

Using Mouse and Keyboard Dynamics to Detect Cognitive Stress During Mental Arithmetic

Yee Mei Lim, Aladdin Ayesh and Martin Stacey

Abstract To build a personalized e-learning system that can deliver adaptive learning content based on student's cognitive effort and efficiency, it is important to develop a construct that can help measuring perceived mental state, such as stress and cognitive load. The construct must be able to be quantified, computerized and automated. Our research investigates how mouse and keyboard dynamics analyses could be used to detect cognitive stress, which is induced by high mental arithmetic demand with time pressure, without using intrusive and expensive equipment. The research findings suggest that when task demand increased, task error, task duration, passive attempt, stress perception and mouse idle duration may increase, while mouse speed, left mouse click and keystroke speed decreased. The significant effects of task demand and time pressure on mouse and keystroke behaviours suggest that stress evaluation from these input devices is potentially useful for designing an adaptive e-learning system.

Keywords Stress detection · Keyboard dynamics · Mouse dynamics · Mental arithmetic · Adaptive e-learning

Y.M. Lim (✉)

Faculty of Applied Sciences and Computing, Tunku Abdul Rahman University College,
Kuala Lumpur, Malaysia
e-mail: ymlim@acd.tarc.edu.my

A. Ayesh · M. Stacey

Faculty of Technology, De Montfort University, Leicester, UK
e-mail: aayesh@dmu.ac.uk

M. Stacey

e-mail: mstacey@dmu.ac.uk

1 Introduction

The research is an investigation of the use of mouse and keyboard dynamics to measure cognitive stress without intrusive special equipment, when experimental subjects do mental arithmetic with different levels of difficulty and time pressure. It is motivated by the desire to develop e-learning systems that can adapt their behaviour and the content they deliver to the needs of their users. Learning is a complex process, which is not simply about acquiring knowledge, but involves mix of working memory organization, attention and cognitive control processes, which could be impacted by motivational and emotional factors. Negative emotion inducing environment and high-stakes situation that generates stress, fear of failure, anxiety or stereotype threat (such as female cannot do programming) could often cause students to perform at their worst. Beilock and Ramirez [1] study the relationship between emotion and cognitive control, and they found that high-pressure and negative emotion-inducing situations reduce student maths performance. On flip side, if they are placed under less emotion-inducing situations, the students are more readily available for executing a more challenging task. This is because negative emotion could inhibit appropriate cognitive resources that are necessary for optimal skill execution to be recruited by human mind. Other factors that affect cognitive load also include causal and assessment factors. Causal factors involve the characteristics of the subject such as skills or expertise possessed, task complexity, environment (such as noise) and their mutual relations. Assessment factors contain mental load, mental effort and performance [2]. To measure cognitive load, one or more assessment techniques can be utilized, which include subjective methods, physiological tests and task-performance based measurement [3]. Subjective methods, such as self-report survey, are done based on the assumption that humans are able to measure their thought (for instance, the amount of mental effort they expended or the level of stress they experienced). Although this method is simple, it is considered unreliable as human thinking is highly subjective and people can easily deny their thoughts. Therefore it is important to have an effective measure to quantify cognitive load. Physiological tests are able to detect changes in cognitive functioning that are reflected in measurable physiological measurements, such as heart rate or eye activity. However they cannot be easily implemented without special equipment (which is normally expensive), so not as part of normal system. Furthermore, physiological tests are invasive to the experimental subjects as the equipments are attached to their bodies, so they may not feel comfortable to carry out the task normally. Task-performance-based techniques measure actual performance of the given tasks. This technique is more reliable than the subjective method, as quantitative data such as success and failure rates of the task could be collected. However, solely relying on task-performance-based techniques may not be good enough as task performance could be affected by other factors such as attitude (e.g. lack of interest or seriousness in work), rather than weak cognitive function. It is better if some of these techniques can be combined to give a relative indication of the acceptable level of cognitive load.

To introduce a cost-effective, non-invasive and computational efficient method, automatic analysis of how users produce mouse and keyboard input during task execution is potentially useful. If mouse and keystroke behaviours are related to task performance and cognitive load, then they can be applied for designing adaptive instructional contents in an e-learning system. Furthermore, if the system can evaluate users' mental states or behaviours by measuring their emotions and stress levels, then the system is able to affect the attitude of the users towards learning and help them overcome learning obstacles [4]. This is because measures of cognitive overload, which leads to difficulty in coping with task demands (overstress in Selye's terminology [5], and underload, leading to boredom and lapses of attention (understress in Selye's terminology), are particularly important. Our research aims to analyse how keystroke and mouse behavioural patterns change according to the task demand, which is varied by mental arithmetic problem complexity and time pressure. We would like to observe how cognitive stress relates to task-performance (such as error rate, the duration spent on a task and the attempt of giving up a task), mouse and keystroke behaviours. If correlations between user's cognitive stress, task performance, mouse and keyboard dynamics can be found, then this information is potentially useful in designing an adaptive e-learning system.

