

Multi-Method Evaluations: Case Studies of an Interactive Tutorial and Practice System

Ruth de Villiers

University of South Africa, Pretoria, South Africa

dvillmr@unisa.ac.za

Abstract

The teaching and learning of a complex section in *Theoretical Computer Science 1* in a distance-education context has been enhanced by a supplementary interactive e-learning system with tutorial and practice functionality in a classic computer-aided instruction (CAI) style. A participative action research process was used to develop, evaluate and refine the application over a longitudinal period. Case studies are presented of evaluations by four different methods – questionnaire surveys, interviews, heuristic evaluation and a post-test. Their respective findings, relating to functionality, usability and contributions to learning, are given and discussed. The findings lead to reflection and lessons are learned regarding the design of e-learning applications. In an overview, the complementary roles of different evaluation techniques are discussed.

Keywords: Action research, computer-aided instruction, e-learning, human-computer interaction, theoretical computer science, triangulated evaluation methods, usability evaluation

Introduction

The objective of this paper is to describe and discuss multi-method evaluation of an e-learning system. It presents case studies of evaluations – by different techniques – of an application called ‘*Relations*’, and gives their findings regarding its functionality, usability and contribution to learning. *Relations* is a supplementary e-learning system that supports learning in a complex section of *Theoretical Computer Science 1* at the University of South Africa (UNISA). It is highly interactive, with tutorial and practice functionality. The composite evaluation methodology involved four methods – a questionnaire survey, interviews, heuristic evaluation, and a post-test. Both qualitative and quantitative data was obtained. Where appropriate, limited statistical analysis was conducted.

The paper introduces *Relations* and its context of use. Action research, which was the development approach and underlying research framework, is outlined. Descriptions and findings of the four separate evaluation studies are given. Conclusions are drawn about *Relations*, which lead to generic reflection on the design of computer-aided instruction and e-learning applications. Con-

cluding discussion overviews the roles of the different evaluation techniques, making a contribution to the discipline of Human-Computer Interaction (HCI).

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

Background

What is e-Learning?

Some definitions equate e-learning solely with use of the Internet in instruction and learning. However the definition of CEDEFOP (2002:5) is broader, describing it as ‘learning that is supported by information and communication technologies [which]...may encompass multiple formats and hybrid methodologies, in particular, the use of software, Internet, CD-ROM, online learning, or any other electronic or interactive media’. This definition suits the present study, which views e-learning as a broad range of learning technologies, including web-based learning and traditional computer-aided instruction (CAI), which was high-profile in the 1980s and 90s.

What is ‘Relations’?

Relations is an interactive multimedia CAI lesson about mathematical relations. It has tutorial functionality (Alessi & Trollip, 2001) and gives the learner extensive practice opportunities. *Relations* is a stand-alone CD-based system and is not Windows-driven. It was custom-built in the School of Computing at UNISA as optional supplementary study material for the first-level module *Theoretical Computer Science 1*, which teaches discrete mathematics relevant to the foundations of theoretical computing and databases. Version 1 (V1) of the lesson was originally produced in 1994, underwent learner evaluation, was positively received, and was used for eight years. By 2003 it was time for a revamp. The general approach and subject matter were re-used, but the content was extended. HCI aspects such as fonts, backgrounds, screen displays, colours, colour-coding, interaction design, etc. were improved and general quality was enhanced. It was developed using the multimedia authoring system, *Quest 7.0*, a tool for the creation of computer-based training courses. *Relations’* navigation is mainly linear in the traditional style of a tutorial, rather than hierarchical. However, there is learner control over which sections to do, the sequence of sections, and whether to omit theory and do exercises only. Following the 2003 redesign (V2.1), it was comprehensively evaluated in 2004, leading to further refinements and evaluations in 2005 (V2.2), including hyperlinks to definitions and elaborations.

Who are the Students and What Is the Course Material?

UNISA is a distance-teaching institution with students worldwide. Most live in South Africa, but there is a focused outreach programme to the rest of Africa. Until recently most students were employed fulltime and studied part-time. The average age was mid-30s. The present learner profile is changing and many are young, fulltime students, mainly from formerly disadvantaged groups. Communication is primarily by conventional mail, because many lack technological access. Computing students, by definition have access to PCs, but not always to the Internet. Some of the young students study, not at home, but in UNISA’s regional learning centres.

Between 800 and 1000 students register annually for *Theoretical Computer Science 1*, of whom about half take the module through to completion and write the exam. The mandatory study material comprises a prescribed book, a study guide and regular tutorial letters. Assignment questions and examinations involve precise and rigorous written mathematical proofs and procedures. Detailed solutions to the assignments are subsequently provided, as well as answers for the previous year’s examination questions. The CAI lesson *Relations* is optional study material that can be purchased at a reasonable price and is available for free use in UNISA’s computer labs throughout South Africa. Between 30% and 40% of the students use *Relations*.

Content of Relations

Relations comprises three sub-lessons: *Background information*, *Properties of relations*, and *Special kinds of relations*. Each has its own subsections and learners select which to do, in what sequence, and how many times to tackle each. Within each section definitions, theoretical concepts and examples are presented, illustrated by graphics, pictures and animations that show step-by-step unfolding of a definition or process. Tutorial segments are interspersed with exercise-and-practice segments. Each major section ends with a revision summary. There are no multiple-choice questions, most of the exercises being fill-in-the-blank questions. Some questions are basic, but most entail a composite answer or entry of a series of mathematical characters and notations. Some of the exercises involve synthesizing a relation to meet a particular set of conditions. Due to the range of possible errors, there is considerable variation in students' input, yet *Relations* provides them with detailed feedback, and second attempts must be made after wrong answers. This diagnostic feedback, which also includes explanations, is one of the lesson's greatest strengths, implemented mainly by pattern matching but also by some elementary artificial intelligence techniques. *Relations* addresses computational concepts from discrete mathematics, i.e. rigid concepts with no scope for personal interpretation. The rules of this 'well-structured domain' (De Villiers, 2005b; Jonassen, 1999) are tightly defined and each problem has a single solution or a finite set of answers that would be judged as correct.

