Usability in Mobile Learning: Results from a study of learning under varying levels of extraneous cognitive load.

Robin Deegan
CIT
Cork
robin.deegan@mycit.ie

Paul Rothwell
CIT
Cork
Paul.Rothwell@cit.ie

ABSTRACT

In this paper we present the findings of an experiment carried out to measure the effect of environmental factors on the usability of mobile devices used for learning. We draw on current research on 'cognitive load theory' to inform the experiment design. Learners, when using a mobile device, are subject to three types of demand on their cognitive resources: one due to the intrinsic complexity of the material being learned, another due to the cognitive demands of the learning activity itself, and finally a demand that is extraneous to these and due to a combination of the use of the physical device itself (the hardware and software interfaces) and the distractions of the environment. This study indicates that using a mobile device while travelling through a changing environment can result in approximately 12% increase in Cognitive load, 3.9% decrease in performance and 4.5% decrease in learning. These effects persist once the environment stabilises. A model of this extraneous cognitive load can be developed and presented that, when applied, can inform future developers of mobile applications on the effects of the environmental distractions that occur when a user is operating a mobile device.

Author Keywords

Mobile usability, Mobile HCI, m-learning, cognitive load.

ACM Classification Keywords

H.1.2 [User/Machine Systems] Human factors, Human information processing, Software psychology.

General Terms

Measurement, Performance, Experimentation, Human Factors, Standardization, Theory.

INTRODUCTION

Mobile devices are used both by a user on the move through changing environments (whilst walking, travelling on a bus, etc.) and by stationary users in a variety of environments (a park bench, cinema lobby, etc). Both scenarios present difficulties for users because of the distracting effect of the uncontrollable environments.

Recent research into the usability of mobile devices focuses heavily on the interaction between the user and the device itself. The design of the software and hardware are the predominate foci. However, little research exists that focuses on the effect that multiple environments have on our cognitive performance. How does the environment affect our cognitive processing as we use mobile devices?

Cognitive load theory [2] attempts to inform the design of instructional material by providing an understanding of the demands of the various cognitive activities involved in learning. In earlier work we have outlined how research in 'cognitive load theory', can be applied to the investigation of the usability of mobile learning (mLearning) applications. [3]

The goal of this paper is to present the results of a study into the effects of changing environments on learners using mobile devices and help suggest a model of cognitive load that can be considered by application designers to counteract these challenges.

Currently application designers will use HCI / usability design guidelines such as usability heuristics [17] to help design suitable applications. It's hoped that this work could help inform similar heuristics specifically related to cognitive load that could add to existing usability heuristics and could be used by the application designers.

This paper reports on research carried out with a view to exploring the effect that multiple and changing environments have on our cognitive performance when using mLearning applications. Here we begin with a brief outline of mobile learning and cognitive load theory and go on to describe and discuss an experiment designed to attempt to measure the effect of environmental factors on cognitive performance when using an mLearning application.

MOBILE LEARNING

Mobile learning is a cognitively demanding task [8]. Technological advances in mobile computing and the

Copyright is held by the author/owner(s).

ubiquitous use of mobile devices have further led to the challenges presented both to learning with the use of mobile devices and to the development of mobile learning applications. Learning can be a difficult activity and as learners use more mobile technology to assist learning, application developers should consider using that same technology to help recognise and limit the effect of distractions (context aware computing).

Sharples, for example, points to the need for learners to control the pace and style of learning when in a mobile environment. He also highlights the need for technology to share control with the learner [5].

COGNITIVE LOAD THEORY

'Cognitive load' refers to demand placed on the resources of working memory in the performance of cognitive tasks. Because working memory has a limited capacity, our capacity to handle cognitive load is limited [1, 2].

Cognitive load theory is predicated on a model of cognitive architecture, outlined by Baddeley [15], where working memory and long-term memory co-operate in conscious cognitive activity. Working memory has a limited capacity such that it can only deal with about seven 'elements' at a time, the more complex these elements, the more complex the cognitive tasks that can be handled. Complex elements are supplied as 'schemas' by long-term memory. Schemas are built by working memory and encapsulate learned knowledge that is stored in long term memory.

Types of cognitive load

Sweller *et al*, [2] describe cognitive load as consisting of 3 parts:

1 – Intrinsic Cognitive Load: The load associated with the inherent difficulty of the task;

2 – Extraneous Cognitive Load: The load associated with all mental activity not directly associated with the task.

