COG Tool: An Automated Cognitive Measurement of Workload for Mobile Event Logging

https://doi.org/10.3991/ijim.v16i20.33363

Ashraf Mousa Saleh[1](✉), Odai Enaizan[2]
[1] Amman Arab University, Amman, Jordan
[2] University of Tabuk, Tabuk, Saudi Arabia
Asaleh@aau.edu.jo

Abstract—Cognitive Workload (CL) has been defined as the cognitive processing amount needed by a user whilst using an application. CL can be seen as being a primary task-based usability evaluation attribute in relation to mobile applications. Typically, subjective measurements of the CL are taken at point following completion of a task and so they are static. Since these methods are static in nature, they are inappropriate for assessment of dynamic changes to the cognitive loads through user interaction. Robust measurement pertaining to the CL of a user is not a simple and trivial undertaking within real-time. Several research studies aimed at assessing the CL of users through the use of various subjective measurements, though it is often the case that those approaches are unsuitable for use with real-life applications due to the high degree of intrusiveness. Automated measurement is seen as being a field that is relatively unexplored, particularly, if used in relation to the CL for mobile applications. The COG tool (automated tool) is proposed in this paper in order to support the interaction logging of the user for derivation of different CL metrics that are based on secondary and primary tasks. Several strengths are provided by the automated measurement of the CL, which fills the gaps of a static nature of its non-objective and automated (subjective) counterpart. In this paper, -objective CL metrics have been identified and used effectively. A total of 68 participants was recruited for the experiment based on the use of an AAU mobile application in relation to the performance of a task that is predefined without any interference of a moderator or researcher. It is found to be proven from the proposed tool that is able to fulfill the expected cognitive load monitoring functions amongst user interaction, collection of cognitive metrics and the logging of different CL metrics that could be analysed further.

Keywords—cognitive load theory, cognitive workload, mobile application, automated tool

1 Introduction

Cognitive Workload (CL) became a component that is considered important because of the observation that mobile device users are often multitasking whilst they use the application; see, for example, [1] and [2]. That stood against the common assumption of different conventional usability studies by which users only perform a single task with the ability of focusing their cognitive resources and attention exclusively on completing this task. Alternatively, a powerful approach that is multi-factorial may be employed [3]. That component stresses the importance of mentioning the effect the use of the mobile device has on performance of additional and simultaneously performed tasks. A user might, for instance, be walking or moving around whilst sending a text message. In this case, the walking speed of the user may be slowed due to them being concentrating upon sending the message. Based in this example, a lower walking speed is considered a mild kind of negative impact [4]. In certain cases, however, such as driving for instance, mobile use can have severely adverse impacts on parallel tasks. The authors argue, therefore, that there might be a lowering of the mobile use impact in order to minimise the adverse impact of mobile use on the primary tasks of the users.

2 Literature search

Cognitive work load was included in PACMAD model of usability the primary contribution for investigating usability [3]. In the context of mobiles, the utilisation of the application may be affected by the capability of users to operate the mobile application and move around. As noted by [5], the mobility of users and the application utilisation of users have to be taken into consideration whilst studying the usability of a mobile. The CL is identified as an attribute of the usability model, PACMAD, without the mentioning of the low-level, related metrics which are representative of each of the attributes. Consequently, it is a requirement that there is extension for low-level, relative metrics to be included as well as attributes of usability. Furthermore, the evaluator subjectively measures the CL once the secondary and primary tasks, that have been assigned to the users within the environment of the mobile application have been performed [6]. In particular, subjective types of measurement are taken at a single point once a task has been completed and so they are static. Since these methods are static in nature, they are not appropriate for assessment of the dynamic changes within CL through the interaction of users. Based upon cognitive load theory, there are two load cases [7]. The first of the cases mainly deals with extraneous load, i.e., demands that the user interface have imposed. The second case can mainly deal with intrinsic load, i.e. the demands that the task is placed through the cognitive resources of a user. The relationship between cognitive demands hailing from a particular task and working memory capacity are described by cognitive load theory [8].

2.1 Measurement of cognitive load

The relationship between cognitive demands hailing from a particular task and working memory capacity, - is described by the CLT (cognitive load theory) [8]. It is founded upon the idea that the cognitive capacity within a working memory has limits, and if there is so much capacity required in a mental task, there will be hindering of reasoning and knowledge acquisition [9]. The CL may be seen as being a theoretical type of construct that describes internal kinds of information processes that may not be directly observed through user interaction. Various CL assessment methods are currently available. However, this may be categorised into two approaches, which comprise; objective or subjective and the linkage relation (indirect or direct) approaches [10] (see Table 1). The relation dimension linkage categorises these approaches according to the sort of relation of the phenomenon in which the measure observes and the attribute that is actually of interest. The extent of the objectivity demonstrates whether the method employs self-reported data, objective behavior observations or subjective rating, performance or user interaction conditions.

Table 1. Assessment (indirect and direct)

	Indirect	Direct
Subjective	Mental effort (self-reported)	Stress level (self-reported)
		Difficulty (self-reported)
Objective	Physiological measures	Brain activity measures
	Behavioral measures	Dual-task performance

As can be seen in Table 1, there is the existence of a clear, direct linkage between the CL and its difficulty associated with the tasks that are being achieved, since that difficulty results directly from the extraneous and intrinsic task load. Additionally, there is the existence of a notable, indirect link between navigational errors and the CL, resulting in the potential for frequent errors to be caused through a mental model for the task environment that is incomplete, and that, in itself, can be due to a cognitive load that is high [11]. Methods of CL assessments that involve performance upon secondary tasks are known as dual-task techniques [12].

2.2 Dual task techniques

Dual Task Techniques (DTT) may be used within two differing approaches. The first approach involves the addition of a secondary task to a primary one so that memory load is induced [13]; [14]; [15]. The secondary cognitive load metrics relates to the primary task performance, which ought to decrease with dual-task conditions when compared with single task conditions, i.e., when there the primary task is solely made available. The second approach involves the use of a secondary task for measuring the memory load that a primary task induces [13]; [14]; [15]. If different memory load amounts are induced by different primary task variants, there will be a corresponding variance in the secondary task performance. Dual task types of technique enable on-task, real-time data collection and are low cost. Instantaneous load is measured by the DDT [16]; which is formally used for calculating the average values during the performance of a task.