Relations caters for different stages of learning, offering learners the option to omit theory and 'dive' straight to sets of exercises or revision screens. Its purpose is learning support; it is non-threatening and does not grade learners or record progress. There is an optional test, covering all aspects, but the score is not stored. Figure 1 portrays a question with an incorrect answer, followed by detailed feedback, which develops the concept in detail. Figure 2 demonstrates a com-

Special kinds of relations: 3.3 Linear orders

Q $B = \{p, q, r\}$

V is a weak partial order on B comprising five members.
Complete this 5-member weak partial order V on B

$V = \{(p,p), (q,q), (q,p), (r,r), (r,q)\}$ ❌

A This was a tricky one! First (r,r) should be included to ensure reflexivity on B . That's the easy part. For the last pair we have five options:
 $(p,q) / (p,r) / (q,r) / (r,p) / (r,q)$
Well (p,q) is out, it violates antisymmetry.
How about (p,r) ? Not on - if we have both (q,p) and (p,r) we would need (q,r) as well to maintain transitivity.
For the same reason (r,q) is out.
Two valid possibilities remain.
The answer is $V = \{(p,p), (q,q), (r,r), (q,p), (q,r)\}$
or $V = \{(p,p), (q,q), (r,r), (q,p), (r,p)\}$

Back Repeat Exit Menu Help Forward

Figure 1. Example of a question with an incorrect answer

Special kinds of relations: 3.1 Equivalence relations

Q $M = \{a, b, c, d\}$
 Consider the partition of M given by the three subsets (a,b), {c} and {d}.

Give the equivalence relation T_{15} that has these subsets as its equivalence classes (cosets).

Give the cosets. Press enter after each coset. You will be marked after doing all four.

$T_{15} = \{(a,a), (b,b), (c,c), (d,d), (a,b), (b,a)\}$ ✓

[a] = {a,b} ✓
 [b] = {b} ✗ [b] = {a, b}
 [c] = {c} ✓
 [d] = {d} ✓

A a lives in a pair with itself and with b;
 b lives in a pair with itself and with a;
 c and d live with themselves only.

Back

Repeat

Exit

Menu

Help

Forward

Figure 2. Example of a completed colour-coded animation

pleted colour-coded animation, along with a question about equivalence relations and cosets, where the concepts entail different styles of brackets: { }; (,); and []. To answer the question, learners enter series of characters and notations, which are not assessed until they have all been entered. The boxed ‘Q’ and ‘A’ are icons indicating question and answer respectively.

Learning Theory Underlying Relations

The learning theory ethos (De Villiers, 2005b; Reeves & Reeves, 1997) underlying an educational software application or a web-learning environment should be explicit in its design. The present researcher is the designer of *Relations* and was the module leader of *Theoretical Computer Science 1* for ten years. She is aware of the domain complexities and problems experienced by students. Her preferred pedagogical position has a social constructivist ethos. However, due to the rigid nature of its subject matter, *Relations* has fundamentally objectivist and instructivist tendencies. It has a classic CAI tutorial structure, alternating teaching and practice segments, and adopts a behaviourist stance in its stimulus-response-reinforcement approach. Yet it also embodies eclecticism, combining paradigms where appropriate. Opportunities were taken to motivate and engage learners, to support true cognition, and to anchor the learning process in authentic contexts. Constructivist multi-perspectives are employed to portray concepts by means of text, figures, graphics, evolving animations, and many examples. It requires synthesis skills as learners generate their own relations. *Relations* does not include learning management or recording of progress, since it is intended for learning gain and not for learning measurement. The feedback provides integrated assessment, and the optional test is a capstone. For innovation and engage-

ment, a blue water theme, comprising informal ‘recreational’ screen displays, is used after complex sections to help students take a break.

Research and Development Process of Relations

An action research (AR) approach (De Villiers, 2005a) was used to develop *Relations*. Action research emanated in the 1940s from action-based social psychology by Kurt Lewin of the University of Michigan, who contended that complex social events cannot be investigated under laboratory conditions. AR encompasses a variety of research techniques and methods, and generates both action outcomes and research outcomes. Its participative, practitioner-researcher approach lends itself to educational research, where developing interventions/products are investigated over several cycles. Commencing with the identification of a problem or situation that calls for action, AR functions as a *change agent*, and is (Baskerville & Wood-Harper, 1996; Cohen, Manion & Morrison, 2000; Dick, 1999; Du Poy & Gitlin, 1998):

- *Cyclic*: iterative steps recur longitudinally, generating information for further action. Zuber-Skerrit (1992) terms its repetitive processes: *plan, act, observe, and reflect*.
- *Participative*: clients, end users and researcher collaborate as co-researchers. AR is often used when practitioner-researchers examine their own work, giving the researcher a central role in both the research and the development processes.
- *Qualitative*: it operates more by verbal aspects than by numbers, although quantitative methods are also used and acceptable.
- *Reflective*: critical reflection on the process and outcomes is vital to each cycle, and is used in designing subsequent steps and events.
- *Responsive*: it reacts and adapts flexibly to the findings from each previous cycle.

An AR model, as shown in Figure 3, was used in the iterative research, development and formative evaluation of *Relations*. The series of cycles was characterized by central, in-depth involvement of the researcher-designer (also the module leader) and participative input by peers and students.

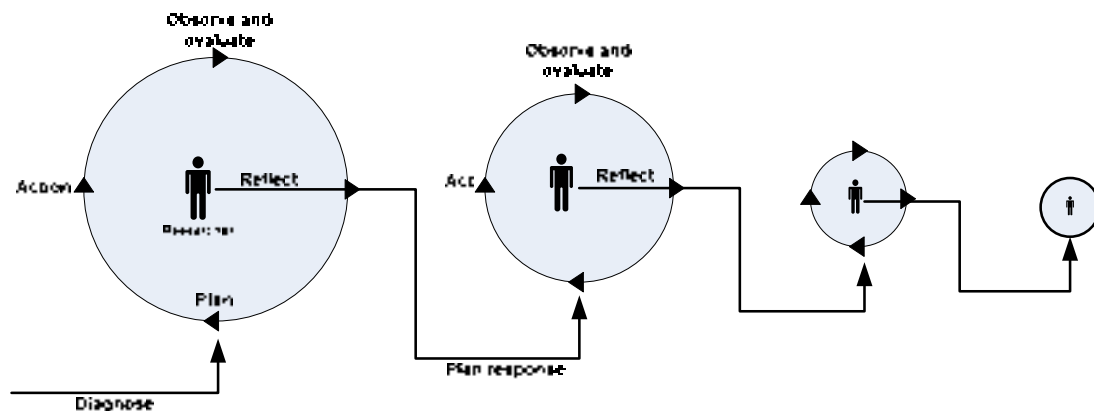


Figure 3: Action research model (De Villiers, 2005a)

Both formative and summative evaluations have been conducted. During design and development *formative evaluation* (Preece, Rogers & Sharp, 2002) was done by the development team, which comprised the course leader (who designed *Relations*), computer programmers, other lecturers, and a graphical designer. All the processes and mathematical exercises were comprehensively tested during development, both when the original V1 was produced in 1994 and again with V2

in 2003. Each set of evaluations led to reflection and responses in the form of corrections and improvements. The programmer has mathematics expertise as well as programming skills, which added value to her contribution. There was a strong designer-developer partnership throughout.

