Qualitative Findings from Usability Testing of an E-learning Tutorial

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

D. Becker
School of Computing
University of South Africa
becked@unisa.ac.za

Abstract: An interactive e-learning tutorial, Relations, was designed for a module in Theoretical Computer Science at a distance-education university. In innovative research, usability-testing technology was employed to investigate the learning behaviours and styles of students using the tutorial. This paper presents qualitative findings from observation, video recordings and post-session interviews. Cognitive and usage aspects are addressed, as well as the combination of electronic and paper resources, think-aloud articulation, and users' preferred learning strategies. Findings highlight the variety of ways in which the different users interacted with Relations and indicate the value of multiple modes of presentation, visualisation and feedback in supporting learning by e-learning.

Introduction

E-Learning encompasses a variety of forms, including among others, traditional computer-assisted instruction (CAI) via interactive tutorials and simulations; web-based learning; constructivist-style collaborative work; and recently, m-learning. An important part of the development and implementation of e-learning, is evaluation of e-learning systems. Many usability evaluation methods (UEMs) exist, such as user-based surveys and interviews; expert evaluation methods such as heuristic evaluation; and experimental evaluation using usability-testing and eye-tracking technologies in the controlled environment of a human-computer interaction (HCI) laboratory.

In innovative work, researchers in the School of Computing at the University of South Africa (Unisa) are applying usability-testing technology not only to evaluate e-learning systems, but also to investigate how users learn with them. In a quantitative study, de Villiers, Becker and Adebesin (under review) used observation, recordings and analytical tools to obtain metrics relating to the interactions of 17 students with an e-learning tutorial called Relations that presents mathematical content for COS101, Theoretical Computer Science 1. Analysis was conducted on the types of errors, error recovery, and distribution of time over learning activities, including cognitive engagement. Scores in an electronic test and the examination were also considered.

The present study focuses on qualitative findings of the usability-testing venture, analysing observations of human-related factors in the learning process and transcripts of post-session interviews to investigate how students learn while using Relations. The context is set by introducing the tutorial and overviewing the technique of usability testing. A literature study describes relevant aspects of adult learning, followed by an outline of the research design, methodology and research question. The body of the paper explains the data collection and analysis techniques and discusses the findings before concluding the study by re-visiting the research question.

Background and Context

Unisa is an open distance-learning institution with close to 300,000 students. In this context, e-learning systems offer added value to isolated learners. Many students are not the classic straight-from-school genre, but older, experienced learners, hence the literature study takes cognisance of adult learning styles. Both authors have had in-depth involvements with Relations and have taught COS101, which presents theoretical concepts for computing. We briefly introduce Relations and usability testing.
Features of Relations

The system used in this study is Relations, a multimedia, CD-based, interactive tutorial on skills in discrete mathematics for COS101. It was designed by a team of academics, programmers and graphic designers at Unisa as supplementary study material. It conforms to Alessi and Trollip’s (2001) definition of tutorial programs as software that teaches concepts by providing interactive sequences of information presentation, elaboration, questions, learner responses, judgment and feedback.

Distinguishing features of Relations are its scaffolding in the form of detailed diagrams, animated step-by-step development of concepts in cumulative visual formats, and use of colour coding (Fig 1). Instruction is interspersed with graphics and exercises in an interactive theory-examples-exercise format. There are a few Yes/No questions, but no multiple-choice questions. In major exercises, learners provide fill-in-the-blank responses by entering mathematical characters and terms. For some of the composite responses, more than one series of characters can be a correct answer, provided it meets specific mathematical conditions. Following incorrect responses, feedback and explanations are provided that diagnose common problems (Fig 2). In line with Boud, Keogh and Walker’s (1996) conscious reflection and recapture, learners can make second attempts after wrong answers. Relevant theory and definitions are available via hyperlinks. As a learning environment in a cognitive domain, Relations requires critical thinking on the analysis, synthesis and evaluation levels of Bloom’s (1998) Taxonomy. Its underlying learning paradigm is a combination of behaviourism and cognitivism (Alessi & Trollip, 2001).