2.3 Cognitive load metrics

As mentioned previously, techniques of cognitive load assessments can be divided into physiological, performance and subjective measures. These measurements are included within the performance measures are performance upon primary tasks and performance related to secondary tasks, which are considered to be objective metrics [17]. The performance measure is identified by [18] as a technique of the CL assessment (accuracy, the task completion time, the number of errors during the task, and the ratio of actual task completion time to ideal task completion time). According to a study undertaken by [19], it is stated that metrics for usability evaluation include performance measurements for task success, number of clicks and task completion time, including subjective metrics for satisfaction [20]. Similarly, based on a study conducted by [20], a task success is used in measuring whether tasks that could be completed successfully by users. The time that elapsed from when the task had started and completion is recorded in seconds, and is considered a system efficiency measure [21]. The assessment of efficiency is also conducted based on the number of clicks that users perform upon the website whilst completing a particular task. The count for number of clicks refers to the time duration from the start of each task along to the associated completion [22].

The CL relates to the mental effort that the user needs in order to perform a task through use of the computer system. If there is a high cognitive workload then there is a likeliness that the user experiences a degree of pressure and stress in the use of the specific user interface in question [23]. If the degree of pressure and stress is at an unacceptably high level, then there needs to be improvement to the user interface with respect to the particular troublesome aspect. The objective cognitive load measurement is also used with metrics including the number of times. Windows Help needs to be accessed by the user, the time spent upon help and/or the user manual and the number of unmatched task executions within the real world [21]. It is concluded from other studies that the reaction time, i.e., that time interval between the request of a task and the response of the subject, can be considered with a response delay leading to a decreased rate of correct response [24]; [25]. The literature has shown that measures for performance can be undertaken through low-level metrics, i.e. performance metrics including task success, task completion time and number of touches [19]. Table 2 illustrates that the CLM yields from the literature with their representative references.

Table 2. Metric of cognitive load based upon DDT

CLM-DTT	Representative references
Task success rate (TS)	[19];[26];[27]
Number of errors (NOE)	[28];[18];[1]
Task time (TT)	[29];[19]; [30]
Number of touches (NOT)	[29];[28]; [31]; [32]
Duration on help (DOH)	[33];[34]; [35]
Help visit count (HA)	[33];[36]; [37]
Total effort (EF)	[39];[40];[39]
Reaction time (RT)	[19];[41];[42];[43];[44]

Based upon a review of measurements of cognitive load and the objective, low-level metrics, the proposed metrics for cognitive load are shown in Table 3. This constitutes proposed logging of metrics of cognitive load (CLOG), which may be suitable for measurement in a way that is automated through the logging of a user's interaction.

Table 3. Metrics of cognitive workload

CLM	Equations for metrics
TS	$\int_{t0}^{tn} (\sum_{1}^{n} success_{task} + (\sum_{1}^{n} partial_{success} * 0.5))/(\sum tasks * \sum users)$
NOE	$\int_{t0}^{t1} \sum E\ (Task_{\text{ Endtime}} - Task_{\text{ Starttime}})$
TT	$\int_{Task\ 1}^{Task\ n} (Task_{\text{ Endtime}} - Task_{\text{ Starttime}})$
NOT	$\int_{Task1}^{Task\ n} \sum TH\ (Task_{\text{ Endtime}} - Task_{\text{ Starttime}})$
DOH	$\int_{u1}^{un} \sum Timestamp_{\ end-start}$
HVC	$\text{HVC(HAP)} = \sum_{SS=0}^{SE} \text{NoV}$
TF	$1/(t - t0) \int_{t0}^{t} (w1\ X\ tch(t) + w2\ X\ p(t))$
RT	$TaskTime_i(Secondary_{response} - Secondary_{request})$

Task success rate. The definition for task success utilises the success task sum during experimentation as defined in the work of [19]; [26] and [27].

The user task success rate {1... n} is defined here can be seen in Formula 1:

$$(\sum_1^n success_{task} + (\sum_1^n partial_{success} * 0.5))/(\sum tasks * \sum users) \tag{1}$$

Where $success_{task}$ relates to number of successful and completed tasks that the user does during time t0-tn. Furthermore, $partial_{success}$ is denoting for tasks that are partially successful during the time interval.

The number of errors. The measurement for the Number of Errors (NOE) is performed through the calculation of the total for various types of errors that happen during task (T), whilst users (U) that interact with the mobile application, through the computation of error numbers during the task time can be calculated through Formula 2:

U={user1...usern}

$$NOE(U.T) = \int_{t0}^{t1} \sum E\ (Task\ \text{Endtime} - Task\ \text{Starttime}) \tag{2}$$

Where E relates to the total number of errors that occurred or was encountered by the user in task time t0-tn. Moreover, the task duration is returned by $Task\ Endtime - Task\ Starttime$.

The number of touches. The measurement of the Number of Touches (NOT) is performed through the calculation of the touch event number upon the mobile gesture whilst the task is being conducted. NOT is defined by for each of the users, U= {user1...usern}, can be seen in Formula 3:

$$NOT(U) = \int_{Task1}^{Task\ n} \sum TH\ (Task \text{ Endtime} - Task \text{ Starttime}) \tag{3}$$

Where TH denotes the number of the motion event pertaining to the pointer over the mobile gesture whilst tasks {Task1…Taskn} are being undertaken by each of the users (U)= {user1...usern} who participate within the experiment.

The duration on help. The measurement of Duration on Help (DOH) is conducted through the calculation of the time spent by users upon the assistance of a mobile application in obtaining a solution to a problem related to their mobile application interaction or obtaining an answer for queries. The calculations for the DOH for each of the users, U= {user1...usern}, that participate within the task experiment (T) are conducted as can be seen in Formula 4:

$$DOH(T) = \int_{u1}^{un} \sum TS\ _{end-start} \tag{4}$$

Where TS denotes the timestamp for calculating the time spent in the mobile application assistance and help page.

