Association for Information Systems
AIS Electronic Library (AISeL)

ACIS 2013 Proceedings | Australasian (ACIS)

2013

Are you smart enough for your smart phone? A cognitive load comparison

Fayez Alshehri
University of Wollongong, fa980@uowmail.edu.au

Mark Freeman
University of Wollongong, mfreeman@uow.edu.au

Alison Freeman
University of Wollongong, afreeman@uow.edu.au

Follow this and additional works at: https://aisel.aisnet.org/acis2013

Recommended Citation

Alshehri, Fayez; Freeman, Mark; and Freeman, Alison, "Are you smart enough for your smart phone? A cognitive load comparison" (2013). *ACIS 2013 Proceedings*. 85.
https://aisel.aisnet.org/acis2013/85

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Information Systems: Transforming the Future

24th Australasian Conference on Information Systems, 4-6 December 2013, Melbourne

Proudly sponsored by

ACIS 2013 Principal Sponsor

Are you smart enough for your smart phone? A cognitive load comparison

Fayez Alshehri
Mark Bruce Freeman
Alison Freeman
School of Information Systems and Technology
University of Wollongong
Wollongong, Australia
Email: fa980@uowmail.edu.au
Email: mfreman@uow.edu.au
Email: afreeman@uow.edu.au

Abstract

The popularity of mobile devices, and particularly smart phones, has been driven by rapid developments in their power, capabilities and features. However, these developments create new and additional demands on users. Increased attention and more complex thought patterns are required to interact with such technology. Testing three popular smart phones, this research applies the principles of Cognitive Load Theory to evaluate the impact of varied operating systems and interaction modes on users' cognitive load. NASA Task Load Index was used to measure participants' perceptions of the cognitive load imposed by interaction with these mobile devices. While different operating systems were found to have a significant impact on cognitive load, interaction modes (sitting or walking) had little impact on users' perceptions of cognitive demands of engagement. Mobile device application designers may conclude that field-based experiments are therefore avoidable, allowing development effort to be focused on improving other aspects of interaction.

Keywords

Cognitive Load, Mobile, NASA Task Load Index, Usability.

INTRODUCTION

The prevalence of mobile device technologies in modern society is underscored by their rapid adoption for personal and business use (Dunn et al. 2013). There were an estimated 487 million global smart phone shipments in 2011, indicating that the sale of new mobile devices surpassed that of sales of new PCs (Bertolucci 2012). The increasing prevalence of mobile technology use internationally is being largely driven by developments in the power, capabilities and features of these new devices. However, the nature of these developments creates new and additional demands on users of these devices, with the additional capabilities and features requiring increased attention and more complex thought patterns in order to interact with the technology. The level of cognitive demand on users is therefore increased, and both researchers and developers are keen to gain a better understanding of "how people interact with their devices in a specific context of use" (Te'eni et al. 2007) so this level of cognitive demand can be managed, and ideally minimized (Alasraj et al. 2011).

This study evaluated user experience on three popular mobile devices (i.e. smart phones) through an investigation of users' performance on real tasks with real devices in simulated contexts of use. To evaluate users' perceived cognitive demands, the NASA Task Load Index (NASA-TLX) instrument was used to measure the elevated or diminished mental workload experienced (Cao et al. 2009) by each user while performing the assigned tasks, to then identify the mobile device with the greatest impact on user performance. Cao et al. (2009) identified that determining a user's mental workload is usually a difficult process; past success has been achieved through the collection of psychophysiological measurements using electroencephalograms (EEGs). However, these machines are not portable and are extremely expensive; the practicalities of their use make them unrealistic for widespread application for this purpose. NASA-TLX is one alternative that can be broadly and consistently applied to provide insight into two key areas of perceived cognitive demand: the demands placed on the user during task performance – as measured by the mental, physical, and temporal demand subscales; and the interaction of the user with the task – as measured through the performance, effort and frustration level subscales. Use of the NASA-TLX instrument also allowed the acquisition of perceived cognitive workload measures after the user's experience of each task while stationary and while walking around a predefined circuit.

BACKGROUND

This section details the current developments in mobile devices, the shift in their functionality and the varied contexts in which users interact with these devices. This is followed by an explanation of the impact of each user's cognitive demands on potential interactions with mobile devices, informed by an understanding of each user's cognitive load via Cognitive Load Theory (CLT). Due to the difficulties of directly assessing CLT, the NASA-TLX is used as a proxy for understanding the perceived cognitive load created by undertaking tasks.

Mobile Devices

Mobile devices and their applications provide a number of advantages to users in relation to "portability, location awareness, and accessibility" (Nayebi et al. 2012, p 1). Users may now access distributed services anywhere, anytime using mobile devices that are increasingly more adaptable to their specific context of use. The field of pervasive computing is concerned with delivering applications that appear natural and unobtrusive within the environment in which the device is being used. Chi (2009) acknowledges that Human-Computer Interaction (HCI) as a field of research and study has progressed far beyond the evaluation setting of "a single user sitting in front of a single desktop computer". As a result, traditional HCI researchers now consider new paradigms that recognize mobile computation and the way that it is impacted by environmental factors (Chi 2009). Indeed, the significant issue for mobile device and application designers is how to address "real-world applications of complex activity" (Lukowicz et al. 2012, p 32) that are particular to the mobile device as it is used in any given context. As a result of these changes in both technology usage and user expectations, mobile application developers must consider the characteristics of many mobile devices as they, and smart phones specifically, become increasingly more complex for users to interact with and use. Through thoughtful consideration and design of these devices, the cognitive demands on users can be minimized.