2 Related Work

2.1 *Mental Arithmetic and Cognitive Load*

Mental arithmetic problems under time pressure are widely used to induce cognitive stress [6–8]. A study by Imbo and Vandierendonck [9] suggested that larger numbers and borrow operations in arithmetic problems, which involve longer sequences of steps and require maintenance of more intermediate products, will place greater demands on human working memory. Once the demand has exceeded the working memory capacity and temporal limitations, then the task is deemed too challenging to be continued [10]. Although much research has investigated how attention, memory and computational processes support arithmetic calculations, but less work has addressed how math performance can be influenced by emotional factors, such as stress. Beilock and Ramirez [1] suggested that stressful and emotion-inducing situations could lead to unwanted performance degradation even for relatively simple calculations in math performance, due to negative emotion could prevent or inhibit the recruitment of the appropriate cognitive resources necessary for optimal skill execution. However, Weinberg et al. [11] argued that human attention to emotion stimuli may not be automatic nor obligatory. When the context of the emotion stimuli is not relevant to the task (such as seeing a picture of a crying face), human may demonstrate little-to-no impact on the emotional modulated arithmetic task. In other words, the effects of the stimuli on cognitive process may depend on both of the attentional demands of the task

and the salience of the stimuli [12]. The impact of negative emotion on performance decrement may be caused by the task demands itself (such as high requirements), or other factors that are related to the task (such as time pressure).

2.2 Mouse and Keyboard Dynamics in Emotion Detection

Mouse and keyboard dynamics analyses shed light on non-invasive emotion detection research since they have generated promising results in biometrics or authentication work. These input devices are not only cheap in cost, but they also provide greater advantages for a solution that can be fully automated and computerized, as compared to physiological methods. Both mouse and keyboard dynamics have been shown to differ according to different emotion, but most previous work has considered them in isolation. Lim et al. [13] investigated the effects of Web menu design on users' emotion, search task performance and their mouse behaviours. Their results showed that the effects of menu design on users' search task performance and their mouse behaviours are statistically significant. Bad setting of menu design generally increases mouse idle duration and occurrences, and reduces mouse speed and mouse click. Tsoulouhas et al. [14] used mouse dynamics to test students' boredom. Their research demonstrates that mouse speeds, mouse inactivity occurrences, mouse inactivity durations and movement directions are significantly different between bored and non-bored users, which they recorded their best results with the intervals of 10 s (false acceptance rate at 2.7586 %). Pusara and Brodley [15] and Shen et al. [16, 17] analysed mouse dynamics by focusing on user behavioural modelling. Shen et al. [18] stated that the user's distinctive mouse operation patterns can be caused by changes to several factors, which include user's emotional states such as anger, despair, happiness, nervous, excitement, pressure and so on, and physical conditions such as tiredness and illness. Vizer [19] analysed keystroke dynamics and linguistic features by detecting changes in typing associated with physical stress and cognitive stress. Their research showed that keystroke features can be significantly changed by cognitive stress, which include keystroke pause length (key latency), time per keystroke (keystroke speed), deletion keys (backspace key and delete key), and use of navigation keys and other keys (such as letter and number keys). However, although using mouse and keyboard dynamics to detect emotion is proven effective, there is very little research done that unifies mouse and keyboard dynamics in emotion detection. The unification of both methods is important as there is a risk of collecting misleading information from only one channel. For instance, if we only analyse keystrokes, the results may be affected by long stops and irregular restarts [20], which could be due to the user's attention being diverted to another activity, or the user using a mouse rather than a keyboard to perform an action (such as drag-and-drop or clicking a button to execute a command). Moreover, in a real application, users may use either mouse or keyboard, or a combination of both for specific tasks.

3 Research Questions and Design

We begin by hypothesizing that an automatic evaluation of cognitive stress can be obtained through acquisition and processing of three datasets, which are task performance ($B(T)$), mouse behaviour ($B(M)$) and keystroke behaviour ($B(K)$). Our research questions are as follows:

1. Do task demand and time pressure affect cognitive stress?
2. Do task demand and time pressure affect user’s task performance, mouse behaviour and keystroke behaviour?
3. Are there correlations between task demand, cognitive stress, task performance, mouse and keystroke behaviours?

We would like to examine the potential significant effects of cognitive stress, which is induced by task demands with time pressure, on the changes of behavioural patterns in $B(T)$, $B(M)$ and $B(K)$. If the answers for the questions above are positive, then a rule-based adaptive e-learning system can be designed. Figure 1 shows our proposed system architecture using model-view-controller design. The models include modelling of keystroke behaviour, mouse behaviour and task performance (see Sect. 4.1). These behaviours are formed based on mouse and keystroke raw data such as mouse locations, time-stamps, keys pressed, etc., which are collected in every 10 ms. Then based on the needs of the system developer, the user behaviour can be analysed for an interval of designated time, t . Due to huge temporal variations of mouse and keyboard dynamics of a user, and also high behavioural differences between individuals, calibration of mouse and keyboard dynamics should be collected during login process, so that baseline condition (non-stressed) can be formed. Furthermore, these huge variations can be sensitive to generate significant difference even small departures from homogeneity and the assumption of normality, hence the collected data should be transformed using appropriate function (such as logarithm and square root).