Help visit count. The Help Visit Count (HVC) refers to the total number of page-views of the mobile application. For counting of the page-views, a filter is used that intercepts all of the requests that arrive at the help page of the mobile application during a particular time from SS (the session start of the user) and SE (the end of session), with each request increasing the number for page-view by 1. Consequently, this can be achieved when a session-listener is used between the SS (session-creation time) and SE (session-destroy) can be seen in Formula 5:

$$HVC(HAP.U) = \sum_{SS}^{SE} NoV \tag{5}$$

Where NoV denotes the number of visits whilst task (T) is undertaken for each of the users ui within the user set {u1…un}.

Total effort. The TF (total mental effort) is determined based on the use of the continual functions as defined by [40] i.e., the physical activity that a person performs in their attempt at accomplishing a particular goal. In practical terms, there is a quantisation of those metrics through the conversion of the integrals into sums. With the assumption that a task T has an interactive experiment starting at time $t0$, the total effort spent at time t is defined can be seen in Formula 6:

$$TF(ti.ui) = \ 1/(t - t0) \int_{t0}^{t} \big(w1 \text{ X } tch(t) + w2 \text{ X } pd(t)\big) \tag{6}$$

Where tch(t) equates to the number of touches in which user ui has produced during the interval of the task time $t{-}t0$. Further, pd(t) equates to the penalty distance factor by measuring the distance among touches that the user has traversed whilst moving through the touch calculated gesture, with the exclusion of the horizontal scrolling, from point $x0$,$y0$ along to point $x1$,$y1$. At this point, the assumption represents $w1$ that equates to $w2$ = 1.

Reaction time. The Reaction Time (RT) is in reference to the amount of time that a user takes in responding to a particular stimulus whilst using a mobile application [45]. Thus, a variable is created which holds up a time when a user has an event popping up to them and the time that a touch reaction is shown by the user to that popped event, with the time difference that lies between those two actions then calculated. For the satisfaction of RT, OnTouchListener () is used so that the mobile application is enabled to respond to the touch of the user. The RT is defined in the manner of the time delay of the user, wgich responds to the particular stimulus can be seen in Formula 7:

$$RT = TaskTime_i(Secondary_{response} - Secondary_{request}) \tag{7}$$

Where TaskTimei is determined by the subtraction method, which is identified by [46] for the calculation of time between the occurring time for the event stimuli (secondary request) and the time of responding to the stimuli (the secondary respond).

2.4 COG tool

This framework of cognitive assessment serves to automatically record interactions between user and target mobile application; it is able to detect more of the objective metrics of usability, though since this study scope is limited, solely cognitive load measurement is offered here. There is an unobtrusive collection of data without the working style of the user being impacted. Additionally, the evaluator or moderator is not able to impact through the participants' interactions. The process mainly comprises time-stamped logs of user interactions and responses to the application. The objective data by which the framework is collected from the mobile application that is used within the study is to have a logged data form with a file for log data containing measurements and metrics of cognitive load as identified within Tables 2 and 3. The log file is made to have a sequential file form with each of the lines representing tracing of the numerical data for user interactions for each of the particular participants, based upon the specific given task. The framework for the COG tool is shown in Figure 1.

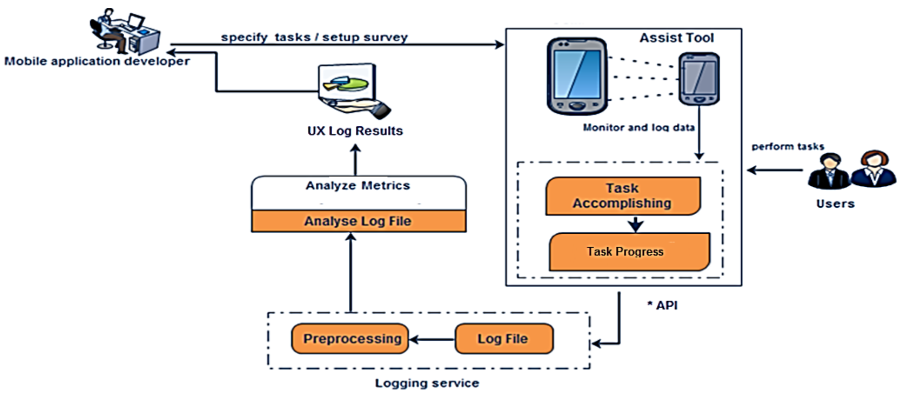

Fig. 1. The Framework of the COG tool

The logging of cognitive metrics. Within this particular phase, there is a preparation for the mobile cognitive tool so that it is enabled to log the data required for the evaluation of usability; the mobile application needs to be set up so that the necessary information can be logged for the assessment of the cognitive load. Because mobile platforms have heterogeneous, there is difficulty in finding a general API (application programming interface) set that could be utilised in the performance of the logging tasks. The development of Android applications requires a modification to the source code through totaling up of API calls, which represents a recompilation of the source code by using the SDK (a software development kit). The API has the function of logging the interaction events of the user.

The design and preparation of the logged data. During the capture and design phase, the necessary code should be added within the source code of the mobile application so that measurement data logging is enabled. This assures to simplify the task's performance where an API set that is provided for Android listeners can efficiently permit measuring the data so that it could be easily logged. Following that, the collection of the metrics for the CL is achieved and the logged data is made ready for being transferred to the logged metrics of the extraction phase.

3 The implementation of the proposed framework

The proposed framework for the COG tool contains common activities that are applied within the AAU (host mobile application). Those common activities can be described in ways based on the use of the host application and its associated evaluation framework. The AAU refers to the Amman Arab University mobile application that is Android-based, and which is under examination for the Amman Arab University represents a development of the tool so that current staff and students of the university can be provided with useful information. The COG tool instrumentation code has the responsibility of being recognised when new tasks have commenced or when there is completion of a particular task that under progress. Whenever the user has launched or terminated the mobile application, it is detected by the tool. Further, the activity in which a user interacts with is detected by the tool. In a length of time, a user remains in an activity as represented by the time duration calculated in seconds and the page number. Since each of the metrics of the cognitive load is detected by the COG tool as noted above, it is able to collect and save logged data within the internal storage of the mobile. A sample code instrumentation procedure is shown below in Table 4.