3 – Germane Cognitive Load: The load associated with schema creation, essentially it is the load necessary to process the intrinsic cognitive load i.e. to learn [4].

Sweller argues that intrinsic cognitive load (ICL) cannot be changed by any means other than changing the task or the knowledge of the participant [4]. As the participant progresses through the task and 'learns', their knowledge increases and so intrinsic load decreases. This is because working memory has access to more complex elements through the schemas created.

Extraneous cognitive load (ECL) is the demand placed on mental resources due to cognitive activity not directly associated with the task. Dealing with aspects such as unnecessary/redundant information, interpretation and navigation of the instructional design and user interface, etc. all contribute to extraneous cognitive load. Also external elements, such as ambient noise or light, beyond the control of the user/task designer can add to extraneous cognitive load.

Germane cognitive load (GCL) is the demand placed on mental resources due to the necessary cognitive activities of schema creation associated with learning [4]. GCL is also responsible for the interpretation, processing, and automation of schemas. Basically GCL is the load applied to working memory that converts data interpreted by our senses into long term memory, and is fundamental to the process of learning.

Changing load

As we learn, both ICL and GCL will change (with knowledge learned) but ECL should remain static (assuming that the environment, equipment and task remain the same). This makes measuring the individual loads quite difficult as we cannot differentiate between extraneous, intrinsic and germane loads [3]. All we can do is measure the total cognitive load.

The challenge to measure ECL and develop a predictive model of ECL became the goal of the experiment. It was decided that a predictive model for extraneous cognitive load should be considered so that a mobile device could identify any potential increases in extraneous load allowing developers to try to adjust the learning content to suit. Essentially a context aware learning application, when aware of the environment, could change the learning content to suit the level of predicted cognitive load that the user is experiencing.

METHODS

In previous work, strategies for measuring and managing cognitive load were explored [3] and CL assessment methods for the experiment were chosen based on this work. Participants in the current study completed a task, within the application Brain Challenge Lite[1], on an apple iTouch.

The experiment participants were organised into three groups. Each member of each group completed three sessions with the application. The three groups were subjected to a different sequence of environments for their three sessions.

The 'Control' group completed all three sessions in a quiet office-like environment. The other two groups completed their first and last sessions in the same office-like environment but had their second session environment manipulated. The 'Simulated' group were subjected to a session two environment created by the experiment designers (the lights were turned low and a distracting video was played in the background[2]). The 'Natural' group were subjected to a normal everyday session two environment where they were required to walk through the college corridors and canteen area (Table 1). After each

[1] http://www.gameloft.com/mobile/brain-challenge/

[2] http://www.youtube.com/watch?v=gPxPC3ZFt4g

session each participant completed a NASA TLX[3] questionnaire. The NASA TLX is a widely used subjective tool that is used to give an approximation of cognitive load.

	Control	Simulated	Natural
Session 1	Application use in office-like environment. Complete TLX	Application use in office-like environment. Complete TLX	Application use in office-like environment. Complete TLX
Session 2	Application use in office-like environment Complete TLX	Application use in office-like environment with artificial environmental Stimuli. Complete TLX	Application used while walking with natural environmental Stimuli. Complete TLX
Session 3	Application use in office-like environment. Complete TLX	Application use in office-like environment. Complete TLX	Application use in office-like environment. Complete TLX

Table 1 - Test groups and sessions

Participants

An *a priori* power analysis was completed to determine an appropriate sample size. The significance criterion was set at the conventional value $\alpha = 0.05$. To achieve the statistical power of 0.80 ($1-\beta = 0.80$) the effect size was defined as large. It was determined that each group should have a minimum of 21 participants [6].

72 people took part in the experiment. No demographic data was recorded. All participants were either students or lecturers. All participants were randomly distributed to one of the 3 groups.

Design and Procedure

Five iPod Touches were used so that five participants could complete the experiment concurrently. To motivate participants to participate each was entered in a draw for an iPod Touch.