Usability Testing

Usability testing (UT) (Barnum, 2002; de Villiers, 2009; Dumas, 2003; Rubin & Chisnell, 2008; Sharp, Rogers & Preece, 2007) is a formal and detailed evaluation technique that measures the performance of participants as they do specified tasks on a target system. The usual purpose is to improve its interface and interaction capabilities, and to identify and diagnose system problems. In this case we had a different purpose, namely to use the UT technology to discover more about the users’ learning processes and impact of errors as they learned with Relations. UT is conducted in specialised HCI laboratories, equipped with sophisticated monitoring and recording facilities and supported by analytical software tools. Researchers observe the participants through one-way glass and on triple-screen monitors, which simultaneously show keyboard interaction, users’ facial expressions, and the screen they are working on. Users work in the participant room (Fig. 3 – used with permission), while researchers sit on the other side.
side of the glass divide, viewing users’ keystrokes and activities (Fig. 4) and hearing what they are saying. UT sessions are video- and audio recorded for re-viewing in subsequent analysis.

Participants are actual users, whose interactions with a product are monitored and recorded to study fine-grained details, e.g. gestures, utterances, and emotions. Typical quantitative metrics are times taken to complete tasks, degree of completion, number of errors, and error recovery. In addition, qualitative data is elicited by careful observation of facial expressions and strategies, study of think-aloud, and post-session interviews. ‘Think-aloud’ relates to users’ verbalisations. At the start of a session, participants are requested to talk out loud as they work through the activities, articulating their reasoning processes. Some do it spontaneously, but others need to be reminded, and some others find it very difficult to do. Although UT is complex and costly, this controlled evaluation is a good way of assessing real-world behaviours of users with a product.

![Figure 3: The participant’s work station](image1)
![Figure 4: The observer’s view](image2)

Previous work by de Villiers (2009) described early research on applying this technology to investigate how learners learn with e-learning systems. The present paper, similarly, is not about UT to evaluate a system per se, but on innovative ways of using the technology to obtain information about students’ learning processes as they interacted with a multimedia tutorial.

**Literature Review**

This review addresses certain characteristics of adult learning, as presented in four classic works on andragogy. We also note how some of these concepts underlie the design of the Relations tutorial.

**Learning Styles**

There are various learning styles: memorisation, understanding, and doing (Cotton, 1995). *Memorising*, i.e. verbatim learning which was once a predominant form of learning, can be daunting for adults. Cotton suggests that visual memory can support recall. This is applied in Relations, which relies heavily on visualisation and diagrams to illustrate concepts. Learning by *understanding* is an individual process. Some learners grasp explanations rapidly, but others take time. Cotton advises different types of explanations and refers to Edward de Bono’s lateral thinking which advises thinkers to step sideways and move beyond the usual approaches. Relations does this by using a graphical perspective and evolving animations to show concept development (Fig. 4). Cotton’s learning by *doing* refers to motor skills, where she advocates learning by watching. In Relations, the aim is the acquisition of
mathematical skills, which are taught first, by demonstration of annotated, colour-coded examples and, second, by exercises for practice.

In Cotton’s (1995) simplified version of Kolb’s learning cycle, she stresses the four iterative stages of experience, reflection, theory and preparation. The theoretically-founded and graphic techniques in Relations are designed to support learners in these activities. Particularly relevant to the use of Relations are reflection, whereby learners ponder on new knowledge, assimilating it in a form of action replay and theory, which is applied when learners consult rules and use working memory to compare new information to previously known theoretical concepts.

