Usability Testing of e-Learning: an Approach Incorporating Co-discovery and Think-aloud

TF (Funmi) Adebesin
School of Computing, UNISA
Meraka Institute, CSIR
P O Box 395, Pretoria, 0001, South Africa
+27 12 841 3936
FAdebesin@csir.co.za

MR (Ruth) de Villiers
School of Computing, UNISA
P O Box 392, UNISA, 0003, South Africa
+27 12 429 6559
dvillmr@unisa.ac.za

Samuel Ssemugabi
School of Computing, UNISA
P O Box 392, UNISA, 0003, South Africa
+27 429 6241
ssemus@unisa.ac.za

ABSTRACT
Computer applications developed to support learning in the cognitive domains are quite different from commercial transaction processing applications. The unique nature of such applications calls for different methods for evaluating their usability. This paper presents the application and refinement of the framework for usability testing of interactive e-learning applications proposed by Masemola & de Villiers [11]. In a pioneering usability testing study, we investigate the effectiveness of the think-aloud method when combined with co-discovery testing.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *evaluation/methodology;*

H.5.2 [**Information Interfaces and Presentation]**: User Interfaces – *evaluation/methodology, Interaction styles, Screen design, User-centered design;*

K.3.1 [**Computers and Education**]: Computer Uses in Education – *computer-assisted instruction (CAI); Distance learning;*

General Terms
Design, Experimentation, Human Factors, Measurement, Performance.

Keywords
CAI tutorial, co-discovery, e-learning, think-aloud, usability evaluation, usability testing.

1. INTRODUCTION
Software applications should undergo usability evaluation to investigate their usability, which is defined in terms of their effectiveness, efficiency and the satisfaction of their users [1]. Various usability evaluation methods exist [8], [14], of which some of the most well known are predictive evaluation, heuristic evaluation, naturalistic observation, user-based methods such as questionnaires and interviews, and usability testing.

Academics at UNISA's School of Computing have designed and refined computer-aided instruction (CAI) tutorials to address complex sections in Computer Science modules. It is particularly important that usability evaluation should be conducted on e-learning applications. Whereas conventional software is frequently used by professionals and business people in the workplace, educational applications are used by learners who must first be able to use them before they can even begin to learn with them. E-learning applications should therefore be easily usable so as to support the learning process. This paper addresses usability evaluation of a CAI tutorial for a 1st level module, using the technique of usability testing (UT).

UT is a formal approach which is usually conducted in a controlled laboratory environment. It involves measuring the performance of typical end-users as they undertake a predefined set of tasks on the system being evaluated to assess the degree to which it meets specific usability criteria [11], [16] and to identify problems in the system. In the sophisticated technological environment of a usability laboratory, evaluators observe and record activities of participants through a one-way mirror and on a computer monitor/TV screen as they carry out the specified tasks. The observational data is usually stored for later review and further analysis and also to get insight into users' emotive reactions [14].

In addition to the classic method of UT, where a single user works independently, other UT approaches encourage participants to think aloud by verbalizing their thoughts, feelings, expectations and decisions while interacting with the application [9]. This can enable evaluators to understand the reasons behind users' actions, as well as explain misconceptions users might have about the system. Some participants find thinking aloud to be unnatural, and need up-front coaching in the form of a pre-recorded video of a session demonstrating effective think-aloud [4].

To address the unnaturalness involved with think-aloud, a variation of this technique, called co-discovery or co-participant testing (both terms used interchangeably), involves two users collaborating with each other while exploring the application being evaluated. The idea is that they verbalize their thoughts as they interact with each other and the application, using a single workstation. In this situation, the verbalizing is more natural, because it involves a conversation between two people [12], [19].

Furthermore, eye tracking can be combined with traditional UT. This involves sophisticated monitoring and recording of eye movements on different screen regions to determine whether important information is perceived by participants [15].

UT has traditionally been applied to task-based business systems. This study relates to UT in the context of evaluating e-learning applications where the emphasis is on the learning process, setting a different scene and approaches.

This paper aims to answer the following research questions:

1. To what extent is the usability testing framework for interactive e-learning applications (see Section 4.1) by Masemola & de Villiers [11] applicable to the present study?
2. How effective is the think-aloud method when combined with co-discovery testing?

2. BACKGROUND

2.1 What is e-learning?

Various definitions exist for e-learning. While some of these take a narrow view of the scope of e-learning, equating it only to Internet and Web-based applications, others are broader in their definitions [11]. Cedefop, cited in de Villiers [6], classifies e-learning as including multiple formats and methodologies. The wider definitions include all electronic learning technologies, whether Web-based or CD-based.

This study subscribes to the broader view of e-learning, recognizing that e-learning incorporates various forms of interactive educational applications, including traditional CAI tutorials and hypermedia,, educational games, simulations, open-ended learning environments, web-based learning, learning management systems, and the use of software tools to support learning.

2.2 The application tested

The target application, *Karnaugh*, is an interactive CD-based e-learning tutorial developed at the School of Computing, UNISA as optional supplementary learning material for the first-level module *Computer Systems: Fundamental concepts* (COS113W).