This paper concentrates on a further cycle, namely, post-release *summative evaluations* (Preece et al, 2002) undertaken in 2004 and 2005 to examine, inspect and improve the finished product. Such evaluations can also be termed *effectiveness evaluations* and *impact evaluations* (Reeves & Hedberg, 2003). These multi-method evaluations are described in the next section.

Evaluations

The redesigned *Relations* was released and evaluated in 2004. Findings from the 2004 student questionnaire survey, along with input from a heuristic evaluation, were used to further correct, improve and slightly extend V2.1, resulting in V2.2, which was released in 2005 and was evaluated by a student questionnaire survey, by interviews among students and by heuristic evaluation. In both 2004 and 2005, students' performance in a particular question in the final exam was used as a post-test. This section presents four evaluation case studies, giving findings from the four different techniques.

Query Method: Questionnaire Survey among Students

Various sets of criteria and principles exist for evaluating usability and effectiveness of educational software and web-based learning environments (Albion, 1999; Tselios, Avouris, Dimitracopoulou & Daskalaki, 2001). The present survey was conducted using criteria of Squires and Preece (1999), who took Nielsen's (1994) classic set of usability heuristics and filtered them through the concepts of cognitive authenticity, contextual authenticity and socio-constructivism. They converted them to eight criteria that integrate usability and learning factors for predictive evaluation of educational software by experts *prior* to selecting systems for use. Seven of their criteria (some slightly amended) were used, along with two new criteria (9 and 10 below) and evaluation questions were generated for questionnaire surveys among learners, *after* use:

1. *Match between designer & learner models*: The software should represent cognitive tasks in ways that foster formation of a learner model consistent with the designer model.
2. *Navigational fidelity*: This involves investigating the structure of navigation, cosmetic authenticity and the effectiveness of the representation of the world within the system.
3. *Appropriate levels of learner control*: This relates to the balance between learner control, self-direction, customization, consistent protocols, and system responsibility.
4. *Prevention of peripheral cognitive errors and usability errors*: There is a relationship between domain complexity and error prevention. Peripheral usability-related errors should be anticipated and avoided, and novices versions could be provided.
5. *Understandable and meaningful symbolic representation*: The representational forms and symbols used are vital. Interfaces should present low cognitive demands and learners should not have to remember the forms of interaction. Symbols, icons and names used for learning objects should correspond with the subject domain and be used consistently.
6. *Personally meaningful forms of learning*: There are multiple methods and representations for supporting varying learning styles. Metacognition should also be supported and software should be used in tandem with other learner support materials.
7. *Strategies for cognitive error recognition, diagnosis and recovery*: While usability-related errors should be avoided (see 4), cognitive errors are part of the learning process,

as substantiated by Mayes and Fowler (1999:485), who stress that in educational applications ‘seamless fluency of use is not necessarily conducive to deep learning ... the software must make learners think’. Techniques such as cognitive conflict, scaffolding, and bridging should be used to promote a recognition–diagnosis–recovery cycle.

8. [*Match with the curriculum*: Irrelevant to the present study, since the software was custom-designed for the curriculum.]
9. *Distinctive features*: Features should be provided that are unique to the environment under evaluation and that support the particular requirements of its content and context.
10. *Capacity of the system to engage learners*: E-learning environments should motivate learners and hold their attention.

A pilot questionnaire survey with 20 students early in 2004 (de Villiers, 2004) elicited highly positive ratings. Two content errors were identified and corrected. In 2004 and 2005 evaluation questionnaires were included in the tutorial material, eliciting 49 returns in 2004 and 37 in 2005. The groups of respondents were thus samples of volunteers; they were not the entire body of students who used Relations. In both years the respondents comprised about 10% of the ‘serious students’ who wrote the final examination.

There are notable differences between the 2004 and 2005 groups of respondents, indicating the rapidly shifting profile of the student population. This may be a reason for the slight decline in quality ratings as assigned by respondents in the two years respectively. The 2004 respondents were a heterogeneous group. While lacking early computing backgrounds, most were comfortably computer literate. Sixty percent of them were 30 or younger. While 38% were fulltime students, 56% were fulltime employed and more than half of these had jobs which involved computing; 33% had a computer at home. However 68% had not used computers at school and 8% were encountering their first computing exposure. Eighty percent were also doing a programming module. More than half were from formerly disadvantaged groups of the South African population: 29% from Asian or Coloured population groups and 22% were Black.

The 2005 group of respondents contained more young, less experienced, fulltime students, in line with tendencies in the student population. To simplify comparison, the corresponding 2004 percentages are given again (in brackets after the 2005 figures). The 2005 group comprised 50% full-time students (38% in 2004); 39% (56%) in full time employment; 80% (60%) 30 or younger; and 86% (68%) had not used computers at school. Of the group 17% (8%) were encountering their first computing exposure; only 14% (33%) had a computer at home, although 69% (80%) were taking a programming module as well; 77% (51%) were from formerly disadvantaged groups: 62% Black (22) and 15% (29%) Asian or Coloured – a distribution in line with the module population. Many of them were ‘on-campus’ students – young people who study fulltime in the university’s regional libraries or learning centres, and have a social experience with peers like that of a contact-learning institution.

Questions in the form of statements investigated aspects of the nine selected criteria on a Likert scale. There were 40 questions in total, some with open-ended sections for elaboration of responses. This paper does not deal with all 40; two or three questions are used to represent each criterion. The Likert scales offered five options, but here too, information is condensed by integrating categories. The findings of 2004 and 2005 are kept separate.

Match between learner and designer models (Criterion 1):

There were highly positive ratings regarding the cognitive effectiveness of the lesson, i.e. how the learners understood the theory of relations. Some open-ended responses were: ‘The teaching content is absolutely brilliant.’ / ‘Practicing makes it easier to learn and understand.’

Doing exercises in the <i>Relations</i> lesson helps me understand the theory of relations:					
Strongly agree or Agree	2004	98%	Maybe or Disagree	2004	2%
	2005	95%		2005	5%
The elaborations and explanations help me understand:					
Strongly agree or Agree	2004	98%	Maybe or Disagree	2004	2%
	2005	84%		2005	16%
Relations helps me understand a difficult section of the module, <i>Theoretical Computer Science I</i> :					
Strongly agree or Agree	2004	78%	Maybe or Disagree	2004	22%
	2005	72%		2005	28%

Navigational fidelity (Criterion 2):

The 2004 responses mainly described *Relations* as easy to use and easy to navigate. The 2005 responses were more conservative; 22% had ‘got lost’, probably due to their novice status.