Kidd (1973) explicates various forms of memory. Recognition entails identifying something that has already been learned, while recall relates to memorising. For memorising, Kidd distinguishes between rote learning and meaningful learning, emphasising that retention can be supported by presenting material in meaningful ways. Similarly, repetitive practice of skills should be done in meaningful ways. This is implemented in Relations, where practice is situated in contextualised exercises. Referring to cognitive learning, Kidd refers to Jerome Bruner, who believed that mere storage of information is inadequate; learners should locate it in memory and retrieve it. Kidd stresses the role of concept formation. As learners grasp a field and its phenomena, they must understand the definitions and concepts that give order and meaning to objects and events, as well as the relationships between concepts. Most important, they should be able to apply problem-solving strategies. Using its features and techniques, as well as reinforcement through multiple presentation and links to theory, Relations aims to support problem-solving by recognition, recall and concrete formation of concepts.

Errors

For error analysis, we used Squires and Preece’s (1999) categories of errors, distinguishing between peripheral usability errors and cognitive errors, the latter being misunderstandings that are part of the learning process in complex domains. Usability-related errors should be avoided, but cognitive errors are helpful if scaffolding mechanisms are provided to support learners in their learning. Mayes and Fowler (1999, p.485) acknowledge the value of such mistakes in educational applications, which must make learners think, and where ‘seamless fluency of use is not necessarily conducive to deep learning’.

Feedback

In classic behaviourism, actions are followed by reward or punishment, termed stimulus-response or operant conditioning (Kidd, 1973). Responses that are rewarded, tend to persist. Boshier’s (2006) research shows that with adults, ‘punishment’ in the form of criticism and lack of encouragement has negative effects on learners. Relations’ feedback is positive and supportive, even for incorrect responses. Although Relations is not an intelligent tutor, it has limited diagnostic functionality that provides customised feedback to some of the answers. In other cases, generic feedback and hints are offered (Fig. 3). The scaffolding and feedback support learners as they aim for recovery, in line with Kay’s (2007) emphasis on error handling by detection, diagnosis and correction.

Research Design and Methods

Research Question

This research applies UT technology in an innovative way, not to evaluate a product, but to investigate participants’ learning processes as they did specified tasks. The quantitative metrics have been discussed in a separate publication (under review). The present research, by contrast, focuses on the qualitative data that emerged from observation and interviews. The research question is:

What do the qualitative findings of usability testing indicate about the learning experiences of university students using an e-learning tutorial with mathematical subject matter?

Methodology

Participants had individual sessions involving: orientation in the laboratory; explanation of the procedures; hands-on interaction with the system; a test on the learning content; and debriefing and interviews. The time was too short to
work through *Relations* in its entirety, so we compiled a task list directing participants to four specific units. Each unit included guidance on theoretical concepts; animated graphical demonstrations; worked examples; and exercises where they entered mathematical characters or terms as answers. Diagnostic feedback was provided and second attempts could be made after incorrect answers.

**Participants**

Due to its time-intensive analysis, the number of participants in a usability-testing study which focuses on finding system problems, is often as few as five to eight (Nielsen, 2000) or based on the ’10 ± 2 Rule’ (Hwang & Salvendy, 2010). Since the present research focuses on learning processes rather than system evaluation, we used more, five in the pilot and 17 in the main study. The experiences of these learners are discussed in the next subsections.

**Pilot Study**

Boshier (2006, p.52) points out the importance of doing a pilot with a sample that corresponds with the profile of the sample used in the main study. Participants in the main study were COS101 students, whereas the five participants in the pilot were not first-year students, but a sample of convenience of fourth- and fifth-year students whose cross-group profile corresponded closely with the main study sample. The pilot was used to try out methods and research instruments, as well as the tasks. The process involved orienting participants into *Relations*. We tested the monitoring and recording processes and assigned responsibilities to each researcher. We established the duration of a full session, involving set-up procedures, participants doing the tasks and exercises, a short post-test, and a post-session interview. The tasks required learners to access units on the mathematical concepts of Reflexivity, Symmetry, Transitivity and Equivalence Relations. The units had segments of guidance and tuition, diagrams, examples, and exercises for learners to try out their skills. The tasks served admirably in covering the full spectrum of tutorial functionalities, and required no changes. However, we needed to improve and extend the interview protocol.