The application teaches learners how to use Karnaugh diagrams to simplify Boolean functions by applying a procedure based on a set of rules. *Karnaugh* combines learning theories from both the objectivist and cognitive paradigms. Its tight computational domain requires the behaviourist stimulus-response-reinforcement learning approach, while the learning content requires cognitive information processing and higher-order thinking skills [5], [6]. It integrates information presentation with question and answer sessions, judgment of learner response, and provision of feedback in line with Alessi & Trollip's [2] specification of what constitutes an e-learning tutorial.

The learning content of *Karnaugh* is divided into six sub-sections, an *Introduction*, *Background knowledge*, materials relating to *Sum-of-minterms*, an explanation of *Karnaugh diagrams*, instructions for the *Simplification of Boolean expressions*, and a sub-section that tests the level of knowledge gained by offering a *Testing game*. Learners can access the sections in any order, although those using the application for the first time are advised to go through the lesson linearly. Figure 1 shows the content of *Karnaugh*.

Figure 1. Content of *Karnaugh*

Most of the questions posed to learners are in the form of fill-in-the-blank, and feedback is provided following learners' responses. Correct answers are acknowledged and clues are provided for the rectification of incorrect ones. The current version of *Karnaugh* (V3.2) incorporates increased learner control, in the form of a multi-option control button that allows learners to work through the lesson in any fashion that suit them as well as selecting the level of difficulty of exercises. *Karnaugh* conforms well to applications classified by de Villiers [6] as full technologies within well-structured domains, where rules are tightly defined, each question has only one correct solution, and learning occurs through active interaction between the learner and the application.

The interface of *Karnaugh* underwent a major redesign in 2005, leading to V3.2, which combines text with new graphics and animations, hot words, and colour coding to emphasise important information. This new version was evaluated using various usability evaluation methods (UEMs), including heuristic evaluation, user-administered questionnaires and interviews [5]. The tutorial is therefore free of serious problems, so the purpose of this latest round of evaluation with UT as a method was more to explore UT in the context of e-learning, than to re-evaluate *Karnaugh* in and of itself with this new technology.

3. LITERATURE SURVEY

3.1 Usability

Usability relates to the development of interactive products that are easy to learn, effective to use, and enjoyable – from the user's perspective [14]. The ISO 9241 standard defines usability as 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 [1]. Using this standard, the usability of an application can be assessed through the three key quality criteria namely: effectiveness – a measure of how well the user is able to use the application to achieve his/her goal; efficiency – measures the speed with which the user can complete tasks; and satisfaction – which is a subjective measure of how pleasant it is to use the system [1], [12].

3.2 Usability in the context of e-learning

Computer applications developed to support learning in the cognitive domains are quite different from commercial transaction processing applications. The latter are developed to support fast completion of tasks and to avoid lengthy and repetitive execution processes. In contrast, e-learning systems aim to support learning through information transfer as opposed to information translation; to manage educational interaction; to provide support for human intellectual thought-processes; to help implement behavioural change; and to reduce focus on the delivery medium (technology) at the expense of the learning content (the message) [6].

The clear distinction between these two types of applications may well call for modification to the traditional interaction paradigm in order to address the unique usability requirements of e-learning [3]. For instance, some form of flexibility could be provided so that learners can omit sections of the learning content that is not relevant to them. The design of commercial task-based applications is concerned with user-centered design (UCD), where designers can assume some elements of homogeneity among users. However, in e-learning, the focus is on learner-centered design (LCD). E-learning applications are targeted at learner groups that are heterogeneous, with different computing backgrounds, learning styles, levels of experience and motivation; hence, the application should be able to address differences in usability needs of learners. The application interface should be well structured so that it provides easy and efficient navigational methods as well as customization of content to suit learners' requirements. In addition to being courseware, e-learning applications are also computer systems. Learners must master their use before effective learning can begin. However, they should not need to spend substantial amounts of time figuring out how to use the application. The interface should be intuitive, so that even novice users can begin meaningful interaction immediately [3].

Apart from conventional usability criteria, e-learning applications should also be evaluated for pedagogical effectiveness. Such pedagogical usability includes the provision of appropriate tools, content, interfaces, and tasks to support different learners within various learning contexts, according to the learning objectives. It is inadvisable to separate usability and pedagogic aspects in such evaluation, since the two are closely related [11]. In other words, when evaluating educational systems, usability cannot be considered in isolation without addressing content and learning functionality.

There are further important differences between UT of conventional task-based applications and UT of e-learning. In e-learning, low completion time for tasks cannot be used to indicate application efficiency, since people have different ways of learning. Rapid completion is thus not necessarily a good measure. In addition, the approach to errors is quite different. In e-learning, a distinction is made between usability errors and cognitive errors; usability errors that distract from quality learning should be prevented but cognitive errors, which form part of the learning process, are permitted because they facilitate higher-order thinking. People learn by making mistakes and trying to correct them. However, learner support should be provided for the recognition, diagnosis and recovery from such errors [11], [17].

3.3 Usability evaluation of e-learning

A number of researchers, including [3], [11], [17], [18], having recognized the uniqueness of e-learning, propose various approaches to evaluating it.

In pioneering work, Squires & Preece [17] developed a set of criteria, called 'learning with software heuristic' that are based on Nielsen's [13] heuristics and the socio-constructivist learning paradigm. This set of heuristics provides educators with an effective tool for assessing the quality of educational software both for interface usability and pedagogical effectiveness.