Gonzalez *et al.* conducted a review of cognitive games for the aging [7]. Based on this review the application 'brain challenge lite' was selected for use. Brain challenge lite is aimed at adolescents, and targets the following cognitive functions; memory, visual attention, executive functions and calculation. The users were presented with simple puzzles by the application and rated for the speed with which they solved the puzzles (fast, normal or slow) and also their accuracy. Fast is completing the subtask within a second or two, and slow is completing the subtask in a time greater than four or five seconds, approximately.

One of the drawbacks to measurements involving cognitive performance is that participants' cognitive abilities vary thus making direct comparisons problematic. To limit this, each participant was asked to complete the task three times. The results from each session were analysed and the differences between the sessions were compared to other participants' differences, rather than comparing participants' performances directly.

Experiment

A pilot experiment, with six participants, was conducted in September 2010. The main experiment ran from November to December 2010. Each session of the main experiment took approximately 6 minutes for a participant to complete. Written permission to use data gathered during the experiment was obtained from all participants and a briefing was also supplied to all participants in advance of the experiment.

Null Hypotheses

1) A change in the physical environment, when the user remains stationary, will not cause an increase in extraneous cognitive load.

2) Moving from one physical environment to another in any environment will not incur a change in extraneous cognitive load.

3) Changing environments and exposure to environmental stimuli will not affect the learning performance.

Analysis

The application kept a record of results and these results were used to indicate performance, learning and cognitive load. Also the participant completed a NASA TLX questionnaire and the results of this were used to indicate cognitive load. The use of both approaches was intended to give a clearer understanding of the effect that the environment has on extraneous cognitive load.

A one way ANOVA [16] was used to determine any statistically significant difference between the 3 groups and if a difference was found a simple t-test confirmed the group causing the difference. Multiple elements were changing (light, sound, motion etc.) but for this experiment these were considered as one variable – the environment.

TASKS

Each participant completed the 'Daily Challenge' within the Brain Challenge application. It consisted of five tasks. And within each task there were 10-20 subtasks approximately. All five tasks were completed in each session. These tasks tested the following cognitive activities: memory, visual attention, executive functions and calculation [7]. [4]

Data elicitation

For this study the following data were obtained from the application:
%correct – this is the percentage of tasks completed

[3] http://humansystems.arc.nasa.gov/groups/TLX/

[4] Space does not permit a full description of the application.

correctly.
%fast – this is the percentage of tasks that were completed correctly at a fast rate.
%slow – this is the percentage of tasks that were completed correctly at a slow rate.
The difference between every session was also noted and is indicated below as follows: e.g. '∆1,2 %slow' denotes the difference between the %slow scores in the first and second sessions.

As already stated, the standard NASA TLX questionnaire was used. In this questionnaire the participant was asked to answer six questions on a rating scale from 0-20 [8]. Each individual score was added and the total for each session was determined. Again the difference between each session was noted and is indicated as follows: e.g. '∆1,3 TLX' denotes the difference between the TLX results in the first and third sessions.

Finally, observations were recorded by the experiment facilitator if anything exceptional occurred.

RESULTS

Overview

From a high level the results demonstrate that a statistically significant difference was identified between the first and second sessions across all groups and this was noted in the TLX and application data. However the application data noted a statistically significant difference between the first and third sessions whereas the TLX data noted a difference between the second and third sessions among all groups.

One statistical outlier and five outliers, from the observation notes, were identified. One individual left abruptly mid session (his sessions were cancelled as along with the other participants' simultaneous sessions). Therefore the results from sixty-six participants are presented, twenty-two in the control group, twenty-one in the simulated group, and twenty-three in the natural group.

The following sections will explain the ANOVA results and then proceed to delve deeper into the actual TLX and application (%correct) results of the ANOVA specifically where a statistically significant difference was found.

ANOVA results

Table 2 shows a summary of the twelve ANOVA tests that were completed. Results from the %fast and %slow data showed no statistically significant difference for either '∆1,2'`(difference between first and second session), '∆1,3'(difference between first and third session) or '∆2,3' (difference between second and third session) between all groups. It appears that the environmental distractions had no effect on the completion time of these specific tasks.

	∆1,2	∆1,3	∆2,3
Sig. Diff. %correct	Yes	Yes	No
Sig. Diff. %fast	No	No	No
Sig. Diff. %slow	No	No	No
Sig. Diff. TLX	Yes	No	Yes

Table 2 Summary of results from ANOVA

Table 2 also shows that a statistically significant difference was noted when analysis was completed on '∆1,2' and '∆1,3' when the results of all the groups' %correct data was compared. Also a statistically significant difference was noted when analysis was completed on '∆1,2' and '∆2,3' when the results of all the groups' TLX data were compared.