Table 4. Samples of code instrumentation

API	Description	Instrumentation sample

User timing	The intent object passes the application information into the user interface assessment framework. This is handled by the OnClick() event handler method in their xml file.	Async function run () {
		  performance.mark("startTask1");
		  await doTask1(); // Some developer code
		  performance.mark("endTask1");
		
		  performance.mark("startTask2");
		  await doTask2(); // Some developer code
		  performance.mark("endTask2");
		
		  // Log them out
		  const entries = performance.getEntriesByType("mark");
		  for (const entry of entries) {
		    console.table(entry.toJSON());
		  }
		}
		run();

Resource timing	Allows JavaScript mechanisms to capture all timing information for mobile resources that are navigated.	function resourceTiming()
		  {
		    var resourceList = window.performance.getEntriesByType("resource");
		    for (i = 0; i < resourceList.length; i++)
		    {
		      if (resourceList[i].initiatorType == "img")
		      {
		        alert("End to end resource fetch: " +
		(resourceList[i].responseEnd - resourceList[i].startTime));
		      }
		    }
		  }

Performance listener	Developing asynchronous callbacks within Android development. Listeners to implement the code to run when a performance event occurs	element.addEventListener("click", e => {
		    const component = getComponent(element);
		    fetch(component.url).then(() => {
		      element.textContent = "Updated";
		      const updateMark = performance.mark("update_component", {
		        detail: {component: component.name},
		      });
		      performance.measure("click_to_update_component", {
		        detail: {component: component.name},
		        start: e.timeStamp,
		        end: updateMark.startTime,
		      });
		    });
		  });

Performance interface	Developing an interface to the performance timeline which extends start time and durations.	[Exposed=Window]
		interface *PerformanceLongTaskTiming* : PerformanceEntry {
		  readonly attribute FrozenArray<TaskAttributionTiming>
		attribution;
		  [Default] object *toJSON*();
		};

Task timing interface	To plan tasks for later executions in a thread, create background threads. Tasks can be run once at random or at regular intervals.	`[Exposed=Window]`
		`interface TaskAttributionTiming : PerformanceEntry {`
		`    readonly attribute DOMString containerType;`
		`    readonly attribute DOMString containerSrc;`
		`    readonly attribute DOMString containerId;`
		`    readonly attribute DOMString containerName;`
		`    [Default] object toJSON ();`
		`};`

4 Empirical study

4.1 Participants and design

Within this research, a quantitative methodology has been applied within the stage of data collection. Since this is a research study that is experimental in nature, observations are undertaken during the UX test so that data can be collected for the measurement of cognitive load metrics. The automated tool of cognitive workload measurement is hosted within the AAU's mobile application (student mobile application) that has been particularly developed for this research as a context that is executable for the automated tool of cognitive load. This execution is performed to provide information regarding e-learning, including chatting through course groups, search of courses, university news and study plans, in addition to other features that lecturers and students that are made available to them. The analysis of the data aims at discovering the cognitive workload metrics in which the automated tool is measured from the mobile application of students. Before implementing the cognitive evaluation, a development of the testing scenario is conducted in order to fit the metrics that are identified within the review of the literature. The participants have to log in, however, by using his (her) personal ID for their university email; as such, there is no requirement for registration. During the process of the experiment, the request will be made available to participants so that they can appropriately achieve the entire tasks that are given within a particular scenario. Moreover, the interaction of users will be monitored by the automated tool by the logging of participants in practice and relevant analysis. The participants were invited by WhatsApp groups to participate in the experiment by downloading the AAU mobile app from Google Play or the App Store and they were instructed on the required steps to performs the experiment tasks. The task instructions were as follow: participants have to log in, however, using her or his ID for their university email; as such, there is no requirement for registration. During the experiment, participants have been requested to achieve all of the tasks that are given within the scenario appropriately. The experiment tasks involve Five tasks as follow: browse the current courses schedule, send a message to any enrolled course groups, view the academic schedule for the current semester of 2021-2022 as a PDF file, take screen shot for the latest training offered by university and respond to the notification message appeared during the experiment.

4.2 The experimental analysis, design and evaluation

There was a creation of five sub-tasks so that full interaction was allowed with the mobile application. Within the experiment test, participants were asked to achieve five tasks by the researcher. Firstly, the experiment starts with a brief description of the study by the experimenter, participants have to read a form for informed consent which the experimenter then collects from them. Secondly, the installation instruction is facilitated by the experimenter, and participants are asked to complete the form of registration on the application page. Then, the participants read instructions regarding the performance of the tasks. The participants had identical tasks, and following that practice run, an inquiry was made by the experimenter to ensure participants were comfortable and whether they had any questions. Following confirmation that the participants understood the task fully, they could then continue to start the experimental stage of the study. Lastly, once the tasks had been completed by participants, the researcher asked them to log out from the mobile application.

Within the experiment, the cognitive load was tested through examination of the sensitivity for each of the eight common cognitive metrics, i.e., TT, TS, NOE, DOH, NOT, TF, RT and HVC, to another secondary task. The participants with IOS or Android mobile phones available to them are able to undertake tasks without settings being adjusted further. A logging of each of the experimental sessions through the use of the automated COG tool is performed. The user interaction is recorded by the tool in a log file, and all of the metrics action is activated upon commencement of the task by the user. Once there is communication of the task to participants by the moderator, no constraints are placed for users within the session. Conditions are established, so that participants are provided within an environment that was conducive for evaluation of the test. There was creation of five sub-tasks so that full interaction was allowed with the mobile application. Within the experiment test, participants were asked to achieve five tasks by the researcher. Last, once the tasks are completed by participants, the researcher is likely to ask them to logout from their mobile application so that bias could be avoided, and the task given to the participants can be undertaken individually without any assistance to minise any effects based on the judgements and interaction of users. The performance of additional tasks by users whilst the mobile application was being used was permitted, such as walking. For that ground, it is significant to take into account the impact that simultaneous use of the mobile application possesses through the performance of users whilst functioning those extra tasks.