**Main Study**

In Unisa’s distance education environment, students are distributed nationally and even globally. A call for local volunteers produced a sample of seventeen students, heterogeneous in age, gender and culture. All were under 30, which corresponded reasonably with the cohort, 65% of whom were in the 20-29 age group and most of the rest were 30 or older. The sessions were held shortly before the final year-end examinations, so that the participants could gain personal benefit. Ethical clearance for the research was obtained from our Faculty’s Ethical Clearance Committee. The procedures and lab environment were explained upfront to the participants, who all signed informed consent.

**Coding Qualitative Data**

Boshier (2006) describes how interview responses emerge in definable groups, very often with individuals using similar terms. Key words and phrases were coded in the analysis of data form observations and interviews, resulting in the emergence of themes and patterns in line with Boshier’s (2006, p.55) so-called ‘response groups’. These groupings around learning patterns and strategies are outlined in the sections that follow.

**Data Collection and Analysis**

*Data Collection*

We allocated between one and two hours per session, based on the task list mentioned under Methodology and using the approach explained in the Usability Testing section.

As stated under Research Design, the quantitative findings were addressed in a separate publication. The present study focuses on qualitative aspects from observation and interviews. We established what participants were doing from their on-screen cursor pointing and from their think-aloud verbalisation. The sessions were video-audio recorded and re-viewed, and each viewing provided further insights. We took notes during real-time observation and the re-viewings. After the sessions, participants were interviewed to ascertain the ways in which they liked to learn; as well as their opinions on technology-supported learning compared with using printed learning material.
Specifically, we investigated their experiences with the Relations environment and the errors (usability- and cognitive errors) they made while using it. We also queried whether working under observation had a negative impact on their performance.

**Data Analysis**

In the quantitative research, data was statistically analysed. In the present qualitative research, analysis of observations was done by manual coding of occurrences and patterns. We identified response groups (Boshier, 2006) and categorized the themes and patterns that emerged. When participants used the think-aloud approach, their recorded verbalisation enhanced analysis by providing information that complemented or confirmed the visual and written data. Similarly, textual transcripts of the interviews were analysed qualitatively to identify groupings and patterns.

**Findings and Discussion**

In mixed-methods research (Creswell, 2009), we triangulated data by using observations and interviews. Qualitative findings from both methods are integrated in this section. Brief mention is made of findings from the earlier quantitative study, and the interpretation of certain findings is related back to the literature review.

**Learning Styles and Strategies**

**Patterns and Preferences**

The interview investigated ways in which participants liked to learn, comprehend and support recall. They were able to mention more than one way. Various preferences and patterns emerged (Table 1), confirming the various forms of remembering and learning suggested by Kidd (1973) and Cotton (1995). For computational subjects, it is understandable that memorising was least preferred, while practicing and understanding were the most preferred, by 12 and 16 students, respectively. The desire to discuss and work collaboratively, cannot be met for many distance learners, but the interaction provided by e-learning can help to fill the gap. In an upcoming section, in which we describe findings of observations, we confirm their need to write by making notes and doing exercises on paper.

<table>
<thead>
<tr>
<th>Style/ Strategy</th>
<th>Memorising</th>
<th>Verbalising or discussing</th>
<th>In groups</th>
<th>Alone</th>
<th>Writing / Summarising</th>
<th>Practicing / Examples</th>
<th>Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of participants</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>16</td>
</tr>
</tbody>
</table>