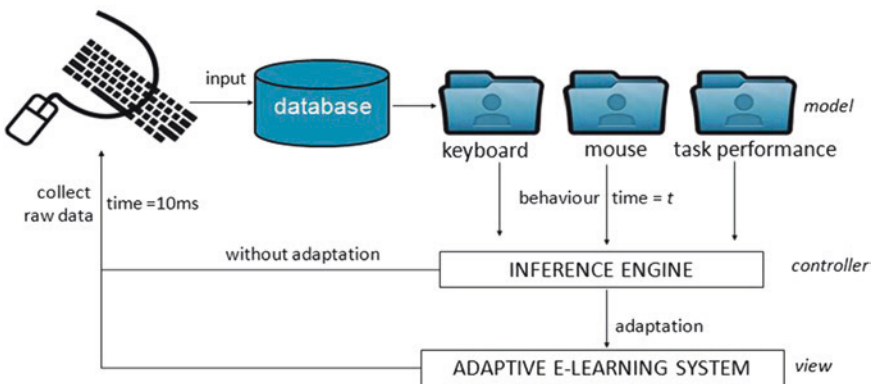


Fig. 1 Proposed system architecture

Privacy must be embedded into the design and architecture of the system, and we must be offering measures as strong privacy defaults, appropriate notice and empowering user-friendly option [21]. Therefore the users should be given an option for not to be observed by the adaptive system. We also need to ensure that at the end of the process, all data are securely destroyed, in a timely fashion, and no data that reveals individual identity would be kept. The actual data of the keys used, which reflect the original content of the text (such as username and password) must not be stored. These data must be encoded for the use of the analysis purpose only (for instance, all number keys or character keys are represented as 'k'). After the necessary transformation and formation of individual user behaviour, the system could then compare the subsequent behaviours with the baseline condition so that the behavioural patterns can be analysed. Once the rule that detects significant increment of stress level is fired, then the instructional content of the e-learning system can be adapted to motivate the learner to continue the task.

4 Methodology

To enable necessary data to be collected for the formation of Task Performance ($B(T)$), Mouse Behaviour ($B(M)$) and Keystroke Behaviour ($B(K)$), a program is written in Java to capture the features of $B(M)$, and another separate program is written in VB.NET to obtain the virtual-key codes by the Windows platform for $B(K)$. For every 10 ms, the mouse location is captured and its respective time in milliseconds is recorded. For every keystroke and mouse click, the key information and the time (in milliseconds) of the event is stored. To simulate an e-learning environment, an imitation of the online assessment system is built. Ten different mental arithmetic problems with diverse complexity are given to the students (see Table 1). Each question is displayed on different individual Web pages. The students must answer all questions by doing mental arithmetic, and must type the answer into a designated textbox on each page. No calculator or calculation on paper is allowed. To force the student to use the mouse, the "Enter" key is disabled, and he or she must click the "Submit" button in order to submit the answer. On each page of the question displayed, a "give up" button is given so that the students can choose to skip the question if they do not wish to continue. Before starting, an instruction page is displayed and their agreements to continue the experiments must be obtained. Once they click the start button, the start time (in milliseconds) will be recorded and the first arithmetic question is revealed. When the participant submitted the page, the end time (in milliseconds) is recorded and the data needed by $B(T)$, $B(M)$ and $B(K)$ will be computed automatically.

Each time after the students completed a question (or skipped the question), a self-report survey will be displayed as follows:

You felt stressed when answering the previous question

Table 1 Mental arithmetic questions. We assume that task demand increased from Question 1 to Question 10 according to the increment of number of digit per number and amount of numbers in the question

No.	Maximum digit per number	Amount of numbers	Arithmetic problem
1	1	2	$6 + 2$
2	1	2	$9 * 4$
3	1	3	$6 * 5 - 1$
4	1	3	$(8 + 9) * 2$
5	2	3	$7 - 8 * 10$
6	2	4	$58 + 20 * (8 - 6)$
7	2	4	$67 - 2 * (4 + 2)$
8	3	5	$(880 + 12 + 50 - 520) * 2$
9	3	5	$105 + 83 * 5 - 3 * 60$
10	3	5	$561 - 81 * 5 + 3 * 610$

This survey enables them to assess their stress perceptions when solving the arithmetic problem, following 7-point Likert scale (1 for strongly disagree, 7 for strongly agree). Therefore, this provides us the subjective measurement of the user's cognitive stress, *SP*.

All participants are required to run the experiments in a computer laboratory of a higher education institution in Malaysia. All the computers in the laboratory were equipped with Windows 7, 3.10 GHz CPU, 4 GB RAM, 17" monitor with the resolution of $1,024 \times 768$ pixels, external standard QWERTY HID keyboard and external HID-compliant mouse. The website runs on Google Chrome by default. Before they started the assessment, instructions are displayed on the screen and they must provide their consensus in order to continue the experiments.