With regard to UT steps, tasks and procedures, Masemola and de Villiers [11] propose a framework, discussed in Section 4, for UT of e-learning applications. Their study involved formal testing in a controlled laboratory environment, with learners carrying out specified tasks using an e-learning tutorial called *Relations*. Valuable metrics in the form of time usage pattern allowed distinction to be made between time spent navigating and time spent on learning activities. Data analysis provided insight into differences in learning styles.

As an extension of the Masemola and de Villiers UT framework for e-learning, the present study applies and refines it using both single user testing and co-discovery testing.

4. RESEARCH METHODOLOGY

4.1 Research design

The design of this research is based on:

1. A variant of action research (AR), involving several cycles of evaluation and hands-on involvement by researchers [5], but undertaken by different researchers at different stages, resulting in modification and improvement of *Karnaugh*.
2. An underlying procedural model, namely, the Masemola & de Villiers framework for the UT of interactive e-learning applications in cognitive domains [11].

AR, by its nature is cyclic, with multiple iterations; it is participative, requiring collaboration between the researcher and other stakeholders with the researcher taking an active and central role in the data collection and analysis. It generates data which is more qualitative than quantitative. Insightful planning, monitoring and analysis are fed back into future iterations and the cycle closes in as a solution is attained [7].

The framework by Masemola & de Villiers involves the following steps:

1. Set up research objectives.
2. Determine which aspects are to be measured and their metrics.
3. Formulate the documents required – these includes the initial test plan, task list, information document for participants, checklist for test administrator, and post-test questionnaire to measure learner satisfaction.
4. Recruit representative participants.
5. Conduct a pilot test.
6. Refine the test plan, task list, and information document, based on the knowledge gained from the pilot.
7. Conduct the main usability test.
8. Determine means of analysis and presentation that address the unique issues of e-learning, as well as the usual aspects.
9. Draw conclusions and provide recommendations.

4.2 Usability metrics

In general, the two forms of usability metrics recorded are quantitative data – aimed at measuring user performance, and qualitative, subjective data – aimed at measuring users' perceptions about the application [4], [12].

For this study, the following performance measures were taken:

1. Number of mouse clicks,
2. Number of errors,
3. Number of repeated errors,
4. Number of calls for help,
5. Time spent reading (determined from think-aloud),
6. Time spent on learning/peer teaching (determined from think-aloud and from co-discovery),
7. Time taken to complete tasks, and
8. Number of correct answers.

In addition to the preceding quantitative metrics, each participant in the study was given a post-test questionnaire to complete. This enabled us to assess their perception of *Karnaugh*.

5. DATA COLLECTION AND ANALYSIS

5.1 Test participants

Participants in this study were drawn from students registered for the module COS113W. As a further requirement, those who took part in co-discovery testing were required to belong to a study group for the module. Acquiring participants for co-discovery testing can present additional challenges, in that it may be necessary to consider the level of expertise of each member of a pair as well as the existing familiarity between such members [10], [12].

A total of sixteen learners participated in the pilot and main studies. Participants were selected such that they were representative of the diverse learner population for COS113W. There were two participants for the pre-pilot test, seven for the pilot, and a further seven for the main test. Both the pilot study and the main study included a set of co-discovery participants.

For the main test, five of the participants were males and two were females. Five were full-time students, while the other two study part-time. All seven students were registered for a programming module. Table 1 provides the profile of the main test participants. (The co-participants in the pilot test are indicated as participants 8 and 9 respectively. This is because their results have been included in the main test findings for co-participant testing – see Section 5.4).

Table 1. Profile of main test participants. (The * indicates a co-discovery participant)

No	Age	Computing experience	Status	Qualification
1	28	Average	Part-time student	Matric
2	31	Average	Full-time student	Diploma
3*	19	Minimal	Full-time student	Matric
4*	20	Minimal	Full-time student	Matric
5	22	Advanced	Part-time student	Matric
6	24	Minimal	Full-time student	Matric
7	19	Minimal	Full-time student	Matric
8*	21	Minimal	Full-time student	Matric
9*	21	Average	Full-time student	Matric

5.2 Ethical considerations

While usability testing might not expose participants to physical danger, the controlled nature of the test environment, together with the presence of recording cameras and sensation of being observed, can be a source of distress to some participants who might feel pressured to perform well, even though they had been told that it was the application being tested and not them [12].

Before starting each testing session, participants were provided with information documentation that detailed the purpose of the study. It was explicitly explained upfront that *Karnaugh* was being tested, not their ability. Participants were assured that their participation was voluntary and they could withdraw at any time during the session without negative repercussion. They were also informed that, even if they withdrew, they would still receive a free copy of *Karnaugh* as promised. Each of our participants signed an informed consent document prior to the commencement of the testing sessions.

5.3 Pilot study

UT involves specialized evaluation techniques. Since the primary author was a novice in UT, two pilots were conducted. The first, a pre-pilot involving two learners, provided first-hand experience in operating the specialized equipment in the usability laboratory. Thereafter, the official pilot was done to try out the framework and see whether participants were able to complete the tasks as set out in the task list. Lessons learnt were incorporated into the planning of the main test. In particular, the pilots enabled us to modify the screening process for participants such that we could identify suitable participants for co-discovery testing.