The context-of-use of mobile devices differs from the traditional stationary computer in the home or business setting in that it is highly dynamic and often involves a range of different software applications (Kjeldskov and Stage 2004). Mobile device interfaces presented to users are markedly different to the interfaces presented for use on PCs (Dunn et al. 2013). Mobile devices also present users with a range of different functionalities and uses. For instance, while textual input and productivity are common uses of both mobile devices and PCs, communications and entertainment uses are far more prevalent on mobile devices than on PCs (Dunn et al. 2013). Designers are responding to the different contextual elements that impact mobile device use through the development of more sophisticated capabilities. One such example is the inclusion of sensing capabilities and applications to leverage basic context information, including the location of users, their motion state and the level of noise in the immediate environment, in modern mobile devices (Al-Hmouz and Freeman 2010; Lukowicz et al. 2012).

As a result, much of the recent research into mobile device usability is focused on the contextual variables impacting HCI from a holistic perspective. Context is part of a user's subjective experience, and central to the exploration of context in mobile device usability evaluations is the question of relevance (Isomäki and Pekkola 2011). Context, in relation to mobile HCI (mHCI) specifically, is thought to include task, physical, social, temporal and technical components (Isomäki and Pekkola 2011). In broader terms: task context refers to the performance of the actual task and any interruptions that may occur; physical context refers to the sensed circumstances of the situation; social context refers to the presence of other people and their impact; temporal context refers to factors pertaining to past or future experiences; and the technical context refers to the device infrastructure and connectivity issues (Isomäki and Pekkola 2011). In relation to these contextual components, many researchers have argued that designers should pay greater attention to contextual and also environmental evaluations on the usability of mobile devices. Dunlop and Brewster (2002) argued that one of the most challenging aspects facing mHCI investigators and designers is mobility. When considering the usability of the design it is important to consider when, where and who is going to use the system. Indeed, understanding the activities people do when interacting with the product is integral to determining the type of activities the systems need to support and optimise the users' interactions with the system.

User Cognition and Cognitive Load Theory

To appreciate the ways in which mobile device sophistication and context-of-use variables may impact a user's experience of mobile devices, a better understanding of user cognition is also required. Models to describe the role of memory in the architecture of human cognitive generally comprise three components: sensory memory (SM); working memory (WM); and long-term memory (LTM). The three components of memory are integral to human cognition: they are where sensory information is either received or stored (Atkinson and Shiffrin 1968). Sensory information is initially processed through different sensory modes in the SM. This information may be forgotten or moved into WM for processing to support the performance of cognitive tasks. Following processing in WM, the information may also be forgotten or coded into LTM. LTM and its huge information storage

capacity is significant to the principles of information storage as it is thought to facilitate most human cognitive activities (Sweller and Sweller 2006).

To manage the vast quantities of retained and new information in LTM, schema structures are constantly reconstructed and altered (Sweller 2003; 2004). Schemas have two primary functions: managing the information stored in LTM and reducing WM load. The reduction of WM load is supported through 'automation' whereby the information is adapted or reconstructed (Sweller and Sweller 2006). New information is impacted by existing knowledge, with schemas constructed and altered according to this previously held information. According to Sweller (2004), LTM operates like a governor over the 'acquired' information, relying on previously established knowledge architectures for its management. With regard to 'new' information, however, the previously formed knowledge architectures are not available and so the management of the new information relies on the formation of new structures.

CLT suggests that human cognitive architectures are produced according to principles similar to those found in the process of natural selection (Sweller 2003; 2004; Sweller and Sweller 2006). According to the assumptions embedded in natural selection, modifications are the result of random mutations, which are then tested for their effectiveness for survival. Those mutations that prove to be effective are retained; whereas those proven to be ineffective are lost. Similar principles are applied to the changes to cognitive architecture in the LTM. Like natural selection, the changes are slow and incremental and are subsequently tested for their effectiveness in different situations that a human finds himself or herself in. Due to the limited capacity of WM, the modifications made in LTM are slow and incremental (Sweller 2003; 2004; Sweller and Sweller 2006). For this reason it is important for designers to limit the load place on WM. Human cognition is thought to operate at a load limit of seven (plus/minus two) chunks of information stored in WM (Miller 1956) and a capacity to handle four items concurrently (Cowan 2001); schemas in LTM may be used to extend this load capacity. The schemas alter and reconstruct multiple items so that they can be managed as a single item in WM, which reduces the effort required from the WM.

