Applying controlled usability-testing technology to investigate learning behaviours of users interacting with e-learning tutorials

M.R. (Ruth) de Villiers
School of Computing, University of South Africa (UNISA)
dvillmr@unisa.ac.za

Abstract: The purpose of this paper is to suggest innovative ways of using the facilities in usability-testing laboratories to find out more about the learning processes and behaviours of users interacting with e-learning applications. The applications investigated in this study are tutorials that present cognitive subject matter. Three added-value techniques are described, namely: visualisation of how time is distributed in the learning process; verbalisation by participants, particularly co-participants; and methods of error analysis, drawing a distinction between usability errors and cognitive errors. The proposals are illustrated by data from three studies of interactive e-learning tutorials, studies which demonstrate the techniques and show their value in providing fine-grained details of the learning experiences. A notable finding is that different users learn from the software in very different ways. The research mechanisms are transferable to other domains.

Introduction

This is a position paper that explains the value of using usability-testing technology, not only to evaluate e-learning applications, but also to explore the process of learner interaction with e-learning applications, in this case interactive tutorials. It is a meta-study in that it is illustrated by salient extracts from two previous studies and a new study. The findings contribute to answering questions such as: What is the added value of usability-testing technology in studying interactive learning behaviour? What techniques and instruments can be applied in future, more extensive research?

In a literature review, e-learning is defined and an overview given of usability evaluation in general and the methods and techniques of usability testing in particular. An explanation is given of how evaluation of e-learning differs from evaluation of conventional systems. This is followed by a description of the context and research methodology of this study. The body of the paper proposes three added-value techniques, namely: visualisation of how time is distributed in the learning process; secondly, the worth of verbalisation by participants - particularly co-participants; and thirdly, methods of error analysis, drawing a distinction between usability errors and cognitive errors. Each of the three techniques is illustrated by aspects of previous studies. The concluding section summarises the findings and suggests avenues of future research.

Literature Review: Usability Evaluation and Usability Testing of e-Learning Applications

We first address definitions and connotations of e-learning and describe the view taken in this study. There are various definitions of e-learning, some that view it only as using the Internet and World Wide Web in teaching and learning, but others (CEDEFOP, 2002; de Villiers, 2005) that are broader and include multiple forms and methodologies: Internet and intranets, Web-based learning (WBL), multimedia CD-ROM, online courses, traditional computer-assisted instruction (CAI), and learning management systems. This paper takes the broad view, including interactive educational software; use by learners of computers as cognitive tools for exploration and construction; and e-learning tutorials, which are the systems on which this usability testing was conducted.

Evaluation of e-learning applications is a vital part of their design and development. Various usability evaluation methods (UEMs) exist, such as model-based methods, user-based surveys (questionnaires and interviews), inspection techniques (systematic criterion-based approaches where experts examine a system, e.g. heuristic evaluation), and observational methods. Observation can be conducted in naturalistic field studies or in formal usability testing, which involves observation, experimentation and data collection in a controlled laboratory environment. Although much general work has been done within the human-computer interaction (HCI) discipline on evaluation, there is scope for evaluation and metrics specifically for formal usability testing (UT) of interactive educational software (Van Greunen & Wesson, 2002).

Usability testing (Jeffries et al, 1991; Rubin, 1994; Dumas & Redish, 1999; Barnum, 2002; Dumas, 2003) is a formal software evaluation technique that involves measuring the performance of real end-users or potential users as they undertake defined tasks on the target application to determine whether it meets stated criteria. The application

tested may be an operational system or a prototype. Since UT elicits detail, it serves well in identifying and diagnosing problems and a main goal is to help improve the usability of products. Since the 1990s, it has been empirically conducted in specialized environments called usability laboratories, equipped with sophisticated monitoring and recording facilities and supported by analytical software tools. Participants are real users, whose actions are rigorously monitored as they interact with the product and are recorded on video – for subsequent re-viewing and analysis; by event logging – down to keystroke level; and by audio – to note verbalisation and emotions. Researchers observe the participants through one-way glass and on split-screen computer monitors, which simultaneously show participants' hands on the keyboard, their facial expressions, and the screen with which they are interacting. The quantitative and qualitative data is analysed in detail, and changes to the system can be recommended. Typical usability metrics include the time taken to complete a task, degree of completion, number of errors, time lost by errors, time to recover from errors, number of subjects who successfully completed a task, etc. (ISO 9241, 1997; Dix, Finlay, Abowd & Beale, 2004; Preece, Rogers & Sharp, 2007). The primary targets of UT are the user interface and other interactive aspects. Academics use UT for research and development, while corporate usability practitioners use it for rapid improvement of interfaces and system usability. However, although formal experimental testing is the best way of assessing real-world user performance, it is a complex and costly UEM.

Various studies report on the usability evaluation of e-learning via the Internet (Storey, Philips, Maczewski & Wang, 2002; Hwang, Huang & Tseng, 2004). This paper, as explained early in this section, relates to usability studies of a different form of e-learning, namely, CD-based tutorials. We describe evaluations that were conducted by the usability-testing UEM, but instead of evaluating a system *per se*, we propose ways of using the UT technology to obtain information about the learning and cognition processes as learners interact with tutorials. This paper does not present major data collection and analysis. Rather, it is a position paper that shows the added value of using the sophisticated technology of a usability lab not only to evaluate educational software and other e-learning systems but also to investigate how learners use them. The concepts are illustrated by extracts from three studies, where observation, recording and analysis tools were applied to study the behaviour of learners during learning experiences and to analyse how they spent time interacting with the applications – navigating through them, learning, and making mistakes as part of the learning experience. Before moving on to these findings, it is important to reflect on how evaluation of e-learning differs from usability evaluation of conventional systems.