In UNISA distance-learning context, it is difficult to obtain participants, so a call-for-volunteer was made in a tutorial letter. From the volunteers, a heterogeneous group was invited, but focused on acquiring co-participants in the form of two sets of pairs who worked together in study group.

5.4 Main study

Following modifications to the testing process as informed by the pilot study, we set out to conduct the main test. Participants were required to carry out the specific tasks in the task list (see Figure 2). No specific task completion time was pre-set, though we allocated a two-hour time block for each session. This acknowledged differences in learning styles and the cognitive nature of *Karnaugh*. Participants were also encouraged to take short breaks in between tasks.

In task 1, participants were required to access the section 'Karnaugh diagrams' (refer back to Figure 1) from the main menu page and study its learning content. They were also required to do the three exercises interspersed between teaching segments. Task 2 required learners to do exercises by answering 'testing' questions until they had filled a total of sixteen squares with 'smiling faces'. They could choose from a set of questions with 'face values' ranging from 2 to 8.

As stated in Subsection 5.1, only one co-paired set of participants was available for the pilot and one set for the main test, respectively. As a result of difficulties in recruiting participants who met the requirement for co-discovery testing, we combined the co-discovery results from the pilot and main tests to enable comparison between the two sets of co-participants. The results from the testing sessions are discussed in Subsection 5.5.

5.5 Findings

The sample of nine participants in total is too small for statistical analysis; therefore qualitative interpretations were made mainly on the co-discovery aspect.

Use by participants of the think-aloud method is a vital requirement for measuring use-of-time patterns, i.e. recording the time spent in navigation and reading screen information and the time spent on actual learning activities. Both sets of co-discovery test participants found thinking aloud and discussion of their joint activities during the sessions to be a comfortable process. This is to be expected, since it emerged naturally as conversation between two people.

In contrast, the single test participants were silent for the greater part of the sessions; they found it difficult to think aloud, despite prior coaching on the process. Various authors, including [12] and [19] have acknowledged this problem that occurs with single-participant think-aloud.

There are two tasks to be completed in this task list; please complete the tasks in the specified order.

Note: Please inform the test administrator after you have completed each task.

Task 1:

- From the title page go to **menu**
- From the **main menu** page go to "**Karnaugh diagrams"** sub-menu
- You are required to study the learning content presented in this section and do the associated exercises.

Task 2:

- From the **main menu** go to "**A testing game"** sub-menu
- Read the instruction on how to select and answer questions
- Do the actual test (**Note**: This version of the **"testing game"** is a subset created specifically for this usability testing study)

Figure 2. Task list for UT of *Karnaugh*

5.5.1 Results of co-participant testing

As indicated in Subsection 5.4, the co-participant results from the pilot and main tests were combined to enable comparison of the performance of the two groups. For easy reference, we label the co-participants in the pilot 'Group A' and the two participants GAP1 and GAP2, i.e. participant 1 in group A and participant 2 in group A. Similarly, we called the set in the main test 'Group B' and the participants GBP1 and GBP2 respectively. Table 2 shows the performance metrics for the activities of the two groups in task 1.

During the testing sessions, we observed a number of learning styles and group dynamics, which, in our opinion, can contribute to collaboration and peer-teaching among learners. Observations include the following:

- In the case of group A, GAP1 took charge of the computer for the two tasks, making all required navigational decisions and typing of answers to exercises. For group B, GBP1 and GBP2 swapped roles, with GBP1 taking charge for task 1 and GBP2 taking charge for task 2.
- Both groups negotiated their progression through the lessons, with the participant who was controlling the computer asking the partner if they could progress.
- In group A, GAP1 appeared to be the 'stronger' student; however, this participant did not move on until the 'weaker' one (GAP2) had fully comprehended a given concept. This is

reflected in Figure 3, which shows the use-of-time patterns for group A as they complete the activities in task 1.

As shown in Figure 3, participant GAP1 used 16% of the total time explaining learning concepts to her partner. This is in contrast to group B, where 8% of the time was spent by GBP2 explaining certain concepts to GBP1 (see Figure 4)

- As shown in Figures 3 and 4, the time distribution patterns between the two groups are fairly similar, with participants in group A using 53% of their total time reading learning-related content and instructions, while those in group B used 60% of their time on the same activity. Two-way discussion of concepts consumed 11% and 13% of time respectively; while 14% and 11% was taken up by answering section exercises.
- Both groups used a combination of mouse and manual touching of the screen, where the participant who did not handle the mouse did the hand-touching.
- Both groups used paper and pencil for rough work and calculations.
- Both groups worked out their own solutions to segment examples in task 1 before calling up the system-provided answers.
- For the exercises in task 2, each of the participants in group B worked on the exercises independently, then compared them before typing or selecting their answers. However, in the case of group A, the 'stronger' participant (GAP1) did most of the work, as well as taking time to explain reasoning behind the answers to her partner. This is reflected in Figures 5 and 6 by the major time difference in the only common question (Q3*) answered by both groups, where the group A pair spent 564.2 seconds on Q3 while their counterparts in group B spent 298.2 seconds on the same question.