In prior research, CLT has been researched in the context of learning environments; discussion in these studies is focused on how to improve a learner's ability to process new information and retain that information for later reuse. Researchers have identified three different types of cognitive load: intrinsic; extraneous; and germane. Prior research has established that developers need to consider the relationship between the three types of cognitive load and their impact on learning when users interact with new systems (Chinnappan and Chandler 2010). Sweller (2006) defined Intrinsic Cognitive Load (ICL) as the mental processes at work when absorbing information elements, with an element being anything that WM processes as a single unit. The load on WM during the learning process is dependent upon the number of elements to be processed simultaneously. This depends on the extent of interactivity between the elements (Sweller 1999). The ICL imposed on WM is low when non-interacting elements are learned in isolation, which is common in basic learning material; more complex material implies a higher degree of interactivity between the learning elements, and therefore a higher ICL on WM (Sweller and Chandler 1994). Learning a new task or solving a complex task requires that the learner process the elements as units into low-order schema. Following the creation of a number of low-order schema, learners then combine these low-order schemas to form a higher-order schema (Sweller and Chandler 1994). In relation to cognitive load, the schema constructed for a complex task contains all the related interactions. This is then managed as a single element by the WM, reducing its overall load. Thus, when the process of learning becomes more automated, a decrease in the interactivity of elements and the ICL takes place (Sweller 1999). According to Sweller et al. (1998), extraneous cognitive load (ECL) is the extra mental effort that is required when instructions are either inadequate or highly complex. There is potential for ECL to hinder the processing and automation of schemas, therefore restricting learning outcomes. As established within the literature, ECL and ICL are additive. Germane Load (GL) (or relevant load) is a capacity in WM that can be redirected from ECL to facilitate the process of schema acquisition (Sweller et al. 1998). On the basis that GL promotes the cognitive resources for schema acquisition and automation, GL has a positive relationship with learning. This distinguishes it from ICL (the actual learning task) and ECL (the features of the system that reduce the ability to learn).

Importantly, the speed and ability of a person to process particular items can be determined by the level of experience and expertise they have in a particular field or domain. If automated schemas are already established there are fewer restrictions encountered in processing items relevant to their field (Sweller 2003; 2004). On the basis of this principle, when novices attempt to process new information there is an increased likelihood of reaching their limits of WM (Sweller 2003; 2004). Therefore, the capacity of the WM becomes restricted when having to process new information that already exists in schemas. These limitations are less evident when pre-stored and pre-organized information is to be processed (Sweller 2004). The organisation of information this way, and the links between the different memory components, allow for the transmission of large quantities of information to be transmitted from LTM to WM when required (Sweller and Sweller 2006). This information is

then processed and activated. LTM may therefore be defined as a highly complex and sophisticated storage technique that can be utilized to meet the needs of WM (Ericsson and Kintsch 1995).

NASA Task Load Index

The NASA Task Load Index (NASA-TLX) was created in the 1980s as a subjective assessment technique to measure cognitive workload. Workload refers to the 'cost' to the individual of carrying out the task. Such costs include fatigue, stress, illness and accidents. An awareness of this individual workload is of particular importance to designers, manufacturers and operators as it is indicative of system performance (Hart 2006). The ability to assess the subjective cognitive load of individuals is deemed to be important as mental workload affects performance. In turn, any impact on performance has implications for an individual's safety, productivity and level of satisfaction (Cao et al. 2009; Rubio et al. 2004). Conceptualizing 'workload' and its relation to task performance can be problematic because individuals will conceptualize workload in their own way (Noyes and Bruneau 2007). As a result, when trying to assess user activity and performance in both experimental and organisational settings in the context of cognitive workload, the researcher is confronted with many challenges. The NASA-TLX tool can help to overcome these challenges by providing individuals with a context in which to deliver their assessment of workload. As a result, data to evaluate system or device usability can be collected (Noyes and Bruneau 2007).

NASA-TLX can be applied to a range of different domains such as technology, medicine and aviation, as well as to a range of different focus points within those domains, such as interface design and model validation (Hart 2006). NASA-TLX testing is often conducted in laboratory or simulation environments and is recognized as an effective tool to provide data on user workload relevant to a specific operational environment (Hart 2006). More specific to the use of NASA-TLX, the tool can be used to gather data on issues relating to interface design or evaluation such as auditory displays, input devices, decision aids and warning systems (Hart 2006). For instance, Yu and Lui (2010) used the NASA-TLX measurement tool in their evaluation of mobile technology visual menu interface usability. The researchers conducted their experiment to assess whether auditory cues for a mobile phone menu improved user performance and experience. Similarly, Hoggan et al. (2008) conducted a study on the effectiveness of tactile feedback for mobile touch screens. NASA-TLX was used to measure the subjective workload of users while typing a particular phrase they had been asked to memorize; users entered this phrase using either a mobile device with a physical keyboard or a mobile device with a touch screen interface (Hoggan et al. 2008). The other dependent variables were speed, accuracy and keystrokes per character.

NASA-TLX and Cognitive Load Theory

An important link has been established between NASA-TLX and CLT. This link is important because the source of the mental workload located within different tasks will impact the workload experienced by the user (Cao et al. 2009). Self-report technologies for measuring cognitive load (such as the NASA-TLX) are just one of the many different categories of measurement (others include physiological measurement and performance measurement). Although there are a number of other similar multi-dimensional subjective workload assessment tools available, including SWAT, Bedford scales, DRAWS and MACE techniques (Stanton et al. 2005), NASA-TLX is the most commonly used tool for the assessment of cognitive load (Haapalainen et al. 2010).