How evaluation of e-learning is different
Conventional usability evaluation assesses the level of usability, based on the formal definition that usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO 9241, 1997). A conventional task-based system has a product as output, e.g. a payroll or a report, or it may be an environment that supports communication, such as an e-mail package. Evaluation of such software judges whether the system operates accurately and rapidly and whether it is easy and satisfying to use. But evaluation of e-learning applications has some notable differences. It is a maturing field that is only now becoming well documented by authors such as Ardito et al (2006), de Villiers (2006), and Masemola and de Villiers (2006).

First, e-learning applications are focused more on a process than on generating products, the process being the learning process. A second difference is that evaluation of e-learning should consider both usability and learning issues as it investigates users' experiences and pedagogic effectiveness (Squires and Preece, 1999). Evaluation criteria should address aspects such as quality of learning, effectiveness of the interface and interaction design. presentation of content, and usage of technology. As well as being computing artifacts, many e-learning applications are also course material. Supportive learning activities are essential to usability as well as to instructional functionality, according to Van Greunen & Wesson (2002).

Third, efficiency can not be determined by low task completion times. The process of working through content is related to a learner's approach, aptitude and personal learning style, as he/she reasons with the subject-matter or does exercises. Rapid closure to a learning task is not necessarily a 'good measurement', particularly where performance on exercises is disappointing. An e-learning application should foster personalized learning. Finally, and most important, it is not always desirable to minimise errors. We learn from mistakes, particularly in complex domains. In evaluating use of an application, evaluators should distinguish between peripheral usability errors and true cognitive errors, which are part of the learning experience (Squires & Preece, 1999). Usability-related errors should be diagnosed as problems and rectified, but cognitive errors should be permitted, provided that support mechanisms exist to promote a recognition–diagnosis–recovery cycle. According to Mayes and Fowler (1999:485), 'seamless fluency of use is not necessarily conducive to deep learning… the software must make learners think'.

Context and Research Methodology

The School of Computing at UNISA has its own usability laboratory, which offers enrichment and synergy to evaluation studies by its quantitative performance measurement and in-depth observation. The lab has video and audio monitoring equipment, event logging and eye-tracking, as well as software tools for analysis and reporting. The administrator/researcher controlling the session and observing participants is invisible behind sound-proof one-way glass and communicates via microphone. All verbalisation by the participant is heard in the control section.

We acknowledge the importance of ethical procedures (Rubin, 1994; Dumas, 2003). Unlike some other forms of evaluation, usability testing does not preserve anonymity, due to the video recordings. Participants are given information about the purpose and processes of the testing; they are asked to provide informed consent, and are assured they are free to withdraw at any time. We explain that the use of the software is being tested and not the participants and, furthermore, that the data will be used for research purposes only. We inform them that although recognition is possible in cases where pictures are shown or published, confidentiality is maintained and names are not disclosed. As a token of appreciation, participants are given a free copy of the e-learning tutorial.

The target systems tested in this study are interactive e-learning tutorials in mathematical aspects of Computer Science, which conform to Alessi and Trollip's (2001) definition of computer-assisted learning tutorials as software that 'present(s) information and guide(s) the student'. They are multi-media CD-based applications in cognitive domains that require critical thinking and rigorous reasoning on the third- (analysis), fourth- (synthesis), and fifth-(evaluation) levels of Bloom's (1956) Taxonomy. The tutorials are structured into topic-specific units, which present theoretical concepts and examples, interspersed with exercises for practice. There are no multiple-choice questions. Learners provide fill-in-the-blank responses by entering mathematical characters or a word/s. Some exercises require a composite series where more than one alternative is correct, provided it meets particular conditions. Diagnostic feedback and explanations are provided, and second attempts must be made after incorrect answers.