4.1 The Control Group and Experimental Group

Seventy-seven year-2 students from Bachelor Degree in Computer Science and Bachelor Degree in Information Technology from the Malaysian higher education institution participated the experiments. However, due to outliers and missing cases, only 60 of them provided valid samples. Among these 60 students, all of them were between 18 and 24 years old, and 90 % were male. These students were divided into 2 groups. For control group, the participants were required to answer the arithmetic questions without time constraint. For the experimental group, they were given 30 s time limit for each question. The page would be submitted automatically if they could not complete the task on time.

4.2 Formulation of Task Performance, Mouse Behaviour and Keystroke Behaviour

Task performance is a dataset that measures activities related to the tasks that a student has completed. Task performance, $B(T)$, is defined as follows:

$$B(T) = \langle TD, Err, PA \rangle \quad (1)$$

- TD The duration to complete one task (milliseconds (ms))
 Err Error of task ($Err = 0$ if no error; $Err = 1$ if the answer is wrong)
 PA Passive attempt ($PA = 999$ if attempt to give up; $PA = 1$ if attempt to wait until the time is up)

We define the mouse behaviour as a dataset that captures the mouse features for each task. The mouse behaviour, $B(M)$, is defined as follows:

$$B(M) = \langle MS, MID, MIO, MC \rangle \quad (2)$$

- MS Average mouse speed (pixels per ms)
 MID Total mouse inactivity duration (ms)
 MIO¹ Total mouse inactivity occurrences
 MC $\langle MCL, MCR^2 \rangle$, which is a dataset that consists of left click rate per ms (MCL) and right click rate per ms (MCR)

Lastly we define the keystroke behaviour, $B(K)$, as a dataset that captures the keystroke features for each task as below:

$$B(K) = \langle KL, KS, EK \rangle \quad (3)$$

- KL Average key latency (ms)
 KS Average typing speed per key (per second)
 EK³ $BSK + DK$, the total occurrences of error keys used (EK), which includes backspace (BSK) and delete (DK) keys.

5 Results

To observe the behavioural patterns of $B(T)$, $B(M)$ and $B(K)$ according to the changes of cognitive stress, SP , we conduct some statistical tests to perform the analyses. First, we use Levene's test to ensure homogeneity between the 2 groups. However due to the fact that Levene's test can be sensitive to detect even

¹ MIO was removed later due to inhomogeneous data.

² MCR was removed later due to no data.

³ EK is removed due to insufficient data for BSK and no data for DK .

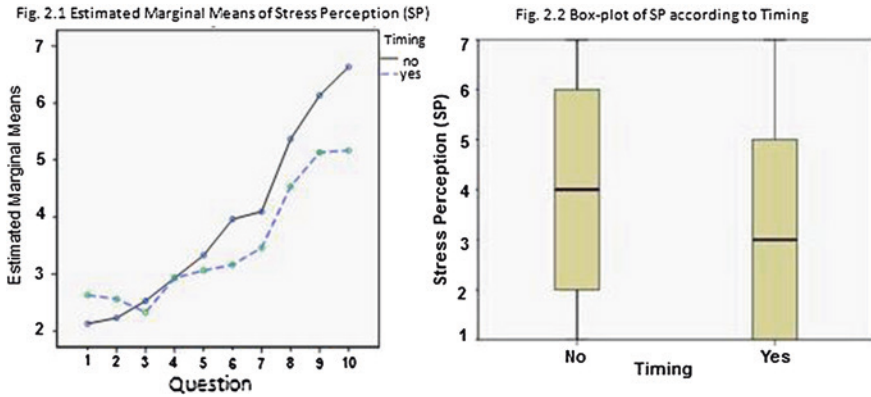


Fig. 2 SP increased according to Demand. The differences between questions are significant ($p < 0.5e^{-46}$). The Timing effect on SP is also significant ($p = 0.0025$)

small departures from homogeneity and the assumption of normality [22], TD, MID and KS are transformed using arctangent function and MCL is transformed using square root function. The following subsections discuss the results of the 3 research questions as shown in Sect. 3.

5.1 The Effects of Task Demand and Time Pressure on Cognitive Stress

After performing necessary data transformation, we first test the main effects of task demand (Demand) and time pressure (Timing) on SP by using Analysis of Variance (ANOVA) [23]. Both effects are significant (see Fig. 2). However, there is no interaction