In the following sections specific ANOVA analysis will be presented in table, 3, 4, 5 and 6. These tables show summary information of the three groups. They display the count (number of participants in each group), the sum of each group's results, the average and the variance of the results. The table also display's the ANOVA specific results between and within the groups. For this experiment only the P-value (highlighted in red) was required.

The P-value is the probability of seeing results as or more extreme as those actually observed if the null hypotheses were true. As '∆1,2 %correct' (table 3) has a p-value of 0.0005, this implies that for another test there is only a 0.0005% probability of observing similar results so as this is less than the set alpha level of 0.05 the hypotheses can be rejected.

%correct analysis

Table 3 and table 4 below display the ANOVA results for the %correct data.

SUMMARY							
Groups	Count	Sum	Average	Variance			
Control	22	33	1.5	25.69048			
Simulated	21	83	3.952381	26.24762			
Natural	23	-55	-2.3913	27.52174			
ANOVA							
Source of Variation	SS	df	MS	F	P-value	F crit	
Between Groups	452.6906	2	226.3453	8.539129	**0.000523**	3.142809	
Within Groups	1669.931	63	26.50684				
Total	2122.621	65					

Table 3 – ∆1,2 %correct

SUMMARY						
Groups	Count	Sum	Average	Variance		
Control	22	92	4.181818	45.77489		
Simulated	21	140	6.666667	41.33333		
Natural	23	-8	-0.34783	25.23715		
ANOVA						
Source of Variatior	SS	df	MS	F	P-value	F crit
Between Groups	560.6008	2	280.3004	7.536382	**0.001163**	3.142809
Within Groups	2343.157	63	37.19296			
Total	2903.758	65				

Table 4 - Δ1,3 %correct

As expected the environment influenced the results between the first and second sessions and also the first and third sessions. ANOVA between the three groups resulted in a P-value of 0.0005 and 0.001 for the 'Δ1,2 %correct' and 'Δ1,3 %correct' respectively.

Further investigation of the results here show that the statistically significant difference comes from the natural group alone. The simulated group did not show any difference when the environment was changed.

Furthermore ANOVA analysis of 'Δ2,3 correct' did not show any statistically significant difference between groups. This seems to imply that the effect that the natural group felt when confronted with the natural environmental stimulation remained after that stimulation was removed. That is, as the participants from the natural group completed the third session it was expected that they would revert to normal levels (their %correct scores would return to the same as their first session results) but they did not. Their results barely changed from the second session, perhaps showing that the negative effect induced by the natural environment remained when they returned to the quite room where they completed the first session.

TLX analysis

Table 5 and 6 display the ANOVA results for the TLX data.

SUMMARY						
Groups	Count	Sum	Average	Variance		
Control	22	162	7.363636	96.90909		
Simulated	21	101	4.809524	177.3619		
Natural	23	434	18.86957	241.8458		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2507.32	2	1253.66	7.243972	**0.001474**	3.142809
Within Groups	10902.94	63	173.0625			
Total	13410.26	65				

Table 5 - Δ1,2 TLX

SUMMARY						
Groups	Count	Sum	Average	Variance		
Control	22	-17	-0.77273	39.42208		
Simulated	21	-95	-4.52381	194.0619		
Natural	23	-389	-16.913	264.2648		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3219.027	2	1609.513	9.636039	**0.000223**	3.142809
Within Groups	10522.93	63	167.0306			
Total	13741.95	65				

Table 6 – Δ2,3 TLX

As expected the environment influenced the results between the first and second sessions and also the second and third sessions. ANOVA between the three groups resulted in P-values of 0.001 and 0.0002 for the Δ1,2 %TLX and Δ2,3 %TLX respectively.

A statistically significant difference was in fact noted between the first and second sessions. Further analysis of the results again confirmed that the simulated group did not have any reaction at all due to the environmental change. Only the natural group showed a statistically significant difference implying that the environmental distractions had a dramatic effect on the perceived difficulty.

However in contrast to the %correct results there was no statistically significant difference between the first and third session. However there was a difference between the second and third session. Again this difference was limited to the natural group.