Table 2. Co-participant performance metrics for task 1

Working through the content of Karnaugh diagrams, including section exercise (time shown in seconds)		
Activity	**Group A**	**Group B**
Number of mouse clicks	44	42
Number of calls for help	0	0
Number of usability errors	0	0
Number of correct answers	3	3
Navigation time	20	25.9
Time spent reading	673.4	952.6
Time spent on discussion	144.6	206
Time spent explaining concepts	211.2	130.8
Time spent working on examples	55	89
Time spent on section exercises	174.4	178.6
Time spent learning/peer teaching	585.2	604.4
Total completion time	1278.6	1582.9

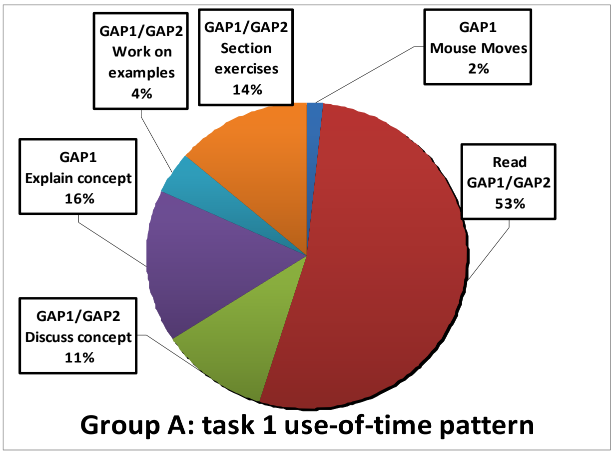

Figure 3. Group A: Time distribution, task 1

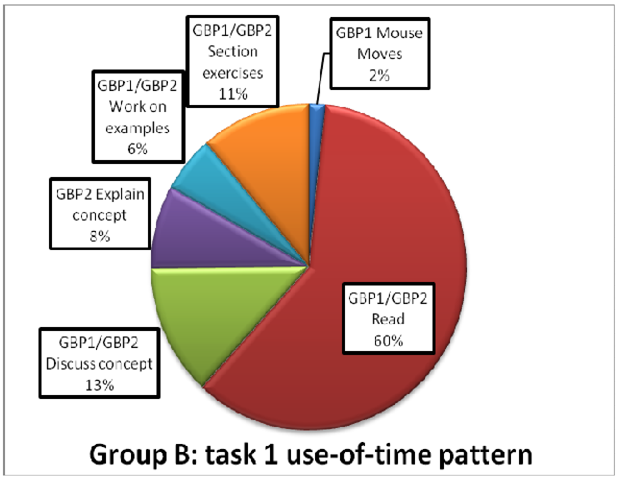

Figure 4. Group B: Time distribution, task 1

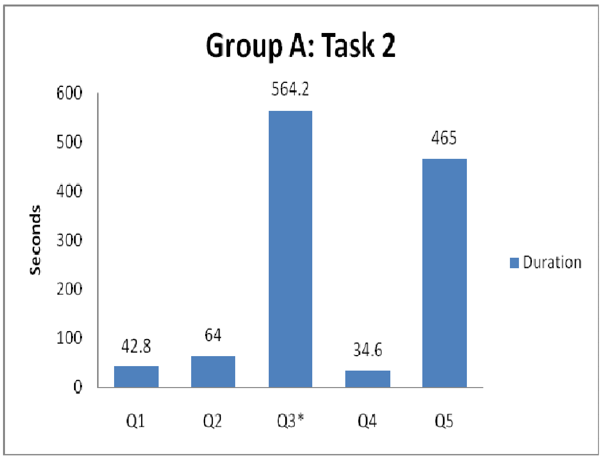

Figure 5. Group A: Time distribution, task 2

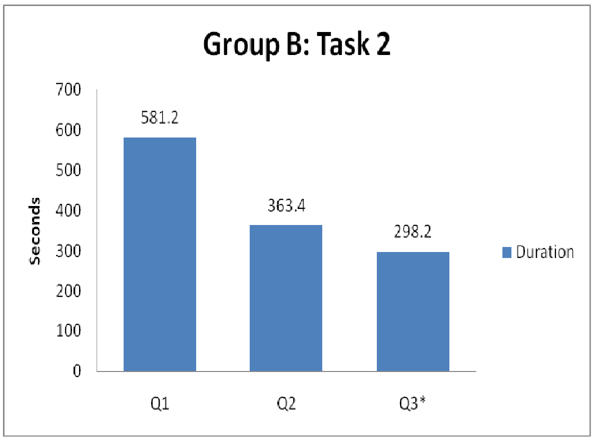

Figure 6. Group B: Time distribution, task 2

As shown in Table 3, group A participants answered 5 questions and used a total of 1170.6 seconds to complete them while the peer in group B answered 3 questions, spending 1242.8 seconds on the questions. It is notable that both groups gave correct answers to the questions on their first attempts.

Table 3. Co-participant performance metrics for task 2

Working through exercises in task 2 (time in seconds)			
Group	**Number of questions answered**	**Number of correct answers**	**Total time spent on all questions**
A	5	5	1170.6
B	3	3	1242.8

5.5.2 Results of single-participant testing

Due to the difficulty experienced by the single test participants in thinking aloud during the testing sessions, we were unable to distinguish between time spent on reading and navigating, and time spent actually learning concepts. Tables 4 and 5 show the basic performance metrics for single participants for activities in tasks 1 and 2 respectively. Of all five participants, only one (participant 6) did not require some form of content-related help from the test administrator (who was the primary researcher) while completing task 1. In each case, time spent on help was deducted from the total completion time. Participants 2 and 5 struggled most with content comprehension, which resulted in considerably more time spent on task 1 in comparison to the others. In task 2, the number of questions answered by the participants ranged from four to ten, with four common questions answered by all. In a similar pattern to task 1, participants 2 and 5 spent the most time on task 2.