However, the application of the NASA-TLX test to measure cognitive load is not without its challenges. Researchers such as Moreno (2006) suggest that because the measuring tool is not based directly on CLT, it may be the case that other cognitive constructs are explicitly assessed rather than the three types identified in CLT (Moreno 2006). In addition, some studies have questioned the reliability of self-report indicators of cognitive load such as the NASA-TLX (Haapalainen et al. 2010). Notwithstanding these challenges, the NASA-TLX can be useful for studies exploring the effects of instructional formats on cognitive load (Paas and Van Gog 2006) and is also highly correlated with other measures of workload (Cao et al. 2009).

Advantages of NASA-TLX

One of the recognized advantages of the NASA-TLX measuring tool is that it is easy to implement. Subjective workload measurement tools such as NASA-TLX provide high 'face' validity and establish ratings that can be applied with relative ease to complex workload situations with multiple demands (Leedal and Smith 2005). Another advantage of the NASA-TLX measuring tool is that it can be adapted to suit different experimental objectives. For instance, in their study of vehicle navigation systems, Park and Cha (1998 cited in Cao et al. 2009) modified the NASA-TLX by replacing four of the subscales and tailoring the descriptions to suit the specific task being assessed. According to Cao et al. (2009), it is the ability to modify the tool and its ease of implementation that make the NASA-TLX tool a popular choice among researchers.

In addition to the relative ease of implementation and adaptability of the NASA-TLX assessment tool, its other advantages include its sensitivity, diagnostic capabilities, selectivity, low intrusiveness and reliability (Cao et al.

2009). Indeed, the reliability, sensitivity and utility of the NASA-TLX have been the focus of a number of independent evaluations. In response to these evaluations, Hart (2006) purports that a particular benefit of the weighting scheme of the measuring tool is that it provides the researcher with the opportunity to allow for an increase in user sensitivity to particular variables and a decrease in between-rater variability.

Application of NASA-TLX in this research

NASA-TLX was used to record the impact of mobile device usage on participants' working memory in this study. The NASA-TLX test typically comprises six dimensions, described by Stanton et al. (2005) as: *Mental demand* – the cognitive and perceptual activity needed to complete the task; *Physical demand* – the physical activity required and its degree of difficulty; *Temporal demand* – the time pressures associated with completing the task; *Effort* – the level of effort required to complete the task; *Performance* – the level of success achieved trying to complete the task; and *Frustration* – the barriers/challenges experienced trying to complete the task

Underpinning these dimensions is the assumption that the way in which the dimensions combine while performing the task will provide a representation of the 'workload' experienced by the user (Cao et al. 2009). Importantly, the dimensions have been specifically identified as the primary factors that define the subjective experience of workload during the performance of a task (Hart 2006). Using the six subscales (i.e. dimensions) to measure overall workload reduces variability among subjects and provides the researcher with diagnostic information in relation to workload sources (Cao et al. 2009).

One advantage of NASA-TLX which is of importance to this study is its reliance on self-report mechanisms for assessing mental workload. A review of the literature reveals that a number of studies cite post-hoc self-reports of cognitive load as a relatively reliable method for assessing mental workload (e.g. Haapalainen et al. 2010). In turn, the weighting and rating 'features' of the NASA-TLX self-report mechanism provide the researcher with the opportunity to take into account the individual differences between users in their perception of workload and performance activity (Hart 2006).

Further discussion about the quantitative subjective data provided by NASA-TLX and the rating scales used is presented in the following section.

PROCEDURE

The following procedure was used for each participant. The participant arrived at the test location and was briefed about the experiment. The person administering the experiment explained that the purpose of the test was to evaluate three different mobile devices (all smart phones) and their different operating systems (Apple iPhone 4S, Samsung Galaxy Nexus and Nokia Lumina 800). The participant was informed that these phones were chosen because they were popular phones and each used a different one of the three major mobile device operating systems. (Note: The Apple iPhone used Apple's iOS 5, the Samsung used an unmodified version of Android 4.0 and the Nokia used Windows 7.5. Apple and Android operating systems had the greatest market share for smart phones, while Windows had the greatest market share for traditional desktop PCs.)

The participant was informed of the order in which they would use the devices and told whether to initially perform all the tasks sitting or walking around a pre-defined circuit. (This was randomly assigned prior to the arrival of each participant. For example, the first participant completed the tasks in the order of Apple, Samsung then Windows; initially all phones were used while sitting and then while moving. There were 12 possible combinations that could have been assigned to participants; all combinations were used throughout the testing.) The participant completed a number of tasks with each mobile device. The same test script was followed for each mobile device, both while sitting and while walking.

The tasks listed in the test script were indicative of tasks that a person would perform as part of daily use of their own mobile device. Based on the tasks specified in the test script, the participant was supposed to:

- Make a call to the work phone of a fictitious colleague named John Smith, to arrange a meeting about a project they were working on together. In response to this meeting request, John (i.e. the person administering the experiment) suggested a time that he was available to meet the following day. The participant needed to check their calendar on the mobile device to determine their own availability. If the participant was able to access the calendar feature while on the phone, they could have identified that the time that John proposed was not possible (they already had a meeting scheduled) and that they needed to identify an alternate time for the meeting.
- Enter the agreed meeting time in the calendar on the mobile device.
- Send a text message to John's mobile phone number to confirm the meeting and suggest a location (in a local café).

