Wang, T., Li, S., & Lajoie, S. P. (2023). The Interplay Between Cognitive Load and Self-Regulated Learning in a Technology-Rich Learning Environment. *Educational Technology & Society*, *26*(2), 50-62. https://doi.org/10.30191/ETS.202304_26(2).0004

The Interplay Between Cognitive Load and Self-Regulated Learning in a Technology-Rich Learning Environment

Tingting Wang[1*], Shan Li[2] and Susanne Lajoie[1]

[1]Department of Educational and Counselling Psychology, McGill University, Canada // [2]Lehigh University, United States // tingting.wang4@mail.mcgill.ca // shla22@lehigh.edu // susanne.lajoie@mcgill.ca

[*]Corresponding author

(Submitted January 19, 2022; Revised July 9, 2022; Accepted August 12, 2022)

ABSTRACT: Cognitive load can be induced by both learning tasks and self-regulated learning (SRL) activities, which compete for limited working memory capacity. However, there is little research on the relationship between cognitive load and SRL. This study explored how cognitive load interplayed with SRL behaviors and their joint effects on task performance (i.e., diagnostic efficiency) in the context of clinical reasoning. Specifically, twenty-seven (N = 27) medical students diagnosed three virtual patient cases in BioWorld, a simulation-based learning environment to improve medical students' clinical reasoning skills. Students' SRL behaviors were automatically recorded in BioWorld log files as they accomplished the tasks. We employed text mining techniques to extract four linguistic features from students' concurrent think-aloud, i.e., *cognitive discrepancy*, *insight*, *causatio*n, and *positive emotions*, which were further used to represent students' cognitive load. The latent profile analysis was then performed to cluster students into high- and low-load group. We also conducted a path analysis to investigate the mediation roles of SRL behaviors in the relationship between cognitive load and diagnostic efficiency (task performance). The results revealed that cognitive load negatively affected diagnostic efficiency, mediated by the ratio of SRL behaviors in the self-reflection phase. This study provides theoretical and methodological insights regarding the measurement of cognitive load and its interplay with SRL. This study informs the design of effective interventions for managing cognitive load in SRL within intelligent tutoring systems.

Keywords: Cognitive load, Self-regulated learning, Technology-rich learning environment, Text mining

1. Introduction

Technology-rich learning environments (TREs), such as multimedia, simulation, virtual reality, and intelligent tutoring systems, have been increasingly employed to foster medical students' clinical reasoning skills (Azevedo & Gašević, 2019). Clinical reasoning refers to a complex reasoning and decision-making process whereby health professionals get familiar with patient information, collect evidence, propose hypotheses, evaluate gathered evidence, and make final diagnostic decisions (Kuiper, 2013; Simmons, 2010). Due to the crucial role of clinical reasoning on patients' health, medical students need to plan, monitor, and control their problem-solving processes to achieve an accurate diagnosis, which is also referred to as self-regulated learning (SRL) (Artino et al., 2014; Brydges & Bulter, 2012; Cleary et al., 2016). SRL is a recursive process by which learners monitor and control their motivational, behavioral, emotional, and cognitive aspects to realize pre-determined goals (Greene & Azevedo, 2007; Panadero, 2017; Pintrich, 2000; Winne & Perry, 2000; Zimmerman, 2000). Researchers have attempted to examine medical students' clinical reasoning process from the perspective of SRL and demonstrated that strategical planning and reflective journal writing significantly promoted their performance in clinical reasoning tasks (Artino et al., 2014; Kuiper et al., 2009).

In addition to SRL, cognitive load is also an explanatory theoretical lens to understand clinical reasoning outcomes (Solhjoo et al., 2019). Cognitive load refers to the amount of working memory capacity (WMC) occupied by solving a specific task (Paas et al., 2003). Decades of research have shown that mental overload induces more negative emotions and leads to poorer academic performance across disciplines (Leutner et al., 2009; Scheiter et al., 2020). Given the complexity of clinical reasoning, including the intricate patient information, uncertainty about the diagnostic decisions, and the detrimental consequences of medical errors, medical students are likely to experience a high cognitive load during the diagnostic process (Durning et al., 2011; Solhjoo et al., 2019). As for the effects of cognitive load on clinical reasoning performance, empirical studies exhibited mixed results. For instance, Solhjoo et al. (2019) demonstrated a negative association between self-reported cognitive load and diagnostic performance, whereas Durning et al. (2011) and Fraser et al. (2012) indicated a positive relationship.

However, the research on cognitive load and SRL is conducted separately, and few studies have investigated the interplay between cognitive load and SRL. Therefore, the current research examines the interplay between cognitive load and SRL and their joint roles in explaining clinical reasoning performance. Specifically, we situated this study in the BioWorld system (Lajoie, 2009), a technology-rich learning environment which simulates virtual patient cases for medical students to improve their clinical reasoning skills. BioWorld keeps track of students' operations in log files, which are necessary to analyze fine-granular SRL behaviors. The following section provides the theoretical foundation and research questions.

2. Theoretical framework

2.1. Cognitive load theory

Cognitive load theory (CLT) is based on a cognitive architecture consisting of working memory and long-term memory (Paas et al., 2003; Sweller, 2011). The long-term memory system provides an infinite capacity to store acquired knowledge in cognitive schemas, a complex unit of interrelated information elements (Bower et al., 1975). Working memory temporarily stores and manipulates novel information (Baddeley, 1992), working as a conduit between external environments and the long-term memory system (Kirschner, 2002). Compared with long-term memory, the working memory system is limited in capacity and duration when dealing with new information. The finite WMC is necessary for mental tasks such as language comprehension, problem-solving, and planning (Cowan, 2011; Wiley & Jarosz, 2012).

Cognitive load refers to the load that performing a specific task exerts on the working memory (Paas et al., 1994; Sweller, 2011). CLT distinguishes three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic load reflects the "necessary load" determined by task complexity and expertise levels. Specifically, the intrinsic load increases with the element interactivity of tasks and decreases with learners' expertise levels (Paas et al., 2010; Park et al., 2015; Sweller, 1988; Sweller, 2011). Charlin et al. (2007) demonstrated that medical experts tended to experience a lower intrinsic load during diagnostic tasks than novices by applying knowledge constructed from prior experiences. However, the extraneous load is ineffective since it is triggered by inappropriate instructional designs and display modes (Paas et al., 1994). For instance, Reedy (2015) indicated that suboptimal designs of simulation-based learning environments, such as redundant information and inappropriate presentation format, were associated with increased extraneous load in clinical reasoning. Lastly, germane load is induced by schema construction and automation and represents a kind of effective load that directly contributes to learning (Sweller, 1988). Moreover, intrinsic, extraneous, and germane load are additive, and the sum is referred to as the overall load (Paas et al., 2003). This study is particularly interested in the overall cognitive load students experienced in the clinical reasoning process.