Table 4. Single participant performance metrics for task 1

Working through the content of Karnaugh diagrams, including section exercise (time shown in second)					
Activity/ Participant	**1**	**2**	**5**	**6**	**7**
Number of mouse clicks	44	85	50	43	58
Number of calls for help	2	2	1	0	1
Number of usability errors	0	0	0	0	0
Navigation time	40.6	70.4	38.6	37.1	48.2
Intervention time by test administrator	25.2	26.6	29.4	0	139.6
Time spent on section exercises	62.2	952.6	270.2	89	437.2
Total completion time	426	3593.4	2437.2	813	1430

Table 5. Single participant performance metrics for task 2

Working through exercises in task 2 (time shown in second)					
Activity/ Participant	**1**	**2**	**5**	**6**	**7**
Number of usability errors	1	2	1	0	1
Number of questions answered	9	10	7	4	5
Number of correct answers	5	2	4	4	4
Total time spent on questions	804.2	2655.5	3378.4	1018.4	1223.2

5.5.3 Comparison between single and co-participant testing

One of the most notable findings was the natural way in which the co-participants were able to think aloud, compared to single participants. It was also of interest that, while the single participants required content-specific assistance from the test administrator, in both sets of co-participants, this type of help was provided by one of the participants. The former highlights an advantage of co-discovery testing, while the latter shows the value of collaborative learning, especially for e-learning situations, where learners work without face-to-face contact with the educator.

Another interesting observation was the time spent completing the activities in tasks 1 and 2. As shown in Table 2, the co-participants in groups A and B spent a total of 1278.6 and 1582.9 seconds respectively on task 1. However, when times spent on discussions with partners and explanations of concepts are deducted, accounting for 355.8 and 336.8 seconds respectively, the total times are reduced to 922.8 seconds for group A and 1246.1 seconds for group B.

Comparison of time spent on tasks by single and co-participants indicates that co-participant testing does not necessarily increase the duration of time. For instance, as shown in Table 6, only participants 1 and 6 spent less time on task 1 than the co-participants, and participants 2 and 5 spent twice as much time as the co-participants on the same task! In fact, co-participant testing may even result in less time as participants assist each other through the learning content. It should be noted however, that no general conclusions can be drawn from these observations, due to the small number of co-participants involved in the study.

Participants who used paper and pencil for rough work performed well in the exercises, both single participants and co-participants.

5.5.4 Learners' perception of Karnaugh (single and co-participants

Post-test questionnaires comprising semi-structured questions were given to each participant to measure their subjective perception of *Karnaugh*. Participants who performed relatively well viewed the structure of *Karnaugh* as 'very easy' or 'easy' while those who did less well found it 'very difficult' or 'difficult'. All the participants liked the use of colours to emphasise important information. Two students found the different representations of Karnaugh diagrams to be confusing (the tutorial provides four different ways of representing Karnaugh diagrams while the printed study guide provides only one representation). All the participants viewed the tutorial as a useful learning tool and would like to have similar tutorials for other modules. Positive comments were made, such as 'I would like to thank the Computing department for such CAI tutorial'.

Table 6. Comparison of time spent on tasks 1 and 2 by participants

Duration of tasks 1 and 2 (time shown in seconds)		
Participants	**Time on task 1**	**Time on task 2**
Group A	922.8 (excluding discussions and explanations)	1170.6
Group B	1246.1 (excluding discussions and explanations)	1242.8
1	426	804.2
2	3593.4	2655.5
5	2437.2	3378.4
6	813	1018.4
7	1430	1223.2
Average time	1552.6	1641.9

5.5.5 Usability problems revealed by the study

Although *Karnaugh* had been previously evaluated using various UEMs such as heuristic evaluation, questionnaire surveys and interviews [5], formal UT involving real users doing real tasks showed some previously unidentified usability problems that could distract from learning. Figure 7 is a simultaneous triple-screen display that illustrates the monitor view seen by the test administrator during UT. The displays, clockwise from left are: (i) current screenshot of the application being evaluated, (ii) the keyboard – at the instant being recorded, the participant's hands are not on the keys, and (iii) the participant's facial expression with his hand near his face. Displays such as this provide a holistic view to the test administrator during the evaluation.

This subsection highlights some of the problems identified and makes recommendations for improvement.