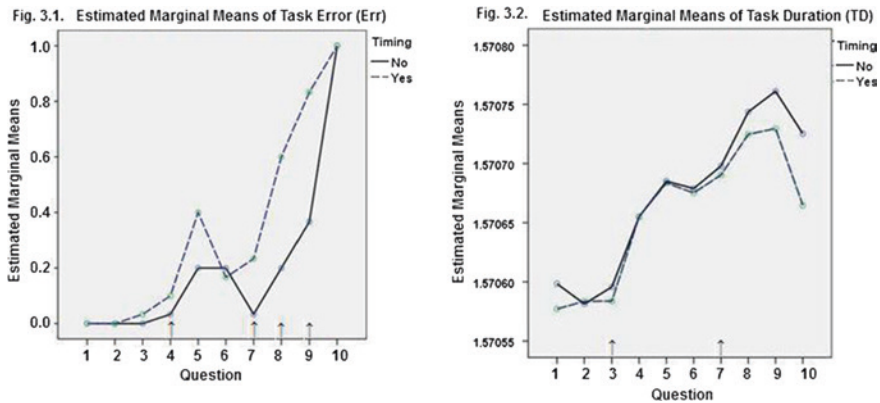


Fig. 3 Mean plots of task performance features

effect between Demand and Timing on *SP*. It is interesting to note that the participants in the control group (who are not given time pressure) in fact perceive higher stress than those in the experimental group. This could be due to some uncontrollable external environmental factors, such as tiredness after attending many classes on the same day, or intrinsic causal factors such as there are possibly more students having math anxiety in the control group than the experimental group. To examine the relationship between Demand, Timing and *SP*, Pearson correlation coefficient test shows that their correlation is significant. The significance of correlation between Demand and *SP* is given as $p = 0.02e^{-48}$, where $r = 0.56$; while the correlation significance between Timing and *SP* is given as $p = 0.0140$, where $r = -0.10$.

5.2 The Effects of Task Demand on Task Performance, Mouse Behaviour and Keystroke Behaviour

Demand significantly changes all the behaviours. Timing affects all features except *MID*. The interaction effect of Demand and Timing is only significant for *Err*, *MID*, *KS* and *KL* (see Table 2). We then performed Tukey Post Hoc Tests to analyse the variations between Demand and Timing effects to the three behaviours. The results are illustrated in Figs. 3, 4 and 5. The arrow markers in each graph indicate the significant difference between classes.

To observe the effects of Demand and Timing on *B(T)*, we first observe the number of users who attempted to give up or could not finish the questions on time (*PA*). Table 3 shows that *PA* started to increase from Question 6 onwards. However at Question 10, *PA* dropped instead of continue to rise although the users scored even higher *SP* at Question 10. This phenomenon shows that there is an anomalous behaviour of *PA* at Question 10. We then observe the number of students who made errors in answering the questions. Figure 3 shows that the students who are not given time pressure are generally making less errors than those who are given 30 s limit. The number of students who made error started to increase from Question

Table 2 MANOVA tests of the between-subjects effects

Demand			Timing			Demand*Timing		
	Feature	<i>p</i> -value		Feature	<i>p</i> -value		Feature	<i>p</i> -value
B(T)	Err	<i>0.0000</i>	B(T)	Err	<i>0.0000</i>	B(T)	Err	<i>0.0000</i>
	TD	<i>0.0000</i>		TD	<i>0.0041</i>		TD	0.2472
	SP	<i>0.0000</i>		SP	<i>0.0025</i>		SP	0.0565
B(M)	MS	<i>0.0000</i>	B(M)	MS	<i>0.0133</i>	B(M)	MS	0.9371
	MID	<i>0.0000</i>		MID	0.3201		MID	<i>0.0033</i>
	MCL	<i>0.0000</i>		MCL	<i>0.0000</i>		MCL	0.0934
B(K)	KS	<i>0.0000</i>	B(K)	KS	<i>0.0036</i>	B(K)	KS	<i>0.0010</i>
	KL	<i>0.0000</i>		KL	<i>0.0031</i>		KL	<i>0.0000</i>

Italicized cell indicates that the difference is significant at the level of $p < 0.05$

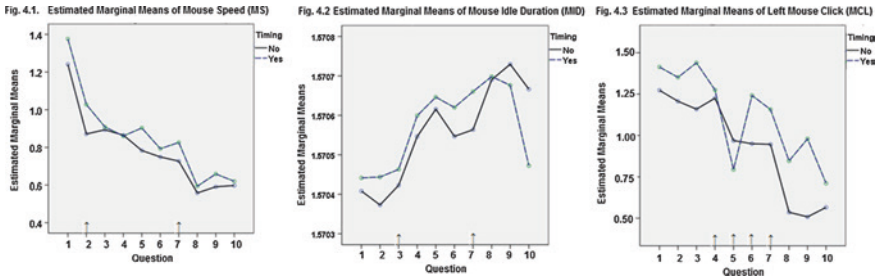


Fig. 4 Mean plots of mouse behaviour features. Generally the means of the experimental group are always higher than the control group, except MID (no significant difference)