I learned to use <i>Relations</i> :					
Very quickly or Quickly	2004	75%	Slowly	2004	25%
	2005	59%		2005	41%
When I use <i>Relations</i> , I know exactly where I am, what parts I have done, and what I should still do:					
Strongly agree or Agree	2004	92%	Maybe or Disagree	2004	8%
	2005	81%		2005	19%
The lesson is easy to operate:					
Strongly agree or Agree	2004	94%	Maybe or Disagree	2004	6%
	2005	86%		2005	14%
I got ‘lost’:					
Strongly agree, Agree or Maybe	2004	6%	Disagree or Strongly disagree	2004	94%
	2005	22%		2005	78%

Some asked for more sensitive navigation facilities: ‘I would rather use the menu to be exactly where I want to be next.’ / ‘It would be useful to navigate to any parts of the lesson, rather than only having the selection between doing the lessons and exercises, or diving straight to the exercises.’ / ‘...frustration when you’ve got to go through all the trial examples to reach the next part.’ While taking cognizance of this, it must be stated that the strict learner control in certain parts deliberately forces users to work through sets of exercises for pedagogical reasons. On the other hand, just before exams, greater flexibility would expedite quick, focused use for revision.

Appropriate learner control (Criterion 3):

Students appeared generally satisfied, but open-ended responses show that some want more control, also exit and re-entry facilities and the option to leave a section without completing it. In the 2004 group, 88% wanted to choose the level of difficulty (a facility not provided), whereas of the more tentative respondents in 2005, only 58% would appreciate that choice, the others preferring the pre-set sequences with increasing difficulty.

I enjoy being in control of the lesson and able to choose what to do next:					
Strongly agree or Agree	2004	90%	Maybe or Disagree	2004	10%
	2005	84%		2005	15%
I appreciate being able to choose whether to do theory and exercises, or go straight to the exercises:					
Strongly agree or Agree	2004	92%	Maybe or Disagree	2004	8%
	2005	87%		2005	13%

I would like to choose the level of difficulty of the exercises:					
Strongly agree or Agree	2004	88%	Maybe or Disagree	2004	12%
	2005	58%		2005	42%

Some requested more sophisticated customization and control: ‘It would be nice to completely disable certain kinds of screens.’ / ‘Allow students to select right at the beginning whether they want these screens or not.’ / ‘Maybe a separate menu for exercises.’

Prevention of usability errors and peripheral cognitive errors (Criterion 4):

Domain complexity should not be simplified. Content-related errors should be permitted as part of the learning process, but situations that cause usability errors and perception errors should be avoided. The problem (third question below) of trying to use operations familiar from commercial systems occurred more with the 2004 group, probably due to their greater familiarity with other software. (This also occurred with two heuristic evaluators – see evaluation case study 3.) In general *Relations* was pleasant to use, e.g. ‘It is visually stimulating and appealing’.

I made mistakes in the way I used Relations, i.e. it was difficult to use the system:					
Agree or Maybe	2004	35%	Disagree or Strongly disagree	2004	65%
	2005	44%		2005	56%
I made mistakes in using the system at a certain stage of my learning:					
The first time I used Relations		The same amount each time		Hardly any at any time	
2004	63%	2004	19%	2004	18%
2005	52%	2005	17%	2005	31%
I made mistakes because I used operations and keystrokes I know from another system:					
Strongly agree or Agree	2004	28%	Maybe or Disagree	2004	72%
	2005	8%		2005	92%

Meaningful symbolic representation (Criterion 5):

Ratings about the symbols, icons and names in *Relations* were positive, although there were a few complaints about typing in brackets and commas. The different bracket styles are vital notations in set theory and relations. In some questions these are pre-positioned, and learners provide content only, but in others brackets must be keyed in as part of the required answer.

The symbols and names that represent mathematical objects are used consistently in the lesson:					
Strongly agree or Agree	2004	83%	Maybe or Disagree	2004	17%
	2005	94%		2005	6%
The screen layouts are easy to read:					
Strongly agree or Agree	2004	85%	Maybe or Disagree	2004	15%
	2005	92%		2005	8%
It is easy to grasp the structure of the lesson:					
Strongly agree or Agree	2004	85%	Maybe or Disagree	2004	15%
	2005	80%		2005	20%
There is too much information on the screens and it confuses me:					
Strongly agree or Agree	2004	19%	Maybe or Disagree	2004	81%
	2005	25%		2005	75%

Personally meaningful ways of learning (Criterion 6):

Socio-constructivist learning theory holds that learners should experience personal satisfaction with the process and style of learning. The tutorial structure of *Relations* tends more to behaviour-

ism than constructivism but, where possible, personal learning styles were supported in the design.

When a concept was taught or illustrated in more than one way, it helped me understand it:					
Strongly agree or Agree	2004	90%	Maybe or Disagree	2004	10%
	2005	91%		2005	9%
The ways that <i>Relations</i> helps me understand are through (participants could choose more than one option):					
Diagrams	Real life pictures		Animations		Practicing
2004	69%	2004	17%	2004	32%
2005	74%	2005	24%	2005	38%
2004		2004		2004	69%
2005		2005		2005	65%
I enjoy approaching studies collaboratively, i.e. working with a fellow-student:					
Strongly agree or Agree	2004	42%	Maybe or Disagree	2004	58%
	2005	40%		2005	60%

The multiple representation approach was used frequently. Concepts were defined, then illustrated with diagrams or animations. This was appreciated: ‘It is easier to understand, since a picture is formed in one’s head.’ There were spontaneous comments about how the examples and diagrams helped them to grasp concepts and concretize the theory, i.e. inductive learning. Collaborative work is part of the social constructivism trend and although it is complex to implement in distance learning, it is occurring naturally with the students who work in learning centres or labs. Several said they had used *Relations* with fellow-students: ‘We can discuss, argue ...’. Some had used it along with the textbook or with their own summaries.

Cognitive error recognition, diagnosis and recovery (Criterion 7):

This relates to domain complexities, misconceptions and pitfalls, and the adequacy of the feedback in helping learners to correct them. The redesign of the lesson had concentrated on HCI features and enhanced control. An aspect not changed was the original feedback, which paid meticulous attention to potential errors. Apart from some requests for ‘harder exercises’, the feedback obtained high praise: ‘It accurately guessed what I did wrong and pointed me in right direction.’ / ‘I could easily pinpoint my mistake the next time.’ / ‘It explains clearly where I went wrong. Those hints really help.’ / ‘The incorrect part was pointed out and they give another chance to try.’ / ‘The feedback uses definitions and helps me to remember definitions and theory.’ / ‘After doing this, it helped me answer the assignment questions correctly.’