- The meaning and use of the multi-option control button, offering the composite functionality of <I>, <E>, <Q_B>, and <Q_C>, representing direct access to information, worked-out examples, selection of a basic question, and a complex question, respectively, are not intuitive. Although this useful feature provides increased learner control and was rated highly in an earlier questionnaire survey [5], none of the seven participants in the main test used it. This is possibly due to insufficient in-context information regarding its functionality. Although an introductory section of *Karnaugh* provides the meanings of the button in a decontextualized way; learners have no easy access to this once they have commenced the lesson. Designers cannot expect learners to remember the meanings throughout their interactions with an application. We advised the designer to use roll-overs to emphasise the meanings within the button.
- A participant was using a screen that offers phased development of a concept. Each click on the <More> button provided further information. He wanted to review the content of that screen and clicked the <Back> button, expecting to return to the beginning of the progression and watch the concept development again. He was, however, returned to the last exercise of the previous segment. The tutorial designer had provided a <Repeat> button for his required purpose which he overlooked, since it created a mismatch between the designer and learner's model, which is in violation of a Squires and Preece [17] heuristic. In addition, the <Repeat> button was greyed out on the screen. We recommended that adequate information should be provided on the intended use of each button in clear and unambiguous terms. The <Repeat> button, which had been erroneously greyed out, should be re-activated to make it accessible.
- In one of the exercises in the 'Testing game' section, four participants provided answers that were longer than the space provided in an answer text box. To check their answers, the participants had to use the arrow key on the keyboard because the incomplete answers were not visible. While we acknowledge that the short space provided for learner response is often an important clue for the required length of

the answer, we recommended that the number of allowable characters should be fixed as a form of forcing function, thus improving the visibility of the answer.

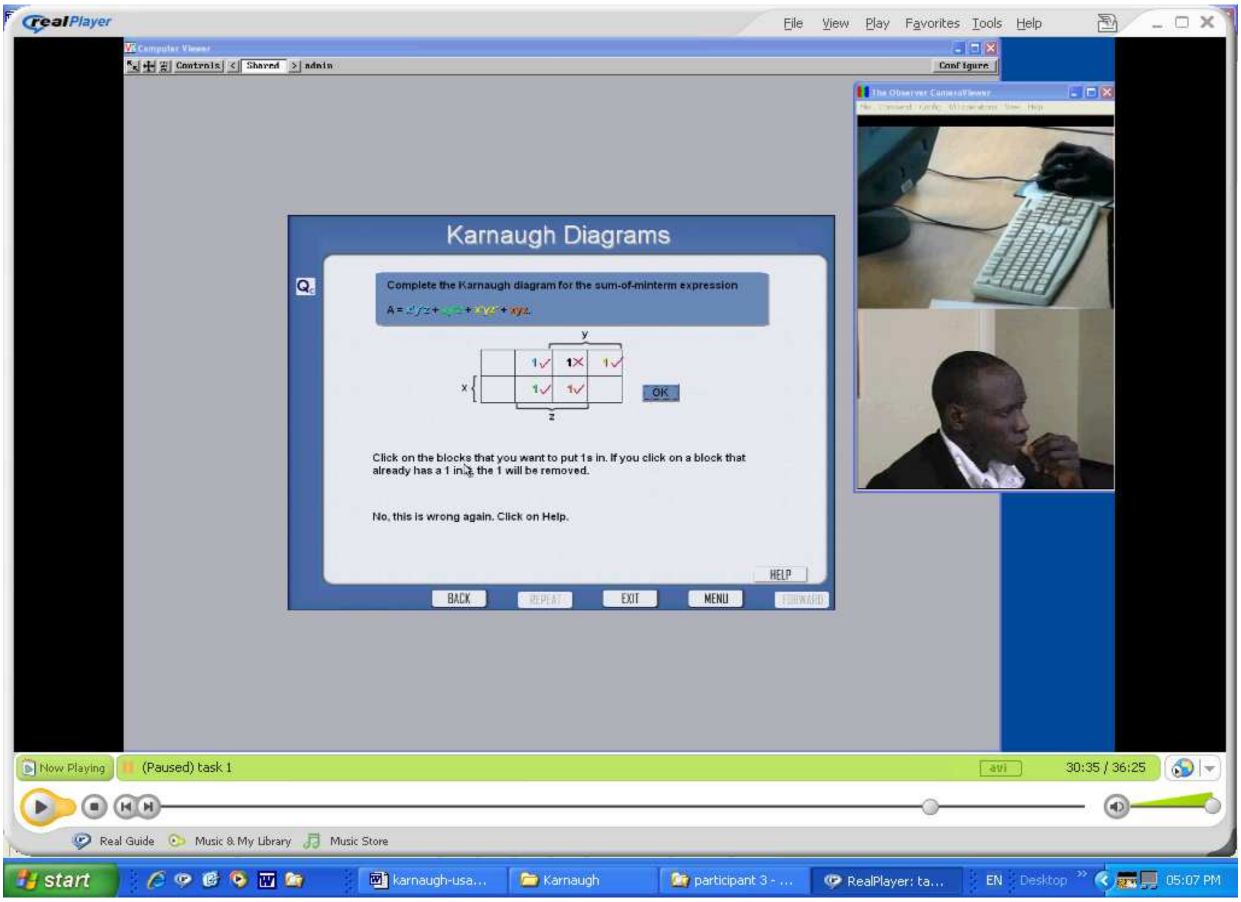

Figure 7: Triple-screen video showing poor feedback to incorrect answer (picture used with permission of participant)

- In certain instances, feedback for incorrect answers were not context-specific; for example, as shown in Figure 7, the learner inserted an extra '1' in the diagram in addition to the four correct '1s', but the feedback, 'No, this is wrong again. Click on Help', was not sufficiently informative. We advised that feedback should be as specific as possible to assist learners in making the necessary corrections.