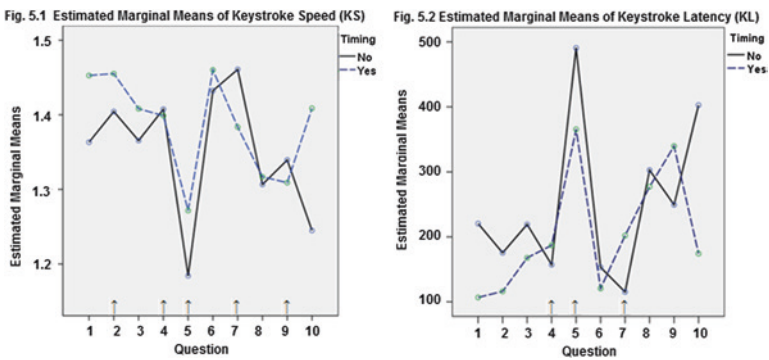


Fig. 5 Mean plot of keystroke behaviour feature. Generally the means of the experimental group are always higher than the control group

Table 3 Number of students who give up or could not finish task on time (PA)

Q.	Timing			Total	Q.	Timing			Total
	No	Yes				No	Yes		
	Give up	Auto submit	Give up			Give up	Auto submit	Give up	
1	0	0	0	0	6	1	0	1	2
2	0	0	0	0	7	0	1	0	1
3	0	0	0	0	8	2	6	1	9
4	0	0	0	0	9	2	14	1	17
5	0	0	0	0	10	5	5	1	11

3 onwards. Generally there are no significant differences among Question 1 to Question 4 in terms of *Err*. However, after Question 4, *Err* significantly increased at Question 5. More students made more mistakes for Question 5, 8 and 9 and all of them answered Question 10 wrongly. The fact that Question 5 achieves higher *Err* than Question 6 shows that Question 5 could be more difficult than Question 6, although they perceive Question 5 is less stressful than Question 6. In terms of

Timing, the users who were given time constraint made more mistakes than those without time pressure. For *TD*, there is no significant difference among Question 1 to Question 3, but *TD* increased significantly at Question 4. Then *TD* dropped slightly at Question 6, and gradually increased again until Question 7. The decrement of *TD* at Question 6 is consistent with the decrement of *Err* at the same question. This indicates that the users spent slightly longer time for Question 5 than Question 6. Again this shows that Question 5 could be more difficult than Question 6. Then there is a significant increment of *TD* at Question 8, but it started to drop at Question 10 instead of continues to rise. This is related to the anomaly at Question 10 as we observed in the behaviour of *PA*. In terms of Timing, those users who are given time constraint completed the task in shorter duration.

To analyze the changes of $B(M)$ according to Demand and Timing, first we examine the effects on *MS*. Figure 4 shows that the highest *MS* falls on Question 1, which indicates that less stressful task would introduce higher mouse speed. There is a gradual decrement of *MS* from Question 3 to Question 7, followed by a significant decrement at Question 8, signify increment of stress perception would lead to lower mouse speed. However after Question 8, there is a slight increase of *MS* from Question 9 onwards. This phenomenon again indicates that the users have again behaved anomalously from Question 9 onwards. With reference to Fig. 4, the changes of *MID* is similar to *TD*, which it gradually increased according to the level of Demand, and it consists of 2 changeover points at Question 3 and 7. It also demonstrates anomalous pattern at Question 9 and 10 (which it started to decrease instead of rise), and a slight increment at Question 5. Figure 4 also shows that generally *MCL* decreased when stress level increased. However, it is notable that there is a drastic decrement of *MCL* at Question 5 and Question 8, but the pattern resumed to its normal behaviour in the subsequent questions. This also indicates that something has changed the pattern of *MCL* significantly at these 2 points. Similar to *MCL*, *KS* and *KL* of $B(K)$ also demonstrate drastic changes at Question 5 and Question 8 as shown in Fig. 5 before they resumed to normal behaviour. Finally, all *MCL*, *KL* and *KS* also demonstrate incongruities between 2 groups of students at Question 10. In terms of Timing, those students in the experimental group demonstrate faster *KS* and lower *KL*, which suggest to us that if the students perceive lower stress, they would demonstrate higher *KS* but lower *KL* in general.

To explain the anomalies occurred, we review the complexity of the questions as shown in Table 1. Question 5 and Question 8 are the starting point of the increment of the digit per number in the arithmetic problems, which require more working memory to be recruited to store the information for further processing. Therefore we could predict that a change of question style, such as bigger numbers used in mental arithmetic, would lead to a temporal anomalous *MCL* pattern, i.e. a significant drop of *MCL* and *KS*, and increase of *KL* at one point. The information obtained from *MCL*, *KL* and *KS* could provide a more accurate measurement of cognitive load than subjective method, to inform the possibility that the question is more challenging than expected. To explain the anomalies happened at Question 10, which achieves the highest *Err* rate and *SP*, we strongly believe that

Table 4 Univariate tests for the effects of demand and timing on *SP*, *B(T)*, *B(M)* and *B(K)*

Factor	Significance, <i>p</i> -value	Wilk’s Lambda, λ
Demand	0.0000	0.1696
Timing	0.0000	0.8561
Demand*Timing	0.0000	0.7654

All factorial effects are significant at the level of $p < 0.5e^{-6}$

the students have reached an ultimate stress point at Question 9 (although it could be also affected by external factors such as fatigue and tiredness), which exceeds their endurance limit and makes them losing motivation to continue the task. As such, besides predicting *SP*, we are able to predict that the students may have experienced the need to cope with the change of question style at a specific point, and the point where they have lost motivation, by observing *B(T)*, *B(M)* and *B(K)*.