I made mistakes in doing the exercises:					
Got most wrong	Got some wrong		Got a few wrong		Hardly any mistakes
2004	0%	2004	42%	2004	46%
2005	8%	2005	31%	2005	42%
2004		2004		2004	12%
2005		2005		2005	19%
The feedback (system responses) to my incorrect answers helped me:					
Strongly agree or Agree	2004	92%	Maybe or Disagree	2004	8%
	2005	90%		2005	10%
When I got an answer wrong, I was able to get it right on the next try.					
Always	Nearly always		Sometimes		Not often
2004	25%	2004	56%	2004	15%
2005	36%	2005	44%	2005	17%
2004		2004		2004	4%
2005		2005		2005	3%

General: distinctive features; motivation and engagement (Criteria 9 and 10)

After complex sections, an innovative water-based recreational theme emerges in the form of displays showing animated water activities – a pool, ducks, boardsailing – accompanied by audio. Supporting the present author's Criterion 10 of motivation and engagement (Hodges, 2004; Keller & Suzuki, 1988), these features were designed in line with dual means of motivation (Alessi & Trollip, 2001), namely, intrinsic motivation – by content that engages learners, and extrinsic motivation – by moments of informality and humour. Responding to the first question below, over 70% chose either the first or second option or both, indicating appreciation of the recreational theme. A very low percentage was 'irritated'.

The recreational screens with the water theme (participants could choose more than one option):							
Are fun		Give me a break		Make no difference		Are irritating	
2004	43%	2004	43%	2004	26%	2004	6%
2005	44%	2005	45%	2005	25%	2005	3%
The shortcut 'Dive' screens to avoid redoing theory are useful.							
Strongly agree or Agree		2004	84%	Maybe or Disagree		2004	16%
		2005	70%			2005	30%

Open responses were mainly positive: 'They were amusing.' / 'It was FUN working with it.' / 'The fun screens attracted me – they make the program friendly' / 'A bit of a surprise! I liked the sounds, they help me relax' / 'It's too stressful to do the whole CD in one go – the screens with water help me take a break.' But: 'When I was tired, the breaks were useful, otherwise they wasted time.' / 'Fun screens slow you down when seriously revising.' A 19-year old criticised our animations: 'Purchasing this CD was a good investment. It helped me grasp concepts much easier. Thanks, guys, for a terrific job. But the animation is quite poor. I know you aren't Hollywood studios, but you are capable of something more exciting. Take it as a challenge.' Another 19-year old said: 'Well done, you guys, it's great! It beats reading the book as it goes straight to the point.' So *us guys* get bouquets and brickbats ... Are any Disney animators looking for instructional design posts in Africa?

Students appreciated the 'dive' option to bypass theory and go straight to exercises. 'If you know the theory, you can just go for the exercises.' / 'It's just great not re-doing the theory after the first time.' / 'It saves time.' / 'It will be useful to dive when revising for exams.' Referring to the neutral character that symbolizes the dive, 'Pictures of real babes diving would be nice!'

There were no explicit questions on the aesthetics of *Relations*, but several students mentioned the attractive, easily-readable screen displays. The 2005 group did not identify any content errors in the system. This is not surprising, given the thorough investigation the system had undergone.

Finally, 'This lesson encourages you' / 'These exercises are much more fun ... compared to the study guide.' / 'More of the same, please!' / 'Keep it up for giving us such a lesson on computer. I think you can do it for other chapters too.' / 'I would like CAI lessons for all modules'.

Note: Difference between the 2004 and 2005 groups of respondents

The findings of the 2004 and 2005 questionnaire surveys were statistically analysed, and indicate that responses to certain questions show significant differences between the groups and some anomalies. This issue is addressed again in the Conclusion. Despite using the improved V2.2, the 2005 group were less positive in some judgements. The statistics in the preceding pages show that they found the system less easy to learn and use and the subject matter more difficult. Despite this, their responses indicate that they apparently made less mistakes than the class of 2004,

which is an anomaly — see the findings under Criteria 4 and 7. A further point of interest is that 58% of the 2005 group would like to choose the level of difficulty in contrast to 88% in 2004 (see Criterion 3). This may indicate less self-confidence and less assertiveness on the part of the 2005 respondents. Another possible explanation of the discrepancies is that some students may not have taken the evaluation seriously. It appeared that a few respondents had rushed through the process. They did not tackle all the exercises or delve deeply into the material.

An interesting fact emerged from investigation of results in the final examinations. Across the board, the average percentages obtained in 2005 exam results in all modules that comprise *Computer Science 1* were 2% to 7% lower than in 2004. This might be related to an increase in the number of young, less-experienced students. This shift in the composition of the student population could also contribute to the more tentative responses to the questionnaire.

Query Method: Interviews with Students

Semi-structured interviews (Preece et al, 2002) were held with five volunteers in 2005 to investigate aspects not explicitly addressed in the questionnaire. The subjects were fulltime students in their first or second year of study, with average age 21. One had a prior diploma qualification in electronics.

They had all used the prescribed book first and then *Relation*, and had experienced a synergy from the mixed model of textbook and CAI lesson. Added value came from the detailed explanations, animated diagrams showing development of concepts, and the diagnosis of errors so that they could be corrected. A quote: ‘The computer is your teacher and you don’t have to make an appointment’. All five found *Relations* easy to learn and use. In line with the motivation and engagement criterion, it engaged them and they experienced flow (Csikszentmihalyi, 1990): ‘I was just enjoying it...’ / ‘I want to do this’.

The relative value of the theoretical teaching segments and the practical exercises was investigated to see if *Relations* fitted students’ personal learning styles. It appears to be on target in addressing more than one style, because the respondents were divided equally in their preferences. Asked which they *liked doing most*, two chose the theoretical teaching, saying respectively: ‘the explanations are different from the book’ and ‘because of those diagrams’, and two chose the exercises: ‘with that second chance to get it right’. The fifth enjoyed both equally. Regarding *learning value*, they did not all make the same choice as they had made for *liking*. One chose ‘the theoretical definitions, because they come with examples’; two selected exercises, because: ‘you learn more when you do things’ and ‘when you just read, you get lazy’. The other two were adamant about both together: ‘the teaching to know and the exercises to prove you know’ and to ‘show how far you have learned from the teaching’.

Collaborative work was addressed. This style was previously impractical for UNISA, since the student population consisted of classic distance-learners working in isolation. But with the advent of fulltime students, working on-campus or at learning-centres, it is feasible. Four liked working with fellow students. Two had used *Relations* jointly with a friend, and two preferred to do the lesson alone and discuss issues with peers afterwards.

With regard to special features, all five spontaneously mentioned the diagrams. The more experienced postgraduate student felt: ‘The most important thing about *Relations* is its usability’. Another stressed the different ways of using it at different times – for exam revision, he would use the Dive facility to skip theory and go direct to practice.

Post-examination investigation of on the five subjects indicated above-average performance. Of the 400 students who wrote the exam in 2005, only 52% actually passed. But these five all passed, obtaining between 50% and 65%. The average mark for the exam question on mathemati-

cal relations was 57.5%, but the interview group averaged 65% for that question. This shows that they were diligent students and enhances their credibility as subjects.

Inspection Method: Heuristic Evaluation by Experts

Heuristic evaluation (HE) (Dix, Finlay, Abowd & Beale, 2004; Nielsen, 1994; Preece et al, 2002; Squires, 1997) is an inspection technique developed by Jakob Nielsen and Rolf Molich. Expert evaluators examine a prototype or an operational system, using a set of guidelines or usability principles called *heuristics*. A recommended number of evaluators is three to five, but HEs of *Relations* were conducted by six evaluators, peer academics who are subject matter experts. Five are involved in teaching various levels of Theoretical Computer Science and the sixth is a young postgraduate student in this area. In a participative action research approach, two had been involved in the initial 1994 evaluation. They mainly used the same criteria as those in the students' questionnaire but, as subject matter experts, the emphasis differed – it was less on whether exercises were 'difficult' and how many they got wrong – and more on system content, rigour, usability and navigation. They worked through the lesson comprehensively, simulating students' activities. Of the six, four did HEs late in 2004 on a 'sanitized' V2.1 with minor improvements fed in after the 2004 student evaluation. No real usability problems were identified. The critiques identified subtle mathematical shortcomings, suggested some additional functionality, and proposed minor modifications to the control structure. These refinements were used to generate V2.2, which was released in 2005, and the other two experts heuristically evaluated this version. Despite the era of hyperlinked web-based learning environments, there was general consensus that the control and usage patterns of *Relations* V2.1 and V2.2 – patterns based on traditional CAI architecture – operate effectively.

The data from all six assessments is integrated under the selected criteria as category headers. Where relevant, distinctions are made between V2.1 and V2.2.

Match between designer and learner models

The evaluators unanimously supported the educational value of *Relations*. All agreed that the exercises, system feedback, elaborations and explanations in the CAI lesson would help students learn the subject matter. It was the designer's intention that certain exercises should be difficult, but opinions varied equally as to whether all were easy or whether some were really complex. (A belief that none were difficult, might serve to establish an evaluator's credentials as a true guru!) Two subtle mathematical inaccuracies were identified in the feedback.

Navigational fidelity

They all learned to use *Relations* quickly and found it easy to operate. Navigation was not experienced as complex; five of the six knew their exact location within the lesson at any time, although one experienced getting 'lost'. Screen layouts were said to be easy to read and one evaluator commented particularly on the simplicity, aesthetics and visibility of the uncluttered screens, which enhanced the message in comparison to other CAI lessons she had encountered.

One made useful suggestions about the introductory *Read-Me* documentation on the CD; these were incorporated into the latest release.

Learner control

The evaluators appreciated the level of control offered to learners in terms of sequence and section control, although the young evaluator stated he was not 'totally in control'. For example, in

places the *Back* button is explicitly deactivated, and both he and a 'techie' student in the learner evaluation objected to this system control. Moreover, three evaluators of V2.1 and both evaluators of V2.2 felt that users should be able to choose the level of difficulty of exercises, rather than working through each entire set. While noting this concern, such a modification was not made in V2.2, because it would involve major structural change. Another problem identified was the lack of exit and re-entry facilities in the sets of exercises.

Usability errors and problems

With relation to mistakes in using the system interface, the experts did not find *Relations* difficult to use. Two evaluators made initial errors by using operations and keystrokes they knew from commercial software systems, but which were invalid in *Relations*, e.g. double clicking on an object when a button in the control area should be clicked. This problem, which also occurred with students in the questionnaire survey, identifies non-adherence to Dix et al's (2004:260) generalizability subprinciple of the 'principles to support usability'.

Symbols, icons and names

No hindrances were identified and the evaluators were unanimous that *Relations* does not make cognitive demands on memory. Their findings indicate that there is adequate visibility, in that it operates on recognition rather than recall, and the meanings of all symbols and icons are clear. No evaluator got 'stuck' at any stage.

Personally meaningful ways of learning

The experts appreciated the multi-perspective approach of demonstrating concepts. They all found that the diagrams and animated developments communicated messages effectively and that the extensive practice opportunities reinforced concepts. It was an evaluator of V2.1 who suggested the feature of obtaining more detailed information and definitions by clicking on hotspots. This was implemented in V2.2 and was appreciated by those who inspected it, as well as by V2.1 evaluators who were shown this new feature.

Recognizing cognitive mistakes and recovering

The experts made very few cognitive mistakes! When they did, the feedback helped them get it right on the next try. However, they deliberately challenged the system with novel or obscure responses, resulting in some refinements to diagnostic feedback.

Special features

Four of the six agreed that the recreational screens were fun; one found them irritating and another felt they made no difference.

Post-test

The role of post-tests in measuring learning is debatable, but they are a traditional quantitative measure of the effectiveness of a learning system/intervention. In an educational context they are also appropriate instruments to assess usability, since without ease of use, functionality is obscured. *Relations* includes a revision test, but due to its role as a non-threatening system for learning gain, scores are not recorded, so this test does not serve as a post-test. As a suitable 'post-test', final examination performance of the questionnaire respondents (intervention group) was compared to the performance of the rest of the students (considered a control group) – see Table

1. One of the exam questions related directly to the subject matter on mathematical relations, as taught in *Relations*. Scores obtained in this question are given in the central row of the table, and the last row indicates (as a percentage) how many students passed that question, irrespective of the actual score (mark) they obtained for it.

Table 1: Examination performance of survey participants

Criterion	2004		2005	
	Participants in the questionnaire survey (study group)	Other students (control)	Participants in the questionnaire survey (study group)	Other students (control)
Average percentage in final end-of-year examination	58%	54%	53%	50.5%.
Average score (mark) for the question on relations	78% (7% more than in control group)	71%	61% (4% more than in control group)	57%
What percentage of students obtained a pass mark for that question?	87% (12% more than in control group)	75%	73% (6% more than in control group)	67%

If effectiveness is measured by grades and performance, then *Relations* resulted in effective learning. Table 1 shows that 12% more of the survey participant group passed the relevant question in 2004 and 6% more in 2005. The average mark obtained by participants was 7% higher for that question in 2004 and 4% higher in 2005, while their averages for the full examination were 4% and 2.5% higher respectively.

For the 2004 data, scores were statistically investigated to determine whether there was a linear association between achievement in the exam as a whole and the mark for the question on relations. Following the exclusion of four outlying cases, the p-value was greater than 0.05, indicating that the correlation was not significant and there was no direct proportional relationship between the two marks. This indicates that several participants in the study attained higher marks for that question than would have been expected from their overall performance, i.e. those who undertook this optional learning experience enhanced their skills in that section of the curriculum. A similar analysis was not done on the 2005 data, since the exam question on relations in 2005 was not as close to the content of the *Relations* lesson as was the case in 2004.

NB: Participation in the questionnaire survey was voluntary, so the respondents do not represent all the students who used *Relations*. Hence the ‘other students’ group includes some who actually were participants in the study (i.e. users of *Relations*) though not participants in the survey. If they had completed the questionnaire, and thus been identified as subjects within the study group, it is probable that the ‘other’ group, then comprising only genuine control students (non-users of *Relations*) would have achieved yet lower performance.

Conclusion and Recommendations

In an action research approach, *Relations* has undergone a continuous process of evaluation, reflection, correction and refinement. Formative and summative evaluations have been conducted; functionality, usability and educational effectiveness have been investigated. Each new cycle reveals usability problems or subtle content flaws – increasingly minor over time.

This paper has presented findings about the functionality and usability of *Relations*, which lead finally to discussions, generic reflections and lessons in two contexts. First, approaches to the design of CAI applications are briefly addressed, several of which are relevant to e-learning in general. The second contribution is a consideration of the different evaluation techniques and their roles in the multi-method evaluation of *Relations*.

Design of CAI Applications: Reflection and Lessons

The role of traditional computer-aided instruction and learning in the WWW era

The development team of *Relations* believes that conventional CAI has a role to play in the milieu of e-learning. Classic CAI can present efficient instruction, motivate and engage learners, challenge them with meaningful exercises, and can support effective learning. The pleas by the students for more such tutorial and practice environments should be noted. *Relations'* greatest strength is its excellent diagnostic feedback, attested to by learners and expert evaluators alike. Courseware authoring systems such as *Quest 7.0*, which was used to develop *Relations*, have powerful facilities that can be used by developers to assess the forms of learner-input and provide appropriate, detailed, tailor-made feedback. This can be done in web-based instruction (WBI) too, using specialized web-programming languages, but it is more complex and time-consuming. A further obstacle to the use of WBI is that many UNISA students lack broadband Internet access.

An expert evaluator and some students requested hi-tech effects – *Flash* animations, 3-D modeling, etc. Such suggestions must be considered, but not all special effects are compatible with conventional authoring tools. A word of caution is sounded not to use special effects, 'bells and whistles', for their own sake, but rather to employ them where they can illustrate concepts in a way that enhances learning and cognition.

Lessons about usability

The concepts of usability and interaction design are receiving increasing attention in the development of commercial and corporate software. It is equally, or even more, important to produce usable applications in educational contexts where the users are not workplace professionals, but learners who must first be able to use and interact with a system before they can even commence learning.

Many learners approach e-learning experiences after exposure to commercial software. As far as possible, learning applications should use operations and keypresses that support the HCI principles of predictability and consistency with familiar systems. The *Quest* authoring tool does not naturally support the generation of software that is identical in appearance and functionality to commercial systems. However, *Relations* was found to be easy to learn and use. It adheres to the fundamental principle of *internal consistency*, where a system's own internal operations are characterized by predictability and visibility.

Modifications for *Relations V3* (if such) and lessons for new CAI environments:

The teaching approaches used in *Relations* presented multi-perspectives on a concept and multiple representations of content. These methods were successful and could be supplemented by multi-modal presentations in contexts where they would add value.

The concerns raised relate mainly to control and navigation. Where pedagogically appropriate, linear sequences should be supplemented with hyperlinked control. Exit and re-entry facilities are required. User-control should be provided over the level of difficulty, so that users can choose particular exercises. This would facilitate different ways of usage at different stages of study.

The hyperlinks incorporated into *Relations* V2.2 were a success. This feature of double-clicking on content hotspots for definitions, elaborations or cross references should be regularly employed.

Some users did not appreciate *Relations*' system control that deactivated certain buttons in particular cases. Although this was carefully planned by the design team for appropriate contexts, some users saw it as inconsistent. Ways of implementing such tactics must be carefully chosen.

Different Evaluation Techniques: Reflection and Lessons

This paper concludes with a review of the different evaluation techniques and their roles in the multi-method evaluation of *Relations*. The complementary use in tandem, of four techniques and triangulated data provided a synergic framework, generating information superior to that which could have been obtained by any single method. Evaluation is currently an important topic in the HCI discipline and these reflections may add some insights.

Heuristic evaluations

Heuristic evaluations, whereby experts inspect (and sometimes analyse) a system, are valuable sources of critique and suggestions. They are cost-effective and can be done in a short time frame.

Experts can move beyond the given evaluation criteria. With critical analysis and lateral thinking, they can identify omissions and suggest additional aspects. Their judgements are very valuable and may be transferable to the development of new systems.

Since the number of participants is small, the data obtained is qualitative rather than quantitative.

The discernment, insight and long-standing expertise of the heuristic evaluators accurately pinpointed the strengths and weaknesses of *Relations*, as well as subtle mathematical inaccuracies. As leading subject matter experts, two tested the feedback to its limits, revealing flaws and compromises of mathematical rigour. For overall evaluation of an application, the team of heuristic evaluators should include experts with subject matter skills and those with usability expertise.

Questionnaire surveys among end-users

The students' responses were very useful with regard to educational aspects. For example, the questionnaire evaluation confirmed the value of the detailed feedback, showing that the time spent painstakingly developing it was worthwhile. Furthermore, the 2004 questionnaire survey elicited particularly useful information and identified some errors that were addressed even before V2.1 underwent heuristic evaluation. The 2005 survey did not provide as much value, mainly because there was not much to add to the exhaustive information obtained in 2004.

Open-ended responses were useful in elaborating responses and motivating the Likert options selected. In general, students used these to express praise and appreciation.

A particular value of the 2005 survey was its role in identifying differences between the groups of respondents. As stated, there were differences between the responses of the groups of 2004 and 2005. Possible indifferent or rushed attitudes on the part of just a few respondents can skew the data and impact on the reliability of the findings. On the other hand, findings should be interpreted in the context of the big picture, such as the general decline in examination performance in 2005. This may be related to the increase in 2005 of young, fulltime, recent school-leaver students. The 2004 group of respondents, on the other hand, contained more mature students, 56% of

them in fulltime employment. This difference might explain some of the more tentative responses in 2005, in spite of the fact that these students were using an improved version.

Questionnaire design is vital for the elicitation of useful information. An inadequate questionnaire can trivialize an evaluation. Likert-scale questionnaires with specified options may exclude some important aspects, but the inclusion of open-ended questions for qualitative responses can be used to compensate and to enrich the information obtained. Unanticipated aspects that emerge can be probed further in interviews or in subsequent questionnaire surveys.

Interviews

Interviews can explore novel avenues and can obtain unexpected information. In the synergy that exists between various methods, such as interviews, questionnaires, and think-aloud observation, aspects that emerge from evaluation by one technique may be investigated in depth using another technique. Examples of aspects that can be followed up in interviews are usability problems and explicit identification of features that foster learning.

Since interviews can be done only with a limited number of subjects, the sample should *ideally* be heterogeneous and scientifically selected. UNISA's distance-teaching context makes personal contact difficult and a sample of volunteers was used in this study, representing a fairly homogeneous group of young fulltime students, typical of the current shift in learner profile. However, contact with a stereotypical group is informative in its own right and, in this case, very rewarding, due to the responsible and serious attitude evidenced by these five young people towards their studies and their subsequent success in the exam.

To optimize on interviews, the structure should be flexible. A good semi-structured interview should be based on a set of common core questions, presented by a researcher who has the discernment to know how to present good follow-up questions and when just to listen. This approach can elicit rich information. If an interview is rigid and can only give rise to specific forms of answers, one might as well stick to questionnaires, which are less labour-intensive.

Post test (a question in the final exam)

Post-tests give quantitative measures that can be statistically analysed, in particular when they can be compared to a similar pre-test.

However, scores in tests remain debatable instruments for measuring the effectiveness of an artefact or other intervention in improving learning. Frequently they show no significant difference between methods of instruction, even when other evaluation methods generate favourable reports of the intervention. When a test, therefore, does indicate improved academic performance, it would appear that the method under investigation indeed enhances learning. This was the case in the present evaluation.

In Closing...

Further evaluation research is planned, entailing formal usability testing in UNISA's new usability laboratory. Performance on specified learning tasks will be monitored by logging, video recording and by capturing usability metrics.

The participative action research process of developing, evaluating and refining the e-learning system, *Relations* has taught the designer and development team a great deal – not just about the application being studied, but also lessons and generic principles for the design of e-learning applications as well as lessons about the rich complementary roles of different evaluation methods and techniques.

References

- Albion, P.R. (1999). Heuristic evaluation of educational multimedia: From theory to practice. Retrieved December 2005, from <http://www.usq.edu.au/users/albion/papers/ascilite99.html>
- Alessi, S.M. & Trollip, S.R. (2001). *Multimedia for learning: methods and development* (3rd ed.). Massachusetts: Pearson Education Company.
- Baskerville, R.L. & Wood-Harper, A.T. (1996). A critical perspective on action research as a method for Information Systems research. *Journal of Information Technology*, 11, 235-246.
- 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.
- Cohen, L., Manion, L. & Morrison, K. (2000). *Research methods in education* (5th ed.). London: RoutledgeFalmer.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- De Villiers, M.R. (2004). Usability evaluation of an e-learning tutorial: Criteria, questions and case study. In G. Marsden, P. Kotze, & A. Adesina-Ojo, *Fulfilling the promise of ICT*. Proceedings of SAICSIT 2004: 284-291. ACM International Conference Proceedings Series.
- De Villiers, M.R. (2005a). Three approaches as pillars for interpretive Information Systems research: Development research, action research and grounded theory. In J. Bishop & D. Kourie, *Research for a changing world*. Proceedings of SAICSIT 2005: 142-151. ACM International Conference Proceedings Series.
- De Villiers, M.R. (2005b). e-Learning artefacts: Are they based on learning theory? *Alternation*, 12 (1b), 345-371.
- Dick, B. (1999). What is action research? Retrieved December 2005, from <http://www.scu.edu.au/schools/gcm/ar/whatisar.html>
- Dix, A., Finlay, J., Abowd, G.D. & Beale, R. (2004). *Human-computer interaction* (3rd ed.). Harlow: Pearson Education.
- Du Poy, E. & Gitlin, L.N. (1998). *Introduction to research: Understanding and applying multiple strategies* (2nd ed.). St Louis: Mosby Inc.
- Hodges, C. (2004). Designing to motivate: Motivational techniques to incorporate in e-learning experience. *The Journal of Interactive Online Learning*, 2, 3.
- Jonassen, D. (1999). Designing constructivist learning environments. In C.M. Reigeluth. (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory Volume II*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Keller, J.M. & Suzuki, K. (1988). Use of the ARCS motivational model in courseware design. In D. H. Jonassen, (Ed.), *Instructional designs for microcomputer courseware*. N.J.: Lawrence Erlbaum Associates.
- 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. (1994). Heuristic evaluation. In J. Nielsen & R.L. Mack, (Eds.), *Usability inspection methods*. New York: John Wiley & Sons.
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design*. New York: John Wiley & Sons.
- Reeves, T.C. & Hedberg, J.G. (2003). *Interactive learning, systems evaluation*. Englewood Cliffs, N.J.: Educational Technology Publications.
- Reeves, T.C. & Reeves, P.M. (1997). Effective dimensions of interactive learning on the World Wide Web. In B.H. Khan (Ed.), *Web-based Instruction*. Englewood Cliffs, N.J.: Educational Technology Publications.

- Squires, D. (1997). An heuristic approach to the evaluation of multimedia software. Retrieved December 2005, from <http://www.media.uwe.ac.uk/masoud/cal-97/papers/squires.htm>
- 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.
- Tselios, N.K., Avouris, N.M., Dimitracopoulou, A. & Daskalaki, S. (2001). Evaluation of distance-learning environments: Impact of usability on student performance. *International Journal of Educational Communications*, 7(4), 355-378.
- Zuber-Skerrit, O. (1992). *Action research in higher education*. London: Kogan Page.

References

Ruth de Villiers is a professor in the School of Computing of the University of South Africa. She has an MSc, an MEd and a PhD (Computer-Integrated Education). Ruth teaches courses in human-computer interaction and theoretical computing. Her cross-discipline research areas – in which she also supervises postgraduate students – are design, usability and evaluation of educational applications; e-learning in general; HCI education; and meta-research.