Summary

These results indicate a distance affect on ECL from the environment factors. ICL and GCL are assumed to remain static so observed changes to cognitive load are seen to be caused by extraneous factors. The TLX data suggests a difference between the first and second session and also the second and third session and the %correct data suggests a difference between the first and second session and the second and third session. The following section will now explore the actual TLX and %correct results and will attempt to determine the approximate performance effects of these factors.

DISCUSSION

Overview

Interpretation of these results implies that there is a statistically significant difference between groups when a participant completes the same task while moving between different environments. Therefore the null hypotheses 2 and 3 can be rejected.

The null hypothesis 1, however, has been proven and it is accepted. There is no difference in extraneous cognitive load when the environment was changed and the participant remained stationary for this experiment. However there still remains the question of whether the participants were in fact under-loaded in this scenario; perhaps it is possible that the tasks may not have been intrinsically complex enough to allow any effect to be noticed when the extraneous load was increased. Future research should address this.

Increased cognitive load

The natural group showed a significant increase in cognitive load when they completed the second session (under natural environmental conditions walking through the college) in relation to the first session (Δ1,2) for both the %correct and TLX data. This implies that the visual, audio, tactile and motion distractions presented to the participant proved too resource hungry for the participants cognitive performance and as such cognitive processing was overloaded.

The natural group showed a very interesting discrepancy between the %correct application data and the TLX data. The application data records the actual performance and the TLX records the participant perceived performance.

Only the %correct results showed a statistically significant difference for the Δ1,3 and the TLX results did not. Also the natural group's %correct Δ2,3 showed no statistically significant difference while the TLX data did.

The Δ1,3 application data implies that the control and simulated group improved equally but the natural group did not. Their performance was lower than the expected rate as determined by the control group.

However the natural group's TLX Δ2,3 shows a difference to the other groups. The participants 'feel' that the third session is less cognitively demanding than the second session, but the actual results from the application suggest that it was not in fact less cognitively demanding. This could also indicate that the induced cognitive load remains after the environmental factors are removed at least from the subjective assessment of the participants if not from the objective results from the application.. For the cognitive load to remain beyond the actual distraction suggest that the cognitive processing dealing with ECL has assimilated information from the distraction to long term memory (i.e. learned from the distraction), this seems to be the only explanation for the lingering effect and this seems to imply that the learner learned from the extraneous cognitive load. If learning can only occur when germane load is present then this suggests a hint of germane learning in the extraneous distractions.

A model for extraneous cognitive load

Each task was different and at the same difficulty to the participant so there was little possibility for the intrinsic cognitive load to change. As the intrinsic load did not change the germane load also remained the same. Of course there could have been some reduction of intrinsic (and associated germane load) due to participant analysis of the presentation of the problems; essentially the participant 'learned' aspects of the application and the presentation of the problem and not the problem itself. However it is assumed that the overall intrinsic load associated with the difficulty of the task was substantial and largely unchanged. Extraneous cognitive load was different for each group so any changes in overall cognitive load would be directly due to extraneous cognitive load.

As the participants completed the second session their TLX result shows an increase in cognitive load in the control group (Δ1,2 = 7.4%) and the simulated group also shows a slight increase (Δ1,2 = 4.8%). It is not clear what is causing these increased cognitive loads; but it is most likely the participants understanding the task better and working harder to solve the tasks, hence reducing ECL and applying more GCL to solving the tasks. This effect is reversed in Δ2,3 with all groups seeing a drop in cognitive load. This can be attributed to the users apparent reaching of an 'expert' level of the application tasks having completed it for the third time.

However the natural group showed a large and statistically significant increase in cognitive load, 19%, for Δ1,2. This shows a direct correlation between the environmental stimuli and the extraneous cognitive load.

Decreased learning

The natural group %correct Δ1,3 showed a statistically significant difference to the other groups indicating that the learning (performance in the task in relation to the first session) increase was lower for the natural group.

Here again the TLX data did not agree with the application data. The natural group TLX Δ1,3 showed no difference to the other groups. It is believed that the participant could only compare the TLX 'rating' to the previous ratings; in other words, they cannot remember the first TLX ratings and so cannot give an adequate rating for cognitive load overall. Further analysis of this phenomenon should be investigated.