Figure 1 shows a screen view as seen by observers in the control room. It is a composite display of three views: the screen on which the participant is working, the participant's facial expressions, and a view of the keyboard.
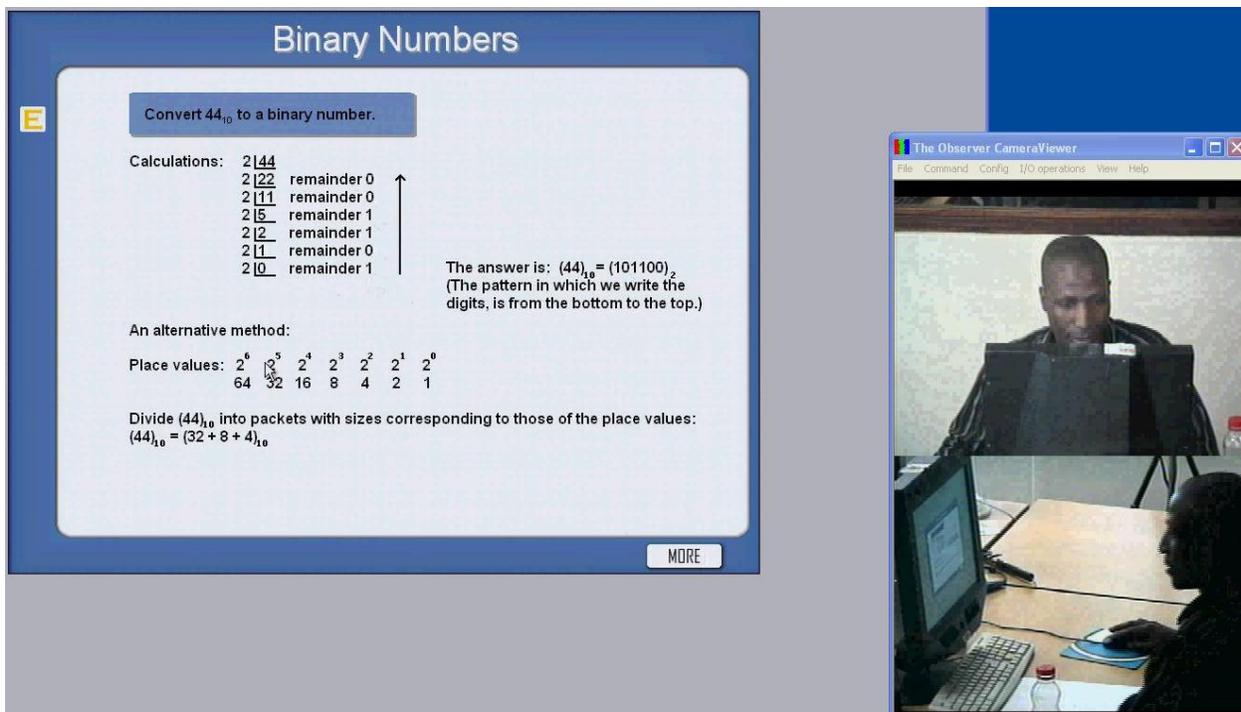

 Figure 1: Triple-view display seen by observers in control room. Clockwise from left: screen being used by participant; face of participant; keyboard and hand of participant (photo used with participant's permission)

The objective of this study is to identify ways in which the technology of a usability laboratory can be used over and above usability evaluation, to find out more about the actual processes of learning with technology. A meta-study in the next section refers to three research ventures that illustrate innovative use of usability-testing technology. Standard and non-standard techniques were used to deduce information about learning behaviours with e-learning applications. In each case participants were a representative sample of the target population.

Data extracted from these three studies is used to illustrate the proposals made and the techniques used. The aspect of time usage during learning has associated complexities. The time taken in accessing a specified task is a measure of usability, but the time spent interactively learning new computational skills is not a usability measure. It is a cognitive measure, indicating the participant's learning style and aptitude. We need to distinguish between ways of using time, and can do so by supplementing quantitative technological metrics with the qualitative approach of think-aloud protocols, whereby participants describe their reasoning processes and what they are doing (Shneiderman, 1998; Boren & Ramey, 2000; Dumas, 2003; Dix et al. 2004). Armstrong, Brewer and Steinberg (2002) sum it up by stating that the 'think-out-loud method' increases the quantity and relevance of data. Regarding functionality and usability, we work on the premise that usability of e-learning includes the effective attainment of learning. Successful cognitive processing is part of both functionality and usability. Therefore this research analyses data relating to learning tasks and exercises as well as investigating navigation through the interfaces.

Before outlining the techniques and illustrating them with extracts from the studies, a description is given of the policies and practical procedures for evaluation of e-learning applications in our laboratory. Usability testing sessions must be carefully planned. According to a framework proposed by Masemola and de Villiers (2006), the aspects to be measured and their associated metrics, as well as all necessary documentation, should be pre-compiled. Documentation includes the task list, an information document for participants, and the administrator's checklist. Representative participants must be obtained. Nielsen (2000) advocates five participants for an optimal tradeoff between use of resources and identification of problems. It can be done with as few as three (Shneiderman, 1998), while Dumas and Redish (1999) advocate six to twelve in subgroups. These sample sizes are too low for meaningful statistical analysis (Preece et al, 2007). A pilot test should be done, not to obtain hard data, but to try out the research design, tasks, measurements, equipment and the software. This gives the researcher or test administrator a chance to practice on trial runs. On occasions, the pilot data may be used for analysis, along with that from the main test. The procedures and documents are refined for the main test in the light of the pilot. Most important, means of analysis and presentation that address the unique, as well as the usual, aspects should be pre-determined (Masemola and de Villiers, 2006).

A single session should not last longer than 40 minutes to an hour to prevent fatigue on the part of the participant. Ethical aspects and the free-to-withdraw policy (they seldom do) must be explained, as mentioned in the previous section. It is advisable to have a short familiarization run before a session. Participants are often required to think aloud as they progress through tasks.

The usability-testing process is time intensive for the researcher. A one-hour lab session involves several hours of processing and analysis. Unlike survey research, where results can be mass-processed or electronically analysed, UT involves individual processing of each set of quantitative and qualitative data. Very seldom does event logging during a session provide sufficient data. The video and audio recordings must be viewed and re-viewed for manual interpretation. The data generated and analysed by the in-built UT software is frequently imported to other systems for further processing. Consequently, our samples are small – around 10 -15 subjects, too few for statistical analysis.

Findings about the interactive learning process – extracts from three usability-testing studies

The sets of data in the cases below, extracted from three different studies, are exactly what the sub-headers indicate, namely: 'illustrations' to demonstrate certain techniques. Studies 1 and 2 involved more data than the data mentioned here. The full studies are described in the original sources (cited in the appropriate subsections below), while Study 3 is a pilot study, yet to be extended to a major study.

