load or effort, although they all contribute to the overall cognitive load (Galy et al., 2012). With multimodal neurophysiological features, machine learning models are promising in capturing these patterns and distinguishing among the different components (Kruger & Doherty, 2016).

A few limitations of this study should be noted. First, performance as another dimension of the assessment framework of cognitive load, was not directly measured during classroom learning. In this study, the very tight break time after each class session made it very challenging to distribute after-class tests. Second, the sample size of participants was relatively small ($n=22$). Considering the challenges of collecting neurophysiological data in real-world learning, the sample size aligns with similar works (Chen et al., 2023; Dikker et al., 2017), which is also close to the size of one class. Future research should combine cognitive load assessments with direct performance metrics in larger and more diverse classroom settings.

CONCLUSION

The re-examination of cognitive load measures emphasizes the crucial issue of the inconsistency among related concepts. While existing methods have been proven effective in assessing one of the concepts (i.e., mental load, mental effort, and performance), they may not indicate the same thing. To address this conceptual inconsistency, neurophysiological data has demonstrated its potential in measuring cognitive load, which is promising for providing more enriched evidence of cognitive load across these diverse perspectives.

AUTHOR CONTRIBUTIONS

Xiaobo Liu: Data curation; software; methodology; writing – original draft; investigation; conceptualization. **Yu Zhang:** Conceptualization; funding acquisition; writing – review and editing; validation; project administration; supervision; resources; methodology.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (62177030) and the Beijing Educational Science Foundation of the Fourteenth 5-year Planning (BGEA23019). The authors appreciate Jun Wei, Ran Liu, Dan Zhang, and Yongqiang Lyu for their support of the project. The authors appreciate Xinqiao Gao, Baosong Li, Zheng Zhang, Zihao Yang, and Lan Fu for their collaboration during the data collection. The authors also appreciate Xin Liu, Ziyang Luo, Yawen Tan, Fei Qin, Yun Long, Haifeng Luo, Zhilin Qu, Guannan Yao, Zheng Dong, Ziyan Xu, Wanyan Sun, Xiaomeng Xu, Xuming Li, Mingxuan Gao, Yanzhuo Zhu, Yujia Luo, Zhaochun Wen, and Yiming Lin for their assistance during the data collection.

CONFLICT OF INTEREST STATEMENT

The authors report there are no competing interests to declare.

DATA AVAILABILITY STATEMENT

The raw data are not publicly available due to underage privacy protection. The de-identified dataset during and/or analysed during the current study are available from the corresponding author on reasonable request.

ORCID

Yu Zhang https://orcid.org/0000-0002-0059-270X

REFERENCES

Amadieu, F., van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, *19*(5), 376–386.

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(3), 287–298.

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, 702538.

Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, *5*(6), 66–70.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, *25*(5), 729–735.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*(2), 109–119.

Callister, R., Suwarno, N. O., & Seals, D. R. (1992). Sympathetic activity is influenced by task difficulty and stress perception during mental challenge in humans. *The Journal of Physiology*, *454*(1), 373–387.

Camp, G., Paas, F., Rikers, R., & van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior*, *17*(5–6), 575–595.

Capa, R. L., Audiffren, M., & Ragot, S. (2008). The interactive effect of achievement motivation and task difficulty on mental effort. *International Journal of Psychophysiology*, *70*(2), 144–150.

Castro-Meneses, L. J., Kruger, J.-L., & Doherty, S. (2020). Validating theta power as an objective measure of cognitive load in educational video. *Educational Technology Research and Development*, *68*, 181–202.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*, 495–518.

Chen, J., Qian, P., Gao, X., Li, B., Zhang, Y., & Zhang, D. (2023). Inter-brain coupling reflects disciplinary differences in real-world classroom learning. *npj Science of Learning*, *8*(1), 11.

Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, *59*(6), e14009.

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods*, *22*(1), 394–420.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–608.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423.

de Bruin, A. B., Roelle, J., Carpenter, S. K., & Baars, M. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review*, *32*, 903–915.

de Cheveigné, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage*, *172*, 903–912.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., van Bavel, J. J., & Ding, M. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375–1380.

Dykstra, J., & Paul, C. L. (2018). Cyber operations stress survey (COSS): Studying fatigue, frustration, and cognitive workload in cybersecurity operations. In *11th USENIX workshop on cyber security experimentation and test (CSET 18)*. USENIX Association.

Endres, T., Lovell, O., Morkunas, D., Rieß, W., & Renkl, A. (2023). Can prior knowledge increase task complexity? – Cases in which higher prior knowledge leads to higher intrinsic cognitive load. *British Journal of Educational Psychology*, *93*, 305–317.

Erinosho, S. Y. (2013). How do students perceive the difficulty of physics in secondary school? An exploratory study in Nigeria. *International Journal for Cross-Disciplinary Subjects in Education*, *3*(3), 1510–1515.

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.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, *63*(4), 797–804. https://doi.org/10.1109/TBME.2015.2474131

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. *52*, pp. 139–183). Elsevier.

Hassin, R. R., Bargh, J. A., Engell, A. D., & McCulloch, K. C. (2009). Implicit working memory. *Consciousness and Cognition*, *18*(3), 665–678.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93.

Hughson, R., Yamamoto, Y., McCullough, R., Sutton, J., & Reeves, J. (1994). Sympathetic and parasympathetic indicators of heart rate control at altitude studied by spectral analysis. *Journal of Applied Physiology*, *77*(6), 2537–2542.

