Enhancing the Learning Experience Using Real-Time Cognitive Evaluation

Maher Chaouachi, Imène Jraidi, Susanne P. Lajoie, and Claude Frasson

***Abstract*—There is increasing evidence that learners' affective and cognitive states play a key role in the learning process. This suggests that systems which are able to detect these states can dynamically use adapted strategies to increase the pace of the learners' skill acquisition and improve their learning experience. In this work, we present a novel approach for automatically adapting the learning strategy in real-time according to the learner's detected mental state. The main goal of the approach is to maintain the learner in a positive state during a lesson by adaptively selecting the best interaction strategy between either using problem solving or worked examples. Two mental indexes, namely, cognitive load and mental engagement were extracted from electroencephalogram (EEG) signals, and used to adapt the system's interaction. The cognitive load index was developped by training and validating a prediction model on various types of memory and logical tasks. The engagement index was directly computed from the EEG signal frequency bands. An experiment with 14 learners was performed in order to evaluate this approach. The obtained results showed that using the learner's mental state to adapt the system's interaction has a positive impact on the learning outcomes, the learning experience and the learners' reported emotional states.**

***Index Terms*—Adaptive system, mental engagement, cognitive load, EEG, affect, learning performance, learning experience.**

I. Introduction

Affect sensitive computerized learning systems have shown considerable promising results in analyzing and improving the learning experience [1], [2]. Methodological, technical and analytical approaches are being constantly explored to help these systems leverage the knowledge of the learners' behavior to analyze and support the learning process. These approaches have also largely benefited from numerous critical advances in the area of physiological computing especially in terms of accuracy, data modeling, portability and scalability. Significant results were obtained specially in 1) relating physiological processes to several cognitive and affective states that occur during learning and 2) analyzing how these states influence positively or negatively the learning outcomes. Several sensors tracking the learners' electro-dermal activity, eye movements, heart rate, posture, facial expressions, or brain activity have been used in a uni-modal or multimodal way to infer states such as engagement, attention, frustration, flow, boredom, etc. [3]-[10].

Nevertheless, despite this increased emphasis on affective computing, there is a paucity of research addressing the impact of designing learning systems that adapt in real-time to the users' affective and cognitive states. In fact, most of these approaches so far rely on offline data analysis procedures to extract information about the learners' states. This is due to several technical issues related to data acquisition, preprocessing, synchronization and classification. In this research, we focus on affect-aware adaptive systems, i.e., how to develop an interactive learning environment able to recognize and act on the learner's affective state. In this paper we present an approach to recognize and respond to two specific cognitive states, namely, 1) engagement which characterizes the level of involvement and interest a learner has during a task, and 2) cognitive load which measures the amount of information processing demands and mental effort imposed on the learner while processing a task.

These states are among the most commonly used indicators to dynamically assess changes in the users' states in several fields such as aviation, robotics and army as they are closely related to the users' performance and experience [11], [12]. According to the Yerkes-Dodson Law, a low cognitive load level (mental *under-load*) as well as high cognitive load level (mental *overload*) are correlated with poor performance [13]. The state of engagement is also considered as a crucial factor during the learning process since it is closely related to motivation, memorization and learning achievement [14], [15].

In this paper we hypothesize that maintaining learners in an appropriate level of cognitive load and engagement throughout a learning session, where they have to acquire and master a new way of writing programming operations, can foster effective knowledge acquisition and a better learning experience. To this end, an adaptive learning environment, MENTOR, was developed to detect the learners' levels of mental engagement and cognitive load in real-time, as a basis for selecting the best approach to deliver the learning content. MENTOR uses two mental indexes as well as the learner's progress during the lesson, to decide the type of problem solving task to administer: a challenging activity that could increase his mental engagement but could likewise highly increase his cognitive load, or a worked example task: a less engaging activity but also less demanding in terms of cognitive load. An experimental study was conducted to evaluate how the learners interacted with the system with a two-fold objective:

1) First, to investigate if the integration of the engagement and the cognitive load brain indexes within an adaptive

Manuscript received May 10, 2019; revised July 11, 2019.

M. Chaouachi, I. Jraidi, and S. P. Lajoie are with the Department of Educational and Counselling Psychology, McGill University, 3700 McTavish Street, Montréal, QC H3A 1Y2, Canada (e-mail: maher.chaouachi@mcgill.ca, imene.jraidi@mcgill.ca, susanne.lajoie@mcgill.ca).

C. Frasson is with the Department of Computer Science and Operations Research, University of Montreal, 2920 Chemin de la Tour, Montréal, H3T-1J8 QC, Canada (e-mail: claude.frasson@umontreal.ca).

doi: 10.18178/ijiet.2019.9.10.1287

learning environment has a positive impact on the learners' results. Our assumption is that if the system considers the learners' level of engagement and cognitive load as part of their teaching strategy, then it will be able to select the most appropriate approach to help the learners better understand the new concepts.

2) Second, to prove that the development of such an adaptive strategy also positively impacts the learner's experience. In other words, the system's awareness and adjustment to the learners' brain indexes will reflect positively on their affective state as well as on their perception of the whole learning process.

The remainder of this paper is organized as follow: Section II reviews previous work on adaptive learning environments. Section III presents the MENTOR system's design and the methodology used to extract the engagement and the cognitive load indexes from the EEG data. Section IV details the adaptive logic used to interact with learners. Section V describes the experimental design and section VI the empirical results. Finally, section VII draws conclusions and proposes future research directions.

II. Related Work

The investigation of existing physiologically adaptive learning environments shows mainly two categories of intervention strategies that are being used to automatically respond to the learners' states, namely: affect-based and problem-based intervention strategies.

A. Affect-Based Intervention Strategy

The objective of this category of interventions is to create a "relational" or "social" dimension between the learning environment and the learner by responding to the learners' state using verbal (through direct messages) and non-verbal (through mimics and expressions showed by animated agents) communication of empathy and encouragement [16]-[20]. Prendinger and Ishizuka [21] proposed one of the first studies in this area. They developed an educational agent that uses skin conductance level and electromyography to extract information about the learners' affective states and to automatically give empathic feedback that help learners preparing for a job interview and managing their stress level. Woolf *et al*. [22] also proposed a multimodal affect interaction approach to help students learn mathematics with an intelligent tutoring system called Wayang Outpost. The approach used a pressure sensitive mouse, a posture analysis seat, a camera and a skin conductance sensor to detect learners' affective states. The system used two animated characters to help learners be aware of their own emotions, either implicitly by mirroring the detected emotional state or explicitly by providing an empathic or encouraging feedback prompt. This same affect-based intervention strategy was employed in Auto Tutor [23]-[25], a conversional tutoring system that helps students master topics in physics, computer science and reasoning. Auto Tutor was able to recognize affective states such as boredom, frustration or confusion by analyzing conversational cues, body movements and facial expressions. The system then automatically interacted with the learners by delivering motivational and empathic messages.

B. Problem-Based Intervention Strategy

This category of interventions aims to attract (increase or recapture) the learners' attention by highlighting specific relevant elements, materials, contents or learning approaches that can help the learners achieve their goals. The idea behind this strategy is that when a negative affective or cognitive state (such as frustration, boredom or mental disengagement) occurs, the system's intervention needs to directly address the source causing this negative state. For example, D'mello *et al*. [26] developed a gaze-reactive intelligent tutoring system that uses attention reorientation strategies when detecting boredom or disengagement from the learner's gaze. The system was based upon a set of rules for delivering direct verbal messages in to reengage the learners into the activity by orienting them towards specific relevant areas of the screen. Szafir and Mutlu [27] also investigated how to avoid learners' drops of attention while listening to lectures delivered by a pedagogical agent. Learners' attention was detected using electroencephalography (EEG). The agent used immediate verbal cues (such increasing its vocal volume) and non-verbal cues (such as gesturing and head nodding) while speaking to learners to increase their attention.