A final comment to clarify the purpose of this paper before proceeding to the data: this paper describes meta-evaluative research to study the role and added value of UT techniques themselves and not actual evaluations of the e-learning tutorials. The two tutorials investigated, had been comprehensively evaluated previously, using a variety

of UEMs including end-user questionnaire surveys and interviews, heuristic evaluation, and pre-and post-testing (de Villiers, 2006, Becker & de Villiers, 2008). Thus they did not contain serious problems that could act as confounds.

This meta-study is based on extracts from usability studies done in our lab at UNISA. Selected data is used to illustrate ways of exploring and presenting sub-activities of learner interaction with e-learning applications. To show how usability-testing technology can be used in innovative ways, the author proposed time-usage pie charts for visualizing and comparing how learners distribute their time. To generate the charts, we total and tabulate time spent on sub-activities of learning with technology, and explore various learning activities by representing them in pie charts (Masemola & de Villiers, 2006; Adebesin, de Villiers & Ssemugabi, 2009). Such charts, shown in the next subsections, highlight factors such as different learners' performance on the same task; time spent navigating through a system versus time spent on actual pedagogic activities; occurrence of different types of errors; and so on. Furthermore, associated qualitative factors, such as emotions, facial expressions and utterances during learning experiences, are evident from the recordings and available for analysis.

We focus on three cases: (i) visualisation of time distribution in learning with technology; (ii) a study of verbalisation by participants - particularly co-participants; and (iii) methods of error analysis, drawing a distinction between usability errors and cognitive errors. These techniques illustrate the proposals for innovative added-value use of UT.

Illustration from Study 1: Time distribution during learning and visualization of such distribution of time

The first study undertaken on testing e-learning applications, used five participants in a pilot test and five in a main test. Its primary goal was the development of a framework for usability testing of tutorials in cognitive domains (Masemola & de Villiers, 2006). A secondary aim was further evaluation by UT on the tutorial, *Relations* that had been previously evaluated by other UEMs (de Villiers, 2006). In addition, as a first move to investigate learning processes in cognitive domains, the time-usage pie charts mentioned above, are used to visually represent the way different learners distributed their time on the same task, as shown in Figure 2 for three selected subjects, Participant 1 (P1), Participant 2 (P2), and Participant 3 (P3) respectively. The segments represent the percentage of time spent on different activities. Observation, both in real time and in re-viewing recordings, determined what participants were doing and identified their emotions: frustration, satisfaction, and the joy of achievement. But observation alone does not capture reasons behind the actions nor does it inform researchers about the internal decision making or cognitive processes. Hence the utility of the think-aloud protocol, which describes users' reasoning processes and activities. Together with mouse-clicks and the characters keyed in to activate animated concept development or to answer questions, the think-aloud protocol helps researchers to distinguish between time spent passively reading the on-screen subject-matter and time spent in active learning. The time spent 'typing' is shown separately but is mainly related to active learning, whereas 'clicking' was primarily used for navigation within the application.

P1 spent 1543 seconds in total, versus the 1119 of P2 and the 1208 of P3, represented by the relative sizes of the circular charts. In the separate test that followed the tasks, P1, P2 and P3 spent 763, 292 and 119 seconds respectively (not shown in the sub-activity pie charts in Figure 2), obtaining test scores of 65%, 80% and 85%. Figure 2 shows that, within the totals, time distributions between P1 and P2 were fairly uniform, with the main difference being P1's 16% on 'Click & Enter', indicating considerable time spent moving between screens. The other participant, P3, spent 1208 seconds, a similar time to P2 but their time-distribution profiles were quite different. This distinction between P2 and P3 is notable. They took similar total times for the session (1119 & 1208 seconds, respectively) and got similar scores in the test (80% & 85%). But time distribution varied dramatically: P3 spent more than half the time on Click & Enter, moving frequently and actively between different screens and instructional concept developments. This process, included under 'Click & Enter' in the figure, can also be viewed as a part of P3's 'Active learning'. P3 then did the test in 119 seconds, versus P2's 292, indicating a confident grasp of the cognitive content and scoring slightly higher than P2. P1 spent 763 seconds on the test and scored 65%.