- Locate a PDF document on the mobile device – this was the schedule for the meeting with John.
- Access a local news service website and read the headlines of the leading news story to be informed of current events.
- Locate a specified book (the title was provided) on the Amazon website and add it to their shopping cart.

Data about each participant's performance of the tasks, in both the sitting and walking modes, was collected using point-of-view video recording. Point-of-view recording allows the capture of the interface of the mobile device from the perspective of the participant during the task. This was accompanied by an audio recording of each participant's comments; audio data was captured using the same device. Participants were encouraged to employ think-aloud protocols during the performance of each task to make both positive and negative proclamations about their experience and actions. The audio data provided rich information about participants' experiences of interacting with the mobile devices.

At the conclusion of each interaction with one of the mobile devices in either sitting or walking at the six tasks (i.e. after the test script was completed on each device in each mode), each participant was asked to respond to the NASA-TLX to measure working memory.

Using NASA-TLX

The NASA-TLX test provided quantitative subjective data about the different workload capacities of the user when performing the tasks with the three different devices. Moreover, the NASA-TLX index allows for specific focus to be given to the amount of mental activity (such as thinking or memory use) that is required to perform the task (Kjeldskov and Stage 2004). As a result, the NASA-TLX test provided useful subjective data on the extent to which there was an increase in the mental workload demands and the attention requirements of the user during the different stages of the experiment. The questions that were asked of participants after they completed using each device were:

- How mentally demanding was the task? (low/high)
- How physically demanding was the task? (low/high)
- How hurried or rushed was the pace of the task? (low/high)
- How successful were you in accomplishing what you were asked to do? (low/high)
- How hard did you have to work to accomplish your level of performance? (good/poor - reversed)
- How insecure, discouraged, irritated, stressed, and annoyed were you? (low/high)

This study adapted the Rubio et al. (2004) method to determine the rating scale for the NASA-TLX dimensions. Each of the six dimensions used to assess mental workload were evaluated by the participant on a 7-point bipolar scale (identified as being between 0 and 6). Based on the six dimensions, a global score (between 0 and 100) was calculated by combining the ratings provided by the participants for each of the six dimensions. For this procedure, a paired comparison task was carried out. The paired comparisons were based on the evaluator's choice about which dimension was the most relevant to workload from all six dimension options in this study (see Table 1 for this comparison). For example, this comparison identified that for this evaluation, Mental Demand (MD) was more relevant than Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E) and Frustration Level (F), therefore MD has a weighting of 5.

Table 1. NASA-TLX Dimension Paired Comparison

	MD	PD	TD	P	E	F
MD	X	0	0	0	0	0
PD	1	X	1	0	1	1
TD	1	0	X	0	1	0
P	1	1	1	X	1	0
E	1	0	0	0	X	0
F	1	0	1	1	1	X
Weighting	**5**	**1**	**3**	**1**	**4**	**1**

For the global score, each dimension score was then multiplied by its weighting, then divided by 15 (the total number of paired comparisons). This was then divided by the maximum response (6), and multiplied by 100 to

obtain the global NASA-TLX workload score (see Figure 1 for the complete equation). Using this method, a score between 0 (least amount of mental load) and 100 (greatest amount of mental load) was generated.

RESULTS AND DISCUSSION

The following section will initially present an overview of the participants that engaged in the study. This is followed by a detailed analysis of the participants' subjective NASA-TLX assessment of their perceived cognitive demands in relation to use of the three devices in both sitting and standing positions.

Participant demographics

The sample size consisted of 41 participants, from 18 to 50 years, with varying levels of education. 66% of participants had a first language other than English, reflecting the global nature of the sample (see Table 2).

Table 2. Demographics of Participants

	Categories	**Frequency**	**Percentage %**
Gender	Male	31	75.61
	Female	10	24.39
Age	18 – 25	19	46.34
	26 – 35	15	36.59
	36 – 50	7	17.07
First Language	English	14	34.15
	Other language	27	65.85
Education	High school	10	24.39
	Bachelors degree	15	36.59
	Masters degree or higher	14	34.15
	Other	2	4.88

All participants in the sample reported daily computer use, with the overwhelming majority (98%) indicating that they accessed the Internet every day. It was compulsory for participants to own a smart phone, and the majority of participants (61%) reported that that had been using their current smart phone for a period longer than six months, providing an indication of each user's level of familiarity with their phone operating system. Existing smart phone users were targeted because the main concern of the study was the cognitive load experienced by existing users; it was not adoption and/or issues faced by new users.

As might be expected in a competitive global market, there was variation in the brands of smart phone owned by participants. The most popular brand was Apple (iPhone), owned by 41.4% of participants, followed by Samsung smart phones (36.6% of participants). Nokia and Blackberry brands were each owned by 7.3% of participants. 4.9% of participants reported owning a brand of smart phone not included in the list of options in the questionnaire, and 2.4% owned a Nokia Lumia (recorded separately to Nokia due to the different operating system). The most popular operating systems were iOS and Android (41.4% each). These results were significantly higher than for participants using either RIM or Symbian (both at 7.3%) or Windows 7.5 (2.4%).