2.2. Linguistic features of cognitive load

There are a variety of methods to measure cognitive load in TREs. Subjective questionnaires have been intensively used to measure cognitive load (Leppink et al., 2013; Paas, 1992). Physiological techniques, including eye-tracking (Joseph & Murugesh, 2020), electroencephalograph (EEG) (Antonenko et al., 2010), and heart rate variability (HRV) (Solhjoo et al., 2019), can objectively trace cognitive load changes. For the performance-based measures, study time, accuracy, and error rate are frequently used to represent students' cognitive load (Brünken et al., 2003; Paas et al., 2003; Sweller, 1988).

Words and language use can also reflect individuals' psychological processes (Darabi et al., 2010), and the advancement of text mining techniques makes linguistic features a promising measurement of cognitive load. In particular, this study employed the text mining program Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015), an automatic tool to count word frequency to extract participants' linguistic features. Linguistic features are more reliable indicators of cognitive load than other techniques since language is the most common and direct way to reflect internal mental states (Pennebaker et al., 2003; Pennebaker et al., 2015). The commonly used linguistic features include speech pause (Khawaja et al., 2014; Müller et al., 2001), use of first/third-person plurals (Sexton & Helmreich, 2000), patterns of personal pronoun (Khawaja et al., 2012), the use of different word categories, repetitive words, and grammatical structures. Researchers have applied linguistic features to infer students' cognitive load in learning contexts. For instance, Konopasky et al. (2020) found that participants used fewer first-person pronouns in high-load contexts because high cognitive load would induce more cognitive processes and thus less attention on the self. Khawaja et al. (2014) revealed that students were more likely to use cognitive words (e.g., think, know, and consider) when they experienced a high cognitive load. Positive emotions words were also confirmed as a negative index of experienced cognitive load (Fraser et al., 2012).

2.3. Self-regulated learning in the context of clinical reasoning

According to Zimmerman (2000), SRL refers to self-generated thoughts, feelings, and behaviors that are monitored and adjusted in three loosely sequential phases to attain learning goals. During the forethought phase, learners prepare their motivation and self-belief, conduct task interpretation, and set goals and strategic plans. In the performance phase, learners take appropriate learning strategies to execute the task and monitor task progress, and they evaluate and reflect on performance during the self-reflection phase.

As aforementioned, the clinical reasoning task requires students to collect patient information, order lab tests, integrate and evaluate evidence, propose and reflect on diagnostic hypotheses, and make final decisions, which process has substantial overlaps with SRL (Artino et al., 2014; Kuiper, 2009; Kuiper, 2013; Simmons, 2010). From the perspective of SRL, medical students review patient information to familiarize themselves with task conditions and plan strategies and efforts to solve the task during the forethought phase (Li et al., 2020). Throughout the performance phase, medical students conduct lab tests, search for external resources, and integrate all gathered evidence to propose diagnostic hypotheses (Artino et al., 2014). During the last self-reflection phase, medical students evaluate and reflect on their diagnostic judgments to see if any additional actions are required (Brydges & Butler, 2012). The emerging literature within medical education showed that medical students did not often exhibit strategic thinking and self-evaluative judgments during clinical reasoning (Cleary et al., 2016). Moreover, several empirical studies have demonstrated that insufficient SRL, such as poor planning, deficits in self-monitoring, and scant self-reflection, led to poor clinical reasoning performance (Artino et al., 2014; Brydges & Bulter, 2012).

2.4. Self-regulated learning and cognitive load

SRL can also be regarded as a set of information processes from the lens of information processing (Panadero, 2017; Winne, 2001; Winne, 2018). That is, learners process task-specific information (e.g., patient symptoms) and psychological information (motivational and emotional states) in the forethought phase and apply tactics, strategies, and schemas to solve the problem during the performance phase (Winne, 2001). Finally, all the information collected in the previous two phases should be integrated and evaluated to reach a final decision in the self-reflection phase.

More importantly, information processes in different SRL phases also demand working memory resources and generate additional cognitive load beyond problem-solving itself (Seufert, 2018). In this regard, SRL behaviors and cognitive load are two contrary forces to competing for the limited WMC, and SRL can be regarded as a function of WMC and cognitive load (de Bruin et al., 2020; Seufert, 2018). As illustrated by Figure 1, adapted from Seufert (2018, 2020), students might be unable to self-regulate their learning when the cognitive load triggered by a task is high since little WMC remains for SRL. However, students can easily achieve learning goals when they experience a low-level cognitive load, even without efficient SRL processes. In other words, conducting SRL activities in less cognitive-demanding contexts is unnecessary. Only when the cognitive load is moderate will the resources sufficient for efficient SRL.

Despite the potential effects of cognitive load on SRL activities from the theoretical perspective (see Figure 1), few scholars provided empirical evidence to support these assumptions, especially in the context of clinical reasoning. The majority of studies explored the separate role of cognitive load and SRL in diagnostic performance (Artino et al., 2014; Durning et al., 2011; Kuiper, 2009; Solhjoo et al., 2019); however, the interactions between cognitive load and SRL and their joint effects on learning have been underexplored (de Bruin et al., 2020; Scheiter et al., 2020; Seufert, 2018; Seufert, 2020). Investigating the associations between cognitive load and SRL can facilitate medical students to achieve an accurate and efficient diagnosis, guaranteeing patients' safety. Therefore, this study addresses this issue by answering the following research questions: (1) Did medical students demonstrate different cognitive load patterns when diagnosing virtual patient cases in TREs? (2) Did cognitive load patterns affect SRL behaviors? and (3) How did cognitive load and SRL behaviors jointly predict diagnostic efficiency?