4.3 Materials and the procedure

The materials related to the tasks comprised a series of steps that had to be followed in the experiment. No type of subjective questionnaire is involved in measuring cognitive load for the task. The design of the task for the experiment managed to present participants with a complete scenario for registration in the mobile application through the student ID, the marks for the enrolled semesters for the inquiry student, view of the student plan analysis, the changing of his or her mobile number, the viewing of the academic calendar, the sending of a public message for the enrolled course group, and the reading of a notification received from the administrator for the dashboard of the mobile application. There is logging of the collected data into a log file following launch of the application by the participant. Instructions are given to participants for following the steps for the task given and for responding to any message encountered without having the experiment moderators provide any assistance. The reaction times of the students are recorded by the mobile application (the RTs, were recorded in milliseconds for later conversion into seconds), whilst the task time is measured in seconds. The recording and logging of other metrics was based upon given computation counters and computation formula as mentioned in Table 2.

5 Results

There is a total of 68 participants in the study (19 academic staff and 49 students). In particular, there are 30 females and 38 males within ages ranging from 18 years old and up to 54 years old. For each participant, cognitive load measurements begin to be taken once participants log onto the mobile application where measurements are taken for all eight metrics of the cognitive load, i.e., TT, TSR, NOT, NOE, HVC, RT, TF and DOH. The characteristic information for the participants are comprised as follows. There are 68 participants in total who are involved in the experiment (38 male participants and 30 female participants). Most of the participants are within the ages of 18-26. According to proficiency of using mobile devices, 14.7% (n=10) were advances users, whilst 51.5% (n=35) of participants were intermediate level users. It is found that most participants who use smart phone devices for a duration of two to six hours represent 83%. Participants' characteristics for the participants are summarised in Table 5.

Table 5. Participants' characteristics

Characteristics		Participants	N (%)
Gender	Male	38	55.9%
	Female	30	44.1%
Age	18-26 year	41	60.3%
	27-35 year	12	17.6%
	36-44 year	10	14.7%
	Above 44	5	7.4%
Proficiency level	Beginner	23	33.8%
	Intermediate	35	51.5%
	Advanced	10	14.7%
Daily duration use of smart phone devices	Less than 1 hour	0	0%
	1-2 hours	9	13. %2
	2-4 hours	27	39.7%
	4-6 hours	29	42.6%
	More than 6 hours	8	11.8%

Once the participants were well-informed over the scenario of the tasks and the steps to be taken, they were invited to undertake the particular tasks within the experiment. Participants were informed that their tasks would be completed on an individual basis without any assistance from the moderator or supervisor and without interference to ensure that the data for user interaction would not be biased. The results that were generated from the log file of the COG tool following the processing of the raw data of the experiment, and its analysis, are summarized within Table 6 below.

Table 6. Results of cognitive load measurement

N=68	(TS)	(TT)	(NOE)	(NOT)	(HVC)	(DOH)	(RT)	(TF)
Metric unit	Percentage	Seconds	Error counter	Touches counter	Visit counter	Seconds	Seconds	Percentage
Average	83.78%	176.94	1.08	38.51	1.29	30.81	4.34	23.53%
Min	-	72	0	19	0	0	1.15	11.41%
Max	-	353	3	74	4	91	9.45	47.13%
Around average of (N,%)	-	12 (17.6%)	9 (13.2%)	7 (10.3%)	26 (38.3%)	5 (7.3%)	12 (17.6%)	14 (20.6%)
Below average of (N,%)	-	26 (38.3%)	26 (38.3%)	20 (29.4%)	13 (19.1%)	25 (36.7%)	12 (17.6%)	19 (27.9%)
Above average of (N,%)	-	30 (44.1%)	33 (48.5%)	41 (60.3%)	29 (42.6%)	38 (56%)	44 (64.8%)	35 (51.5%)
Total of participants (N)	-	68 (100%)	68 (100%)	68 (100%)	68 (100%)	68 (100%)	68 (100%)	68 (100%)

Firstly, TS (task success rate) equates to the rate at which participants could successfully achieve tasks in accordance with a predefined navigation page through the mobile activities and pages. A total of 83.78% of participants (n=57) correctly completed the task since they navigated the path properly. It was considered that a task had been completed when the task was ended by the participants by logging out from the account of the mobile application. The computation of the TT represents the time that elapses from the beginning of the task till its end, and this is calculated in seconds. The average task time TT came to 176.94 seconds. NOE counted the occurrence of error whilst the task was being done, and the average for NOE was equal to one error. A total of 13.2 % (n=9) of participants had around the number of error average in committing one error whilst doing the experiment. The average for NOT (number of touches) was around 38 touches for each task. A total of 10.3% (n=7) of the participants made 40 touches in completing the task. The results also showed that average number for the help visit counter came to 1.29; this could be rounded down to 1 since that was an average that was reasonable in representing the frequency of visits for assistance or help in completion of the task. A total of 42.6% of the participants exceeds the average for visits when scoring more than one visit per task.

As mentioned above, the participants are allowed within the experiment to go to a help page so that unsupervised assistance could be obtained for completion of a task if an issue was encountered by them whilst the task was being done. The average DOH (duration on help) was 30.81 secs, the minimum score being 0 and the maximum for DOH being 91 seconds. Most of the participants, i.e., 56%, consumed over the average for time spent on help. The result shows that the maximum time logged in reference to RT stood at 9.45 secs, whilst the minimum time logged was 1.15 secs and the RT average was 4.34 secs. With a 1.15 second for minimum RT, there is an indication that some users have quick reactions and, therefore, had not required a substantial degree of mental workload processing. Indeed, 64.8% of participants (n=44) reacted lately, with an RT time that was above average, and 17.6% of participants (n=12) spending a time that was below average. Finally, the estimation of the total effort by the participants was also measured in terms of a percentage. The TF average for the experiment was measured as 23.53%, with a maximum TF of 47.13% and minimum TF of 11.41%. A total of 51.5% of participants (n=35) required over the TF average whilst 27.9 % of participants (n=19) required a level of effort that was below TF average.