The vast majority of participants were either 'often' satisfied (54%) or 'always' satisfied (44%) that their smart phone met their needs. Only 2% indicated that their smart phone only 'sometimes' met their needs. Most participants made a call on their smart phone 'every day' (76%), or at least 'a few times a week' (22%). Only 2% of participants made only one phone call per week. Similarly, the majority of participants (85%) sent a message via their phone every day, with the remaining 15% indicating they sent messages 'a few times a week'.

There was greater variation among participants regarding their use of smart phone for Internet access. 78% used the Internet on their smart phone every day, and 15% 'a few times per week'. This behaviour contrasted with the 5% of participants who reported using their smart phone to access the Internet never or less than once per week. Only 2% of participants used their smart phone to access the Internet once per week. 41% indicated rarely or never playing games on their smart phone, whereas 24% indicated playing games as frequently as once a week.

NASA-TLX analysis

The ordering of the three devices when used while sitting, as determined by participants' subjective rankings using the NASA-TLX, was iOS (M=17.66, SD=18.85) followed by Windows (M=40.76, SD=20.26) and finally Android (M=40.80, SD=12.66). To identify whether there were significant differences between each of the results, a one-way ANOVA within-subjects was conducted. Mauchly's test confirmed the assumption of sphericity ($\chi 2(2)=2.38$, $p=0.305$), therefore the within-subject effects was calculated without adjustment. There was a significant main effect within-subjects ($F(2, 80)=30.04$, $p<0.00$). From a post-hoc analysis (see Table 3,

with Bonferroni adjustment on the significance), there were significant differences between the cognitive load using the iOS and Android devices, and between the iOS and Windows devices. However, there was no significant difference between the Android and Windows devices.

The ordering of the three devices when used while walking, as determined by participants' subjective rankings using the NASA-TLX, was iOS (M=18.89, SD=17.24) followed by Windows (M=44.56, SD=25.35) and finally Android (M=47.05, SD=25.85). To identify whether there were significant differences between each of the results, a one-way ANOVA within-subjects was conducted. Mauchly's test confirmed the assumption of sphericity ($\chi 2(2)=1.86$, $p=0.395$), therefore the within-subject effects was calculated without adjustment. There was a significant main effect within-subjects ($F(2, 80)=41.14$, $p<0.00$). From a post-hoc analysis (see Table 3), there were significant differences between the cognitive load using the iOS and Android devices, and between the iOS and Windows devices. There was no significant difference between the Android and Windows devices.

Table 3. Post-hoc Analysis, Using the Devices while Sitting and Walking

Pair (i, j)	Mean difference (i-j)	Significance	Mean difference (i-j)	Significance
	Sitting		Walking	
iOS, Android	-23.146	0.000	-28.073	0.000
iOS, Windows	-23.098	0.000	-25.585	0.000
Android, Windows	0.049	1.000	2.488	1.000

An overall comparison of usage of all devices in both sitting and walking positions reveals no significance in the results ($F(1, 40)=2.84$, $p=1.00$). This result indicates that the participants did not perceive that there was any increase in the cognitive demands of the tasks when they were interacting with the mobile devices in different positions (i.e. sitting or walking). However a review on how the cognitive load was assessed (Table 1) ranked physical demand at a very low level. Previous mobile device field evaluations involving engagement with tasks requiring a high level of cognitive load has identified that participants typically either narrowed their focus to only the device (i.e. did not consider their environmental surroundings) or stopped moving and moved to somewhere quiet to complete the tasks (Kaikkonen et al. 2005). These previous findings contrast with the findings of this study. However, during the walking task, participants were constrained to walking around a fixed circuit in a laboratory environment. Further research should be conducted to evaluate the impact of an expanded walking task, where actual environments representative of typical device use are employed.

Upon completion, participants were asked to rank the order of the devices based on their evaluation experience. 38 participants (93%) ranked the Apple iPhone (iOS) as the preferred (i.e. first) device with regard to its usability (two participants ranked the Android device first; one participant ranked the Windows device first). Of the participants that ranked the iOS device first, 27 participants ranked the Android device second, followed by the Windows device. These results are of interest as 40 participants (98%) stated that they were 'often' or 'always' satisfied with their current device. Of these participants, 24 (58.5%) owned an alternate device.

From a CLT perspective, it could be argued that iOS has the greatest ability for the user to apply information that they learnt through their interactions with the device in subsequent tasks, and store these learnings in their long-term memory (achieved through a process of schema automation). This would occur by taking inputs from the user's sensory memory and coding them into working memory. Apple's detailed 'Human Interface Guidelines' include discussion about the importance of a focus on the user's primary task. The benefits of applying these guidelines are demonstrated in the results of this study: the user's intrinsic cognitive load of the task becomes the main concern of development, thereby reducing the extraneous cognitive load. As users become familiar with a particular device (in this study, all participants has previously used a smart phone), they build schemas that allow operations with the devices to occur without a high level of cognitive effort on their working memory.