**Table 1: Preferred learning styles and strategies**

**Time spent**

The greatly varying speeds that had been recorded in the quantitative time metrics became tangible as we watched participants interact with Relations. Some worked rapidly, others slowly carefully. Although performance and achievement are more a function of aptitude and intellect than of preferred styles, the students who were more confident about the subject matter, tended to move faster, revise less, and tackled the exercises enthusiastically. Eleven completed the tasks in less than 40 minutes, six of them gaining distinctions (≥75%) in the post-test, while one obtained 71% and the others 54, 57 and 58%. It is notable that the two who worked fastest, 23 and 24 minutes, respectively, had grasped the subject matter and obtained 86% and 79%. The latter concentrated and consolidated by explicitly reading the theory out loud and touching her focus point on the screen. The others who spent around 38-39 minutes to obtain high marks, used the menus fast and fluently, yet without rushing the theory and were able to spend time, 2-5 minutes when necessary, on doing the complex exercises and re-visiting their reasoning. A participant who rushed and finished in 27 minutes, showed little interest and scored only 25%. But in general, the participants who worked rapidly, paid attention to the graphical representations, and were bold in venturing answers. They were risk-takers who were not intimidated by mistakes and used the feedback to help them recover.

The three participants who took the longest, an hour or more, obtained 83%, 72% and 25% respectively, illustrating that slow-and-systematic can either be a successful strategy or indicate lack of direction. These three outliers were interesting, in that the two high scorers were systematic and meticulously careful, spending up to four minutes on single exercises and re-doing answers to correct mistakes. The third (who scored 25%) showed discomfort; he worked in a traditional style, consulting printed hand-outs even when the relevant definition was onscreen. He liked to ask the facilitator questions. He was one of the three participants who needed personal contact. The other two who needed human contact, constantly communicated with us. They both passed the test; one (57%)
hardly used the electronic content but enjoyed 1-on-1 contact and asked us questions that supported learning, while
the other (79%) made sure he understood the on-screen theory and asked questions just to confirm. Another style
evidenced by the three who were uncomfortable with e-learning was use of printed material. The student with 25%
has been mentioned; the others similarly consulted the hand-outs; one used them instead of on-screen definitions
(57%) and the other as supplementary to the screen (86%). In total, five participants preferred the printed definitions
over the on-screen ones.

Learning with Technology versus Print and Paper
The interviews showed that nine were having their first exposure to e-learning. Nevertheless, none of them
expressed negativity about the experience. When asked for their perceptions, ten gave highly positive responses and
seven positive. Factors in support of these impressions were similar to those in the comparison between e-learning
and paper-based: Easier to learn/learn more (10); Friendly/engaging/enjoyable/excellent (14); Faster learning (3);
Supported understanding-guides you (11); Feedback (4); Interactivity/good for practice (10); Added value over book
(7); Relations emphasises the most important points (5); Variety in presentation media/variety in content (e.g
visuals, animations, theory, etc. (9); Learner-control (2). The responses indicated that Relations offered both
affective and academic value, as well as giving distance-learners hands-on experiences with multimedia material.

We discussed how learning with technology differed from textbooks and notes. Four of them said it was
fun, an engaging and attractive way to learn. Twelve stressed the value of the animated diagrams and step-wise build
up: ‘You see a line growing from A to B’; ‘When it drew lines, it helped me understand’; ‘You see things
happening’; ‘With electronic material I understand and remember.’ Four emphasized the value of visualisation and
colour coding.

Another factor was the crisp presentation and highlighting of key issues, which helped six to learn quickly:
‘More concise’; ‘Keywords stand out’. For three participants, the navigation, easy access and links were the
distinguishing aspect: ‘You can go straight to the things you don’t understand’. All of the above relate to the rich
interaction within Relations’ environment. However, five participants explicitly referred to interactivity in their
characterisation of e-learning. Two charming quotes: ‘It’s like you are talking to the computer’ and ‘…with
technology, I am doing it’. When asked in interviews from which medium of learning they recalled information
better, thirteen participants stated e-learning, two chose printed matter, and two required both together.

Although joint use of the printed task list, screen, mouse and keyboard had been daunting, when it came to
personal activity, we saw from observations that paper and pen still play important roles. Thirteen wrote notes,
sketched diagrams, and/or explored solutions by trying out exercises on paper. Eleven did this to a great extent and
two very little. When referring back to theory, it was clear that five preferred printed definitions to those on the
screen.