According to the results shown above, the productivity of the participants may be measured as (Symonds, 2011) defined. As such, the productive period can be measured by the equation that follows:

$$\textit{Unproductive time (UT)} = DOH + RT \tag{5.1}$$

$$\textit{Productive time (PT)} = TT - UT \tag{5.2}$$

$$\textit{Productivity} = (PT/TT) * 100\% \tag{5.3}$$

Metrics	Average
TT	176.49 seconds
PT	141.34 seconds
UT	35.15 seconds
Productivity	80.1%

The estimated productivity average amongst the participants during the experiment was equal to 80.1% indicating slack time for the duration of the help visit and the time required for reacting to error, the hints, warnings and error messages whilst the experiments were being conducted.

The research objective within this study has been to conduct an evaluation of cognitive load measurement in user interactions with mobile applications through the use of an automated cognitive tool. The proposed tool could fulfil the functions expected for the monitoring of cognitive load amongst user interactions, with collection of cognitive metrics and the logging of metrics for use in further analysis. There was establishment of user interaction monitoring based upon the mobile application being synchronized with the automated tool in order to allow monitoring of the users whilst they were interacting with the selected mobile application. The objective of the monitoring process is to compute the metrics for cognitive loads and their collection in the form of numerical values (counters, time and percentages) to be stored with the buffer of the mobile device and to be transferred to a locally stored log file within the memory of the mobile device. In the log file context, the logged data includes; TT, TS, NOE, NOT, DOH, HVC, TF and RT.

6 Conclusion and discussion

This research paper has provided a number of insights into progress that is currently being made for automation of human-mobile applications through use of automated tools for metrics of interaction to serve as measures for the cognitive load metrics that humans' experience. The data for average cognitive load has been summarised based on the results for each of the metrics by considering all users. In alignment of the study conducted by [47] and [39], positive relationships are observed for tasks with highest effort for cognitive load with those assigned for the following categories: for highest task time [48]; [27], the number of touches [32] error rate [1], count of help visit [36]; [37], reaction time [48]; [44] and duration on help [34]. Firstly, the tool is able to support researchers within human-computer interactions in having a framework that automates cognitive load measurement metrics with mobile applications. Second, data acquired from experiments for measuring cognitive load for different users offers information regarding associated total effort for each for each of them. It can serve as an initial framework in the evaluation over time of their cognitive performance, associating them to cognitive measurements. The tool is able to support researchers of human—computer interactions in having a tool framework capable of collecting the sorts of cognitive metrics for mobile applications [49];[50]. The data acquired through the experiment for analysing cognitive load for assigned tasks gives information to researchers regarding the total related effort for each of them. It serves as starting framework for evaluation of the performance over time, enabling mapping with cognitive difficulties. It is found to be estimated from the obtained results that the automated tool of the COG could fulfil the metrics expected for the measurement of users' interaction within the cognitive load measurement context, with the collection of the metrics for user workload and their logging for later analysis. The proposed tool by this research implies user experience designers, usability evaluators, and the human-computer interaction (HCI) research domain. This work support researcher in HCI in having a tool framework capable of collecting the sorts of cognitive metrics for mobile applications. The focus of the future research is to further enhance the experiment with a metrics' formula that is more tuned along with better mathematical formations [51]; [52], by classifying them into types and cognitive load categories[21], and through increasing the task set to be performed and the number of the population.

7 References

[1] B. Schildbach and E. Rukzio, "Investigating selection and reading performance on a mobile phone while walking," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 2010, pp. 93–102. https://doi.org/10.1145/1851600.1851619

[2] B. Curum and K. K. Khedo, "Cognitive load management in mobile learning systems: principles and theories," *J. Comput. Educ.*, vol. 8, no. 1, pp. 109–136, 2021. https://doi.org/10.1007/s40692-020-00173-6