No generalisations about learning can be made from this limited data, but it highlights the major differences between ways that individuals use the same e-learning application.
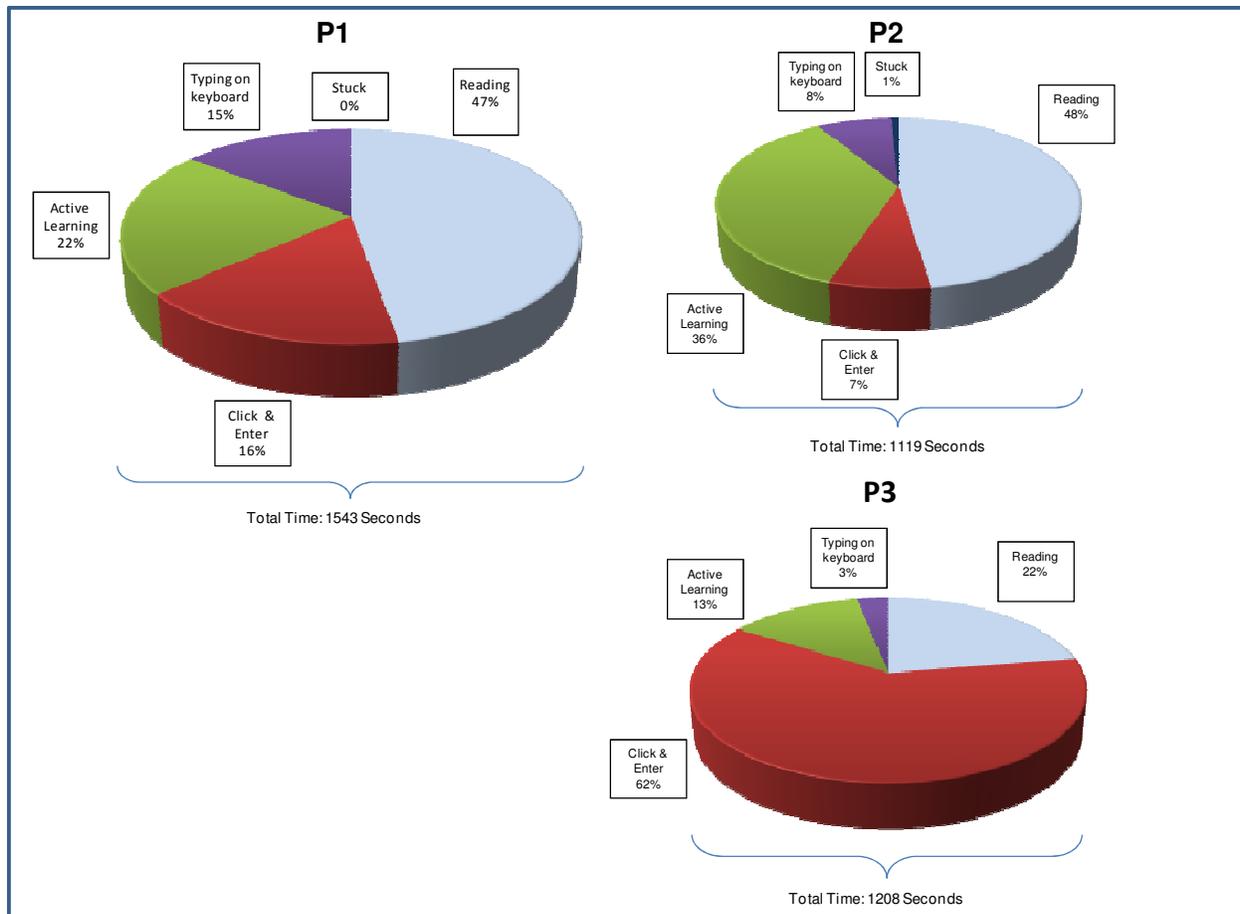
Figure 2: Time distribution on sub-activities of tasks by three participants using an e-learning tutorial

Illustration from Study 2: Time-distribution by pairs; enhanced verbalisation by the use of co-participation

According to Armstrong et al (2002), verbalisation informs the observer what participants wish to accomplish and their impressions at each step. This is particularly useful in analysing learning, which involves periods of non-interaction with technology. Many participants find it unnatural to think out loud, so up-front coaching with a video of effective think aloud is helpful. Administrators can intervene to remind them to think aloud and to re-direct them if they become totally 'stuck'.

An alternative approach is co-discovery (Wilson, 1998), which pairs two participants in a joint UT session. As they work through tasks collaboratively, natural conversation and interaction occur. One partner often helps the other by explaining a concept. Moreover, they are less inhibited about expressing opinions. Co-discovery is also useful in identifying causes of the problems encountered and strategies used to solve them. Co-discovery was used in a testing sessions on the e-learning tutorial, *Karnaugh* (Adebesin et al, 2009) that had been previously evaluated by other UEMs (Becker and de Villiers, 2008). Figure 3 depicts the collaborative activities of two pairs of co-participants, using new categories such as 'Explain' and 'Discuss'. G1 represents Group 1 with members, g1-1 and g1-2. G2 is Group 2 with members, g2-1 and g2-2. The sub-activity data shows fairly similar time distributions, but with some notable differences in learning styles and group dynamics. In both cases, peer-teaching occurred. In G1, g1-1, the 'stronger' learner took charge throughout: explaining, taking decisions and typing answers, but negotiating actions and not moving on until g1-2 understood. In G2, g2-1 and g2-2 took turns to be in charge. 16% of G1's time was spent on explanations which, even considering G1's shorter total time (1279 secs) than G2's (1583 secs), was still longer (211 seconds) than G2's bi-directional explanation time (131 seconds). G2, by contrast, spent considerable time reading and studying: 953 seconds (60%), versus G1's 673 seconds (53%). There is an interesting difference