Think-aloud and Reflection
Think-aloud does not come naturally to everyone (Barnum, 2002) and, although users can be prompted and
discerningly reminded, it should not be forced. There were three distinct response groups.
1. Six verbalised clearly and consistently. Two were reminded to talk aloud, but the others did it naturally, non-
stop.
2. Eight did it sporadically. Some thought aloud while actively reasoning through examples and exercises; others
verbalised while reading the screen content, repeating definitions, or studying the task list.
3. Three moved their lips continually, but nothing was audible.

There were various types of articulation and at varying speeds, confirming Cotton’s (1995) four stages of learning:
1. Some students reflected and negotiated with themselves: ‘Yeah...OK, Oh, alright’; ‘Now let us see...’; ‘That
was a tricky one’: ‘I do not think I can do that’; ‘Wow, alright’. [Cotton’s (1995) reflection].
2. Another pattern was explaining the reasoning through the academic processes: ‘What happened to (p,q)?’; ‘
What is missing is (r,q) and (q,r)’; ‘That is what I expected’; ‘I do have this, so it must be...’. And on reading
feedback: ‘This is so clear’; ‘Great’. [Cotton’s (1995) theory].
3. There were outbursts of excitement on enlightenment: ‘I see! I get it!’; ‘Wow, I did not know this!’; ‘Yes, yeah,
I get it now!’; ‘Good!’ And distress: ‘OK, I am lost here’; ‘It is hard... but... OK, got it. Sharp!’; Some
verbalised actively when getting interactive feedback after their answers: ‘Got it right, just took too long’; ‘Wait
a minute; I have the reverse of that’. On right answer, ‘yes, yes it is good’. On wrong answers, ‘Why not?’; ‘It is
bad...’ and ‘Aish..!’ [Cotton’s (1995) experience].
4. One of those who wrestled until he grasped it, expressed his thinking: ‘Yes, I will need to.... I was confused a
little bit, but now....’. [Cotton’s (1995) preparation].
**Body Language**

Distinct behavioural patterns were evidenced, but with overlaps. Participants displayed different behaviours as the session progressed or different reactions as situations changed. Their reactions were articulated by emotions and body language: Relaxed/smiling (13); Using cursor as a pointer (12); Using finger to touch screen (6); Laughing, enthusiastic or delighted (5); Serious/focused/concentrating (6); Anxious/insecure/uncertain (2); Uninterested (2).

Some were clearly in control. One used both mouse and finger as pointers, and pointed simultaneously at two places using two hands. There were interesting gesticulations: one threw hands up in delight; another, a visual thinker, moved hands graphically in 3-D motion to simulate mathematical elements. A puzzled participant tapped a pencil: tick, tick... while another habitually sucked fingers. The intent students handled it in different ways. Some leaned forward at vital moments, while others worked steadily, consistently and motionlessly. Two nodded vigorously when they agreed. In general, participants were pleased when answers were correct, and smiled. Attitudes to incorrect answers varied. Some were frustrated. One laughed. Most were comfortable with mistakes, but as part of the learning process, with supportive feedback (Boshier, 2006) to help them progress. They just focused intently on the screen and their next attempt. Only one seemed bored and keen to finish.

**Use of Hyperlinked Theory and Definitions**

In live observation and in viewing videos, we investigated certain aspects that had not been covered in the quantitative study. One of these related to hyperlinks that allowed users to access relevant theory and definitions when necessary. The links had been built into the most recent redesign of Relations to include features of web-based learning in the systematic and structured tutorial. Participants were told upfront about this scaffolding but, disappointingly, little use was made of these hyperlinked pop-up definitions. Only four used them consistently and intently, of whom two were high performers in the post-test. Three students grasped the theory so well (getting 83%, 86% and 93% respectively) that they did not need to refer to definitions at all. Eight others did not use them; while two used them once. This was a disappointment, since the links had been constructed with high expectations.