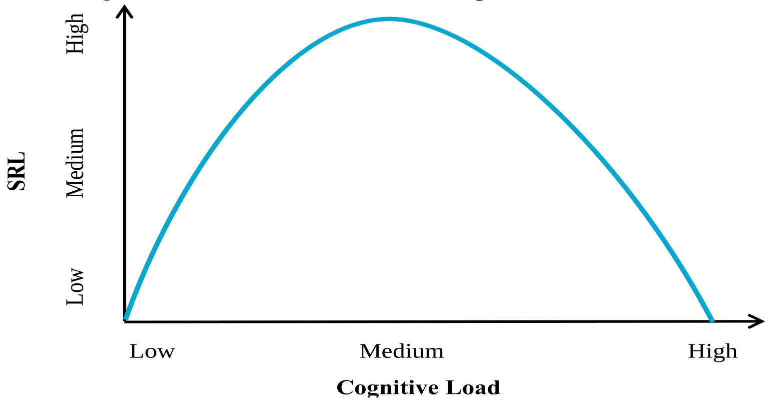


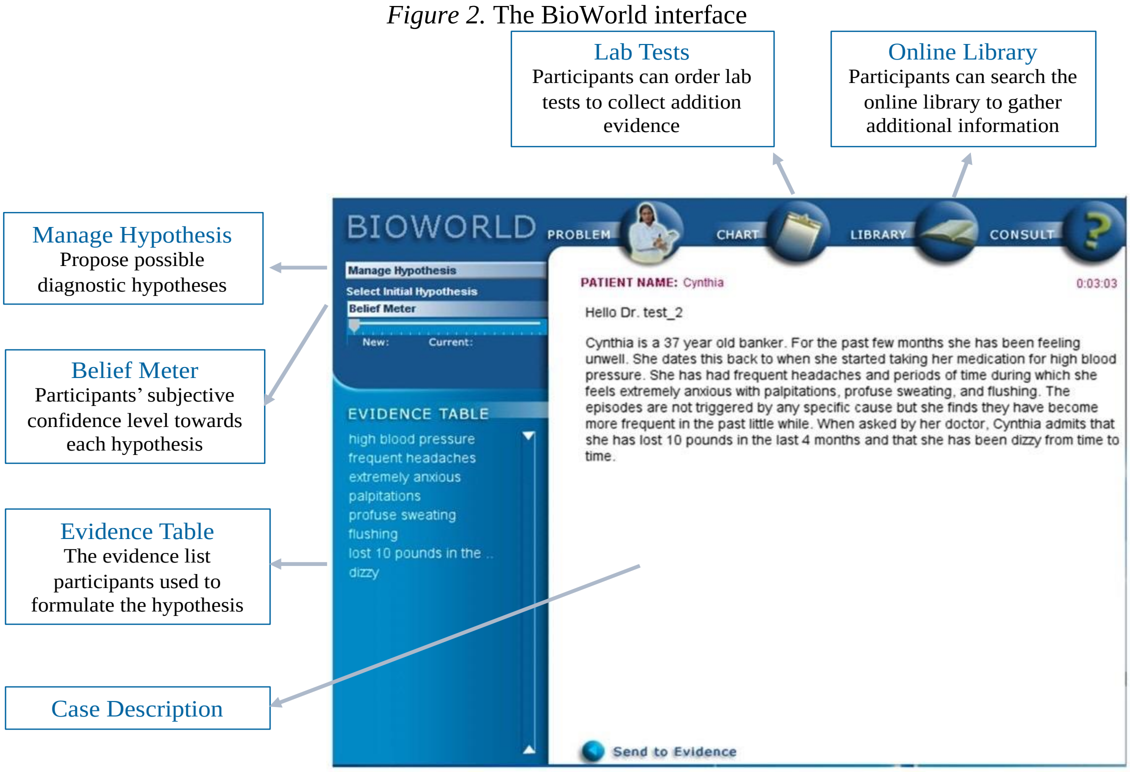
Figure 1. The interaction between cognitive load and SRL

3. Method

3.1. Participants, procedures, and learning context

Twenty-seven (*N* = 27) participants were recruited from a large North American university. They consisted of 14 first- and 13 second-year medical students. Among them, seventeen (63%) were male students, and ten (37%) were female students, with a mean age of 23 (*SD* = 2.66). All students had completed a prerequisite course module on endocrinology, metabolism, and nutrition. Therefore, they mastered the necessary knowledge to complete the tasks.

Figure 2. The BioWorld interface



Participants were first provided with a sample case to familiarize themselves with the BioWorld system (Lajoie, 2009), a simulation-based learning environment developed to help medical students practice clinical reasoning skills. During the formal experiments, each student was tasked to solve three virtual patient cases in BioWorld. The correct diagnoses for the three tasks are *Diabetes (Type1), Hyperthyroid (Grave's disease), and Pheochromocytoma*, respectively. Therefore, there were a total of 81 (27 x 3) different cases in this study. Participants were also instructed to concurrently speak out whatever comes to their minds without modifying their feelings and thoughts. The think-aloud protocols were audio-recorded and transcribed afterward for further analysis. Each of the three cases lasted about 40-50 minutes, and the whole process took approximately 2-2.5 hours.

As shown in Figure 2, students initiated each diagnostic task by reading patient history and extracting critical symptoms. To collect additional information and evidence, students ordered appropriate lab tests and searched the online library. Students could propose one or more clinical hypotheses and manage hypotheses throughout the processes. They also linked gathered evidence with specific hypotheses and classified it into three categories, i.e., useful, neutral, and useless. In the end, students were also required to prioritize their hypotheses based on their subjective judgments. After submitting the final hypothesis, the BioWorld system provided individualized performance feedback for students.

3.2. Measures

3.2.1. Linguistic indicators of cognitive load

Cognitive load was inferred from transcribed think-aloud using the LIWC (Pennebaker et al., 2015). Based on an embedded dictionary that defines categories of word lists, LIWC automatically calculates the percentage of specific function words and provides a practical approach to detecting individuals' psychological processes from their linguistic patterns (Robinson et al., 2012). The current study extracted four markers, i.e., *positive emotion*, *cognitive discrepancy*, *insight*, and *causation*, from the LIWC output variables to represent cognitive load for each case. *Positive emotion* was a negative indicator of cognitive load and was calculated based on the percentage of words indicating pleasant perceptions. *Cognitive discrepancy* refers to the students' perceived inconsistency between prior knowledge and current task contexts. For instance, the words *should* and *would* were counted as cognitive discrepancies. *Cognitive insight* and *causation* indicated the intensity of cognitive efforts in solving the tasks. The three variables in the cognitive category were regarded as positive indicators of cognitive load. In sum, we used the four cognitive load indicators to infer students' cognitive load levels by the following latent profile analysis and calculated an individual's cognitive load value by the equation: *Cognitive load = Discrepancy + Insight + Causation – Positive Emotion*.