Immordino-Yang, M. H., & Gotlieb, R. (2017). Embodied brains, social minds, cultural meaning: Integrating neuroscientific and educational research on social-affective development. *American Educational Research Journal*, *54*(1), 344S–367S.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*(8), 1395–1399.

Kanoga, S., Kanemura, A., & Asoh, H. (2019). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing*, *347*, 240–250.

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior*, *27*(1), 99–105.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997.

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45–77.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88.

Kraft, P., Rise, J., Sutton, S., & Røysamb, E. (2005). Perceived difficulty in the theory of planned behaviour: Perceived behavioural control or affective attitude? *British Journal of Social Psychology*, *44*(3), 479–496.

Kruger, J. L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology*, *32*(6), 19–31.

Larmuseau, C., Vanneste, P., Cornelis, J., Desmet, P., & Depaepe, F. (2019). Combining physiological data and subjective measurements to investigate cognitive load during complex learning. *Frontline Learning Research*, *7*(2), 57–74.

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*, 1058–1072.

Liu, Y., Wang, T., Wang, K., & Zhang, Y. (2021). Collaborative learning quality classification through physiological synchrony recorded by wearable biosensors. *Frontiers in Psychology*, *12*, 674369.

Lyu, Y., Luo, X., Zhou, J., Yu, C., Miao, C., Wang, T., Shi, Y., & Kameyama, K. I. (2015). Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 857–866). ACM.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*(1), 49–63.

Medeiros, J., Couceiro, R., Duarte, G., Durães, J., Castelhano, J., Duarte, C., Castelo-Branco, M., Madeira, H., de Carvalho, P., & Teixeira, C. (2021). Can EEG be adopted as a neuroscience reference for assessing software programmers' cognitive load? *Sensors*, *21*(7), 2338.

Miller, R. (2016). Neuroeducation: Integrating brain-based psychoeducation into clinical practice. *Journal of Mental Health Counseling*, *38*(2), 103–115.

Minkley, N., Xu, K. M., & Krell, M. (2021). Analyzing relationships between causal and assessment factors of cognitive load: Associations between objective and subjective measures of cognitive load, stress, interest, and self-concept. *Frontiers in Education*, *6*, 632907. https://doi.org/10.3389/feduc.2021.632907

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, *23*, 1316–1340.

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618.

Orru, G., & Longo, L. (2019). The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: A review. In *Human Mental Workload: Models And Applications: Second International Symposium, H-Workload 2018* (pp. 23–48). Springer International Publishing.

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

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*(1/2), 1–8.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. (2016). Cognitive load measurement as a means to advance cognitive load theory. In *Cognitive load theory* (pp. 63–71). Routledge.

Paas, F., Tuovinen, J. E., van Merrienboer, J. J., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*, 25–34.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paas, F. G., & van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371.
Qi, X., Wu, J., Shen, J., Ji, X., Lyu, Y., & Zhang, Y. (2020). Assessing cognitive load in adolescent and adult students using photoplethysmogram morphometrics. *Cognitive Neurodynamics*, *14*(5), 709–721.
Riddle, J., Scimeca, J. M., Cellier, D., Dhanani, S., & D'Esposito, M. (2020). Causal evidence for a role of theta and alpha oscillations in the control of working memory. *Current Biology*, *30*(9), 1748–1754.
Sarailoo, R., Latifzadeh, K., Amiri, S. H., Bosaghzadeh, A., & Ebrahimpour, R. (2022). Assessment of instantaneous cognitive load imposed by educational multimedia using electroencephalography signals. *Frontiers in Neuroscience*, *16*, 744737.
Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410–417.
Seufert, T. (2020). Building bridges between self-regulation and cognitive load—An invitation for a broad and differentiated attempt. *Educational Psychology Review*, *32*, 1151–1162.
Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*(4), 307–314.
Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J., Schuwirth, L., Artino, A. R., Jr., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. *Scientific Reports*, *9*(1), 14668.
Stafford-Brizard, K. B., Cantor, P., & Rose, L. T. (2017). Building the bridge between science and practice: Essential characteristics of a translational framework. *Mind, Brain, and Education*, *11*(4), 155–165.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71–85). Springer.
Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292.
van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26.
van Gog, T., & Paas, F. (2012). Cognitive load measurement. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 599–601). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_412
von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Javier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3–17). Lawrence Erlbaum Associates.
Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, *3*(8), e002847.
Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review*, *35*(3), 69.
Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573.
Xie, K., Heddy, B. C., & Vongkulluksn, V. W. (2019). Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology*, *59*, 101788.
Xu, J., & Zhong, B. (2018). Review on portable EEG technology in educational research. *Computers in Human Behavior*, *81*, 340–349.
Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., & Jacob, R. J. (2016). Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 5372–5384). ACM.
Zhang, J., Wang, K., & Zhang, Y. (2021). Physiological characterization of student engagement in the naturalistic classroom: A mixed-methods approach. *Mind, Brain, and Education*, *15*(4), 322–343.
Zhang, Y., Qin, F., Liu, B., Qi, X., Zhao, Y., & Zhang, D. (2018). Wearable neurophysiological recordings in middle-school classroom correlate with students' academic performance. *Frontiers in Human Neuroscience*, *12*, 457.
Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., & Zhang, D. (2021). Cognitive workload recognition using EEG signals and machine learning: A review. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(3), 799–818.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu, X., & Zhang, Y. (2024). Re-examining cognitive load measures in real-world learning: Evidence from both subjective and neurophysiological data. *British Journal of Educational Psychology*, *00*, 1–18. https://doi.org/10.1111/bjep.12729