**Errors and problems**

**Use of Printed Task list**

Participants would have preferred to view the screen only, without directions from a printed document. Most of them struggled when initially using the tasklist, misreading it or hardly consulting it. On starting the session, 12 of the 17 selected units from Relations’ menu consecutively, instead of accessing only the designated units. Most of them self-corrected, but some had to be prompted by researchers. A more explicit orientation upfront might have prevented the problem. The other five worked accurately and systematically throughout. Despite the difficulties noted in using the tasklist to select specified units from the menus, when it came to using paper for their own support, it was a different story and most made notes or practiced, pen in hand.

**Usability Problems**

Relations had been evaluated by several UEMs previously and corrected, and was largely error-free. However, it is difficult to eliminate all the latent or subtle errors and this study unveiled further usability errors. Issues surfaced under UT, some in the quantitative study of participants’ mistakes and others in this qualitative study. Observation led to identification of a design weakness that had not been pin-pointed by other UEMs. On using mouse-overs, four students thought the overlays which appeared, were menus, and tried to select them. This confusing aspect was also mentioned in an interview, and warrants correction. One of the design aims of Relations was to ensure that users covered all the content in a unit. Yet in certain places, we saw participants going forward, omitting some exercises, which should not have been permissible. There should be constraints throughout, forcing users to attempt all the exercises. On the other hand, there should have been opportunities to take official breaks. An interviewee made a justifiable request for exit and re-entry facilities. This would have been a good compromise and its absence indicated another weakness.

Some surprising errors occurred, in that five participants did not use ‘Backspace’ to correct mistakes, nor press ‘Enter’ after typing responses. These must be viewed as a consequence of technological illiteracy on the part of participants. Nine of the 17 were using e-learning for the first time.

**Cognitive Issues**
When surveys are used to evaluate technological artifacts, the responses are usually broad and general, particularly when Likert scaling is used. However, observation and listening are excellent UEM for studying students’ cognitive errors (Squires and Preece, 1999) and for determining details of their learning processes and perceptions. We observed them making minor slips in the formats and mathematical syntax used in doing exercises. They struggled with rigorous accuracy in fine-grained details. Some of the slips were so minor that they might have gone unnoticed in grading hand-written answers, but the computer is a stickler for syntax and students frequently had to re-do answers involving brackets, commas, and the difference between ‘for all’, ‘for some’, ‘if’ and ‘whenever’. Two verbalizations:

‘Now I understand! I lost it, because I took whenever as being all’;
‘I need to know if this is reflexive, and for that, all the possible members must be considered’.

After our major design efforts, it was gratifying to see the features being used as intended. The value of the feedback for error handling (Kay, 2007) was evident. Most students studied it intently after making content-related mistakes and frequently recovered after conscious reflection and recapture (Boud et al, 1996). Some read it intently even after they got an answer right, reinforcing their understanding. The diagrams and animations were well received with think-aloud comments such as ‘OK, I get the connection’; ‘This is helpful!’; ‘I see that the circle loop is also symmetric’. It was evident that visualization enhanced cognition and contributed to insight.

We saw various individual cognitive styles, showing the value of flexible learner-control. Systematic users used the ‘Repeat’ function to revise and consolidate. One enjoyed re-doing the open-ended exercises to explore alternative answers. He had a classic ‘learning-by-mistakes’ style. Another spent four minutes meticulously doing a short exercise, but got it right first time. Another rushed on, getting answers wrong, but using the feedback to learn and to correct mistakes. He did well in the post-test and, in the interview, acknowledged this deliberate tactic. In other cases, we saw ‘rush-ahead’ learners changing their approach. After starting fast and making mistakes, they slowed down, read the feedback carefully, and their accuracy improved. Three of them mentioned this in their interviews.

As it did with the usability issues, UT unveiled latent problems in aspects of cognitive support. We found some academic inadequacies in Relations. In transitivity, certain ordered pairs play multiple roles, which students find complex. We had addressed this fundamental problem well in the printed study guide but, strangely, had not tutored it as adequately in the interactive e-tutorial. Moreover, we found two academic errors; in one, the system graded an unusual, but correct, answer as incorrect. Most students would have thought they were wrong, but insight and objection by a student and observation by a researcher, identified this issue.