3.2.2. SRL behaviors

A total of 22 different operations were recorded in the BioWorld log files. Among them, ten operations were extracted as SRL behaviors (Table 1), and the remaining activities, such as *switch area* and *unlink evidence*, were excluded from our analysis. As shown in Table 1, this study conducted both a macro- and micro-level analysis of SRL. The forethought phase consisted of one SRL behavior, i.e., *Task Analysis*, whereby students review and interpret patient information. The performance phase of SRL included four behaviors: *Orientation*, *Execution*, *Help-Seeking*, and *Link Evidence*. Readers can find a detailed description of these behaviors in Table 1. In the self-reflection phase, students evaluated and reflected on the problem-solving process by the Evidence Evaluation and Hypothesis Evaluation behaviors. Moreover, we calculated the relative ratio of each SRL behavior to represent SRL behavior frequencies.

Table 1. SRL behaviors extracted from log files

Macro-level	Micro-level	Description	Sample
Forethought	Task Analysis	Collect patient symptoms to construct an overall view of the problem	Add evidence
Performance	Orientation	Propose or mange hypotheses based on the collected information and prior knowledge	Add hypothesis
	Execution	Order lab tests to collect additional evidence	Add tests
	Help-Seeking	Seek for external help from online library embedded in BioWorld	Search library Search library category
	Link Evidence	Link evidence with specific hypotheses to claim progress	Link evidence
Self-Reflection	Evidence Evaluation	Evaluate and classify the evidence into supportive group and against group	(Re)categorize
	Hypothesis Evaluation	Evaluate possibilities of each hypothesis	(Re)prioritize

Note. The coding scheme was adapted from Lajoie and Lu (2012).

3.2.3. Diagnostic efficiency

Diagnostic efficiency was automatically measured by the BioWorld system. Diagnostic efficiency refers to the matching degree between medical students' clinical reasoning process and experts' steps to reach the diagnosis. For instance, students would obtain a score of 50 if 50% of their clinical reasoning steps were matched with experts' solution steps (that were embedded in the BioWorld system). Diagnostic efficiency was designed to range from 0 to 100.

3.3. Data analysis

This study employed the latent profile analysis (LPA) to model different cognitive load patterns, *t*-tests to examine the effects of cognitive load patterns on SRL behaviors, and path analysis to explore the joint effects of cognitive load and SRL behaviors on diagnostic efficiency. The following session described how we used LPA and path analysis to address our research questions.

3.3.1. Latent profile analysis

Latent profile analysis, a person-centered mixture modelling method, can detect homogeneous clusters from observed variables through a probabilistic framework. In contrast to traditional cluster analytical techniques, LPA is model-based, whereas the hierarchical and *K*-means clustering methods are not (McLachlan et al., 2019; Pastor et al., 2007). Remarkably, this study had 81 different cases, and we treated each case as a sample due to the small sample size. To reach a stable solution, we constrained the variances of cluster indicators to be equal, but the means can vary across clusters (Scherer et al., 2017).

Using the "tidyLPA" packages in R (Rosenberg et al., 2019), we applied the maximum likelihood (ML) algorithm to estimate the model parameters and generate fit statistics for six candidate models with k values ranging from 1 to 6. Because the number of clusters k is unknown priori in LPA (Nylund et al., 2007), multiple alternative models with different k values should be compared. There are several well-established model fit indices to determine the goodness-of-fit of specific models. First, Akaike Information Criteria (*AIC*), Bayesian Information Criteria (*BIC*), and Sample-Size-Adjusted Bayesian Information Criteria (*SSA-BIC*) were utilized in LPA to decide the number of clusters. The lower the values of these three indices, the better the model fit (Schwarz, 1978). Second, the Bootstrapped Likelihood Ratio Test (*BLRT*) revealed whether a model with k clusters was significantly better than $k-1$ clusters (Lo et al., 2001). The significant result (i.e., $p < .05$) of BLRT implied that adding a cluster increased the model fit. Third, the estimate of classification certainty was also essential in LPA, and Entropy values > .70 indicated an acceptable accuracy (Celeux & Soromenho, 1996). In addition, the appropriate size of each cluster should be no less than 5% of the sample, which guarantees latent profiles to be theoretically significant and generalized (Pastor et al., 2007).

3.3.2. Path analysis

A path analysis was performed to examine the mediating role of cognitive load in SRL behaviors and diagnostic efficiency, using PROCESS Macro in SPSS (Hayes, 2012). The PROCESS Macro can automatically execute computation, run the analysis, and generate meaningful mediation output (Uchechukwu Onu et al., 2020). Specifically, this study employed the "Model 4" embedded in PROCESS Macro, which allows researchers to examine the significance of parallel mediators. In this study, Cognitive load served as the independent variable, the Performance-phase behavior Ratio (PR) and the Self-reflection-phase behavior Ratio (SR) were entered as two parallel mediator variables, and diagnostic efficiency was the dependent variable. The Forethought-phase behavior Ratio (FR) was excluded to avoid multicollinearity. To increase inference accuracy, bootstrapping with 10000 bias-corrected bootstrap samples was conducted to depict the sampling distribution of direct and indirect effects.

4. Results

In this section, we first employed the LPA to classify medical students as high- and low-load clusters. Then we compared the differences in macro- and micro-level SRL behaviors between the high- and low-load cluster to examine the effects of cognitive load on SRL. To further investigate the joint predictive functions of cognitive load and SRL, we performed the path analysis.

4.1. Did students demonstrate different cognitive load patterns when they diagnosed virtual patient cases in a technology-rich learning environment?

As aforementioned, we utilized the LPA to identify medical students' profiles of cognitive load based on four linguistic markers extracted by the LIWC, i.e., *positive emotion*, *cognitive discrepancy*, *insight*, and *causation*. The model fit indices for LPA with clusters ranging from 1 to 6 are shown in Table 2. The 2-cluster solution was deemed superior to the one-cluster solution due to the lower *AIC*, *BIC*, and *SSA-BIC* values and the significant result of *BLRT* (p = .01). However, the *BLRT* demonstrated that the 3-cluster solution did not have a significantly better fit than the 2-cluster solution (p = .70). The 4-cluster solution fitted better than a 3-cluster solution with decreased *AIC* and *SSA-BIC* values and significant *BLRT* results (p = .02). In terms of the 5-cluster and 6-cluster solutions, the nonsignificant results of *BLRT* revealed that they did not fit better than a 4-cluster solution (p = .62 and .31, respectively). As for the comparison between the 2-cluster solution and 4-cluster solution, we deemed the former was a better fit than the latter, considering the higher Entropy of a 2-cluster solution (*Entropy* = .81) than that of the 4-cluster solution (*Entropy* = .74). Overall, the 2-cluster solution is optimal for differentiating medical students' cognitive load profiles, with 21% and 79% of the students being labelled as high- and low-load cluster.