Even though the affect-based intervention strategy is meant to strengthen the social interaction between the learners and the system and to help learners handle their emotions, little evidence supports the effectiveness of such an approach. For instance, Prendinger *et al*. [28] showed that the empathic companion did not have any significant impact on the learners' outcomes. Besides, the empathic strategies can generate the opposite of their intended effect on learners and rather increase their negative emotions especially if they are not timely and properly delivered [22], [29], [30]. In this work, we decided to focus on the problem-based intervention strategy. Our objective is to develop a system that recognizes the learners' engagement and cognitive load states and adapts the teaching content accordingly to help them reach an optimal learning state.

III. System Design

MENTOR (MENtal tuTOR) is a tutoring system that uses two brain indicators, namely engagement and cognitive load extracted from the EEG physiological data to adjust the learning strategy according to the learner's mental state [31], [32]. The overall objective of the system is to maintain the learners in an appropriate state.

The system was designed to interface directly with an Emotiv EEG wireless headset[1] to collect EEG raw signals. The Emotiv headset contains 16 electrodes located according to the 10-20 international standard [33]. It allows recording simultaneously 14 regions (O1, O2, P7, P8, T7, T8, FC5, FC6, F3, F4, F7, F8, AF3 and AF4). Two additional electrodes are used as references, which correspond respectively to the P3 and P4 regions. The system's sampling rate is 128 Hz.

Two different approaches are used by the system to compute the two brain indexes. The first one, which is used to calculate the engagement index, is based on a direct extraction and processing of specific frequency bands from the EEG signal. The second approach, which is used to

[1] www.emotiv.com

compute the cognitive load index, relies on training a machine-learning model.

A. Engagement Index Extraction

The term mental engagement refers to the level of alertness and attention allocated during a task. The engagement index used in this work draws on the findings of Pope and colleagues [34] at the National Aeronautics and Space Administration (NASA). This index, which is based on neuroscientific research on attention and vigilance [35]-[37], was studied and used as a criterion for switching between manual and automated piloting modes and showed a positive impact on the pilots' performance when it was used to activate the autopilot mode or to control the level of task automation in the cockpit [38].

Since its development, this engagement index has become a very important and popular parameter for real time or offline tracking and analysis of individuals' engagement in several laboratory studies. Within educational settings for example, this index was used for monitoring learners' engagement during problem solving and listening activities [9], [27], [39], [40]. This engagement index was also selected as a criterion for adapting a game's difficulty according to the player's level of engagement [41].

The engagement index is computed using three EEG frequency bands, namely: θ (4-8 Hz), α (8-13 Hz) and β (13-22 Hz) as follows:

$$Eng\_Index = \frac{\beta}{\theta + \alpha}$$

Since the EEG signal is very sensitive to all kinds of artifacts such as eye blinks or body movements, MENTOR system uses an artifact rejection heuristic before the computation of the EEG ratio. The procedure developed by Freeman *et al*. [42] is applied to each incoming 1-second EEG epoch. This procedure consists of examining whether the signal amplitude exceeds a fixed threshold in 25% of the epoch data points. In such a case, the epoch is considered as contaminated and rejected. The extraction of the θ, α and β frequency bands is performed first by filtering each 1-second of the non-rejected EEG signal by a Hamming window to reduce the signal discontinuities at the epoch edges. Then, a Fast Fourier Transform (FFT) is applied to each windowed epoch to convert it to the frequency domain and to extract the needed frequencies. As the Emotiv headset measures 14 regions at the same time, we used a combined value of the θ, α and β frequency bands by summing their values over all the measured regions. The engagement index is computed each second from the EEG signal. In order to reduce the fluctuation of this index, MENTOR uses a moving average on a 40-second mobile window. Thus, the value of the index as the time *t* corresponds to the total average of the ratios calculated on a period of 40 seconds preceding *t*.

B. Cognitive Load Index Extraction

The term cognitive load (also referred to as mental cognitive load or simply workload) is the amount of information processing demands placed on an individual by a task [43]. Unlike the engagement index, there is no common established method to directly assess mental cognitive load from the EEG data. However, the development of EEG indexes for cognitive load assessment using machine learning algorithms is a well-developed research topic, which was investigated in various application domains. Linear and non-linear classification and regression models were used to measure this state in different kinds of cognitive tasks such as memorization, language processing, visual, or auditory tasks. These models rely mainly on a frequency processing approach using either the Power Spectral Density (PSD) or Event Related Potential (ERP) techniques to extract relevant EEG features [38], [44]-[46]. For instance, Wilson [47] used an Artificial Neural Network (ANN) to classify the operators' workload level by taking the users' EEG data as well as other physiological features as an input. The reported results showed up to 90% of classification accuracy. Gevins and Smith [48] used spectral features to feed a neural network classifying the user's workload while performing various memorization tasks. Kohlmorgen *et al*. [49] used a Linear Discriminant Analysis (LDA) on EEG features extracted and optimized for each user for workload assessment. The authors showed that decreasing the driver's workload (induced by a secondary auditory task) improves reaction time. Berka and colleagues developed a workload index using Discriminant Function Analysis (DFA) for monitoring alertness and cognitive load in different learning environments [11], [50], [51].

In this work, we propose to build an individual predictive cognitive load model for each learner before the interaction with MENTOR. The main idea is that this model is trained using data collected from a first training phase, during which the learner performs a set of brain training exercises while his EEG signals are recorded. In this training phase, the learner performs different sets of cognitive exercises with different levels of difficulty. Three different types of cognitive exercises are used to collect the EEG data used to train the predictive model namely: digit span (DS), reverse digit span (RDS) and mental computation (MC). The objective of these training exercises is to induce different levels of cognitive load on the learner.

In the DS and RDS exercises, the learner is asked to memorize and recall a series of simple digits successively presented on the screen. The MC consists in mentally performing addition and subtraction operations. The manipulation of the level of cognitive load imposed on the learner is made by varying the difficulty level of the exercises, i.e. by increasing the number of the digits in the sequence to be recalled for DS and RDS, and by increasing the numbers to be added or subtracted for the mental computation exercises [51], [52]. In total each learner performs 54 DS exercises, 54 RDS exercises and 36 MC exercises with different difficulty levels. After performing each exercise, the learner is asked to report his cognitive load level using the subjective scale of NASA Task load index (NASA_TLX) [53].

Once this exercise phase is completed, the training of the cognitive load predictive model begins. Fig. 1 summarizes the model building steps. The collected EEG raw data are cut into 1-second segments and filtered by a Hamming window. The same artifact rejection procedure applied for the engagement index is used for the cognitive load index. A FFT is used to transform each EEG segment into a spectral frequency. A set of 40 bins of 1 Hz ranging from 4 to 43 Hz

(EEG pretreated vectors) is generated. These data are then reduced using a Principal Component Analysis (PCA) to 25 components (the score vectors). Next, a Gaussian Process Regression (GPR) algorithm with an exponential squared kernel and a Gaussian noise [54] is run in order to train a mental cognitive load predictive model (the EEG workload index) from the normalized score vectors. Normalization is done by subtracting the mean and dividing by the standard deviation of all the vectors. In order to reduce the training time of the predictive model, we used the local Gaussian Process Regression algorithm, which is an optimized (faster) version of the GPR algorithm [55].