Our recommendations for improvement were adopted and the appropriate corrections have been made to *Karnaugh.*

6. CONCLUSIONS AND FUTURE RESEARCH

This section re-visits the research questions posed in Section 1 and provides answers:

1. To what extent is the UT framework for interactive e-learning applications by Masemola & de Villiers [11] applicable to the present study?

The framework provided a useful and effective working structure for the study. To adapt it to the situation, two pilot studies were conducted instead of one, so that the test administrator could acquire the necessary expertise. In addition to the testing of single participants, two sets of pairs were involved in the testing process as co-participants. Certain positive group dynamics were observed during the pair interactions (see Subsection 5.5.1).

2. How effective is the think-aloud method when combined with co-discovery testing?

One of the primary aims of this study was to investigate the ease with which single and co-participants in UT could think aloud. Thinking aloud is an essential requirement for distinguishing between time spent reading-and-navigating and time spent studying/processing content. Single test participants struggled with think-aloud and remained mainly silent during testing sessions, making it difficult to make such distinctions. However, for the co-participants, thinking aloud came naturally because it involved two people having a conversation. In addition, co-participant testing does not necessarily result in increased amount of time spent on activities.

The two observations of co-discovery revealed that co-participant testing has the potential to reduce the level of intervention by the test administrator, and is especially relevant for testing e-learning, where collaboration is currently promoted as a useful form of learning. Although the number of pairs is too low for reliable generalizations, both cases demonstrated that it is possible for learners to assume the role of peer-teacher. Extrapolating beyond the immediate context of usability testing to the situation of learners in a distance-teaching context such as UNISA, collaborative learning and peer-teaching could play valuable roles.

The limited number of participants involved in the study did not allow for statistical analysis of learner performance. Future empirical research involving a greater number of participants, both single- and co-participants (but particularly the latter) could be undertaken to validate the findings of this study.

7. ACKNOWLEDGMENTS

We express our sincere appreciation to the COS113W module leader, Mrs. Daphne Becker, for her support and assistance in this study and to the students who participated in the research.

8. REFERENCES

[1] Abran, A., Khelifi, A., Suryn, W. and Seffah, A. 2003. Usability Meanings and Interpretations in ISO Standards. *Software Quality Journal*, 11(4): 325-338. http://www.springerlink.com/content/t01h116568q48352/.

[2] Alessi, S.M. and Trollip, S.R. 2001. *Multimedia for learning: methods and developments*. 3rd edition. Massachusetts: Allyn & Bacon.

[3] Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. and 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/.

[4] Barnum, C.M. 2002. *Usability Testing and Research*. Massachusetts: Allyn & Bacon.

[5] Becker, D. and 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.

[6] de Villiers, M.R. 2005a. e-Learning artefacts: are they based on learning theory? *Alternation* 12.1b: 345-371.

[7] de Villiers, M.R. 2005b. Three approaches as pillars for interpretive Information Systems research: development research, action research and grounded theory. *Proceedings of the 2005 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)* 2005:142-151. http://portal.acm.org/citation.cfm?id=1145675.1145691.

[8] Dix, A., Finlay, J., Abowd, G.D. and Beale, R. 2004. *Human-Computer Interaction*. 3rd edition. Harlow Essex: Pearson Education Ltd.

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

[10] Kahler, H. 2000. Constructive Interaction and Collaborative Work: Introducing a method for testing collaborative systems. *ACM Interactions*, May/June 2000:27-34. http://portal.acm.org/citation.cfm?doid=334216.334226.

[11] Masemola, S.S. and de Villiers, M.R. 2006. Towards a Framework for Usability Testing of Interactive E-Learning Applications in Cognitive Domains, Illustrated by a Case Study. In: J. Bishop and D. Kourie. 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/citation.cfm?id=1216283.

[12] Nielsen, J. 1993. *Usability Engineering*. Boston: Academic Press, Inc.

[13] Nielsen, J. 1994. Heuristic Evaluation. In Nielsen, J. and Mack, R. L. (Eds.) *Usability Inspection Methods.* New York: John Wiley & Sons.

[14] Preece, J., Rogers, Y., and Sharp, H. 2007. *Interaction Design: Beyond Human-Computer Interaction*. 2nd edition. Chichester: John Wiley & Sons, Ltd.

[15] Pretorius, M.C., Calitz, A.P. and Van Greunen, D. 2005. The added value of eye tracking in the usability evaluation of a network management tool. *Proceedings of the 2005 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)* 2005:1-10. http://portal.acm.org/citation.cfm?id=1145675.1145676.

[16] Rubin, J. 1994. *Handbook of Usability Testing: How to plan, design, and conduct effective tests*. New York: John Wiley & Sons.

[17] Squires, D. and Preece, J. 1999. Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11(5): 67- 483.

[18] Ssemugabi, S. and de Villiers, R. 2007. A Comparative Study of Two Usability Evaluation Methods Using a Web-Based E-Learning Application. *Proceedings of the 2007 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)* 2007:132-142. http://portal.acm.org/citation.cfm?id=1292491.1292507.

[19] Wilson, C. 1998. *Usability techniques: Pros and Cons of co-participation in usability studies*. http://www.stcsig.org/usability/newsletter/9804-coparticipation.html