between the ways the groups did exercises. In G1, the stronger partner did most of the work, while in G2, both partners did exercises independently then compared the answers before deciding what to type in. Actual times spent on exercises by the pairs were virtually identical: 174 seconds and 179 seconds respectively.
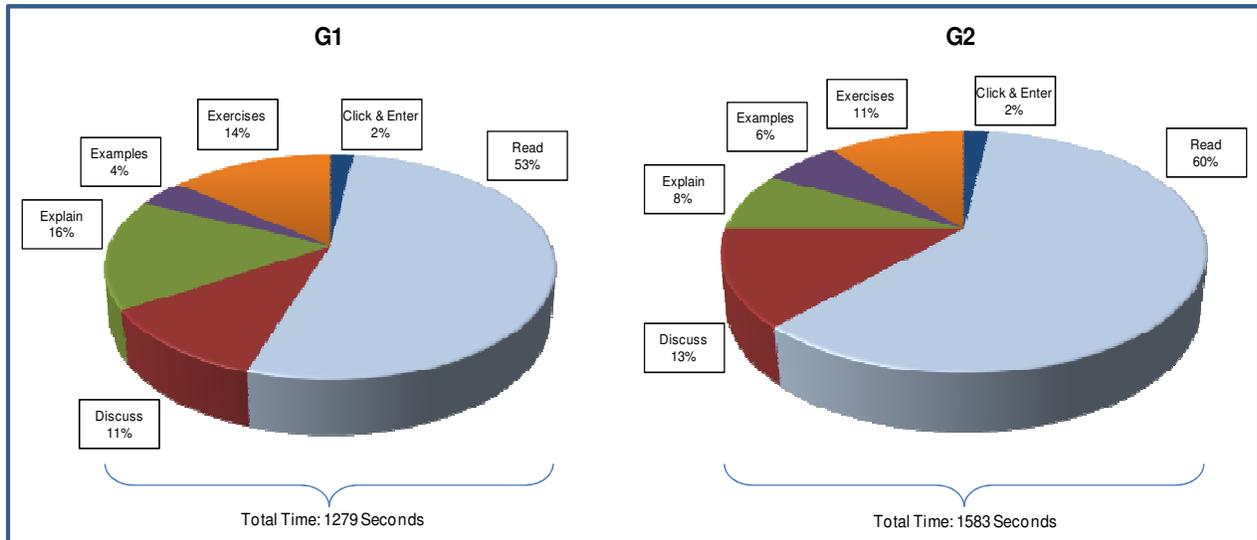

Figure 3: Time distribution on sub-activities by co-discovery pairs interacting with an e-learning tutorial

The purpose in this paper is to describe and illustrate novel ways of using UT to find out about learning processes, not to present and analyse the entire data of each study. But it is worth mentioning that five single participants also took part in Study 2, two of whom took less time on the assigned task than the co-participants, while three spent more time. As in the extract from Study 1, no general conclusions are made.

Illustration from Study 3: Error analysis: usability errors and cognitive errors

The final study involves work at an early stage on representing different kinds of errors. As was stressed earlier, Squires and Preece (1999) distinguish between usability errors and cognitive errors that are part of learning. In Studies 1 and 2, the time spent on errors was not separately noted, but was recorded within the containing activity, e.g. cognitive errors were included in active learning or exercises. The target application in Study 3 was, once again, the *Karnaugh* tutorial. Testing was done on a section that was a good choice for research on errors, because it had not undergone prior usability evaluation or refinements.

The metrics shown in Figure 4 highlight usability errors and cognitive errors, but show the cognitive errors alongside, and in the same colour as, the exercises because that is where they occurred. The participants represented in Figure 4 are termed P4 and P5 respectively. Their total times (920 and 2050 seconds, respectively) were very different, due mainly to four usability errors by that took up 27% of P5's time, one of them taking P5 off track for 400 seconds until the dead end was terminated by an administrator-intervention (one of six interventions for P5, taking 280 seconds in total). This particular error occurred because P5 selected the second sub-unit of a task instead of the first, due to the location of the first at a higher level on the screen. This was a usability error due to non-predictability. Without the required content as a foundation, P5 could not comprehend the material and did not progress   By contrast, P4 made two usability errors, recovered quickly and used 40 seconds altogether, only 4% of the time. The actual times spent in active learning, 310 and 260 seconds respectively, were similar, but the percentages vary due to the difference in total time. Times spent on exercises were separated into time working correctly and time making cognitive errors, as identified during re-viewing of videos. The percentages of time on exercises were: 13 + 21 = 34% and 17 + 9 = 26%, for P4 and P5 respectively, with total times on exercises correspondingly being 310 and 550 seconds. The reasons for the high total time P5 spent on exercises are the periods of uncertainty, and the time spent making mistakes. As in Studies 1 and 2, the facts and figures are not used to draw conclusions or generalizations. Rather, they are presented as illustrations of the analyses that can be done, the details that can be obtained, and distinctions between the two kinds of errors.
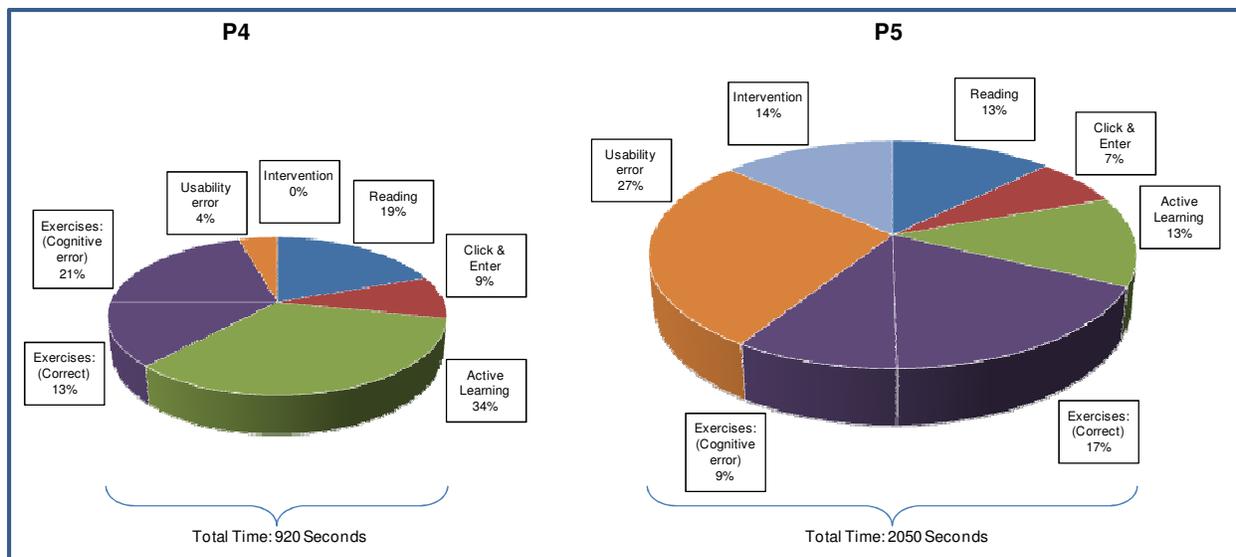
Figure 4: Time distribution on learning sub-activities, in particular cognitive and usability errors