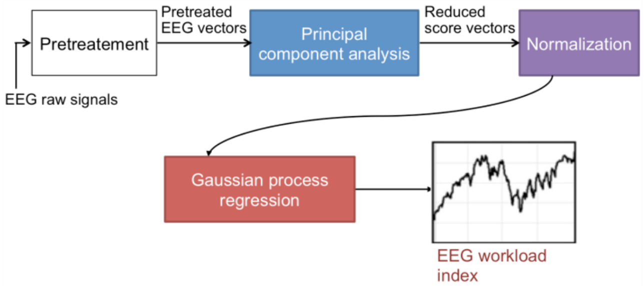

Fig. 1. Mental cognitive load predictive model.

C. Analysis of the Computed Indexes

In order to evaluate the learner's mental state, the system analyzes the behavior of the engagement and cognitive load indexes throughout the learning session. A slope of each index is computed using the least squared error function of the indexes' values from the beginning of the activity. For the engagement index, if the slope value is postive, then the learner is considered mentally engaged. Otherwise, the learner is considered as mentally disengaged. For the cognitive load index, if the slope value is between - 0.03 and + 0.03, than the cognitive load is considered as positive. Otherwise, if the slope value is above 0.03, the learner is considered as overloaded, and if the slope is below -0.03 the learner is considerd as under-loaded.

IV. Learning with the Adaptive Learning Environment

MENTOR tutoring system is designed to help learners understand the Reverse Polish Notation (RPN), which is also known as the postfix notation. The lesson presented by the system includes four successive parts. The first part presents a set of formal definitions of the algebraic expressions as well as their structures and constituent elements. The second part explains how to determine the priorities between the operators and how to evaluate an algebraic expression without parentheses. The third part focuses on the concept of the RPN; the basics of the postfix notation are introduced and explained. The fourth part details the techniques used for the assessment of an RPN expression.

After the learner finishes each part of the lesson, the system presents four pedagogical activities so that the learner puts into practice the concepts seen in this previous part of the lesson and enhances his or her understanding. Each activity uses one of the two following pedagogical resources:

- Questions: each question presents a problem that the learner has to resolve. Hints are provided with each problem to help the learner find the solution and improve his or her knowledge acquisition. At the end of each question, the system informs the learner whether his or her answer was correct or not. In case of a wrong answer, the solution of the problem is given without presenting any explanation of the resolution process.
- Worked examples: a worked example describes a problem statement with the detailed steps and explanations leading to the solution. The learner is simply asked to read and understand these examples.

A. MENTOR's Adaptive Rules

MENTOR's decisional process lies mainly in the selection of the type of the pedagogical resource (a question or a worked example) to be provided as a next activity. In summary, 16 decisions (4 parts × 4 activities) are made by the system according to the learner's mental state. This choice between worked examples or problems has often been discussed in educational psychology. On one hand, worked examples tend to have a lower mental load impact compared to problems [56]. Indeed, a worked example provides all the required steps of the problem resolution process. The only effort that a learner has to produce is to understand these steps. On the other hand, problems are more demanding in terms of mental effort as the learner has to resolve the problem and in case of a wrong answer, he must also understand the solution.

Providing only worked examples to the learners can have a negative impact. The learner may not identify the relevant information pertaining to the worked example, and focus rather on useless or secondary information. Another phenomenon that frequently occurs when the learning activities are only based on worked examples is the phenomenon of the illusion of understanding. This phenomenon arises when learners thinks they understood the example while they did not. This generally occurs when the learner browses the elements of the example superficially without producing a minimum effort to understand the goal of each step of the resolution process [57]. Besides, presenting a worked example does not guarantee that the learner will be able to generalize from the shown example. Indeed, some learners do not spontaneously make an effort to analyze, reproduce and compare the resolution steps of the example, as compared to the efforts that they would have made if they had to reslove the problem by themselves.

The advantage of using problem solving activities in a learning session is therefore to avoid these risks. The questions are always considered as an efficient educational instrument to assess the learners' knowledge and help them efficiently acquire new skills. However, using a pedagogical approach based on solving problems exclusively can also hinder the learning process. In fact, as the mental effort is considered stronger compared to worked examples, the learner can become easily tired and overloaded. Moreover, if the learner fails to solve the problems, he or she can be frustrated, demotivated and even disengaged from the task.

The decision of presenting a worked example or a problem within MENTOR is based on a continuous analysis of the learner's mental engagement and cognitive load. The goal is to select the pedagogical resource that maintains the learner in a positive mental state. Particularly, the system has to keep the learner mentally engaged and avoid both overload and under-load. If the system detects a negative mental state caused by an engagement drop, an overload or an under-load, then it will try to correct this state by switching the type of the next pedagogical activity.

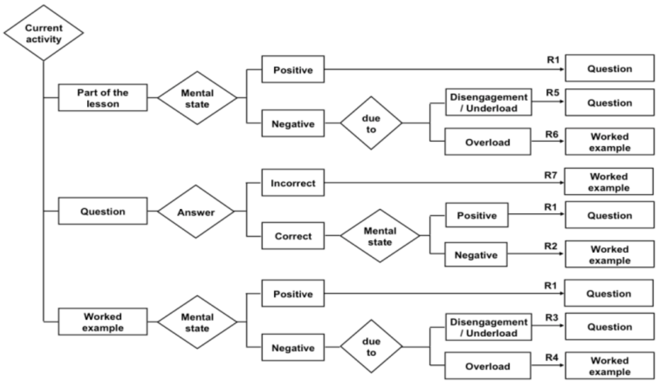

Fig. 2. MENTOR's adaptive logic for selecting the next pedagogical resource.

A total of seven adaptive rules are used by MENTOR as shown in Fig. 2:

(R1) If the learner's mental state is positive (mentally engaged and neither overloaded nor under-loaded), then the system selects a question for the next activity. This rule is applied whatever the current activity is (question, worked example or reading a part of the lesson).

(R2) At the end of a question, if the learner's mental state is negative (disengaged, overloaded or under-loaded), then the system provides a worked example in the next activity.

(R3) At the end of a worked example, if the system detects a negative mental state due to disengagement or under-load, then it provides a question as a next activity.

(R4) At the end of a worked example, if the system detects a negative mental state due to overload, then it provides a worked example in the next activity.

(R5) After reading a part of the lesson, if the system detects a negative mental state due to disengagement or under-load, then it provides a question as a next activity.

(R6) After reading a part of the lesson, if the system detects a negative mental state due to overload, then it provides a worked example for the next activity.

(R7) Whatever the learners' mental state is, if he answers a question incorrectly, then the system provides a worked example in the next activity.

The idea behind the use of these rules is given hereafter:

Decision after reading a part of the lesson. The system uses the questions as a main pedagogical approach to help the learner understand the presented concepts. The rule (R1) makes that the system automatically provide a question if the learner's state is positive. The hypothesis behind this rule is that if the learner reads a lesson while maintaining a positive state, then he or she likely did not have difficulty understanding the presented concepts. So, by choosing a question as a subsequent activity, the system checks the learner's knowledge. However if the learner's mental state is negative, the system analyzes the cause of this state. If this negative state is due to a mental overload, then the rule (R6) makes the system choose a worked example in the next activity. The hypothesis behind this rule is that a mental overload signals generally a cognitive difficulty with regards to the presented concepts. The learner produces then a high level of mental effort to understand what he or she was reading. So the decision of presenting a worked example after this activity can help the learner better understand the presented part without producing further mental effort. In this case, we think that giving a problem to solve while the learner is overloaded can worsen his or her cognitive load level and disturb the learning process. However, if the learner's negative state is due to a disengagement or a mental under-load, the system selects a question using the rule (R5). In this case, we assume that this lack of mental investment is either due to the fact that the learner was perfectly mastering what he or she was reading, or was rather disinterested and neglecting the lesson. In both cases, a question can be a more stimulating and challenging activity for the learner and can probably enhance his mental investment.