Lastly, we performed Multivariate Analysis of Variance (MANOVA) [22] to verify the effects of Demand and Timing on *SP*, *B(T)*, *B(M)* and *B(K)*. Table 4 shows the degree of confidence of the two factors’ effects. Although both Demand and Timing give significant impacts to *B(T)*, *B(M)* and *B(K)*, however the Wilk’s Lambda values for Timing and its interaction with Demand are high. This shows that the between-groups dispersions of Timing and its interaction with Demand are small. In other words, Demand is the main effect that affects stress perception and all three behaviours, but the impact of Timing is small.

5.3 Correlations Between Task Demand, Cognitive Stress, Task Performance, Mouse Behaviour and Keystroke Behaviour

To examine the relationships between Demand, *SP*, *B(T)*, *B(M)* and *B(K)*, we conducted Pearson Correlation Coefficient tests. The results are shown in Table 5. The highlighted cells indicate negative correlation coefficient. When Demand

Table 5 Correlation between features

	Demand	SP	TD	Err	MS	MID	MCL	KS	KL
Demand		✓	✓	✓	✓	✓	✓	✓	✓
SP	0.0000		✓	✓	✓	✓	✓	✓	✓
TD	0.0000	0.0000		✓	✓	✓	✓	✓	✓
Err	0.0000	0.0000	0.0000		✓	✓	✓	✓	✓
MS	0.0000	0.0000	0.0000	0.0000			✓	✓	✓
MID	0.0000	0.0000	0.0000	0.0000			✓	✓	✓
MCL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0023		✓	✓
KS	0.0000	0.0002	0.0000	0.0000		0.0000	0.0000		✓
KL	0.0000	0.0000	0.0000	0.0000	0.0314	0.0000	0.0000	0.0000	

Significant correlation exists between two features at $p < 0.05$ (2-tailed) level if it is ticked (✓). Highlighted cell indicates negative correlation coefficient.

increased, *TD*, *SP*, *MID* and *KL* also increased, but *MS*, *MCL* and *KS* decreased. Therefore we could predict that if *TD*, *Err*, *MID* and *KL* increased but *MS*, *MCL* and *KS* decreased, then *SP* should increase.

6 Discussions

From the statistical analyses, task demand is the main factor that influences student's stress perception, task performance, mouse and keystroke behaviours. Although time pressure effect is also significant, however its impacts on the changes of stress perception and behaviours are relatively small. There is also an interaction effect between task demand and time pressure, but the combined effect of these 2 factors is also small. Correlation tests results suggest that prediction of cognitive stress increment is possible. When task demand increased, students' stress perceptions, duration spent to complete a task, error rate, passive attempt, mouse idle duration and key latency increased; while on flip side, mouse speed, mouse click rate and keystroke speed decreased. When task difficulty increased, but task performance, mouse and keystroke behaviours do not behave in a way that is expected, then anomaly can be detected. Anomalous behaviours indicate three possibilities: (i) there is either a wrong assumption about the independent factors (e.g. Question 5 appeared to be more challenging than Question 6 although it consists of fewer terms and operators); (ii) qualitative difference in task demands (e.g. the number of digits per number in the task increased would require more working memory to process the task), which can be observed through *MCL*, *KS* and *KL*; or (iii) the user is either understress or overstress, which is beyond their motivation limits (e.g. Question 10 contributes highest error rate and stress perception). After this ultimate stress point, prediction of *SP* using the production rules could probability become invalid, as the users have lost motivation to continue the task. Therefore it is important to activate the adaptive content to motivate the students to continue the task.

7 Conclusion

Our research shows that an automated evaluation of cognitive stress can be obtained through acquisition and processing of task performance, mouse behaviour and keystroke behaviour. When typing task demand increased, task error, task duration, passive attempt, stress perception and mouse idle duration may increase, while mouse speed, left mouse click rate and keystroke speed decreased. This is consistent with the findings by Lim et al. [13], which they found that when the users feel uncomfortable with the bad setting of menu design, generally their mouse idle duration and mouse idle occurrences would increase, but mouse speed and left mouse click rate would drop, although there is no significant correlation

between users' stress perception and search task performance. This research findings also show that task demand is the main factor that affects all three behaviours. The correlations between mouse behaviour and keystroke behaviour suggest that unifying mouse and keyboard dynamics analyses could be more useful than utilizing them separately. Anomalies of mouse and keystroke behaviours, such as mouse click, key latency, and keystroke speed, could be also be observed where there is a change of question style or the students may have lost motivation. Adaptive system can then be activated to motivate the students to continue the task. However our research has a few limitations. The results such as time pressure effects may be influenced by external environmental factors, such as external time pressure, and the attitude and motivation of the participants. It is also difficult to ensure all students are having comparable mental arithmetic skills. Therefore homogeneity among all students is not guaranteed. We also excluded mouse movement direction in the analyses, which can be an important variable that can be affected by cognitive stress. Besides, the sample size is small, and we may not be able to generalize our findings to represent the actual population. More rigorous experiments need to be conducted to validate the stress evaluation model.