As shown in Figure 3, the two clusters represent distinct cognitive load patterns. Cluster 1 consisted of 17 (21%) cases in which students experienced more cognitive discrepancy (M = 3.25), insight (M = 3.27), and causation processes (M = 2.09), but less positive emotions (M = 3.10). In contrast, the 64 cases (79%) in Cluster 2 demonstrated less cognitive discrepancy (M = 1.68), insight (M = 2.93), and cognitive causations (M = 1.69) bur more positive emotions (M = 2.97). Therefore, we labelled Cluster 1 and Cluster 2 as high- and low-load group, respectively.

Table 2. Fit indices for models with number of clusters ranging from 1 to 6

Models	AIC	BIC	SSA-BIC	Entropy	BLRT_p	n_min
1 Cluster	857	876	851	1.00		1.00
2 Clusters	840	871	830	.81	.01	.21
3 Clusters	845	888	831	.63	.70	.11
4 Clusters	835	890	817	.74	.02	.14
5 Clusters	838	905	817	.74	.62	.14
6 Clusters	840	919	815	.76	.31	.04

Note. BLPT_*p* refers to the *p* values for the Bootstrapped Likelihood Ratio Test, *n_min* refers to the ratio of respondents in clusters with the smallest sample size.

Figure 3. Two levels of cognitive load identified by LPA

4.2. Did cognitive load patterns affect SRL behaviors?

A series of independent t-tests were conducted to examine how high- and low-load groups differed in macro-level (i.e., forethought, performance, and self-reflection) and micro-level SRL behaviors. Table 3 showed that the ratio of SRL behaviors in the forethought phase (FR) did not significantly differ between high- and low-load cases. The ratio of SRL behaviors in the performance phase (PR) was significantly higher in high-load cases (M = 52.53, SD = 14.99) than that in the low-load cases (M = 41.23, SD = 12.92), $t(79) = 3.10$, $p = .003$. However, high-load cases (M = 32.06, SD = 15.14) led to a significantly lower ratio of SRL behaviors in the self-reflection phase than the low-load cases (M = 41.69, SD = 32.06), $t(79) = -2.54$, $p = .013$. In addition, the effect sizes of cognitive load levels on PR and SR were large (Cohen's d = .85 and -.69, respectively).

The results of the micro-level analysis are illustrated in Table 4. Students in the high-load case showed significant differences from those in the low-load cases in two micro-level SRL behaviors: Execution (performance phase) and Hypothesis Evaluation (self-reflection phase). The ratio of Execution behavior was higher for the high-load cluster (M = 14.41, SD = 7.13) compared to the low-load cluster (M = 10.48, SD = 5.83), t = 2.35, p = .21, Cohen's d = .59. In contrast, the ratio of Hypothesis Evaluation behavior was significantly higher in low-load cases (M = 24.80, SD = 14.14) than that in high-load cases (M = 16.53, SD = 12.95), t = -2.15, p = .035, Cohen's d = -.59.

Table 3. The predicative role of cognitive load level on macro-level SRL behavior ratios

	High-load cases (n = 17)		Low-load cases (n = 64)		t	*Sig.*	Cohen's d
	M	SD	M	SD			
FR	15.29	3.74	17.14	4.88	-1.45	.152	-.40
PR	52.53	14.99	41.23	12.92	3.10	.003**	.85
SR	32.06	15.14	41.69	13.56	-2.54	.013*	-.69

Note. FR = Forethought-phase behavior Ratio, PR = Performance-phase behavior Ratio, SR = Self-reflection-phase behavior Ratio. $^{*}p < .05$; $^{**}p < .01$, $^{***}p < .001$.

Table 4. The predicative role of cognitive load level on micro-level SRL behavior ratios

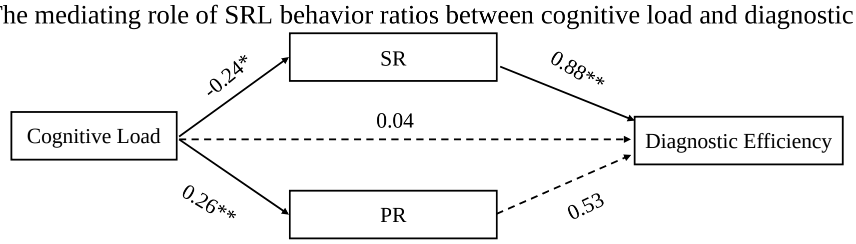
	High-load cases (n = 17)		Low-load cases (n = 64)		t	*Sig.*	Cohen's d
	M	SD	M	SD			
TAR	15.29	3.74	17.14	4.88	-1.45	.152	-.40
ORR	16.88	7.11	16.26	5.66	.37	.714	.10
EXR	14.41	7.13	10.48	5.83	2.35	.021*	.64
HSR	5.88	6.70	4.34	5.50	.98	.331	.23
LER	15.35	10.46	10.28	9.81	1.87	.065	.52
EER	15.47	3.81	16.91	5.46	-1.02	.312	-.28
HER	16.53	12.95	24.80	14.14	-2.15	.035*	-.59

Note. TAR = Task Analysis Ratio, ORR = Orientation Ratio, EXR = Execution Ratio, HSR = Help-Seeking Ratio, LER = Link Evidence Ratio, EER = Evidence Evaluation Ratio, HER = Hypothesis Evaluation Ratio. $^{*}p < .05$; $^{**}p < .01$, $^{***}p < .001$.

4.3. How did cognitive load and SRL behaviors jointly predict diagnostic efficiency?

As aforementioned, path analysis was conducted to test the hypothesis that SRL behavior ratios mediated the relationships between cognitive load and diagnostic efficiency. This study did not include the Forethought-phase behavior Ratio (FR) in the model to avoid the issue of collinearity. The results in Table 5 showed that the direct effect of cognitive load on diagnostic efficiency was not significant (R^2 = .25, $F(3, 77)$ = 1.42, p > .05). Cognitive load positively predicted the Performance-phase behavior Ratio (PR) (β = .26, p < .05), but it was a negative predictor of the Self-reflection-phase behavior Ratio (SR) (β = -.24, p < .05). In addition, SRL behaviors in the self-reflection phase positively predicted diagnostic efficiency (β = .88, p < .01), whereas SRL behaviors in the performance phase had no significant predictive effect on diagnostic efficiency (β = .53, p = .11). Although the direct effect of cognitive load on diagnostic efficiency was not significant, the indirect effect through SRL was significant (β = -.21, 95% CI [-.46, -.01]). Overall, the effect of cognitive load on diagnostic efficiency was exclusively mediated by SR. The direct and indirect paths are shown in Figure 4.

Figure 4. The mediating role of SRL behavior ratios between cognitive load and diagnostic efficiency



Note. PR = Performance-phase behavior Ratio, SR = Self-reflection-phase behavior Ratio. The dotted lines represent insignificant effects. $^{*}p < .05$; $^{**}p < .01$, $^{***}p < .001$.

Table 5. Mediation model

Model	β	*SE*	*t*	*p*	*CI* (lower)	*CI* (upper)
$X \rightarrow M_1$ *(a1)*	.26	.66	2.27	.026*	.18	2.80
$X \rightarrow M_2$ *(a2)*	-.24	.68	-2.00	.049*	-2.71	-.01
$M_1 \rightarrow Y$ *(b1)*	.53	.53	1.62	.109	-.20	1.91
$M_2 \rightarrow Y$ *(b2)*	.88	.51	2.73	.008**	.38	2.43
$X \rightarrow Y$ *(c)*	-.03	1.08	-.24	.808	-2.42	1.89
$X \rightarrow Y$ *(c')*	.04	1.03	.35	.725	-1.68	2.41
$X \rightarrow M_1 \rightarrow Y$ *(a1*b1)*	.14	.10			-.02	.34
$X \rightarrow M_2 \rightarrow Y$ *(a2*b2)*	-.21	.12			-.46	-.01

Note. X = cognitive load, M_1 = Performance-phase behavior Ratio, M_2= Self-Reflection-phase behavior Ratio, Y= diagnostic efficiency; *a1* and *a2* represent the direct effect of X on M_1 and M_2, respectively; *c* means the total effect of X on Y; *c'* refers to the direct effect of X on Y. $^{*}p < .05$; $^{**}p < .01$.

5. Discussion

We applied text mining techniques to extract cognitive load indicators from students' concurrent think-aloud protocols as they diagnosed virtual patient cases in BioWorld. Specifically, four indicators, including *cognitive discrepancy*, *insight*, *causation*, and *positive emotions*, were selected in this study. We then employed the LPA on cognitive load indicators to cluster medical students to see if any patterns of cognitive load emerged in solving the tasks. The results from LPA demonstrated that medical students could be identified as high- and low-load group when addressing the tasks. Notably, medical students with a high cognitive load experienced fewer positive emotions and exerted more cognitive effort (i.e., more cognitive discrepancies, insight, and causation activities) compared to those who experienced a low-level cognitive load.

Specifically, students may encounter large knowledge gaps in high cognitive load situations and find more inconsistencies between their acquired knowledge and the ongoing learning task (Reiser, 2004). Therefore, they tended to use words such as *should* and *would*, to express their *cognitive discrepancies*. As well, a high-level cognitive load required students to conduct more cognitive operations, such as thinking, evaluation, and analyses, to achieve predetermined learning goals (Baddeley, 1992; Khawaja et al., 2014; Sweller, 2011). In this study, medical students demonstrated more *insights* (words such as *think* and *know*) and performed more *causal inferences* (words such as *because* and *so*) in high-load contexts. Consistent with Fraser and McLaughlin (2019), we also found that the proportion of positive emotion words decreased with increased cognitive load. According to Pekrun's (2006) control-value theory, individuals' appraisal of perceived controllability over the diagnostic task would be weakened by the high cognitive load level; thus, students were expected to generate more negative emotions in cognitive-demanding contexts. Noticeably, linguistic features developed by the LIWC indicate the overall load instead of distinguishing between three types of cognitive load. For instance, linguistic features cannot differentiate the intrinsic load caused by task complexity from the extraneous load induced by the interface and presentations of the BioWorld. However, the interactions between multidimensional load and SRL matter in learning (Seufert, 2020) and are worthy of further investigation.

The interplay between cognitive load, SRL, and diagnostic efficiency is of primary interest to this study. We found that students with a higher cognitive load had a significantly higher ratio of SRL behaviors in the performance phase but a significantly lower ratio of SRL behaviors in the self-reflection phase. Cognitive load did not affect the ratio of SRL behaviors in the forethought phase. As aforementioned, the high-level cognitive load led students to experience more cognitive discrepancies and uncertainties; thus, it was essential for them to try more operations in the performance phase to collect additional evidence for diagnoses. In this regard, students' mental efforts in the performance phase would occupy a vast of limited working memory resources, suggesting that few cognitive capacities remained for self-reflection behaviors (Seufert, 2018; Sweller, 2011; Winne, 2001). Therefore, students in the high-load cluster had a lower ratio of SRL behaviors in the self-reflection phase compared to the low-load cluster.

Furthermore, this study found that the micro-level SRL behaviors in the performance and self-reflection phases were affected by cognitive load levels. Specifically, students conducted relatively more *Execution* behaviors and fewer *Hypothesis Evaluation* behavior when experiencing a higher cognitive load. Students with a high cognitive load might activate all relevant cognitive schemas stored in the long-term memory; thus, they were inclined to conduct more lab tests (Execution) to reduce the feeling of uncertainty and to ease the cognitive load. The SRL behavior of *Hypothesis Evaluation* required students to integrate all information obtained from the forethought and performance phases to make a judgment about the proposed hypotheses. This SRL behavior imposed enormous cognitive burdens on the working memory system (de Bruin & van Merriënboer, 2017). However, a high cognitive load indicates limited working memory resources, which prevent students from performing many *Hypothesis Evaluation* behaviors.