Decision after a question. At the end of a question, if the system does not detect a negative state, it chooses another question as a next *activity* using the rule (R1). We suppose in this case that the learner reacts well mentally and that the strategy based on the questions is currently well suited to the learner's state. It is important to note that the use of the rule (R1) is limited by the rule (R7). So, in case of a wrong answer, the system automatically switches the next activity to a worked example even though the learner is in a positive mental state to prevent the occurrence of a negative state due to a succession of wrong answers.

This same switch is also performed using rule (R2), if the system detects a negative mental state even though the learner's answer is correct. The assumption is that if the learner shows a negative state following the resolution of a problem, then changing the type of the activity can be in any case beneficial. More precisely, if the learner is overloaded, switching to a worked example in the next activity can correct or prevent this state from getting worse (this would probably be the case if the system follows-up with another question). If the negative state is caused by a disengagement or an under-load, changing the type of the activity can be stimulating for the learner and may correct this negative state.

Decision after a worked example. After presenting a worked example, the system opts *for* a question as a next activity if the learner's mental state is positive using the rule (R1). The reason of using this strategy is to target an effect known as the problem completion effect [58], which is generally obtained by providing a worked example followed immediately by a problem. This type of strategy is used to increase the learning performance and enhance the learner's motivation [59]. For this reason, we decided to choose a question as a subsequent activity to the worked example even if the learner's state is positive, rather than pursuing with another worked example.

Finally, if the system detects a negative mental state caused by an overload, the system continues to present a worked example in the next activity using rule (R4). The assumption behind this rule is that if learners have some cognitive difficulties to understand the example, or if they are simply tired, it would not be suitable to give them a problem to solve since this can worsen the overload. Therefore, another worked example can support their knowledge without being more demanding mentally.

V. Experimental Study

In order to highlight the impact of using the learners' mental indicators as an adaptive criterion to manage the system's pedagogical resources, our experimental study relied on two different versions of MENTOR (V1 and V2). The difference between these versions lies only in the adaptive logic of the decision module. The first version (V1) leaves intact the adaptive logic with the seven intervention rules described previously. The selection of the resource to be provided is done according to the assessment of the learner's mental state. In particular, the system tends to privilege the questions in case of a positive mental state. In the opposite case, the selection of the type of the resource is made following different heuristics that aim to correct the learner's mental state.

The second version of the system (V2) does not take into account the brain indexes of engagement and cognitive load in selecting the type of the resource to be provided. Only the rule (R7) is preserved in the adaptive logic of MENTOR, and the six other rules are ignored. The principle of this version is quite simple: after reading each part of the lesson, the system chooses to ask a question to the learner. As the learner answers correctly, the system continues to adopt the same strategy: asking questions. However, if an incorrect answer is given, the system switches immediately to a worked example to fix the learner's reasoning. Once the learner finishes reading the example, the system automatically follows-up with a question to increase motivation and elicit a problem completion effect. Thus, the unique parameter that can trigger an adaptation action in this version is an incorrect response of the learner.

The two used versions share a common point in their operation: if the adaptation parameters are positive, both opt for a question as a next step. The mental state sensitive version of the system (the first) is then an augmented version of the second, insofar as in addition to considering the accuracy of the response (through the 7th rule), it also applies other adaptive actions based on the mental parameters.

In summary, we will compare two versions of the system; both versions have the same pedagogical content. However, V1 uses in its adaptive logic, an analysis of the mental indexes in addition to the response of the learner to decide the appropriate timing of the content. V2 gives the same content but based solely on the response of the learner. Both versions use, in the same order, exactly the same pedagogical resources. That is the system will have to choose between the same pair of resources including a question and a worked example. The difference between them will therefore lie in the choice of the type of resource to be selected; the two versions can opt for the same resource or for two different resources.

A. Experiment Protocol

14 participants were recruited and took part to our study. All were students of the University of Montreal in the same certification program in applied computer science. Each participant was randomly assigned to one of the two following groups. 1) The experimental group ($N = 7$) used the adaptive version of MENTOR (V1): the learning activities are actively adapted to both the learners' brain indexes and answers. 2) The control group (N=7) used the second version of MENTOR (V2) that considers only the learners' answers.

For each participant, the experiment was conducted on two successive days. On the first day, the participant used the training module of MENTOR in order to create his individual cognitive load model. In this phase, which lasted about an hour, participants performed a series of 40 brain training exercises including digit span, reverse digit span and mental computation as described earlier. The cognitive load predictive model was trained for each participant on randomly selected 80% of the dataset and tested on the remaining 20%. The mean value of the root mean squared error (RMSE) of each cognitive load model across the 14 participants was equal to 0.144 on the testing set. Fig. 3 details the RMSE for each individual model. The mean correlation coefficient between the values predicted by the models and the target values was equal to 0.53 with a minimum correlation coefficient value equal to 0.2 and a maximum correlation coefficient equal to 0.88.

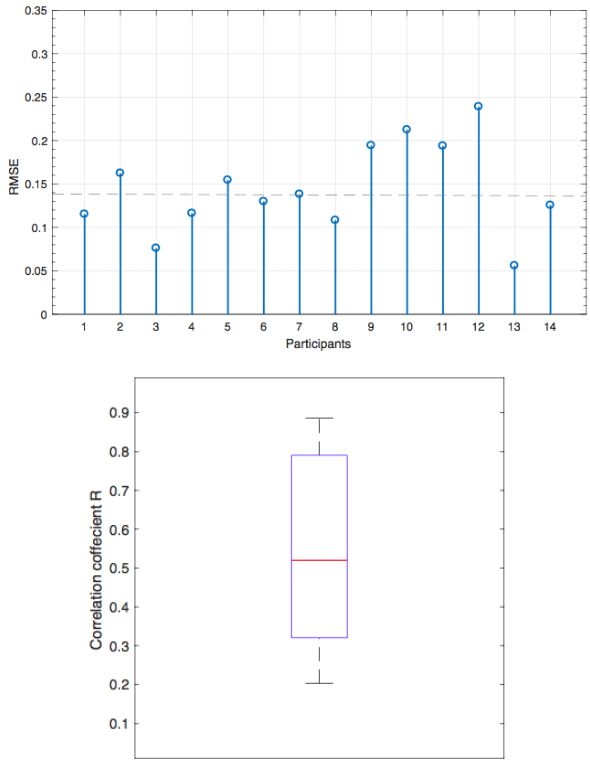

Fig. 3. Stem plot of RMSE relative to each individual cognitive model. Box plot of the Mean correlation coefficient between the predicted values and the targets for all the participants.

On the second day of the experiment, participants used the learning module of MENTOR. The duration of this phase was approximately one hour, including 20 to 30 minutes to learn the four parts of the Reverse Polish Notation lesson. The session starts with a pre-test followed by the lesson, then a post-test, and ends with a debriefing phase. Two 5-minute breaks were taken between the pre-test, the lesson and the post-test.