A model for learning

As the participant completed each session their performance (%correct) increased slightly in the control group (Δ1,3 = 4.2%) and the simulated group (Δ1,3 = 6.7%). The natural group showed a 0.3% performance decrease for Δ1,3 thus implying a performance decrease in relation to the control group, and therefore a degradation of learning in relation to the other groups. The simulated group showed no statistically significant difference to the control group but there was a small improvement in performance. This could possibly be due to the white noise effect [12]. This possibility is discussed below in the section on speculative findings.

Speculative findings

Some results of this experiment were surprising. Although not the focus of the experiment, the following aspects were observed. These may help to aid interpretation of the main work and they may help to direct future research.

Subjective assessment vs. application assessment

The application's data can be used by the experimenter to make a specific calculation in relation to a participant's performance (e.g. results from first, second and third sessions can be compared), but a participant has difficulty recalling a feeling or level previous to the current one (can compare from first to second, and second to third but perhaps not from first to third).

It was noted that TLX data indicates that the natural group participants noticed an increase in load while completing the second session but a decrease in load while completing the third session. While the $\Delta 1,2$(%correct) from the application data supported their feeling, the $\Delta 1,3$(%correct) did not. Perhaps this is because when the participants completed the third session they rated the load in relation to the second session alone - they simply found the overall load 'easier' in relation to the second session. Literature does not exist on such an application of TLX data on multiple sessions; perhaps the participants 'learned' to use the TLX questionnaire in the same way they 'learned' to use the application but it is also surprising that the TLX data returned to a similar level after the distraction to what was reported before. Again this should be further studied.

Conversely the application data shows that the cognitive load level actually remained the same for both the second and third session. This raises the question: if the extraneous cognitive load due to the walking environment has been removed, what has taken its place? It seems likely that a new extraneous load has replaced it or it has in fact lingered in another guise after the event. This seems to support the theory that learning has in fact taken place due to extraneous load and this is why it remains when the load is removed. Perhaps the learner is still thinking about what they saw or experienced on the walk. This could infer problems with the current model of cognitive load as learning is only believed to take place when germane or intrinsic load is applied.

There is also a possibility that the extra cognitive load has induced false memories [14] and so disturbed the participants' ability to accurately rate the TLX. Future research will critically investigate CLT further with a view to refining it.

The control group vs. the simulated group: no difference?

One of the more surprising outcomes of this study was the findings of no statistically significant difference between the control group and the simulated group. The simulated group's second session was subjected to a video designed to distract them but it did not seem to have that effect.

Visually the participants were focused on the screen of the device and they did not focus on/see the projector screen playing the video so it is quite likely that there was little to no visual distraction and therefore no extraneous cognitive load due to the visual distraction. Furthermore, Mangipudy found, in her Doctoral theses, that for successful distraction, causing extraneous cognitive load, both audio and visual effects must be experienced by the participant [13]; i.e. visual distractions must interfere with the task. This did not happen in this experiment as the user could simply look away from the projector screen displaying the video.

In terms of audio, distraction to participant's auditory system comes from pitch/frequency changes [11]. Four audio clips were sourced that in themselves were distracting. They were then combined and added to a video clip to make the video that was played to the second group to simulate a distracting environment. Perhaps this 'combination' made 'white noise' which can actually improve performance in some instances [12].

Individual Improvement vs. Individual result

Cognitive load depends on the task, the environment and the participant's knowledge/experience. Therefore it is extremely difficult to compare one individual's load to another when they have differing knowledge / experience in the task domain and equipment use. This is one of the reasons why each individual's improvement was compared and not the individual result.

Speed vs. accuracy

Cognitive overload should affect efficiency as well as effectiveness; that is the task should take longer and the participant should make more errors.

There was no difference in speed throughout the experiment. The extra load did not seem to affect the %slow or %fast data in the slightest and so it seems to go against the belief that cognitive load slows down efficiency but does not affect accuracy [10]. Also there was a statistically significant difference in the natural group for %correct and so it was confirmed that cognitive load affects accuracy by making the participant more prone to errors.

Germane load decreased with every session?

It was originally hoped that the overall cognitive load would decrease for the control group on each session thus showing that as they performed the task more their performance would increase i.e. the task became easier and the load reduced (indicating a lowering of germane cognitive load in relation to learning).