Since this is a pilot study using only two participants, there is no quantitative data and findings cannot be generalized. However, it offers an opportunity for in-depth qualitative study of the two participants and their respective cognitive behaviours, approaches to errors, and recovery from errors. The researcher re-viewed the videos to note details of how time was spent and to study the body language and verbalizations of P4 and P5.

P4 learned easily and naturally via the electronic learning medium. Initially, he was unfamiliar with the content, but he engaged with the instructions and learning material, bending close over the screen when deeply engrossed. The position of the cursor showed his focus of attention. He handled think aloud well, reasoning effectively as he worked through computational content. P4's tone of voice varied, indicating pleasure, uncertainty, and confidence. His body language and vocalization explicated his systematic approaches, thoughts, actions and problems, for example: 'I am now trying to do this exercise'; 'I think the objective of this question is to…'; 'I read the instruction that I must…'. He also questioned himself rhetorically as he worked. After an incorrect attempt: 'What did "he" (himself!) do wrong?' Later, 'How do I get back?'. After he had grasped a process: 'OK, I think I now understand'.

P4 made a cognitive error and entered a wrong answer. He attempted to overwrite it, thus making a usability error, due to not noticing the instruction to clear wrong answers by using <Enter>. Frustration was evident on his puzzled face and from his comment, 'It seems to be impossible!'. However, he was able to solve all problems – both usability errors and cognitive errors – within a short time, maximum 20 seconds.

With respect to actual learning, P4 effectively used the scaffolding provided by the tutorial. After overcoming the usability error described above, he tackled the initial cognitive error by consulting the system's help and support. His systematic reasoning supported his problem-solving processes: 'And how do I get the result? Do I read from here to there or there to here…? Let's see… OK – from the bottom to the top'. He solved the problem quickly and exclaimed, 'I got it. Eh heh!'.

By contrast, P5 appeared to learn better by personal interaction with a facilitator than by e-learning support. The pie chart shows that considerable time was spent recovering from usability errors and cognitive errors, and that the test administrator and the researcher had to intervene. P5 started the session by moving rapidly and unsystematically between screens, reading screen text out loud, but not accessing the menu items needed for the required task.

The administrator intervened, referring to the main menu and advising P5 to start with <Background Info>. P5 did so, then used buttons and icons to cycle between the menu, a worked example and the required exercise, still reading out loud. When P5 began the exercise in earnest, the verbalization decreased and he/she made notes on paper, returning regularly to the example which could be controlled in a self-paced way. P5 tended to rush, not engaging deeply with the academic content. Progress was slow, but eventually, he/she obtained the correct answer, greeting the achievement with 'Yeah!' and a smile. By then, the time allocated to the full set of teaks was up. Nevertheless,

P5 started the second exercise. Progress was disappointing, so the researcher intervened with personal tuition and leading questions, referring him/her to the appropriate theory in the tutorial. After these prompts P5 began to engage better with the learning content, and moved a pen over material on the screen. The researcher intervened a third time and discussed the material with P5, who grasped the concepts rapidly and achieved the correct answer with delight.

In this case, verbal explanations and prompting led to rapid comprehension, and it appears that this participant is more comfortable with a conventional contact-learning approach. It is also notable that when P5 began to understand a complex section, he/she used a pen on the screen as though it was a book.

Conclusion and Future Research

The objective of this study was to suggest innovative ways in which usability testing technology can be applied to investigate learning processes, illustrating the techniques by extracts from three different studies. In conclusion, one may ask, at a meta-level, what was the added value of the technology in studying interactive learning behaviour? What techniques and instruments can be applied in future, more extensive research?

First, the sub-studies are integrated by their common use of time-usage pie charts but distinguished by the different emphases of these charts. Study 1 portrayed time distribution between active learning and supportive activities as identified by think aloud, while in Study 2, think-aloud verbalisation was extended to collaborative co-discovery. Study 3 focussed on error analysis and recovery, classifying mistakes as usability issues related to design flaws or as useful cognitive errors assisting comprehension. Throughout the studies, differences were noted between the ways in which participants used the applications for learning.

Second and more important, usability laboratory testing can include real-time event logging, but the detailed type of analysis required in studying human behaviour demands intricate processing of visual data and utterances. In this regard, the lab's triple-screen recording facilities enhanced the subsequent re-views during which fine-grained details emerged and were noted by researchers. The analytical software served well in processing the data and offered initial results that were improved by importing them into other presentation packages. Finally, because the applications investigated were e-learning tutorials in cognitive domains with computational subject matter, findings were made about appropriate and inappropriate approaches, activities and interaction, which can inform future design of such products.

Although no general conclusions about learning processes and activities can be drawn from the data presented here as illustrations, the techniques showed their value as research mechanisms in the application of controlled usability-testing technology. In line with the small number of participants and time-intensive analysis of usability testing, the three studies did not have sufficiently large samples to make generalisations.