Pre-test and post-test. The objective of the pre-test is to determine a priori the level of knowledge of the learner on the subjects covered by the course. The post-test determines the level of knowledge acquired after the learning session. This allows us to assess the learner's progress (between the pre-test before the lesson and the post-test after the lesson).

These tests use a set of 16 questions relative to the concepts of the lesson. Each of the four parts of the lesson is concerned with four different questions. The same questions are asked in the pre-test and in the post-test. For each question, the learner can answer true or false, or may choose not to respond. A typical example of a question is to check whether two postfix expressions are equivalent. The score in each test is calculated as follows: a correct answer is worth 1 point, while a wrong answer (or a non-response) is worth 0.

Debriefing. During this phase, the learner is first asked to give his opinion regarding his interaction with the learning environment by rating his *satisfaction level.* A scale of seven grades ranging from 1 (strongly disagree) to 7 (strongly agree) is used to rate how much he agrees with the following statement: "Overall, I am satisfied with of my learning experience with the system".

Then, the learner evaluates the quality of the tutoring provided by the system by rating his perceived level of *relevance* of the system's proposed activities, using another scale of seven grades ranging from 1 (strongly disagree) to 7 (strongly agree), on how much he agrees with the following statement: "Overall, I am satisfied with the learning activities selected by the system. The examples and questions are presented at the right time and helped me understand the lesson. The choices made between asking a question or presenting an example fits my level of understanding". This scale is therefore an evaluation of the relevance (or the perspicacity) of the tutor's decisions.

B. Recording Emotions

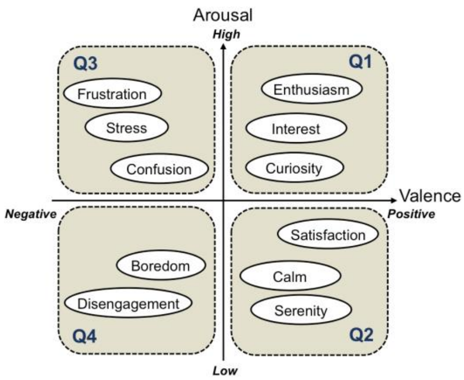

Fig. 4. The valence/arousal emotional model with the four quadrants.

At the end of each activity, the learner was asked to describe his emotional state. The two-dimensional valence/arousal model shown in Figure 4 (adapted from [60]) was used. This model classifies emotions into four quadrants: Q1, Q2, Q3 and Q4 in terms of valence ranging from unpleasant or negative emotions to pleasant or positive, and arousal ranging from low intensity or activation to high. The learner was asked to choose the quadrant that involves the emotions which best match his current emotional state.

- Q1: includes positive emotions with high intensity such as interest, curiosity and enthusiasm.
- Q2: includes positive emotions with low intensity like calm, satisfaction and serenity.
- Q3: includes negative emotions with high intensity like confusion, stress and frustration.
- Q4: includes negative emotions with high intensity like boredom and disengagement.

VI. Results

The experimental results are presented in the following subsections. First, we analyze the behavior of the engagement and the cognitive load indexes. Then, we assess the impact of using the EEG indexes as an adaptive criterion on learning: we compare the learners' outcomes and progression in the two considered groups (experimental vs. control) between the pre-test and the post-test. Next, we present a comparative study of the emotional responses between the two groups of participants. Finally, we analyze the impact of using the two versions of the system on the learners' satisfaction level.

A. EEG Indexes

As a first step, we are interested in analyzing the behavior of the learners' brain indexes when the system detects a negative mental state. Specifically, we aim to validate the effectiveness of our strategy in analyzing the learners' mental state to trigger MENTOR's interventions. Thus, we aim to compare how these indexes behave before and after signaling a negative state caused by a considerable engagement drop or an important cognitive load decrease (or increase). For this purpose, we were specifically interested in the learners of the control group who interacted with the second version of the system (V2): with an intervention strategy based only on the accuracy of the response given by the learner. The question we were asking was: how do the learners' mental engagement and cognitive load indexes behave in such cases? That is, how do these indexes vary when the analysis module of MENTOR would have detected a negative mental state, and that this is not taken into account by this second used version of the system? In other words, we want to analyze the behavior of these two mental indexes, if there is a divergence between the two intervention strategies concerning the selection of the type of the pedagogical resource to be provided for the next activity.

First, we started with the cases where the adaptive system (V1) would have detected an engagement drop causing a negative mental state, and would have proposed a pedagogical resource different from that actually selected by the non-adaptive version of the system used in the control group (V2). These cases occurred 15 times on the 112 possible decisions taken by this version of the system (7 participants * 16 choices). A repeated measure ANOVA was performed, with the time variable as a within-subject variable (before and after the selected activity), the participants' ID as a between-subject variable, and the mean value of the engagement index as the dependent variable, i.e. the repeated measure. The results showed that the mean value of the engagement index was significantly higher before the intervention points detected by MENTOR, compared to the mean value of the engagement index after these intervention points: $F(1, 8) = 21.156$ $p < 0.05$. This reveals that the analysis module correctly identifies the engagements drops, and that in the absence of an adequate adaptation, this index continues to fall.

The second cases concern the divergences of decisions for the learners of the control group when the system detects a mental overload. These cases occurred 11 times during our experimentations. The same repeated measure ANOVA was conducted, but this time with the mean value of the workload index as a dependent variable. The ANOVA revealed a statistically significant difference in the mean values of the index before and after the activity proposed by the system: $F(1, 5) = 40.866$, $p < 0.05$; with a mean value significantly higher after the activity selected by the system. Hence, if the system does not select a pedagogical strategy that takes into consideration this mental overload, the value of this index will continue to increase, and the learner's negative state may worsen.

The third case concern the situations of mental under-load. Similarly, we wanted to see how the workload index behaves if the analysis module detects a cognitive under-load, and that the activity proposed by the non-adaptive version of the system is different from that which would have been proposed by the mental state sensitive version of MENTOR. These cases occurred 8 times during the experimentations. Another repeated measure ANOVA revealed that there was a significant difference in the mean values of the workload index before and after the activity proposed by the system: $F(1, 2) = 33.597$, $p < 0.05$; with a mean value even lower after the proposed activity.

This analysis confirms that: first the analysis module of MENTOR can correctly detect the engagement drop, the overload and the under-load critical mental states; and secondly, it is indeed necessary to undertake adaptive actions that correct these states.

B. *Learning Performance*

A 2 (group: experimental vs. control) × 2 (time: pre-test vs. post-test) mixed-model analysis of variance (ANOVA) was conducted to compare the learners' outcomes of the two groups in terms of scores achieved in both tests. The group variable is a between-subject factor that compares the scores between the two experimental conditions, whereas the time variable is a within-subject factor that analyzes for each participant, individually, the score variation (i.e. changes) between the pre-test and the post-test.

First, the analysis yielded a main effect of the time variable, showing a significant difference of the learners' scores in both groups between the pre-test and the post-test: $F(1, 12) = 2253.353$, $p < 0.001$. Thus, there was a significant learning gain regardless of the group (experimental vs. control). Second the analysis yielded a significant interaction effect of both factors (group × time) on the learners' outcomes: $F(1, 12) = 29.824$, $p < 0.001$. The results revealed that over time (i.e. between the pre-test and the post-test), the learners of the experimental group have got significantly better learning performance compared to the control group. The means of scores obtained in the pre-test and the post-test for both groups are listed in Table I.