It now seems that the overall cognitive load increased for each session and a model of change in germane cognitive load could not be realised. This was due in part to the application's design but most likely to other unknown elements.

Summary

The results of this experiment clearly indicate that environmental factors can and do effect performance and learning and they also incur an elevated level of extraneous load. Also there were several unsubstantiated observations that were noted and can help to inform future experiments and research.

CONCLUSION

In this study the intrinsic load appeared to remain the same and the germane load also appeared to remain the same. The extraneous cognitive load appeared to vary based on different environments.

When the participant moves in the natural environment there is an increase in extraneous cognitive load. In the control environment and simulated environment there is also a smaller increase in extraneous cognitive load. It is unclear why this was the case and further research should be carried out to investigate this.

This study accepted the null hypotheses 1: there is no change in extraneous cognitive load when the environment is changed for a stationary learner.

However it rejected the null hypotheses 2: there is no change in extraneous cognitive load when the environment is changed for a learner traversing environments and also the null hypotheses 3: learning performance is affected by changing environmental stimuli.

In the control group, during the second session, there was a participant-perceived natural cognitive loading (from the TLX data) of approximately 7.4% along with a performance increase (%correct) of 1.5%. In the control group the final performance increase after the third session was 4.2% meaning that all things considered an average user should show a performance increase of 4.2% on their third session on the application.

If we consider the effect the natural world has on the mobile learning participant we can see the following. Using a mobile device while traversing through an environment causes an increase of 12% (the natural group TLX $\Delta 1,2$ - the control group TLX $\Delta 1,2$) extraneous cognitive load *(TLX scores being a direct representation of overall cognitive load)* and a drop of 3.9% in performance (the natural group %correct $\Delta 1,2$ - the control group %correct $\Delta 1,2$).

In terms of learning the control group displayed a 1.5% increase in performance (%correct) in their second session compared to their first session, a 2.7% increase in performance in their third session in relation to their second session therefore giving an overall performance increase (due to learning) of 4.2% from first to third session.

Again if we consider the effect the natural world has on the mobile learning participant we can see the following: Using a mobile device while traversing through an environment causes a 4.5% reduction in learning (the natural group %correct $\Delta 1,3$ - the control group %correct $\Delta 1,3$).

Impact for mobile usability

Future mobile web application developers could use the above results to inform their design of relevant applications.

A heuristic called ‘Consider the changing environment’ could be created for mobile devices. This heuristic could inform the application developers that changing environments can affect the cognitive load, performance and learning associated with the application and it should encourage the application developers to consider how and where the device and application will be used.

FUTURE WORK

These results confirm that there are considerable concerns with mobile learning while the environment is changing. But this effect is not only limited to learning, it affects all tasks completed on our mobile device, from work to games.

Mobile software developers can use this model of extraneous cognitive load to develop suitable strategies to predict changes in human performance due to changes in the environment. Modern devices are already built with technology that can detect changes; microphones can detect frequency and pitch changes, onboard photoreceptors or built in cameras can detect changes in lighting and visual distractions, accelerometers can detect motion and GPS systems can detect location and movement. When these changes are detected the content should attempt to reduce the ICL or GCL to counteract the changes to ECL so the user does not become overloaded. There are many different methods for this such as restructuring ICL [3].

When the device detects that the participant is in fact overloaded with extraneous cognitive load they may implement any of the strategies for managing cognitive load identified in earlier work such as restructuring the intrinsic cognitive load [3].

A lingering effect was identified where actual cognitive load remained after the distraction was removed. This was validated by the results from the application; however perceived cognitive load seemed to diminish. This suggests that learners learned from the distraction (ECL) and suggests that there are some elements of our accepted model of cognitive load theory which are confusing especially when considering learning and germane cognitive load. This area will need to be researched further but it suggests a more comprehensive model of cognitive load is needed.

Future research should also attempt to develop a model of germane cognitive load change when learners are exposed to environment conditions. This is not a simple task; as intrinsic cognitive load changes so does the germane cognitive load: they are directly linked. Also as extraneous cognitive load reduces germane cognitive load may increase and conversely as extraneous cognitive load increases it causes both germane cognitive load to decrease (affecting learning) and perhaps an overall cognitive overload (affecting performance). Furthermore, as identified above, extraneous and germane cognitive load may be more involved than is currently considered.