References

1. Beilock, S.L., Ramirez, G.: On the interplay of emotion and cognitive control: implications for enhancing academic achievement. *Psychol. Learn. Motiv. Adv. Res. Theory* **55**, 137 (2011)
2. Paas, F.G.W.C., Van Merriënboer, J.J.G.: Instructional control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* **6**, 351–371 (1994)
3. Kirschner, P.A.: Cognitive load theory: implications of cognitive load theory on the design of learning. *Learn. Instr.* **12**, 1–10 (2002)
4. Chatzara, K., Karagiannidis, C., Stamatis, D.: Students attitude and learning effectiveness of emotional agents. 2010 IEEE 10th international conference on advanced learning technologies (ICALT), pp. 558–559 (2010)
5. Selye, H.: *The Stress in Life*. McGraw-Hill, New Jersey (1956)
6. Setz, C., Amrich, B., Schumm, J., La Marca, R., Troster, G., Ehlert, U.: Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans. Inf. Technol. Biomed.* **14**, 410–417 (2010)
7. Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E.: N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* **25**, 46–59 (2005)
8. Sloan, R.P., Korten, J.B., Myers, M.M.: Components of heart rate reactivity during mental arithmetic with and without speaking. *Physiol. Behav.* **50**, 1039–1045 (1991)
9. Imbo, I., Vandierendonck, A.: Do multiplication and division strategies rely on executive and phonological working memory resources? *Mem. Cognit.* **35**, 1759–1771 (2007)
10. Sweller, J., Ayres, P., Kalyuga, S.: *Cognitive Load Theory*. Springer, Berlin (2011)
11. Weinberg, A., Ferri, J., Hajcak, G.: Interactions between attention and emotion. *Handb. Cogn. Emot.* **35** (2013)
12. Hajcak, G., Dunning, J.P., Foti, D.: Neural response to emotional pictures is unaffected by concurrent task difficulty: an event-related potential study. *Behav. Neurosci.* **121**, 1156 (2007)

13. Lim, Y.M., Ayesh, A., Stacey, M.: The effects of menu design on users' emotions, search performance and mouse behaviour. In: Patel, S., Wang, Y., Kinsner, W., Patel, D., Fariello, G., and Zadeh, L.A. (eds.) IEEE 13th international conference on cognitive informatics & cognitive computing, pp. 541–549. IEEE, London (2014)
14. Tsoulouhas, G., Georgiou, D., Karakos, A.: Detection of learner's™ affective state based on mouse movements. *J. Comput.* **3**, 9–18 (2011)
15. Pusara, M., Brodley, C.E.: User Re-authentication via Mouse Movements. In: Proceedings of the 2004 ACM workshop on visualization and data mining for computer security, pp. 1–8, ACM, New York (2004)
16. Shen, C., Cai, Z., Guan, X., Du, Y., Maxion, R.: User authentication through mouse dynamics. *IEEE Trans. Inf. Forensics Secur.* **8**(1), 16–30 (2012)
17. Shen, C., Cai, Z., Guan, X.: Continuous authentication for mouse dynamics: a pattern-growth approach. *Dependable systems and networks (DSN)*. 42nd annual IEEE/IFIP international conference on 2012, pp. 1–12 (2012)
18. Shen, C., Cai, Z., Guan, X., Sha, H., Du, J.: Feature Analysis of Mouse Dynamics in Identity Authentication and Monitoring. In: Proceedings of the 2009 IEEE international conference on communications, pp. 673–677, IEEE Press, Piscataway (2009)
19. Vizer, L.M.: Detecting Cognitive and Physical Stress Through Typing Behavior. In: Proceedings of the 27th international conference extended abstracts on human factors in computing systems CHI EA 09. 3113 (2009)
20. Filho, J.R.M., Freire, E.O.: On the equalization of keystroke timing histograms. *Pattern Recogn. Lett.* **27**, 1440–1446 (2006)
21. Cavoukian, A.: Privacy by design: origins, meaning, and prospects for assuring privacy and trust in the information era. In: Yee, G. (ed.) *Privacy protection measures and technologies in business organizations: aspects and standards*, pp. 170–208. IGI Global, Hershey (2012)
22. IBM: multivariate general linear modeling, http://pic.dhe.ibm.com/infocenter/spsstat/v21r0m0/index.jsp?topic=/com.ibm.spss.statistics.cs/glmm_intro.htm
23. IBM: GLM univariate analysis, http://pic.dhe.ibm.com/infocenter/spsstat/v20r0m0/index.jsp?topic=/com.ibm.spss.statistics.help/idh_glm.htm