The comparison of the learners' scores between the experimental group and the control group revealed that there were no statistically significant differences between the two groups in the pre-test: $F(1, 12) = 4.190$, p = n.s. The overall mean score in the pre-test was $M = 4.21$ ($SD = 1.31$). In contrast, the comparison of the learners' scores in the post-test showed that the scores achieved in the experimental group were significantly higher than the control group: $F(1, 12) = 50.069$, $p < 0.001$.

TABLE I: LEARNERS' OUTCOMES IN BOTH GROUPS BEFORE AND AFTER THE TUTORING SESSION

	Pre-test	Post-test
Experimental group		
M	4.86_a	13.86_b
SD	1.07	0.70
Control group		
M	3.57_a	10.71_c
SD	1.27	0.95

Values with different subscripts differ significantly.

These results confirm our first hypothesis, that is using the cognitive load and the engagement indexes as a main criterion to control the learner's activities can have a positive impact on his learning performance. The learners' whose pedagogical resources were selected according to their mental states were able to provide an average of 86.6 % correct answers after the tutoring session. An increase of 22.7 % in terms of learning outcomes was achieved using this adaptive strategy.

C. *Emotional Responses*

In order to evaluate the impact of our approach on the learners' emotional state during their interractions with the two versions of MENTOR, we have calculated the proportions (percentages) of occurrence of each quadrant of the two-dimensional valence/activation model during the tutoring session. A multivariate analysis of variance (MANOVA) was conducted to compare the learners' experienced emotions between the experimental group and the control group. The group factor was used as an independent variable and the proportions of each quadrant (Q1, Q2, Q3 and Q4) as a dependent variable.

The results showed that there is a significant difference between the two groups in terms of proportions of quadrants: $F(3, 10) = 8.665$, $p < 0.05$. The analysis of each specific quadrant, using four distinct ANOVAs with a Bonferroni correction, showed that the two groups were statistically different emotionally. The mean proportions of emotions are given in Table 2. The following results were found for each quadrant:

- Q1 (positive valence and high arousal): $F(1, 12) = 5.945$, $p < 0.05$; the proportions of Q1 in the experimental group were significantly higher than those in the control group.
- Q2 (positive valence and low arousal): $F(1, 12) = 5.37$, $p < 0.05$; the proportions of Q2 in the experimental group were also significantly higher compared to the control group.
- Q3 (negative valence and high arousal): no significant difference was found for Q3: $F(1, 12) = 4.101$, p = n.s. However, the proportions of Q3 in the experimental group were lower.
- Q4 (negative valence and low arousal): $F(1, 12) = 10.8$, $p < 0.05$; the proportions of Q4 in the experimental group were significantly lower than the control group.

The analysis of the learners' emotions experienced in the two experimental conditions confirms the positive impact of the use of the mental state-based adaptive version of MENTOR on the learners' emotional experience. With a large dominance of the quadrant Q1 (including emotions such as interest, curiosity and enthusiasm) and Q2 (calm, satisfaction and serenity), the learners tended to report more positive emotions when the system takes into account their mental state in the activity sequencing (experimental group). Similarly, the negative emotions of the quadrant Q4 (boredom and disengagement) were significantly more prevalent in the control group, with the non-adaptive version of the system.

TABLE II: DESCRIPTIVE STATISTICS OF THE PROPORTIONS OF EMOTIONS IN EACH EXPERIMENTAL CONDITION

	Control group		Experimental group	
Quadrant	M	SD	M	SD
Q1	0.25	0.19	0.51	0.20
Q2	0.20	0.05	0.33	0.13
Q3	0.20	0.15	0.07	0.06
Q4	0.35	0.19	0.09	0.07

D. Subjective Measures

An ANOVA was conducted in order to compare the learners' satisfaction levels between the experimental group and the control group. This ANOVA showed an almost significant difference between the two groups: $F(1, 12) = 4.545$, $p = 0.054$. The learners of the experimental group reported higher satisfaction ($M = 5.71$, $SD = 1.604$) compared to the control group ($M = 4.29$, $SD = 0.756$).

A second ANOVA was performed to compare the learners' ratings regarding the relevance of the activities proposed by the tutoring system in both groups. These ratings were significantly higher in the experimental group ($M = 5$, $SD = 1.414$) versus ($M = 2.43$, $SD = 0.787$) in the control group: $F(1, 12) = 17.673$, $p < 0.05$.

These results confirm thus that increasing the system's adaptive logic with the EEG engagement and cognitive load indexes has a positive effect on the learners' opinion regarding the relevance of the decisions taken by the system in the selection choice of the pedagogical resources more specifically.

VII. CONCLUSION

In this paper we have presented an intelligent tutoring system that adapts its tutoring strategy according to the learner's brain activity. The goal was to show that the use of mental indicators of the learners' state such as the engagement and the cognitive load levels can have a positive impact on the learning outcomes as well as on the learners' interaction experience. The approach is based on recording EEG data and inferring the two indexes using two different methods respectively. (1) The system calculates the mental engagement index by computing a ratio of specific frequency bands extracted from the EEG signal. (2) The system applies a machine-learning algorithm to compute the learners' cognitive load using brain-training exercises to record learners' EEG data and infer their workload index.

In our experimental study, a learning session was conceived during which a group of learners interacted with the system to learn a new lesson about the postfix notation. The learning module of the system provides a tutoring environment that adapts its teaching strategy actively according to the learner's brain indexes. Two different versions of the system were tested. In the first version (the experimental group), the system evaluates the learner's mental state, and selects between a problem solving and a worked example, the activity that best suits the learner's mental state as well as his current performance. In the second version (the control group), the system passively computes the mental indexes, and only the learners' performance is used as a criterion to switch the activity.

First, the analysis of the EEG data showed that the interventions undertaken by MENTOR to correct the engagement drops, overloads and under-loads were indeed required, otherwise, the learner's mental state gets worse. Second, it was found that augmenting the adaptive logic of the system with the cognitive load and the engagement indexes has a positive impact on the learners' performance in terms of learning gains before and after using the tutoring environment as compared to the control group. Third, it was found that using the learners' mental state as a criterion to sequence the tutoring activities also has a positive impact on the learners' interaction experience in terms of positive emotional responses and higher ratings regarding the relevance of the system's activities.

In this paper the proposed cognitive load model is based on a training phase where a model was calibrated based on each individual learner. That is an individual machine-learning model was created as each learner executed various cognitive tasks. Even though many researchers in brain-computer interface embrace this idea of individualized predictive models of cognitive load as they provide highly accurate results [51], the time and the computing processing required to build these models is clearly an obstacle for the application of such an approach within non-laboratory and operational contexts. In our study, the training phase of the model was performed during the first day of the experiment whereas the interaction with the learning system was realized on the second day. This experimental setup was utilized for two reasons: first the Gaussian Process Regression model has a cubic complexity, which requires some time to train the models. The second reason is that this training phase, which used various cognitive tasks with different difficulty levels, was mentally demanding for the participants. Hence, the learning activity had to be performed on a different day. An alternative solution to make this EEG cognitive load modeling approach more practical outside the laboratory context is to use a generalized approach where a single predictive model is trained and validated using gathered EEG data once and for all. This unique model would be then used for any new participant without requiring a training phase. However, even though the generalized model could save the training time of the individualized models, the accuracy of the resulting model could be reduced, and the system's behavior could be hindered.

In our future work, we will focus on comparing individualized and generalized predictive cognitive load models. We will investigate and compare the use of different machine-learning techniques. Moreover, we will compare the performance of this EEG cognitive load models with other cognitive load indicators extracted from different sensors (such as heart rate and skin conductance response). Regarding the learning system, we will also integrate other kinds of adaptive strategies such as actively selecting hints or automatically switching the format of the content of the lecture (for instance from text to video).