[3] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *J. Interact. Sci.*, vol. 1, no. 1, pp. 1–16, 2013. https://doi.org/10.1186/2194-0827-1-1
[4] N. B. Fabil, A. Saleh, and R. B. Isamil, "Extension of PACMAD model for usability evaluation metrics using Goal Question Metrics (GQM) Approach," *J. Theor. Appl. Inf. Technol.*, 2015.
[5] L. Cabañero, R. Hervás, I. González, J. Fontecha, T. Mondéjar, and J. Bravo, "Analysis of cognitive load using EEG when interacting with mobile devices," *Multidiscip. Digit. Publ. Inst. Proc.*, vol. 31, no. 1, p. 70, 2019. https://doi.org/10.3390/proceedings2019031070
[6] J. Buchner, K. Buntins, and M. Kerres, "The impact of augmented reality on cognitive load and performance: A systematic review," *J. Comput. Assist. Learn.*, vol. 38, no. 1, pp. 285–303, 2022. https://doi.org/10.1111/jcal.12617
[7] P. Chandler and J. Sweller, "Cognitive load theory and the format of instruction," *Cogn. Instr.*, vol. 8, no. 4, pp. 293–332, 1991. https://doi.org/10.1207/s1532690xci0804_2
[8] E. W. Anderson, "Evaluating visualization using cognitive measures," in *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization*, 2012, pp. 1–4. https://doi.org/10.1145/2442576.2442581
[9] T. De Jong, "Cognitive load theory, educational research, and instructional design: Some food for thought," *Instr. Sci.*, vol. 38, no. 2, pp. 105–134, 2010. https://doi.org/10.1007/s11251-009-9110-0
[10] G. Hossain and M. Yeasin, "Understanding effects of cognitive load from pupillary responses using hilbert analytic phase," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 375–380. https://doi.org/10.1109/CVPRW.2014.62
[11] L. Cabañero, R. Hervás, I. González, J. Fontecha, T. Mondéjar, and J. Bravo, "Characterisation of mobile-device tasks by their associated cognitive load through EEG data processing," *Futur. Gener. Comput. Syst.*, vol. 113, pp. 380–390, 2020. https://doi.org/10.1016/j.future.2020.07.013
[12] H. Rheem, V. Verma, and D. V. Becker, "Use of mouse-tracking method to measure cognitive load," in *Proceedings of the human factors and ergonomics society annual meeting*, 2018, vol. 62, no. 1, pp. 1982–1986. https://doi.org/10.1177/1541931218621449
[13] M. Jeon, T. M. Gable, B. K. Davison, M. A. Nees, J. Wilson, and B. N. Walker, "Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload," *Int. J. Hum. Comput. Interact.*, vol. 31, no. 1, pp. 1–16, 2015. https://doi.org/10.1080/10447318.2014.925774
[14] S. C. Castro, D. L. Strayer, D. Matzke, and A. Heathcote, "Cognitive workload measurement and modeling under divided attention.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 45, no. 6, p. 826, 2019. https://doi.org/10.1037/xhp0000638
[15] S. Maldonado, J. López, A. Jimenez-Molina, and H. Lira, "Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment," *Expert Syst. Appl.*, vol. 143, p. 112988, 2020. https://doi.org/10.1016/j.eswa.2019.112988
[16] S. Zhao, D. P. Brumby, M. Chignell, D. Salvucci, and S. Goyal, "Shared input multimodal mobile interfaces: Interaction modality effects on menu selection in single-task and dual-task environments," *Interact. Comput.*, vol. 25, no. 5, pp. 386–403, 2013. https://doi.org/10.1093/iwc/iws021
[17] B. Donmez, P. E. Pina, and M. L. Cummings, "Evaluation criteria for human-automation performance metrics," in *Performance evaluation and benchmarking of intelligent systems*, Springer, 2009, pp. 21–40. https://doi.org/10.1007/978-1-4419-0492-8_2

[18] C. D. Wickens, "Multiple resources and performance prediction," *Theor. issues Ergon. Sci.*, vol. 3, no. 2, pp. 159–177, 2002. https://doi.org/10.1080/14639220210123806
[19] S. Roy, P. K. Pattnaik, and R. Mall, "A quantitative approach to evaluate usability of academic websites based on human perception," *Egypt. Informatics J.*, vol. 15, no. 3, pp. 159–167, 2014. https://doi.org/10.1016/j.eij.2014.08.002
[20] H. Sørum, K. N. Andersen, and R. Vatrapu, "Public websites and human–computer interaction: an empirical study of measurement of website quality and user satisfaction," *Behav. Inf. Technol.*, vol. 31, no. 7, pp. 697–706, 2012. https://doi.org/10.1080/0144929X.2011.577191
[21] D. Mutlu-Bayraktar, V. Cosgun, and T. Altan, "Cognitive load in multimedia learning environments: A systematic review," *Comput. Educ.*, vol. 141, p. 103618, 2019. https://doi.org/10.1016/j.compedu.2019.103618
[22] N. Debue, N. Ou, and C. van de Leemput, "An investigation of using a tablet computer for searching on the web and the influence of cognitive load," *Tutor. Quant. Methods Psychol.*, vol. 16, pp. 226–239, 2020. https://doi.org/10.20982/tqmp.16.3.p226
[23] W. Huang, J. Li, and L. Alem, "Towards preventative healthcare: a review of wearable and mobile applications," *Data, Informatics Technol. An Inspir. Improv. Healthc.*, pp. 11–14, 2018.
[24] M. Barifah and M. Landoni, "Emotions associated with failed searches in a digital library," 2020. https://doi.org/10.47989/irisic2027
[25] P. Ramakrishnan, B. Balasingam, and F. Biondi, "Cognitive load estimation for adaptive human–machine system automation," in *Learning Control*, Elsevier, 2021, pp. 35–58. https://doi.org/10.1016/B978-0-12-822314-7.00007-9
[26] A. M. Saleh, R. Ismail, N. Fabil, N. M. Norwawi, and F. A. Wahid, "Measuring Usability: Importance Attributes for Mobile Applications," *Adv. Sci. Lett.*, vol. 23, no. 5, pp. 4738–4741, 2017. https://doi.org/10.1166/asl.2017.8879
[27] K. M. Richardson, S. D. Fouquet, E. Kerns, and R. J. McCulloh, "Impact of mobile device-based clinical decision support tool on guideline adherence and mental workload," *Acad. Pediatr.*, vol. 19, no. 7, pp. 828–834, 2019. https://doi.org/10.1016/j.acap.2019.03.001
[28] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, 2015. https://doi.org/10.1080/00140139.2014.956151
[29] J. Gwizdka, "Distribution of cognitive load in web search," *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 11, pp. 2167–2187, 2010. https://doi.org/10.1002/asi.21385
[30] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington, "Understanding the impacts of mobile phone distraction on driving performance: A systematic review," *Transp. Res. part C Emerg. Technol.*, vol. 72, pp. 360–380, 2016. https://doi.org/10.1016/j.trc.2016.10.006
[31] L. M. Mazur *et al.*, "Toward a better understanding of task demands, workload, and performance during physician-computer interactions," *J. Am. Med. Informatics Assoc.*, vol. 23, no. 6, pp. 1113–1120, 2016. https://doi.org/10.1093/jamia/ocw016
[32] W. Van Winsum, "The effects of cognitive and visual workload on peripheral detection in the detection response task," *Hum. Factors*, vol. 60, no. 6, pp. 855–869, 2018. https://doi.org/10.1177/0018720818776880
[33] P. W. Braddy, A. W. Meade, and C. M. Kroustalis, "Online recruiting: The effects of organizational familiarity, website usability, and website attractiveness on viewers' impressions of organizations," *Comput. Human Behav.*, vol. 24, no. 6, pp. 2992–3001, 2008. https://doi.org/10.1016/j.chb.2008.05.005