Finally future research should investigate the discrepancy between the TLX results which measure the participant's perceived cognitive load and the application data measuring the actual effect of cognitive load. Why do the participants feel the cognitive load has been removed when the results show no change in performance?

REFERENCES

1. Feinberg, S. and Murphy, M. Applying cognitive load theory to the design of web-based instruction. In Proceedings of IEEE Professional Communication Society international Professional Communication Conference and Proceedings of the 18th Annual ACM international Conference on Computer Documentation: Technology & Teamwork (Cambridge, Massachusetts, September 24 - 27, 2000), 353-360. Accessed January 2011

2. Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. Cognitive architecture and instructional design. Educational Psychology Review, (1998).10 (3), 251–296. http://www.springerlink.com/content/vj4917q523256673/. Accessed Jan 2011

3. Deegan. R. and Rothwell, P. The application of Cognitive Load Theory to the Usability of m-learning applications: First steps. In Proceedings of IHCI 2010, Dublin Ireland. (2010) http://www.clarity-centre.org/iHCI2010/pdf/iHCI2010Proceedings.pdf Accessed January 2011

4. Sweller, J. Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*. (2010). DOI: 10.1007/s10648-010-9128-5 Accessed January 2011

5. Sharples, M., Taylor, J., Vavoula, G. *A Theory of Learning for the Mobile Age*, In R. Andrews and C. Haythornthwaite (eds.) The Sage Handbook of Elearning Research. London: Sage, pp. 221-47 (2007) Accessed January 2011

6. Cohen, J. A power primer. *Psychological Bulletin*, *112* (1), 155-159 (1992). Accessed January 2011

7. Gonzalez, MF. Facal, D, Buiza, C, Urdaneta, E. Koffel, C. Geven, A. Petsatodis, T. Soldatos, J. Cognitive Training Exercises (2009) http://www.fp7-hermes.eu/uploads/media/deliverables/HERMES_D61_CognitiveTrainingExercises_final.pdf Accessed January 2011

8. Sharples M., Corlett D. & Westmancott O. The design and implementation of a mobile learning resource. Personal and Ubiquitous Computing 6, 220–234. (2002) Accessed January 2011

9. Deegan, R. Rothwell, P. A Classification of M-Learning Applications from a usability perspective. Journal of the Research Center for Educational Technology, Vol 6, No 1 (2010) http://www.rcetj.org/index.php/rcetj/article/viewArticle/49 Accessed January 2011

10. Sharon Oviatt. 2006. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th annual ACM international conference on Multimedia* (MULTIMEDIA '06). ACM, New York, NY, USA, 871-880. DOI=10.1145/1180639.1180831 http://doi.acm.org/10.1145/1180639.1180831 Accessed January 2011

11. Banbury SP, Macken WJ, Tremblay S, Jones DM. Auditory distraction and short-term memory: phenomena and practical implications. *Human Factors* 2001; **43:** 19–29. http://www.cf.ac.uk/psych/home2/papers/jones_dylan/2001%20Auditory%20distraction.pdf Accessed January 2011

12. Hughes R, Jones DM. The intrusiveness of sound : Laboratory findings and their implications for noise abatement. Noise Health (2001);4:51-70. http://www.noiseandhealth.org/text.asp?2001/4/13/51/31802 Accessed January 2011

13. Mangipudy, R. The Impact of Eliminating Extraneous Sound and Light on Students' Achievement An Empirical Study Doctoral Theses, Florida International University (2010)http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1333&context=etd Accessed January 2011

14. Mather M, Henkel LA, Johnson MK. 1997. Evaluating characteristics of false memories: remember/know judgments and memory characteristics questionnaire compared. *Mem. Cogn.* 25:826–37 http://www.usc.edu/projects/matherlab/pdfs/MatherHenkelJohnson1997.pdf Accessed January 2011

15. Baddeley, A. 1992. Working memory: The interface between memory and cognition. J. Cognitive Neuroscience 4, 3 (Jul. 1992), 281-288. DOI= http://dx.doi.org/10.1162/jocn. 1992.4.3.281

16. Miller Jr., R. G. 1986. *Beyond ANOVA, Basics of Applied Statistics.* John Wiley & Sons, New York. Accessed January 2011

17. Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proc. ACM CHI'94 Conf.* (Boston, MA, April 24-28), 152-158. Accessed January 2011