CONFLICT OF INTEREST

The authors do not have any conflicts of interest to report.

AUTHOR CONTRIBUTIONS

M. Chaouachi conducted the data collection and processing, and co-wrote the paper. I. Jraidi conducted the data analysis and interpretation, and co-wrote the paper. S. P. Lajoie co-supervised this work and co-wrote the paper. C. Frasson co-supervised this work.

ACKNOWLEDGMENTS

This research was supported by funding from the Natural Sciences and Engineering Research Council (NSERC) and the Social Sciences and Humanities Research Council (SSHRC) of Canada.

REFERENCES

[1] N. Bosch *et al*., "Using video to automatically detect learner affect in computer-enabled classrooms," *ACM Trans. Interact. Intell. Syst.*, 2016, vol. 6, no. 2, pp. 1-26.

[2] J. M. Harley, A. Jarrell, and S. P. Lajoie, "Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation," *Instructional Science*, 2019, pp. 1-30.

[3] S. M. Khan, "Multimodal behavioral analytics in intelligent learning and assessment systems, in innovative assessment of collaboration," *Methodology of Educational Measurement and Assessment*, 2017, Springer: Cham, pp. 173-184.

[4] S. Oviatt, J. Grafsgaard, L. Chen, and X. Ochoa, "Multimodal learning analytics: assessing learners' mental state during the process of learning," *The Handbook of Multimodal-Multisensor Interfaces*, 2019, Association for Computing Machinery and Morgan & Claypool., pp. 331-374.

[5] A. B. Khedher, I. Jraidi, and C. Frasson, "Tracking students' mental engagement using EEG signals during an interaction with a virtual learning environment," *Journal of Intelligent Learning Systems and Application*, 2019, vol. 11, no. 1, pp. 1–14.

[6] A. B. Khedher, I. Jraidi, and C. Frasson, "Local sequence alignment for scan path similarity assessment," *International Journal of Information and Education Technology*, 2018, vol. 8, no. 7, pp. 482–490.

[7] I. Jraidi and C. Frasson, "Student's uncertainty modeling through a multimodal sensor-based approach," *Journal of Educational Technology & Society*, 2013, vol. 16, no. 1, pp. 219-230.

[8] A. B. Khedher, I. Jraidi, and C. Frasson, "Static and dynamic eye movement metrics for students' performance assessment," *Smart Learning Environments*, 2018, vol. 5, no. 1, p. 14.

[9] M. Chaouachi, P. Chalfoun, I. Jraidi, and C. Frasson, "Affect and mental engagement: Towards adaptability for intelligent systems," in *Proc. 23rd International FLAIRS Conference*, 2010, AAAI Press: Daytona Beach, Florida, USA.

[10] I. Jraidi, M. Chaouachi, and C. Frasson, "A hierarchical probabilistic framework for recognizing learners' interaction experience trends and emotions," *Advances in Human-Computer Interaction*, 2014.

[11] C. Berka *et al*., "Evaluation of an EEG workload model in an Aegis simulation environment, in defense and security," *International Society for Optics and Photonics*, pp. 90-99, 2005.

[12] M. B. Sterman and C. A. Mann, "Concepts and applications of EEG analysis in aviation performance evaluation," *Biological Psychology*, 1995, vol. 40, no. 1- 2, pp. 115-130.

[13] K. H. Teigen, "Yerkes-dodson: A law for all seasons," *Theory & Psychology*, 1994, vol. 4, no. 4, pp. 525-547.

[14] M. Arguedas, T. Daradoumis, and F. Xhafa, "Analyzing how emotion awareness influences students' motivation, engagement, self-regulation and learning outcome," *Educational Technology & Society*, 2016, vol. 19, no. 2, pp. 87-103

[15] C. Su and C. Cheng, "A mobile gamification learning system," *Journal of Computer Assisted Learning*, 2015, vol. 31, pp. 268-286.

[16] G.-D. Chen *et al*., "An empathic avatar in a computer-aided learning program to encourage and persuade learners," *Educational Technology & Society*, 2012, vol. 14, no. 2, pp. 62-72.

[17] S. K. D'Mello *et al*., "I feel your pain: A selective review of affect-sensitive instructional strategies, in design recommendations for adaptive intelligent tutoring systems," *Army Research Laboratory: Orlando*, pp. 35-48, 2014.

[18] I. Jraidi and C. Frasson, "Subliminally enhancing self-esteem: Impact on learner performance and affective state," *Intelligent Tutoring Systems*, 2010, Springer Berlin Heidelberg, pp. 11-20.

[19] I. Jraidi, P. Chalfoun, and C. Frasson, "Implicit strategies for intelligent tutoring systems," *Intelligent Tutoring Systems*, 2012, Springer Berlin Heidelberg, pp. 1-10.

[20] Y. Kim and A. L. Baylor, "Research-based design of pedagogical agent roles: A review, progress, and recommendations," *International Journal of Artificial Intelligence in Education*, 2016, vol. 26, no. 1, pp. 160-169.

[21] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users' affective states," *Applied Artificial Intelligence*, 2005, vol. 19, pp. 267-285.

[22] B. Woolf *et al*., "Affect aware tutors: Recognising and responding to student affect," *International Journal of Learning Technology*, 2009, vol. 4, no. 3/4, pp. 129-164.

[23] S. D'Mello, R. W. Picard, and A. Graesser, "Toward an affect-sensitive autotutor," *IEEE Intelligent Systems*, 2007, vol. 22, no. 4, pp. 53-61.

[24] S. D'Mello and A. C. Graesser, "Feeling, thinking, and computing with affect-aware learning technologies," *The Handbook of Affective Computing*, Oxford University Press: Oxford, pp. 419-434, 2015.

[25] A. C. Graesser, "Conversations with autotutor help students learn," *International Journal of Artificial Intelligence in Education*, 2016, vol. 26, no. 1, pp. 124-132.

[26] S. D'Mello, A. Olney, C. Williams, and P. Hays, "Gaze tutor: A gaze-reactive intelligent tutoring system," *Int. J. Hum.-Comput. Stud*., 2012, vol. 70, no. 5, pp. 377-398.

[27] D. Szafir and B. Mutlu, "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, 2012, ACM: Austin, Texas, USA, pp. 11-20.

[28] H. Prendinger, S. Mayer, J. Mori, and M. Ishizuka, "Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces," in *4th International Working Conference on Intelligent Virtual Agents (IVA-03)*, Springer: KlosterIrsee, Germany, 2003, pp. 283-291.

[29] B. P. Bailey and S. T. Iqbal, "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management," *ACM Trans. Comput.-Hum. Interact*, 2008, vol. 14, no. 4, pp. 1-28.

[30] C. A. Monk, D. A. Boehm-Davis, and J. G. Trafton, "The attentional costs of interrupting task performance at various stages," *Human Factors and Ergonomics Society Annual Meeting*, 2002, SAGE Publications, pp. 1824-1828.

[31] M. Chaouachi, I. Jraidi, and C. Frasson, "Adapting to learners' mental states using a physiological computing approach," in *Proc. 28th International Florida Artificial Intelligence Research Society Conference (FLAIRS-28)*, 2015, AAAI Press, pp. 257-262.

[32] M. Chaouachi, I. Jraidi, and C. Frasson, "MENTOR: A physiologically controlled tutoring system," in *Proc. 23rd International Conference in User Modeling, Adaptation and Personalization*, UMAP, 2015, Springer International Publishing: Cham, pp. 56-67.