As for the joint effects of SRL behaviors and cognitive load on diagnostic efficiency, the path analysis revealed that cognitive load negatively predicted diagnostic efficiency by influencing the ratio of SRL behaviors in the self-reflection phase. Specifically, cognitive load positively predicted students' efforts in performance-phase operations but negatively predicted the ratio of SRL behaviors in the self-reflection phase due to the limited cognitive capacity. However, the ratio of SRL behaviors in the self-reflection phase was a positive indicator of diagnostic efficiency. The self-reflection behaviors facilitated students to construct a more comprehensive and deeper understanding of the task and led them to elaborate on their problem-solving processes (Lew & Schmidt, 2011). Thus, SRL behaviors in the self-reflection phase were beneficial to students' diagnostic efficiency, which was in line with the findings of Zheng et al. (2020). Moreover, we found that SRL behaviors in the self-reflection phase completely mediated the relationship between cognitive load and diagnostic efficiency. These results suggested that effective allocation strategies of working memory resources matter to diagnostic efficiency.

6. Conclusion

In conclusion, this paper is the first to explore how cognitive load interacts with SRL behaviors and their joint roles in predicting diagnostic efficiency in the context of clinical reasoning. Theoretically, findings from this study provide empirical evidence for integrating cognitive load and SRL frameworks. Furthermore, this study made a methodological contribution to the measurement of cognitive load by applying text mining techniques to extract cognitive load indicators from students' think-aloud protocols. Moreover, this study has educational implications in that it provides educators with insights regarding how to facilitate students' self-regulated learning and academic performance when they experience a high cognitive load. For example, educators should pay particular attention to students' self-reflection behaviors. When students experience a high-level cognitive load, they may conduct fewer self-reflection behaviors (a positive performance indicator). Under this condition, instructors can provide metacognitive scaffoldings to foster students' metacognitive awareness and self-reflective activities. Moreover, this study informs TREs developers and instructors to design optimal instructional activities to minimize the extraneous load. Otherwise, the high extraneous load would occupy limited cognitive resources and limit SRL behaviors.

While the present study has theoretical, methodological, and practical significance, it is not without limitations. First, this study has a small sample size. Therefore, additional research is needed to verify the findings of this study with a larger number of participants. Second, we did not differentiate between the cognitive load caused by problem-solving tasks and SRL activities due to the limitation of linguistic features. Given that individuals only have limited WMC, it is vital to balance the cognitive load caused by problem-solving and SRL activities to avoid mental overload. Further research should employ various techniques to differentiate these two cognitive load sources. Lastly, the measurement of cognitive load relies entirely on the linguistic features of students, and the validity and reliability of this approach have not yet been verified in the literature. Scholars should combine linguistic features and other cognitive load measures, such as self-ratings, to address this issue in future research.

There are several areas that we will pursue as future research directions. First, we will use multimodal data, such as physiological signals and self-rating, to measure students' cognitive load. A second research direction is to examine the temporal interplay between cognitive load and SRL during the dynamic learning process. Third, this study emphasized the overall load students experienced in a task, and it did not explore the cognitive load in more fine-grained SRL behaviors, such as during planning, monitoring, and evaluation behaviors. The investigation of these research topics will facilitate an integrative framework incorporating cognitive load and SRL theories and guide educators to design more effective instruction activities.

Acknowledgement

This research was made possible by the Social Sciences and Humanities Research Council of Canada (SSHRC) and the China Scholarship Council (CSC). This study was approved by the Research Ethics Board of McGill University. The authors declare that they have no competing interests.

References

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

Artino Jr, A. R., Cleary, T. J., Dong, T., Hemmer, P. A., & Durning, S. J. (2014). Exploring clinical reasoning in novices: A Self-regulated learning microanalytic assessment approach. *Medical Education, 48*(3), 280-291. https://doi.org/10.1111/medu.12303

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition, 3*(2), 216–220. https://doi.org/10.3758/BF03212900

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

Brydges, R., & Butler, D. (2012). A Reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education, 46*(1), 71-79. https://doi.org/10.1111/j.1365-2923.2011.04100.x

Celeux, G., & Soromenho, G. (1996). An Entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212. https://doi.org/10.1007/BF01246098

Charlin, B., Boshuizen, H. P., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education, 41*(12), 1178-1184. https://doi.org/10.1111/j.1365-2923.2007.02924.x

Cleary, T. J., Durning, S. J., & Artino, A. R. (2016). Microanalytic assessment of self-regulated learning during clinical reasoning tasks: Recent developments and next steps. *Academic Medicine, 91*(11), 1516-1521.

Cowan, N. (2011). The Focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia, 49*(6), 1401–1406. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2011.01.035

Darabi, A., Arrastia, M. C., Nelson, D. W., Cornille, T., & Liang, X. (2010). Cognitive presence in asynchronous online learning: A Comparison of four discussion strategies. *Journal of Computer Assisted Learning, 27*(3), 216–227. https://doi.org/10.1111/j.1365-2729.2010.00392.x

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*(4), 903-915. https://doi.org/10.1007/s10648-020-09576-4

de Bruin, A. B., & van Merriënboer, J. J. (2017). Bridging cognitive load and self-regulated learning research: A Complementary approach to contemporary issues in educational research. *Learning and Instruction, 51*, 1–9.

Durning, S., Artino Jr, A. R., Pangaro, L., van der Vleuten, C. P., & Schuwirth, L. (2011). Context and clinical reasoning: understanding the perspective of the expert's voice. *Medical Education, 45*(9), 927-938. https://doi.org/10.1111/j.1365-2923.2011.04053.x

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & Mclaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education, 46*(11), 1055–1062. https://doi.org/10.1111/j.1365-2923.2012.04355.x

Fraser, K., & McLaughlin, K. (2019). Temporal pattern of emotions and cognitive load during simulation training and debriefing. *Medical Teacher, 41*(2), 184-189. https://doi.org/10.1080/0142159X.2018.1459531

Greene, J. A., & Azevedo, R. (2007). A Theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334–372. https://doi.org/10.3102/003465430303953

Hayes, A. F. (2012). *PROCESS: A Versatile computational tool for observed variable moderation, mediation, and conditional process modeling* [White paper]. http://www.afhayes.com/public/process2012.pdf

Joseph, A. W., & Murugesh, R. (2020). Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. *Journal of Scientific Research, 64*(01), 168–175. https://doi.org/10.37398/jsr.2020.640137

Khawaja, M. A., Chen, F., & Marcus, N. (2012). Analysis of collaborative communication for linguistic cues of cognitive load. *Human Factors, 54*(4), 518–529. https://doi.org/10.1177/0018720811431258