While most users had equal experience with the iOS device (41.4%) and the Android device (41.4%), the differences in cognitive load reported in this study cannot be attributed to the users who used other devices; 93% of participants reported lower cognitive load with the iOS device. The cognitive load reported with the Windows device was similar to the Android device, however only one participant (2%) used this device as their smart phone normally. It could be argued that this device has a lower level of intrinsic cognitive load compared to the Android device, as 39% of participants could use their previously created schemas (germane load) when using the Android device to improve their working memory cognition.

CONCLUDING REMARKS

Through an understanding of CLT, the development and conduct of mobile device user testing allows for greater understanding of users' cognitive abilities. With this in mind, users can be pushed to the limits of their working memory and this is when greater understanding of issues with the devices can be seen. This study employed NASA-TLX as a proxy for understanding the users' perceptions of their cognitive demands to explore how

devices differed in their cognitive demands and whether physical actions increased cognitive demand when using these devices. This research has identified that further research is needed to explore how the principles of CLT can be applied in environments outside of traditional learning environments; as with all tasks, device users must use the automation of previously developed schemas to reduce the load on their working memory.

This study examined the cognitive load of participants using three different mobile devices in two different positions (sitting and walking). By utilizing NASA-TLX, the order that participants subjectively ranked the devices based on their cognitive load was assessed. To gain a comprehensive picture of device interaction, six different tasks that are typical of interaction with such devices were completed by participants. This work has demonstrated that, when users conduct tasks on mobile devices, the choice of sitting or walking positions does not create a significantly different cognitive load. This finding is valuable to mobile device application designers, as it indicates that the results obtained through traditional usability study methods can be applied to the actual use of real devices in a range of contexts. The results indicate that the difficulties of creating field-based experiments may be avoidable; participants in this study did not believe that they had increased cognitive demands placed on them whilst interacting with the devices when moving. This finding suggests that one beneficial avenue for future research is identification of ways to improve the flow of interaction so users can interact with both the devices and their surroundings simultaneously.

One significant finding from this study was users' ranking of the cognitive demands created by each of the devices whilst performing the tasks. With most users considering that their own device 'often' or 'always' met their needs, it is interesting to note that 93% of participants reported the same device as creating the lowest cognitive demand. The results from this study have identified that the Apple iPhone with iOS generated the lowest cognitive demand by a significant amount, and was also considered to be the most usable; this finding is particularly interesting given that such a high percentage of participants were satisfied with their current device.

Only one participant currently used the Windows device; this was a limitation of the study. Future work requires a greater number of participants with diversity in the devices that they use. Participants should also be asked about their device history. Another area for future research is evaluation of third-party applications that run on these devices; this study focused on tasks and applications that were built-in to the devices by the manufacturer.

Given the overall increasing complexity of mobile devices (and their applications), context-of-use variables may impact on device usability. One pressing issue to consider is the application of CLT (founded in the field of formal education) to the concept of more informal, socially-impacted user training. Lukowicz et al. (2012) argues for the need to increase training options for device users, suggesting that for the user of the mobile device to achieve 'good results' from the complex functionalities of the device, there is the need for 'user- and environment-specific training'. User-specific training is linked to more reliable functional outcomes and Lukowicz et al. (2012) discuss the potential for social network connections to overcome the challenges of user training. This issue is something that is not typically considered for consumer devices, which are typically designed for the user to purchase and start using straight away. This creates a potential area for future work in considering the process that a user employs in learning to interact with different mobile devices.

REFERENCES

Al-Hmouz, A., and Freeman, A. 2010. "Learning on location: An adaptive mobile learning content framework," in: *IEEE International Symposium on Technology and Society*. Wollongong: IEEE, pp. 450-456.

Alasraj, A., Freeman, M., and Chandler, P. 2011. "Considering Cognitive Load Theory Within E-Learning Environments" in: *15th Pacific Asia Conference on Information Systems*. Brisbane, Australia. http://aisel.aisnet.org/pacis2011/14

Atkinson, R.C., and Shiffrin, R.M. 1968. "Human Memory: A Proposed System and Its Control Processes," in *The Psychology of Learning and Motivation,* K.W. Spence and J.T. Spence (eds.). New York: Academic Press, pp. 89-195.

Bertolucci, J. 2012. "Smartphone Sales Boom -- Who Needs a Laptop?" Retrieved 3 May, 2013, from http://www.pcworld.com/article/249313/smartphone_sales_boom_who_needs_a_laptop.html

Cao, A., Chintamani, K.K., Pandya, A.K., and Ellis, R.D. 2009. "NASA TLX: Software for Assessing Subjective Mental Workload," *Behavior Research Methods* (41:1), pp. 113-117.

Chi, E. 2009. "A Position Paper On 'living Laboratories': Rethinking Ecological Designs and Experimentation in Human-Computer Interaction," in *Human-Computer Interaction. New Trends, LCNS 5610,* J.A. Jacko (ed.). Berlin: Springer, pp. 597-605.

Chinnappan, M., and Chandler, P. 2010. "Managing Cognitive Load in the Mathematics Classroom," *Australian Mathematics Teacher* (66:1), pp. 5-11.