[33] H. H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, 1958, no. 10, pp. 371-375.

[34] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological Psychology*, 1995, vol. 40, no. 1-2, pp. 187-195.

[35] R. J. Davidson, "EEG measures of cerebral asymmetry: Conceptual and methodological issues," *International Journal of Neuroscience*, 1988, vol. 39, no. 1-2, pp. 71-89.

[36] K. Offenloch and G. Zahner, *Computer Aided Physiological Assessment of the Functional State of Pilots during Simulated Flight,* NATO Advisory Group for Aerospace Research and Development-Conference Proceedings, 1990, vol. 490, pp. 1-9.

[37] J. F. Lubar, “Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders,” *Biofeedback and Self-regulation*, 1991, vol. 16, no. 3, pp. 201-225.
[38] L. J. Prinzel, F.G. Freeman, and M. W. Scerbo, “A closed-loop system for examining psychophysiological measures for adaptive task allocation,” *International Journal of Aviation Psychology*, 2000, vol. 10, no. 4, pp. 393-410.
[39] I. Jraidi, M. Chaouachi, and C. Frasson, “A dynamic multimodal approach for assessing learners' interaction experience,” in *Proc. the 15th ACM on International Conference on Multimodal Interaction*, 2013, ACM: Sydney, Australia, pp. 271-278.
[40] I. Jraidi, A. B. Khedher, M. Chaouachi, and C. Frasson, “Assessing students’ clinical reasoning using gaze and EEG features, in intelligent tutoring systems,” *Lecture Notes in Computer Science*, 2019, Springer: Cham.
[41] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, “Emotion assessment from physiological signals for adaptation of game difficulty,” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2011, vol. 41, no. 6, pp. 1052-1063.
[42] F. G. Freeman, P. J. Mikulka, L. J. Prinzel, and M. W. Scerbo, “Evaluation of an adaptive automation system using three EEG indices with a visual tracking task,” *Biological Psychology*, 1999, vol. 50, no. 1, pp. 61-76.
[43] R. Parasuraman and D. Caggiano, “Mental workload,” *Encyclopedia of the Human Brain*, 2002, vol. 3, pp. 17-27.
[44] M. A. Just, P. A. Carpenter, and A. Miyake, “Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work,” *Theoretical Issues in Ergonomics Science*, 2003, vol. 4, no. 1-2, pp. 56-88.
[45] M. Fabiani, G. Gratton, and M. G. Coles, “Event-related brain potentials,” *Handbook of Psychophysiology*, Cambridge University Press: Cambridge, England, pp. 53-84, 2000.
[46] A. T. Pope, E. H. Bogart, and D. S. Bartolome, “Biocybernetic system evaluates indices of operator engagement in automated task,” *Biological Psychology*, 1995, vol. 40, no. 1, pp. 187-195.
[47] G. F. Wilson, “An analysis of mental workload in pilots during flight using multiple sychophysiological measures,” *Int J Aviat Psychol*, 2004, vol. 12, pp. 3-18.
[48] A. Gevins and M. E. Smith, “Neurophysiological measures of cognitive workload during human-computer interaction,” *Theoretical Issues in Ergonomics Science*, 2003, vol. 4, no. 1-2, pp. 113-131.
[49] J. Kohlmorgen *et al*., “Improving human performance in a real operating environment through real-time mental workload detection,” *Toward Brain-Computer Interfacing*, 2007, MIT Press: Cambridge, MA, pp. 409–422.
[50] C. Berka *et al*., “Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset,” *International Journal of Human-Computer Interaction*, 2004, vol. 17, no. 2, pp. 151-170.
[51] C. Berka *et al*., “EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks,” *Aviation, Space, and Environmental Medicine*, 2007, vol. 78, no. 5, pp. B231-B244.
[52] M. Chaouachi, I. Jraidi, and C. Frasson, *Modeling Mental Workload Using EEG Features for Intelligent Systems*, in *User Modeling, Adaption and Personalization*, Springer Berlin Heidelberg, p. 50-61, 2011.
[53] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (task load index): Results of empirical and theoretical research,” *Human Mental Workload*, 1988, vol. 1, no. 3, pp. 139-183.
[54] C. E. Rasmussen, *Gaussian Processes for Machine Learning*, 2006.
[55] D. Nguyen-Tuong, J. R. Peters, and M. Seeger, “Local Gaussian process regression for real time online model learning,” *Advances in Neural Information Processing Systems*, 2008.
[56] S. Kalyuga, P. Chandler, J. Tuovinen, and J. Sweller, “When problem solving is superior to studying worked examples,” *Journal of Educational Psychology*, 2001, vol. 93, no. 3, p. 579.
[57] D. Scott, “On engendering an illusion of understanding,” *The Journal of Philosophy*, 1971, pp. 787-807.
[58] F. G. Paas, “Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach,” *Journal of Educational Psychology*, 1992, vol. 84, no. 4, p. 429.
[59] J. Sweller, “Evolution of human cognitive architecture,” *Psychology of Learning and Motivation*, 2003, vol. 43, pp. 215-266.
[60] P. J. Lang, “The emotion probe: Studies of motivation and attention,” *American Psychologist*, 1995, vol. 50, no. 5, pp. 372-385.

Maher Chaouachi holds a MS degree in business informatics specialized in the field of machine learning from the University of Tunis and a PhD in computer science specialized in the field of artificial intelligence obtained from the University of Montreal. He worked as a postdoctoral researcher at the advanced technologies for learning in authentic settings laboratory (ATLAS lab) at McGill University. Dr. Chaouachi research interests focus on how to use artificial intelligence techniques, process and data mining, data analytics and predictive modeling to develop adaptive learning systems and to optimize learning outcomes and experience.

Imène Jraidi is a research associate in the Department of Educational and Counselling Psychology at McGill University. She received her MS degree in business informatics at the University of Tunis and completed her PhD in computer science at the University of Montreal. Dr. Jraidi is now working in the advanced technologies for learning in authentic settings (ATLAS) laboratory. Her main research interests are in artificial intelligence, human and machine learning, multimodal interaction and cognitive and affective modeling.

Susanne P. Lajoie is a Canadian research chair in advanced technologies for learning in authentic settings in the Department of Educational and Counselling Psychology and a member of the Institute for Health Sciences Education at McGill University. She is a fellow of the Royal Society of Canada, the American Psychological Association as well as the American Educational Research Association (AERA). She received the ACFAS Thérèse Gouin-Décarie Prize for social sciences along with the AERA-TICL outstanding international research collaboration award. Professor Lajoie directs the learning environments across disciplines partnership grant funded by the social sciences and humanities research counsel in Canada. She explores how theories of learning and affect can be used to guide the design of advanced technology rich learning environments in different domains, i.e. medicine, mathematics, history, etc.

Claude Frasson is a professor in the Computer Science and Operations Research Department at the University of Montreal. He is the head of the Higher Educational Research ON Tutoring Systems (HERON) laboratory. Professor Frasson founded the Intelligent Tutoring Systems (ITS) international conference. He also created the GRITI group (Inter-university Research group in Intelligent Tutors) a multidisciplinary research group involving seven universities in Quebec including researchers in database, artificial intelligence, multimedia, cognitive psychology, and education. His interest is related to intelligent tutoring Systems, artificial and emotional intelligence, education, pedagogy, learning techniques, knowledge engineering, e-learning, and neural techniques. His goal is to understand how the human brain functions and produces knowledge, how this knowledge is stored, retrieved and deployed.