Interviews further confirmed the observations regarding participants’ cognitive experiences. In answering a question about errors and recovery, they mentioned: Value of feedback and second attempts (9 participants); The value of theory/definitions in recovery (8); Features of Relations that facilitated their understanding, e.g. step-wise build-ups, colour-coding, visualisation, concise presentation (4); Learning from mistakes (quote ‘The more mistakes I make, the better I understand’) (4); The need to work slower (3).

In particular, we monitored how they consulted and applied the theoretical concepts presented in the units. By listening to think-aloud and observing points of focus, we determined that 13 diligently studied the theory before applying it in exercises. They read definitions, interacted with graphics, and verbalised their reasoning as they applied it, some using self-made terms that aptly described the concepts. Nine of those 13 scored more than 70% in the post-test, while three scored mid-50’s and only one failed. Of the four who did not appear to study the theory, two passed the test with low scores and the other two failed.

Human and Personal Aspects

Self Consciousness
In the interviews, participants were asked whether working in a controlled environment and being observed by researchers, had put them off. Only three had felt somewhat daunted: ‘Because of this, I said less out loud, which is the opposite of my natural learning style’; ‘When somebody watches me I get uncomfortable ’; ‘Once I had settled I forgot you were there’. The other fourteen were not intimidated. Two comments were particularly pleasing: ‘I had already built relationships with the lecturers before I started the session’; ‘It was fine, you guys were friendly!’.

Interaction between Participants and Researchers
There were two forms of human interaction, namely student-initiated queries and interventions by researchers. In student-initiated communication, eight participants asked questions, to which we responded. Three needed human contact and continually asked questions on numerous aspects. Three had meaningful academic queries, while two asked basic questions about system usage. As researchers, we intervened or prompted at times. On ten occasions, we
briefly corrected participants, particularly at the beginning, to remind them to use the task list. Sometimes we briefly interjected with praise, ‘Well done’, and on three occasions, added a short explanation or gave advice.

**Conclusion**

The worth of employing usability-testing equipment in this qualitative study was confirmed, in particular the use of re-viewing to elicit fine-grained details. It was an excellent experience for the designer-educators to observe users interacting with Relations and note how the different types of content and features contributed to meaningful learning. We now revisit the research question: *What do the qualitative findings of usability testing indicate about the learning experiences of higher-education students using an e-learning tutorial with mathematical subject matter?*

All 17 participants had positive perceptions of the learning experience. The over-arching finding was that Relations supported different learners in different ways. Due to user-control and flexibility, participants used varying styles for time usage, re-viewing of worked examples, and consultation of theory. Furthermore, they applied their own preferred strategies in using the animations and visualizations and doing the exercises. In particular, students indicated that the scaffolding and detailed feedback supported recovery from errors and encouraged them when their answers were right.

Use of two methods, observation and interviews, triangulated and confirmed the findings. The effectiveness of Relations was also confirmed in the separate quantitative study where the positive subjective attitudes were ratified by the result that six of the 17 achieved notably higher scores in Relations tasks than their natural level in the examination, indicating the value of the additional learning resource. Although the performance of 11 of the 17 is in line with Clark’s (1994) classic ‘no significant difference’, the positive perceptions that emerged from observation and interviews indicate that the e-learning tutorial motivated students and supported them in using personal learning strategies in an enhanced learning experience.

In the interviews most participants asked spontaneously for more such tutorials in COS101 and other BSc modules. We acknowledge that, for studies that investigate human behavior and learning strategies, a sample of 17 is too small to generalise findings on the processes of learning with technology. However we can confidently advocate the use of well-designed user-controlled e-learning tutorials to support learners in flexible ways that are adaptable to their preferred learning styles and strategies.

The findings have relevance and implications for the design of e-learning. In future design efforts they can be customised to the development of a variety of e-learning technologies. Moreover, they can be applied in the redesign of existing systems and tools.

**References**