Cowan, N. 2001. "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity," *Behavioral and Brain Sciences* (24:1), pp. 87-114.

Dunlop, M., and Brewster, S. 2002. "The Challenge of Mobile Devices for Human Computer Interaction," *Personal and Ubiquitous Computing* (6:4), pp. 235-236.

Dunn, B.K., Galletta, D.F., Hypolite, D., Puri, A., and Raghuwanshi, S. 2013. "Development of Smart Phone Usability Benchmarking Tasks," in: *46th Hawaii International Conference on System Sciences*. Hawaii: IEEE, pp. 1046-1052.

Ericsson, K.A., and Kintsch, W. 1995. "Long-Term Working Memory," *Psychological Review* (102:2), pp. 211-245.

Haapalainen, E., Kim, S., Forlizzi, J.F., and Dey, A.K. 2010. "Psycho-Physiological Measures for Assessing Cognitive Load," in: *12th ACM International Conference on Ubiquitous computing*. Copenhagen, Denmark: ACM, pp. 301-310.

Hart, S.G. 2006. "NASA-Task Load Index (NASA-TLX); 20 Years Later," in: *50th Annual Meeting of the Human Factors and Ergonomics Society*. San Francisco: pp. 1-5.

Hoggan, E., Brewster, S.A., and Johnston, J. 2008. "Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens," *Computer Human Interaction*, Florence, Italy, pp. 1573-1582.

Isomäki, H., and Pekkola, S. 2011. "Introduction: Reframing Humans and Information Systems," in *Reframing Humans in Information Systems Development,* H. Isomäki and S. Pekkola (eds.). London: Springer, pp. 1-14.

Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., and Cankar, M. 2005. "Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing," *Journal of Usability Studies* (1:1), pp. 4-16.

Kjeldskov, J., and Stage, J. 2004. "New Techniques for Usability Evaluation of Mobile Systems," *International Journal of Human-Computer Studies* (60:5-6), pp. 599-620.

Leedal, J.M., and Smith, A.F. 2005. "Methodological Approaches to Anaesthetists' Workload in the Operating Theatre," *British Journal of Anaesthesia* (94:6), pp. 702-709.

Lukowicz, P., Pentland, S., and Ferscha, A. 2012. "From Context Awareness to Socially Aware Computing," *Pervasive Computing, IEEE* (11:1), pp. 32-41.

Miller, G. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review* (63), pp. 81-97.

Moreno, R. 2006. "When Worked Examples Don't Work: Is Cognitive Load Theory at an Impasse?," *Learning and Instruction* (16:2), pp. 170-181.

Nayebi, F., Desharnais, J.M., and Abran, A. 2012. "The State of the Art of Mobile Application Usability Evaluation," in: *25th IEEE Canadian Conference on Electrical & Computer Engineering*. Montreal: IEEE, pp. 1-4.

Noyes, J.M., and Bruneau, D.P.J. 2007. "A Self-Analysis of the NASA-TLX Workload Measure," *Ergonomics* (50:4), pp. 514-519.

Paas, F., and Van Gog, T. 2006. "Optimising Worked Example Instruction: Different Ways to Increase Germane Cognitive Load," *Learning and Instruction* (16:2), pp. 87-91.

Rubio, S., Díaz, E., Martín, J., and Puente, J.M. 2004. "Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods," *Applied Psychology: An International Review* (53:1), pp. 61-86.

Stanton, N., Salmon, P., Walker, G., Baber, C., and Jenkins, D. 2005. *Human Factors Methods: A Practical Guide for Engineering and Design*. Hampshire, England: Ashgate.

Sweller, J. 1999. *Instructional Design in Technical Areas*. Melbourne, Australia: Australian Council for Educational Research Press.

Sweller, J. 2003. "Evolution of Human Cognitive Architecture," *Psychology of Learning and Motivation* (43), pp. 215-266.

Sweller, J. 2004. "Instructional Design Consequences of an Analogy between Evolution by Natural Selection and Human Cognitive Architecture," *Instructional Science* (32:1), pp. 9-31.

Sweller, J. 2006. "How the Human Cognitive System Deals with Complexity," in *Handling Complexity in Learning Environments: Theory and Research,* J. Elen and R.E. Clark (eds). Amsterdam: Elsevier, pp. 13-25.

Sweller, J., and Chandler, P. 1994. "Why Some Material Is Difficult to Learn," *Cognition and Instruction* (12:3), pp. 185-233.

Sweller, J., and Sweller, S. 2006. "Natural Information Processing Systems," *Evolutionary Psychology* (4), pp. 434-458.

Sweller, J., Van Merrienboer, J.J.G., and Paas, F.G.W.C. 1998. "Cognitive Architecture and Instructional Design," *Educational Psychology Review* (10:3), pp. 251-296.

Te'eni, D., Carey, J., and Zhang, P. 2007. *Human Computer Interaction: Developing Effective Organizational Information Systems*. New York: Wiley.

Yu, Y., and Liu, Z. 2010. "Improving the Performance and Usability for Visual Menu Interface on Mobile Computers," *International Conference on Advanced Visual Interfaces*, Rome, Italy: ACM, pp. 369-372.

COPYRIGHT

Alshehri, Freeman & Freeman © 2013. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.