[34] A. Hussain and M. Kutar, "Usability evaluation of SatNav application on mobile phone using mGQM," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 4, no. 2012, pp. 92–100, 2012.
[35] A. Saleh, R. Ismail, and N. Fabil, "Evaluating usability for mobile application: a MAUEM approach," in *Proceedings of the 2017 International Conference on Software and e-Business*, 2017, pp. 71–77. https://doi.org/10.1145/3178212.3178232
[36] W. Junmei and W. Jihong, "Research on Mobile Application Automation Testing Technology Based on Appium," *2019 Int. Conf. Virtual Real. Intell. Syst.*, pp. 247–250, 2019.
[37] J. D. Mator *et al.*, "Usability: adoption, measurement, value," *Hum. Factors*, vol. 63, no. 6, pp. 956–973, 2021. https://doi.org/10.1177/0018720819895098
[38] C. Wang, T. Fang, and R. Miao, "Learning performance and cognitive load in mobile learning: Impact of interaction complexity," *J. Comput. Assist. Learn.*, vol. 34, no. 6, pp. 917–927, 2018. https://doi.org/10.1111/jcal.12300
[39] H.-Y. Chang, H.-F. Wu, Y.-C. Chang, Y.-S. Tseng, and Y.-C. Wang, "The effects of a virtual simulation-based, mobile technology application on nursing students' learning achievement and cognitive load: Randomized controlled trial," *Int. J. Nurs. Stud.*, vol. 120, p. 103948, 2021. https://doi.org/10.1016/j.ijnurstu.2021.103948
[40] C.-C. Chen and P.-H. Huang, "The effects of STEAM-based mobile learning on learning achievement and cognitive load," *Interact. Learn. Environ.*, pp. 1–17, 2020. https://doi.org/10.1080/10494820.2020.1761838
[41] B. Verschuere, N. C. Köbis, Y. Bereby-Meyer, D. Rand, and S. Shalvi, "Taxing the brain to uncover lying? Meta-analyzing the effect of imposing cognitive load on the reaction-time costs of lying," *J. Appl. Res. Mem. Cogn.*, vol. 7, no. 3, pp. 462–469, 2018. https://doi.org/10.1016/j.jarmac.2018.04.005
[42] K. P. Wijayaratna, M. L. Cunningham, M. A. Regan, S. Jian, S. Chand, and V. V Dixit, "Mobile phone conversation distraction: Understanding differences in impact between simulator and naturalistic driving studies," *Accid. Anal. Prev.*, vol. 129, pp. 108–118, 2019. https://doi.org/10.1016/j.aap.2019.04.017
[43] J. G. Frederiksen *et al.*, "Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial," *Surg. Endosc.*, vol. 34, no. 3, pp. 1244–1252, 2020. https://doi.org/10.1007/s00464-019-06887-8
[44] P. Liu and Y. Li, "Response Time Evaluation of Mobile Applications Combining Network Protocol Analysis and Information Fusion," *Inf. Softw. Technol.*, p. 106838, 2022. https://doi.org/10.1016/j.infsof.2022.106838
[45] R. Goenarjo *et al.*, "Cerebral oxygenation reserve: The relationship between physical Activity level and the cognitive load during a stroop task in healthy young males," *Int. J. Environ. Res. Public Health*, vol. 17, no. 4, p. 1406, 2020. https://doi.org/10.3390/ijerph17041406
[46] D. F. Hultsch, S. W. S. MacDonald, and R. A. Dixon, "Variability in reaction time performance of younger and older adults," *Journals Gerontol. Ser. B Psychol. Sci. Soc. Sci.*, vol. 57, no. 2, pp. P101–P115, 2002. https://doi.org/10.1093/geronb/57.2.P101
[47] J. Wang, P. Antonenko, M. Celepkolu, Y. Jimenez, E. Fieldman, and A. Fieldman, "Exploring relationships between eye tracking and traditional usability testing data," *Int. J. Human–Computer Interact.*, vol. 35, no. 6, pp. 483–494, 2019. https://doi.org/10.1080/10447318.2018.1464776
[48] E. J. Nilsson, M. L. Aust, J. Engström, B. Svanberg, and P. Lindén, "Effects of cognitive load on response time in an unexpected lead vehicle braking scenario and the detection

response task (DRT),” *Transp. Res. part F traffic Psychol. Behav.*, vol. 59, pp. 463–474, 2018. https://doi.org/10.1016/j.trf.2018.09.026
[49] H. Y. Abuaddous, A. M. Saleh, O. Enaizan, F. Ghabban, and A. B. Al-Badareen, “Automated User Experience (UX) Testing for Mobile Application: Strengths and Limitations.,” *Int. J. Interact. Mob. Technol.*, vol. 16, no. 4, 2022. https://doi.org/10.3991/ijim.v16i04.26471
[50] A. M. Saleh, H. Y. Abuaddous, O. Enaizan, and F. Ghabban, “User experience assessment of a COVID-19 tracking mobile application (AMAN) in Jordan,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1120–1127, 2021. https://doi.org/10.11591/ijeecs.v23.i2.pp1120-1127
[51] K. Hochberg, S. Becker, M. Louis, P. Klein, and J. Kuhn, “Using smartphones as experimental tools—a follow-up: cognitive effects by video analysis and reduction of cognitive load by multiple representations,” *J. Sci. Educ. Technol.*, vol. 29, no. 2, pp. 303–317, 2020. https://doi.org/10.1007/s10956-020-09816-w
[52] K. Zhampeissova, A. Gura, E. Vanina, and Z. Egorova, “Academic performance and cognitive load in mobile learning,” 2020. https://doi.org/10.3991/ijim.v14i21.18439

8 Authors

Ashraf Mousa Saleh is with Amman Arab University, Jordan (email: Asaleh@aau.edu.jo).

Odai Enaizan is with University of Tabuk, Saudi Arabia (email: oenaizan@ut.edu.sa).

Article submitted 2022-06-15. Resubmitted 2022-07-26. Final acceptance 2022-08-12. Final version published as submitted by the authors.