Khawaja, M. A., Chen, F., & Marcus, N. (2014). Measuring cognitive load using linguistic features: Implications for usability evaluation and adaptive interaction design. *International Journal of Human-Computer Interaction, 30*(5), 343–368. https://doi.org/10.1080/10447318.2013.860579

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and instruction, 12*(1), 1-10. https://doi.org/10.1016/S0959-4752(01)00014-7

Konopasky, A., Durning, S. J., Artino, A. R., Ramani, D., & Battista, A. (2020). The Linguistic effects of context specificity: exploring affect, cognitive processing, and agency in physicians' think-aloud reflections. *Diagnosis, 7*(3), 273–280. https://doi.org/10.1515/dx-2019-0103

Kuiper, R. A. (2013). Integration of innovative clinical reasoning pedagogies into a baccalaureate nursing curriculum. *Creative Nursing, 19*(3), 128–139. https://doi.org/10.1891/1078-4535.19.3.128

Kuiper, R., Pesut, D., & Kautz, D. (2009). Promoting the self-regulation of clinical reasoning skills in nursing students. *The Open Nursing Journal, 3*, 76. https://doi.org/10.2174/1874434600903010076

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.), *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments* (pp. 61–83). Cambridge University Press.

Lajoie, S. P., & Lu, J. (2012). Supporting collaboration with technology: Does shared cognition lead to co-regulation in medicine? *Metacognition and Learning, 7*(1), 45-62.

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior, 25*(2), 284–289. https://doi.org/10.1016/j.chb.2008.12.010

Lew, M. D., & Schmidt, H. G. (2011). Self-reflection and academic performance: Is there a relationship? *Advances in Health Sciences Education, 16*(4), 529-545.

Li, S., Zheng, J., & Lajoie, S. P. (2020). Efficient clinical reasoning: Knowing when to start and when to stop. *Education in the Health Professions, 3*(1), 1–7.

Lo, Y., Mendell, N. R., & Rubin D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application, 6*, 355–378. https://doi.org/10.1146/ANNUREV-STATISTICS-031017-100325

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., & Wittig, F. (2001, July). Recognizing time pressure and cognitive load on the basis of speech: An Experimental study. In *International Conference on User Modeling* (pp. 24-33). Springer.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569. https://doi.org/10.1080/10705510701575396

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A Cognitive load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (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

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review, 22*(2), 115-121. https://doi.org/10.1007/s10648-010-9133-8

Paas, F., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1 Pt 2), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Panadero, E. (2017). A Review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductive-details-effect? A Cognitive-load-study using eye-tracking. *Educational Technology & Society, 18*(4), 24–36.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A Latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*(1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Pekrun, R. (2006). The Control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and psychometric properties of LIWC2015*. University of Texas at Austin. https://doi.org/10.15781/T29G6Z

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577. https://doi.org/10.1146/ANNUREV.PSYCH.54.101601.145041

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaert, P. R. Pintrich, & M. Zerdner (Eds.), *Handbook of self-regulation* (1st ed., pp. 451-502). Academic Press. https://doi.org/10.1016/b978-012109890-2/50043-3

Reedy, G. B. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing, 11*(8), 355-360. https://doi.org/10.1016/j.ecns.2015.05.004

Reiser, B. J. (2004). Scaffolding complex learning: The Mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2

Robinson, R. L., Navea, R., & Ickes, W. (2012). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology, 32*(4), 469–479. https://doi.org/10.1177/0261927X13476869

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2019). tidyLPA: An R package to easily carry out Latent Profile Analysis (LPA) using open-source or commercial software. *Journal of Open Science Software, 3*(30), 978. https://doi.org/10.21105/joss.00978

Scheiter, K., Ackerman, R., & Hoogerheide, V. (2020). Looking at mental effort appraisals through a metacognitive lens: Are they biased? *Educational Psychology Review, 32*(4), 1003–1027. https://doi.org/10.1007/s10648-020-09555-9

Scherer, R., Rohatgi, A., & Hatlevik, O. E. (2017). Students' profiles of ICT use: Identification, determinants, and relations to achievement in a computer and information literacy test. *Computers in Human Behavior, 70*, 486–499.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Seufert, T. (2018). The Interplay between self-regulation in learning and cognitive load. *Educational Research Review, 24*, 116–129. https://doi.org/10.1016/j.edurev.2018.03.004

Seufert, T. (2020). Building bridges between self-regulation and cognitive load—An Invitation for a broad and differentiated attempt. *Educational Psychology Review, 32*(4), 1151-1162. https://doi.org/10.1007/s10648-020-09574-6

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: The Links between language, performance, error, and workload. *Journal of Human Performance in Extreme Environments, 5*(1), 63–68.

Simmons, B. (2010). Clinical reasoning: Concept analysis. *Journal of Advanced Nursing, 66*(5), 1151–1158.

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J., Schuwirth, L., Artino, A. R., 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), 1-9. https://doi.org/10.1038/s41598-019-50280-3

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (pp. 37-76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Uchechukwu Onu, D., Iorfa, S. K., & Ugwu, D. I. (2020). Negative centralisation of HIV/AIDS trauma and health-related quality of life: do post-traumatic stress symptoms explain the link? *African Journal of AIDS Research, 19*(3), 206-213. https://doi.org/10.2989/16085906.2020.1797842

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychology Science, 21*(4), 258-262. https://doi.org/10.1177/0963721412447622

Winne, P. H. (2001). Self-regulated learning viewed from models of information Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives* (pp. 145–178). Routledge.

Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology, 88*(1), 9-20. https://doi.org/10.1111/bjep.12173

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaert, P. R. Pintrich, & M. Zerdner (Eds.), *Handbook of self-regulation* (1st ed., pp. 531-566). Academic Press. https://doi.org/10.1016/B978-012109890-2/50045-7

Zheng, J., Li, S., & Lajoie, S. P. (2020). The Role of achievement goals and self-regulated learning behaviors in clinical reasoning. *Technology, Knowledge and Learning, 25*(3), 541-556.

Zimmerman, B. J. (2000). Attaining self-regulation: A Social cognitive perspective. In M. Boekaert, P. R. Pintrich, & M. Zerdner (Eds.), *Handbook of self-regulation* (1st ed., pp. 13-39). Academic Press. https://doi.org/10.1016/b978-012109890-2/50031-7