Numerous avenues of future work suggest themselves. A large-scale study should have at least twenty participants in order to make tentative suggestions about learning behaviour and time-usage patterns. The participants in Studies 1 and 2 were all at a similar stage of study in the courses for which the tutorials were custom-designed, but their personal competence prior to the UT session was not assessed; correlations with prior knowledge should therefore be investigated in future studies. The research described in this position paper relates specifically to e-learning tutorials but is transferable beyond. Similar research could be undertaken in other areas, with the categories and activities being matched to the context. Such studies could be conducted on applying usability testing techniques and measurements to different kinds of e-learning, investigating, for example, time-usage and learning behaviours in web-based learning and open-ended constructivist environments. In addition, the type of research outlined in this paper could also be conducted in the domains of e-business, e-health and e-government systems to find out, in detail, how users interact with them.

References

Adebesin, F., de Villiers, M.R. & Ssemugabi, S (2008). Usability testing of e-learning: an approach incorporating co-discovery and think-aloud. *Proceedings of SACLA Conference 2009*. In press.

Alessi, S.M. & Trollip, S.R. (2001). *Multimedia for Learning: Methods and Developments*. 3rd edition, Allyn & Bacon.

Armstrong, S.D., Brewer, W.D, & Steinberg, R.K. (2002). Usability testing. In: S.G.Charlton & T.G. O'Brien (Eds), *Handbook of Human Factors Testing and Evaluation.* Mahwah: Lawrence Erlbaum Associates.

Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. & Rossano, V. (2006). An approach to usability evaluation of e-learning applications. *Universal Access in the Information society*, 4(3):270-283. http://www.springerlink.com/content/755507r7144m3845/

Barnum, C.M. (2002). *Usability Testing and Research*. Allyn & Bacon.

Becker, D. & de Villiers, M.R.. (2008). Iterative design and evaluation of an e-learning tutorial: a research-based approach. *South African Computer Journal, Special edition* 42: 38-46.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain*. New York: David McKay.

Boren, M.T. & Ramey, J. (2000). Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3): 261-278.

CEDEFOP (European Centre for the Development of Vocational Training) (2002). E-learning and training in Europe. Luxembourg: Office for Official Publications of the European Communities.

de Villiers, M.R. (2005). e-Learning artefacts: Are they based on learning theory? *Alternation* 12.1b: 345-371.

de Villiers, M.R. (2006). Multi-method evaluations: Case studies of an interactive tutorial and practice system. In: *Proceedings of the 9th Annual Informing Science + IT Education (InSITE) Conference.* Manchester, UK. University of Salford.

Dix, A., Finlay, J., Abowd, G.D. & Beale, R. (2004). *Human-Computer Interaction*. Pearson Education, Ltd, Harlow.

Dumas, J.S. (2003). User-based evaluations. In: J.A. Jacko & A. Sears (Eds), *The Human-Computer Interaction Handbook*. Mahwah: Lawrence Erlbaum Associates.

Dumas, J.S. & Redish, J.C. (1999). *A Practical Guide to Usability Testing*. Exeter: Intellect.

Hwang, G-J., Huang, C.K. & Tseng, J.C.R. (2004). A group-decision approach for evaluating educational web sites. *Computers & Education* 42:65-86.

ISO 9241. (1997). *Draft International Standard: Ergonomic requirements for office work with visual display terminals (VDT). Part 11: Guidance on Usability,* ISO.

Jeffries, R., Miller, J.R, Wharton, C. & Uyeda, K.M. (1991). User interface evaluation in the real world: a comparison of four techniques. Proceedings *ACM CHI'91 Conference*: 119-124. New Orleans, LA, April 1991.

Masemola, S.S. & de Villiers, M.R. (2006). Towards a framework for usability testing of interactive e-learning applications in cognitive domains. In: J. Bishop & D. Kourie (Eds), Service-Oriented Software and Systems. *Proceedings of the 2006 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)* 2006:187-197**.** http://portal.acm.org/results.cfm?coll=GUIDE&dl=GUIDE&CFID=23448572&CFTOKEN=68849601**.**

Mayes, J.T & Fowler, C.J. (1999) Learning technology and usability: a framework for understanding courseware. *Interacting with Computers* 11(5): 485–497.

Nielsen, J. (2000). Why you only need to test with five users. http://www.useit.com/alertbox/20000319.html.

Preece, J., Rogers, Y. & Sharp, H. (2007). *Interaction Design: Beyond Human-Computer Interaction*. 2nd ed. John Wiley & Sons.

Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests.* New York: John Wiley.

Shneiderman, B. (1998). *Designing the User Interface* (3rd ed.). Reading, MA: Addison Wesley Longman.

Squires, D. & Preece, J. (1999). Predicting quality in educational software: evaluating for learning, usability and the synergy between them. *Interacting with Computers* 11(5): 467–483.

Storey, M. A., Phillips, B., Maczewski, M. & Wang, M. (2002).Evaluating the usability of web-based learning tools. *Educational Technology & Society* (3)

Van Greunen, D. & Wesson, J L. (2002). Formal usability testing of interactive educational software: A case study. *World Computer Congress (WCC): Stream 9: Usability*. Montreal, Canada, August 2002.

Wilson, C. (1998). Usability techniques: pros and cons of co-participation in usability studies. http://www.stcsig.org/usability/newsletter/9804-coparticipation.html

Acknowledgements: