

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## The Design, Use and Effectiveness of Different Forms of Content in e-Learning Tutorials

### Thesis

How to cite:

Ellis, Jim (2011). The Design, Use and Effectiveness of Different Forms of Content in e-Learning Tutorials. EdD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2011 The Author

Version: Version of Record

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)



The Design, Use and Effectiveness  
of Different Forms of Content  
in e-Learning Tutorials

Jim Ellis

(X8486729)

Doctor of Education (EdD)

October 2010

DATE OF SUBMISSION : 27 OCTOBER 2010

DATE OF AWARD : 9 AUGUST 2011

# Contents

<b>Chapter 1 Introduction</b> .....	1
<i>Evolution of e-Learning</i> .....	2
<i>The Nature of CAL</i> .....	8
<i>Scope of Research</i> .....	10
<i>Target audience</i> .....	11
<i>Personal background</i> .....	12
<b>Chapter 2 Multimedia Design Principles</b> .....	14
<i>Cognition and Learning Design</i> .....	18
<i>Dual coding</i> .....	20
<i>Mix of Media</i> .....	23
<i>Text</i> .....	23
<i>Proposition 1</i> .....	27
<i>Static imagery</i> .....	28
<i>Animation</i> .....	29
<i>Audio</i> .....	31
<i>Video</i> .....	33
<i>Media richness</i> .....	36
<i>Proposition 2</i> .....	39
<i>Summary</i> .....	39
<b>Chapter 3 The Nature of Tutorial Content</b> .....	41
<i>The Nature and Effects of Interest</i> .....	41
<i>Situational interest</i> .....	44
<i>Concreteness</i> .....	45
<i>Seductive Details</i> .....	46
<i>Issues with research into seductive details</i> .....	49
<i>Rethinking seductive details</i> .....	54
<i>Supporting the Learning Process</i> .....	55
<i>Catalytic Content</i> .....	58
<i>Analysing and identifying catalytic content</i> .....	59
<i>Theoretical basis for catalytic content</i> .....	64
<i>Summary</i> .....	70
<i>Proposition 3</i> .....	72
<i>Conclusions</i> .....	72
<b>Chapter 4 Research Methodology</b> .....	73
<i>Research Approach</i> .....	74
<i>Qualitative research</i> .....	76
<i>Quantitative research</i> .....	79
<i>Pragmatism</i> .....	81
<i>Chosen research approach</i> .....	82
<i>Development of Experiments</i> .....	83
<i>Design variants</i> .....	84
<i>Learning outcomes</i> .....	86
<i>Sample</i> .....	86
<i>Method of Data Collection</i> .....	88
<i>Participant testing</i> .....	88
<i>Questionnaires</i> .....	89
<i>Interviews</i> .....	90
<i>Procedure</i> .....	91
<i>Ethical considerations</i> .....	92
<i>Research Integrity</i> .....	92
<i>Validity</i> .....	93
<i>Methods of Data Analysis</i> .....	94
<i>Methodological Limitations</i> .....	97
<i>Summary</i> .....	98

<b>Chapter 5 Presentation of Results</b> .....	100
<b>Results</b> .....	101
<i>Initial checks</i> .....	101
<i>Proposition 1a</i> .....	111
<i>Proposition 1b</i> .....	111
<i>Proposition 2a</i> .....	112
<i>Proposition 2b</i> .....	113
<i>Proposition 2c</i> .....	114
<i>Proposition 2d</i> .....	115
<i>Proposition 3a</i> .....	115
<i>Proposition 3b</i> .....	116
<i>Interview Outcomes</i> .....	117
<i>Summary of Results</i> .....	117
<b>Chapter 6 Discussion of Results and Method</b> .....	119
<b>Discussion of Results</b> .....	121
<i>Effects of wordage on appeal and effectiveness</i> .....	121
<i>Effects of different media on appeal and effectiveness</i> .....	122
<i>Effects of catalytic text on appeal and effectiveness</i> .....	124
<b>Validity of Data</b> .....	125
<i>Internal validity</i> .....	125
<i>Statistical conclusion validity</i> .....	127
<i>External validity</i> .....	128
<b>Evaluation of the Methodology</b> .....	129
<i>Experiment format</i> .....	129
<i>Tutorial designs</i> .....	130
<i>Rigour versus realism</i> .....	131
<b>Chapter 7 Conclusions and Recommendations</b> .....	132
<b>Recommendations</b> .....	134
<i>Recommendations for CAL designers and practitioners</i> .....	134
<i>Recommendations for further research</i> .....	134
<b>Specification for a Future Study</b> .....	136
<i>Redefining catalytic content</i> .....	136
<i>Experiment design</i> .....	137
<i>Sample</i> .....	140
<i>Experimental procedure</i> .....	141
<i>Analysis</i> .....	141
<i>Epilogue</i> .....	142
<b>References</b> .....	143
<b>Appendix A: Glossary of Terms</b> .....	159
<b>Appendix B: Pack for Experiment Participants</b> .....	160
<b>Appendix C: CAL Tutorial Screen Samples</b> .....	168
<b>Appendix D: Analysis of Text-Based Design Variants</b> .....	169

## Figures

<b>Figure 2.1. Three stage model of memory</b> .....	19
<b>Figure 2.2. Dual channel cognitive model of multimedia learning</b> .....	21
<b>Figure 2.3. PET scans of neural activity in response to different stimuli</b> .....	22
<b>Figure 3.1. Compilation of some causes and effects of interest</b> .....	42
<b>Figure 3.2. Examples of seductive details placement</b> .....	52
<b>Figure 3.3. Conversational Framework: activities in the learning process</b> .....	56
<b>Figure 3.4. Potential categories of CAL content</b> .....	61
<b>Figure 3.5. Cosmetic and catalytic imagery</b> .....	64
<b>Figure 5.1. Histogram and Q-Q Plot: video post-test scores</b> .....	103
<b>Figure 5.2. Histogram and Q-Q Plot: Text 1 post-test scores</b> .....	103
<b>Figure 5.3. Histogram and Q-Q Plot: Text 2 post-test scores</b> .....	104
<b>Figure 5.4. Histogram and Q-Q Plot: anticipated appeal of video</b> .....	105
<b>Figure 5.5. Histogram and Q-Q Plot: anticipated appeal of audio</b> .....	105
<b>Figure 5.6. Histogram and Q-Q Plot: anticipated appeal of text</b> .....	106
<b>Figure 5.7. Histogram and Q-Q Plot: anticipated effectiveness of video</b> .....	106
<b>Figure 5.8. Histogram and Q-Q Plot: anticipated effectiveness of audio</b> .....	107
<b>Figure 5.9. Histogram and Q-Q Plot: actual enjoyability of video</b> .....	108
<b>Figure 5.10. Histogram and Q-Q Plot: actual enjoyability of Text 1</b> .....	109

## Tables

<b>Table 2.1. Complementary aspects of three cognitive theories</b> .....	23
<b>Table 2.2. Recall of knowledge about the formation of lightning</b> .....	26
<b>Table 2.3. Second lightning experiment with different length captions</b> .....	26
<b>Table 2.4. Effects of different audio types on learning</b> .....	32
<b>Table 3.1. Recall of biographical text based on interest and importance</b> .....	47
<b>Table 3.2. Effect of seductive details on recall and problem-solving scores</b> .....	51
<b>Table 4.1. Features of positivist and phenomenological paradigms</b> .....	75
<b>Table 4.2. Experiment sample details</b> .....	87
<b>Table 4.3. Propositions, methods of analysis and design variants</b> .....	95
<b>Table 5.1. Experiment sample details</b> .....	101
<b>Table 5.2. Effect of different word counts on test scores (P1a)</b> .....	111
<b>Table 5.3. Effect of word count on tutorial appeal (P1b)</b> .....	112
<b>Table 5.4. Expectations of media appeal (P2a)</b> .....	112
<b>Table 5.5. Expectations of media effectiveness (P2b)</b> .....	113
<b>Table 5.6. Effect of media on tutorial appeal (P2c)</b> .....	114
<b>Table 5.7. Effect of media on tutorial effectiveness (P2d)</b> .....	115
<b>Table 5.8. Effect of catalytic wordage on tutorial effectiveness (P3a)</b> .....	116
<b>Table 5.9. Effect of catalytic wordage on tutorial appeal (P3b)</b> .....	116
<b>Table 6.1. Summary of propositions, analysis and outcomes</b> .....	120

# Chapter 1

## Introduction

In a few years, multimedia computers will be an anachronism. All computers will readily integrate images, sounds, and motion video ... In several years, these computers will be connected to worldwide networks ... to make real-time sharing of high-quality images, sounds and video practical.

Vaughan (1994, p.xxv)

From its first recorded use in 1801 at the West Point Academy, twenty-two years passed before the blackboard was patented by Samuel Reed Hall and its design and use remained largely unchanged for more than a century<sup>1</sup>. Contrast this with personal computing: the PC barely existed 22 years ago and yet its evolution and uptake has been astonishing, as Vaughan's words above demonstrate, and always with educators seeking and finding opportunities to make the most of its ever advancing capabilities. Intel affirmed this rate of advancement when releasing its latest processor chip in 2000, noting that, had automobiles developed at the same rate as transistors, it would be possible to drive from New York to San Francisco in 13 seconds (Intel, 2007).

The educational use of computer-based activities, resources and systems is generally referred to as *e-learning*: a term which was first coined by US commentator Jay Cross in 1998 (Cross, 2004) and is now in widespread use (e.g. a June 2011 *Google* search returned 262,000,000 hits). Definitions vary, particularly regarding whether or not it should apply *only* to online delivery. However, a broader view is helpful to encourage a focus on the *learning* rather than specific technologies. According to the Joint Information Systems Committee (JISC, 2007), e-learning is "Learning facilitated and supported through the use of information and communications technology (ICT)". But it is not sufficient to simply deliver materials or provide support by electronic means, JISC requires that these be "pedagogically sound, learner-focused and accessible". Hence, e-learning must be - not unreasonably - usable, apposite, clear and effective, according to the needs of its intended target audience.

---

<sup>1</sup> From a distillation of various internet sources, but principally Ergo (2007) and eNotes (2007).

This research examines one particular form of e-learning – Computer Assisted Learning (CAL) – with the emphasis on effective design in order to achieve JISC’s requirement for *pedagogic soundness*. This chapter begins with an overview of e-learning and its antecedents, looking briefly at the evolution of terminology, technology, usage and user perceptions, before focussing more specifically on CAL, its sub-genres and how they can satisfy different pedagogic scenarios and requirements. Finally, the scope of this research is described, together with its intended audience.

## ***Evolution of e-Learning***

Early computers of the 1950s and 60s were designed essentially to perform rapid sequences of calculations, often sending some form of alphanumeric output to a teleprinter rather than a screen. It did not take long for some educators to recognise that many procedural or fact-based teaching tasks could be defined in terms of sequences of logical conditions, actions and outputs. Applications such as the Self-Adaptive Keyboard Instructor (SAKI) were developed by cybernetician Gordon Pask and colleagues in the 1950s to teach punch card operators (Pask, 1960). This emerging computer technology also fitted well with the principles of programmed learning which was popular at the time with behaviourists such as Skinner and Rogers (1956). The term *programmed learning* referred not specifically to a computer program but to the general process of delivering information in small chunks, at a pace controlled by the learner, with knowledge-check questions and the provision of immediate feedback (Ravenscroft, 2001). These chunks of information were often referred to as *frames* and, whereas a frame had tended to equate to a page in a workbook, it was a natural extension to convert this into an individual screen of information in any computerised version. However, simply using the computer as an ‘electronic workbook’, presenting information on screen instead of in print, did nothing to take advantage of the potential power of even these early devices and programs, and this, according to Rowntree (1990), was one of the main reasons for its demise.

The branching capability of computer programs offered potential to provide some variation in *what* was taught and *how*. This might be as a result of user preferences – such as via the domain map in Pask and Scott’s 1970s Course Assembly System and Tutorial Environment (CASTE) – or through the computer

attempting to hold some form of discourse with the student, which was the basis for Intelligent Tutoring Systems (ITS) of the 1970s and 80s. However, the subtleties of such discourse and the computer's relatively limited ability to 'build a picture' of the growing body of knowledge in the learner and adapt its approach as the activity progressed, limited what may initially have seemed to be a perfect individually-tailored solution (Ravenscroft, 2001).

Lessons learned from these early computer-based applications were often as much about working within the restrictions of early technology – based on 'dumb' terminals with character-based monochrome displays and no video or spoken commentary – as much as pedagogic issues. These limitations, plus the relatively high costs of equipment and scarcity of developers with specialist skills, meant that much early work tended to be confined to academia and the military (Cross, 2004). The emergence of integrated circuits and the arrival of desktop machines in the 1980s brought computing to home enthusiasts but, even though early Acorns, Amstrads, Ataris and other devices appeared to represent a huge step forward, they were still quite rudimentary. The medium we recognise today took off in earnest around 1991 with the availability of affordable PCs with CD-ROM drives, sound cards and high colour graphics.

Multimedia PCs also enabled programmed learning, which had persisted through to the early 1980s, to make a resurgence in the guise of computerised Integrated Learning Systems (ILS). One of the most difficult aspects of classroom tuition is to provide effective instruction that suits the needs of students with different abilities, prior knowledge and learning preferences (Slavin, 1987); teachers therefore normally aim for some middle ground which can end up suiting few students' exact needs. ILS offered potential to overcome this by providing individualised learning according to the abilities and required pace of learning of each student. However, Bailey (1993) refers to four components that Slavin believed were essential for successful instruction (level, time, incentive and quality) and comments, "It is my perception that ILS systems are weakest in the last of these – instructional quality – largely as a consequence of attending so dutifully to the first" (p19).

Bailey's (1993) remarks might have been levelled not just at ILS, but to other genres of computer-based learning because, with the exception of rudimentary knowledge-centric tuition based on facts, figures, lists or process steps, the computer appeared to be a poor substitute for a skilful and adaptive teacher.



Improved support for richer media such as audio and video did start to give computers slightly more ‘personality’ but, at their heart, remained relatively inflexible, hard-coded programs. However, not only was hardware becoming cheaper, more powerful and readily available, but so too were authoring tools, offering users the chance to develop interactive multimedia with ‘*no programming skills required*’. Potentially this was true but, as sardonic commentator John Barker (1993) observed:

It is all well and good to introduce a technology that enables the man in the street to give vent to his creative ability, but what if he does not have any? (p15).

Authoring tools meant that teachers, trainers, academics, human resources staff, graphic designers, video producers, and others, could all *have a go* at developing e-learning. Unfortunately, the box they bought from their authoring tool supplier contained software and manuals but not talent, so the results were a pedagogic bell curve, ranging from inspiring to despairing, with a peak of mediocrity in between. This did not always hamper adoption or impact but, as Najjar’s (1996) analysis of nearly 40 studies comparing multimedia with classroom lectures concluded, success may often have been due more to the relative novelty of this new medium than effective design.

Those novelty days are long gone: 76% of UK households now own a home computer and 71% of UK households have a fixed-line broadband connection (Ofcom, 2010a) with an average connection speed of 6.2 Mbps (Ofcom, 2011). Furthermore, 32% of users accessing the internet do so via a mobile device or laptop with a plug-in ‘dongle’ (Ofcom, 2010b). People of quite diverse ages and backgrounds are comfortable with basic computing: over half of UK over-55s now have broadband access at home (Ofcom, 2010a) and there is a growing number of ‘digital natives’ (Prensky, 2001) who simply can’t imagine life *without* a computer, the internet, social networking, games consoles and mobile phones. An OECD survey (2010) found that, “...less than 1% of 15-year-old students in [the 30] OECD countries declared that they had never used a computer” (p12) and e-Learning has long since become an expected component of many university courses (Larreamendy-Joerns & Leinhardt, 2006). Smith et al’s (2009) Educause study of US undergraduate students found that 88.3% owned a laptop, 44.0% owned a desktop computer and 51.2% owned an internet-capable handheld device. Forecasters Ambient Insight (2010) report a global

market for e-learning products and services across all sectors estimated at \$27.1bn in 2009, rising to \$49.6bn by 2014.

So how does the potentially rich pool of users perceive e-learning? In their annual survey of current and emerging trends in learning and development, the Chartered Institute of Personnel and Development surveyed 859 practitioners across the private/public/voluntary sectors (CIPD, 2009). They found that 42% had used e-learning more in their organisation in the previous two years and yet, paradoxically, only 7% considered it to be one of the most effective learning and development practices. Clayton et al (2010) surveyed the attitudes of 132 psychology students to online learning and found only 2% preferred it over traditional (73%) or hybrid (25%) teaching methods. The American Society for Training and Development found, in its annual trends survey of 284 learning and development professionals (ASTD, 2009), that organisational concerns about adopting e-learning included:

- Costs (46.9%);
- Technical skill requirements (45.3%);
- Clash with organisational culture (42.2%);
- Management (37.5%) and employee (35.9%) buy-in; and
- “e-Learning is not perceived as ‘real’ training” (34.4%).

Different perspectives of the same situation also vary. In a CDW-G (2009) survey, 74% of HE lecturers in the US believed they incorporated technology into every or nearly every class, but only 38% of students thought their instructors “understand technology and fully integrate it into their classes” (p9).

This brief snapshot of the marketplace paints a mixed picture, but the 93% (by exception) of sceptics in the CIPD (2009) survey above and the 98% in the Clayton et al (2010) research should give cause for concern. Given these reported obstacles and impediments, there *must* be some compelling grounds for persisting in its use. One example reported by Sloman and Rolph (2003) comes from the Royal Bank of Scotland. RBS introduced e-learning to retrain 26,000 employees in 650 locations. The bank found that the medium gave them speed to market and a consistent training model across the business; they also reported a successful return on investment (ROI) of around 7:1.

'Success', however, may mean different things to a bank, a manufacturer and a university. In a survey conducted for the Chartered Management Institute by Scott-Jackson et al (2007), when employers were asked to identify the key advantages of online learning, the clear winner was 'flexibility' (83%), ahead of 'accessibility', 'cost-effectiveness' and 'scalability' (all at around 40%). *Learning effectiveness* might be assumed to be a given; however, exactly what this means will vary according to organisation and scenario, but is likely to include some combination of post-test scores, retention of knowledge and/or transferability into the workplace

Whilst the foregoing data represent professional *opinion*, more empirically based research into e-learning comes from DeRouin et al's (2005) examination of e-learning in organisations. However, their findings are varied, often contradictory, and so inconclusive; for example:

...higher failure rates in a Web section compared with a classroom section of an introductory psychology course.

...no difference in final course grades for ... Web-based and classroom sections of a graduate Research Methods in Social Work course.

...a 2-day Intranet delivered problem-solving course ... indicated that trainee knowledge scores increased significantly from pre- to post-test.

(p928)

They decided that it was difficult to conclude whether e-learning was more, less, or equally effective than traditional face-to-face teaching. This is not particularly helpful, nor is the mix of references to 'web-based', 'e-learning', 'distance learning' and 'telelearning' in their study, meaning it is not always clear whether like is being compared with like. The teaching and learning scenarios were also quite varied: their *poor* psychology example above was based on work by Waschull (2001) that included the requirement for students to engage in the relatively passive activities of reading four to five online lectures a week. However, the *good* problem-solving example from Brown (2001) included high degrees of learner control, choices and opportunities for practice: all factors which are widely accepted (e.g. Anderson, 2008) to induce favourable performance and outcomes.

It is reasonable to attribute at least some of DeRouin et al's (2005) inconclusiveness to the age of their study and the variability inherent in e-

learning and online connectivity at the time. A more recent meta-analysis by Means et al (2009) examined 51 studies representing a range of learner types and learning content. Their findings indicated that e-learning quality, consistency and benefits have improved, such that online students consistently perform better than those receiving equivalent face-to-face instruction.

It is clear from these studies that e-learning, in common with any other single teaching method, is rarely a panacea. It needs to be applied in situations where it has the potential to deliver learning value and then developed and implemented in ways that give it the greatest chance of achieving this. As a first-hand example, I was part of a Royal Air Force study into inefficiencies in teaching Morse code to trainee telegraphists to a required standard of 18 words per minute (Ellis, 1994). A computer-based drill-and-practice Morse trainer was developed which allowed students to practice for as long, and as many times, as they required, always receiving consistently delivered Morse and consistently judged feedback to their own inputs. Once fully integrated into the programme, this saw the average course failure rate fall from around 20% to 4% and allowed the overall course length to be reduced from 21 to 13 weeks. The computerised solution was never intended to replace the whole course, but it delivered impressive tangible benefits in those key areas for which it was targeted and designed.

In a different sphere, Mitchell (2000) reports that a financial services regulatory course with an original duration of 7 classroom days was replaced by a blend of 12 short web tutorials followed by a self-assessment workbook and, finally, classroom tutoring for consolidation. The workbook shrunk from 300 to 120 pages, the classroom period reduced from 7 to 2 days and the pass rate rose from around 60% to almost 90%.

Learning outcomes and pass rates will always be important, but sometimes the benefits manifest themselves in different and perhaps less quantifiable ways. Maag (2004) conducted a study at two American universities to determine the effectiveness of (i) interactive multimedia, (ii) traditional text, and (iii) text and image learning on nursing students' mathematical knowledge, self-efficacy and learning satisfaction. There were no particular differences in post-test and retest knowledge or self-efficacy scores between students in any of the groups, but those in the interactive multimedia group reported higher satisfaction with this method of learning and considered it to be enjoyable and more interesting.

learning and online connectivity at the time. A more recent meta-analysis by Means et al (2009) examined 51 studies representing a range of learner types and learning content. Their findings indicated that e-learning quality, consistency and benefits have improved, such that online students consistently perform better than those receiving equivalent face-to-face instruction.

It is clear from these studies that e-learning, in common with any other single teaching method, is rarely a panacea. It needs to be applied in situations where it has the potential to deliver learning value and then developed and implemented in ways that give it the greatest chance of achieving this. As a first-hand example, I was part of a Royal Air Force study into inefficiencies in teaching Morse code to trainee telegraphists to a required standard of 18 words per minute (Ellis, 1994). A computer-based drill-and-practice Morse trainer was developed which allowed students to practice for as long, and as many times, as they required, always receiving consistently delivered Morse and consistently judged feedback to their own inputs. Once fully integrated into the programme, this saw the average course failure rate fall from around 20% to 4% and allowed the overall course length to be reduced from 21 to 13 weeks. The computerised solution was never intended to replace the whole course, but it delivered impressive tangible benefits in those key areas for which it was targeted and designed.

In a different sphere, Mitchell (2000) reports that a financial services regulatory course with an original duration of 7 classroom days was replaced by a blend of 12 short web tutorials followed by a self-assessment workbook and, finally, classroom tutoring for consolidation. The workbook shrunk from 300 to 120 pages, the classroom period reduced from 7 to 2 days and the pass rate rose from around 60% to almost 90%.

Learning outcomes and pass rates will always be important, but sometimes the benefits manifest themselves in different and perhaps less quantifiable ways. Maag (2004) conducted a study at two American universities to determine the effectiveness of (i) interactive multimedia, (ii) traditional text, and (iii) text and image learning on nursing students' mathematical knowledge, self-efficacy and learning satisfaction. There were no particular differences in post-test and retest knowledge or self-efficacy scores between students in any of the groups, but those in the interactive multimedia group reported higher satisfaction with this method of learning and considered it to be enjoyable and more interesting.

These latter three examples fall into the e-learning sub-category of *Computer Assisted Learning* (CAL); however, even this is quite broad and so some further examination of this genre is worthwhile.

## ***The Nature of CAL***

CAL is a quite well used term but it has been difficult to find a recent and succinct definition, even though more lengthy expositions are available. Weller (2002) observes, "As with many such terms, it is difficult to find a definition that is widely agreed upon" (p135). There is also much overlap between references to, and descriptions of, CAL and related terms such as Computer Based Training/Learning (CBT/CBL), Computer Assisted (or Aided) Instruction (CAI) and Technology Enhanced (or Enabled) Learning (TEL). Whilst it is possible to find authoritative definitions of each of these, they are actually of limited value because different researchers, practitioners or organisations often use the terms interchangeably and inconsistently. The problem is exacerbated by the fact that terminology has evolved quite quickly, so the same entity or medium may have attracted a different label dependent upon when that reference was made.

It is perhaps more useful to define CAL by means of its sub-genres although, once again, there is no absolute agreement between authors over how many of these there are and what they should be called. Rushby (1989) suggests four broad sub-genres of CAL and, although exact titles and definitions vary from author to author, these are typical and have stood up well against recent advances in technology:

- **Drill-and-practice.** This is linear tuition and practice with the emphasis on learning through repetition, the primary aim being for the students to *remember*. That may sound unsophisticated but it can be a prudent way to teach sequences, lists, procedures, facts, etc, before moving on to more meaningful tasks.
- **Tutorial.** Students are taught through combinations of some or all of: explanations, examples, problems, rich media, interactivity, exercises, assessment and feedback. The aim is usually to go beyond remembering,

so they *understand* or *apply* their knowledge, and sometimes *analyse* different scenarios.

- **Simulation.** Alessi and Trollip (2001) define simulation as a representation of “some phenomenon or activity that users learn about through interaction with the simulation” (p213). Interaction in the context of simulation is defined by Blake and Scanlon (2007) as allowing students “...to change some of the parameters in the program and observe what happens as a result” (p491). The aiming point here is for students to *apply*, *analyse* or *evaluate* by ‘having-a-go’ and seeing the outcomes in a controlled and risk-free environment.
- **Information Seeking.** This category aims not to teach directly but to provide resources (or links to) from which students learn or undertake a subsequent task. In using these resources and considering their meaning, relevance, validity, etc, students may be operating anywhere from *remember* to *evaluate* according to their set task.

Of these four, the focus of this research is on the one which seeks to ‘teach’ in a way that is most comparable to a didactic, face-to-face lesson: *Tutorial CAL*. As described earlier, it is widely used, and often to good effect, in the training and FE sectors although it is less prevalent in HE, often being confined to science and technology subjects, more vocational courses (e.g. healthcare) or entry-level material.

The italicised terms in the foregoing descriptions suggest some differences in the cognitive constituents of each genre, which in turn will influence factors such as method of presentation, type of media used, functionality of the program and the nature of any assessment. This can be contextualised by a taxonomy from Anderson and Krathwohl (2001), based on earlier work by Bloom (1956), which classifies cognitive activities into six categories of increasing sophistication:

1. **Remember.** E.g. define, list, name, recognise;
2. **Understand.** E.g. classify, explain, summarise, translate;
3. **Apply.** E.g. calculate, demonstrate, perform, solve;
4. **Analyse.** E.g. compare, examine, explore, test;
5. **Evaluate.** E.g. critique, estimate, justify, review;
6. **Create.** E.g. compose, design, generate, synthesise.

This taxonomy is a useful yardstick for learning designers because it encourages them to consider, for learning activities, content *and* assessment, whether they are pitching these at the required level. In some cases, it is perfectly reasonable to be learning and recalling facts, figures or process steps but often, particularly in educational as opposed to training scenarios, it is desirable to promote a deeper level of learning (Marton and Säljö, 1996). Anderson and Krathwohl's (2001) categories relate to teaching and learning in general, but the costs of designing, developing and testing more complex program functionality, interactivity, analysis of user inputs/actions and the generation of appropriate and accurate responses/feedback can be prohibitive, leaving many stand-alone CAL programs typically addressing levels 1-3.

Whilst tutorial CAL and a face-to-face lesson *are* equivalent in terms of trying to achieve learning outcomes through some form of *beginning-middle-end* teaching process, the student experiences are markedly different. CAL is frequently used in isolation and the most notable absentee is a teacher, whose role is not simply to present information, but to explain it clearly, gauge the student response and respond/adapt as appropriate. Therefore, what will differentiate passable CAL from good CAL is not *what* information is presented, but *how effectively* it is presented.

## ***Scope of Research***

The overall aim of this thesis is to investigate the pedagogic design of Tutorial CAL intended for use by adults studying at a distance as independent learners (i.e. without face-to-face or online interaction with other learners or tutors). More specific goals are:

1. With respect to text-plus-imagery-based CAL, to investigate the effects of different amounts of written wordage on appeal and effectiveness.
2. To determine whether learners' expectations expressed beforehand about the relative effectiveness of different media types are supported by their actual post-test scores.
3. To determine whether learners' expectations expressed beforehand about the relative appeal of different media types are supported by corresponding opinions expressed after completing a CAL tutorial.



4. With respect to text-plus-imagery-based CAL, to investigate the effects of supportive or 'catalytic' content (described later) on appeal and effectiveness.

This might seem like reasonably well-trodden ground, but what will become clear from the review of literature that follows is that there are a great many theories, pedagogic approaches, media choices and other variables to be taken into account when selecting, designing and implementing CAL. Some of these can be applied with a fair degree of predictability but others have differing effects for reasons that are not always clear or consistent.

### ***Target audience***

One section of the target audience for this research is the academic community investigating and researching the design and effectiveness of tutorial CAL, for whom it is intended that the outcomes reported here will add to existing knowledge. The other section, many of whom are often less well supported by formal evidence-based guidance, is the diverse community of practising CAL designers, developers and commissioners in the adult learning sector. These often have backgrounds and experience in learning and development, but far fewer have an equivalent and relevant academic qualification. MacLean and Scott (2007) questioned 307 learning designers worldwide about their professional training/education and found that 21.2% of British practitioners had no specific training or qualifications in learning design (or some close equivalent). The United Kingdom came out consistently worse than other geographic areas (e.g. 3.5% had a Bachelor's degree versus 27.3% in Australia; 32.9% had a Masters degree versus 50.7% in the US; and 4.7% had a Doctorate versus 33.3% in Canada). Unlike teaching, learning design in the UK is something that many people 'get into' later in their career and *Learning (or Instructional) Designer* is not a widely recognised profession, as was confirmed by MacLean and Scott's (2007) analysis of 337 UK Higher Education prospectuses:

It was found that only nine could provide learning design related programmes that would be likely to prepare individuals to work in the profession. (p196)

Despite this, there is no particular evidence to suggest that UK learning design is poor, but it is clearly not so infused with theory or empirical evidence as

might be the case in some other countries. In considering how practitioners pick a design approach, my own experience suggests there is no definitive must-have textbook, web site, or methodology. Instead, practice is based much more on intuition, previous successes and a dash of folklore. However, this is not just an e-learning phenomenon; Hirst (1974) expressed his opinion that, often, "...teaching methods new and old are advocated or defended on little more than a hunch or personal prejudice" (p101).

## ***Personal background***

My own background in the field dates back to 1979, as a hobbyist using a Tandy TRS-80 home computer with just 4kB of memory and my first program – an Ohm's Law calculator – saved to cassette tape. I have been involved more formally in CBT, multimedia and e-learning since 1992, in capacities including learning analyst, designer, developer, project manager and consultant. I have experienced much of the IT revolution first-hand: from the emergence of the PC, Windows, CD-ROMs, multimedia and dial-up internet through the 1980s and 90s, to DVDs, broadband, the community/collaborative world of Web 2.0 and mobile devices of the 2000s. I have experienced life as an educational technology manager in the RAF, a commercial e-learning supplier and, since 2003, from a Higher Education perspective at the Open University. In earlier days there was a pioneering spirit and I have known and been part of projects that were held together by the electronic equivalents of string and sticky tape. However, attitudes have matured, awareness has improved and expectations increased, such that those who commission and use e-learning are more astute and demanding, with a shared desire to *do the right thing* and *do the thing right*.

## ***This Thesis***

This chapter has presented some of the background factors and issues that define how tutorial CAL has evolved and is used and regarded today. Chapter 2 comprises a review of the literature pertaining to e-learning and tutorial CAL with a specific focus on principles, theories and evidence relating to the use of different media. Chapter 3 then considers aspects of the design of those media in greater detail, leading to a new notion of *catalytic content* as that which supports the process of learning. Chapter 4 describes the methodology used to

carry out the research and Chapters 5 and 6 present, analyse and discuss the research findings and methods. Chapter 7 then forms the conclusion to the thesis and offers recommendations for practice and future researchers. Included within this, as the outcomes were not entirely as expected, is a specification for a more rigorous future study into catalytic content.

The diverse nature of e-learning means that some of the writers and researchers cited may differ in their use of terminology (or it may simply have evolved over time). For the sake of clarity, I will treat the term *teacher* as being synonymous with any similar face-to-face role such as lecturer, tutor, trainer, facilitator or instructor; similarly, *teaching* also includes lecturing, tutoring, training, etc; and *learner* refers to any recipient of teaching such as student, trainee or pupil.

A glossary of specialist terms is provided at Appendix A.

# **Chapter 2**

## **Multimedia Design Principles**

There is a lack of extant research to address the critical issue of how to develop effective multimedia that leads to desirable learning performance and satisfaction.

Sun & Cheng (2007, p662)

There is no shortage of anecdotal evidence and folklore amongst practitioners about what makes good e-learning and there has also been considerable research into various aspects of pedagogic design. One name that frequently crops up in the literature is that of Richard Mayer who has, alone and with colleagues, experimented and written extensively about the attributes, presentation and effectiveness of different media used in multimedia learning materials. This initially resulted in a set of seven multimedia design principles which formed the basis for his 2001 book, *Multimedia Learning*, a convenient synopsis of which can be found in Mayer and Moreno (2002, p94). Four years later, the seminal *Cambridge Handbook of Multimedia Learning* (Mayer (Ed), 2005) presented additional evidence from himself and various co-contributors, resulting in some revised and additional design principles. By the second edition of *Multimedia Learning* (Mayer, 2009) there were 12 principles, although five of the originals remain the most enduring and widely investigated; these five lie at the heart of the conceptual framework that set the initial direction for this research:

- **Multimedia** principle. People learn better from words and images than from words alone (Mayer, 2001; Fletcher & Tobias, 2005).
- **Contiguity** (or **Split Attention**) principle. Words and corresponding images should be presented spatially and temporally as closely as possible (Mayer 2001; Ayres & Sweller, 2005).
- **Modality** principle. Visual *and* auditory information is more effective than either visual or auditory alone (Mayer, 2001; Low & Sweller, 2005).
- **Redundancy** principle. Animation and narration is more effective than animation, narration, and on-screen text (Mayer, 2001; Guan, 2009).

- **Coherence** principle. Extraneous material should be excluded rather than included (Mayer, 2001; Muller et al, 2008).

Evidence will be considered here that supports, complements, underpins and sometimes refutes Mayer's principles; the relationship between the principles, cognition, cognitive load and dual coding are also examined. The pedagogic features and limitations of individual media types are considered, leading to the first research proposition, followed by a more detailed examination of richer media leading to the second proposition and the conclusion of the chapter.

Mayer's five key principles above not only make sense from a purely logical perspective but they are also supported by a broad body of research evidence. For example, Ginns (2005) conducted a meta-analysis of 43 separate studies of the *modality* effect, finding:

Across a broad range of instructional materials, age groups and outcomes, students who learned from instructional materials using graphics with spoken text outperformed those who learned from a graphics [sic] with printed text. (p326)

Two experiments conducted by Moreno and Mayer (1999) were based on teaching psychology students about the formation of lightning using different media presented in different ways. The first experiment ( $N=132$ ) used three methods of presentation (animations with integrated (or adjacent) text, animations with separate text and animations with concurrent narration); using recall, transfer and matching tests<sup>2</sup>, they confirmed a *spatial contiguity* effect and *modality* effect. The second ( $N=127$ ) used six methods of presentation (animation with text, animation with narration, narration followed by animation, narration followed by text, narration preceded by animation and narration preceded by text) which clearly showed *modality* and *temporal contiguity* effects. However, Tabbers et al (2004) found evidence that there is no *modality* effect when learners are able to control the pace of e-learning and replay media files. Mayer (1997) reviewed eight studies into a *multimedia* effect and found that students who received coordinated explanations in verbal and visual format generated a median of over 75% more creative solutions on transfer tests than those who received verbal explanations alone.

---

<sup>2</sup> Recall tests measured recollection of factual information; matching tests measured the application of knowledge to label artefacts on a diagram; and transfer tests measured ability to transfer theoretical knowledge into solving written problems.

Support for a *redundancy* effect comes from Kalyuga et al (1999), who conducted experiments with 34 first-year apprentices using a computer-based tutorial to teach the three states of tin/lead solder at different temperatures, and the effects of varying the relative proportions of each metal. They show that a static diagram and narration are more effective than the same diagram and text; however, a diagram, narration *and* text are least effective. Mayer et al (2001) presented 78 psychology students with an animation and concurrent narration explaining the formation of lightning. Those who also received concurrent on-screen text performed worse on tests of retention and transfer than did students who received no on-screen text, thus upholding a *redundancy* effect. Further experiments were then conducted in which interesting but irrelevant details were added, either to the narration or in the form of video clips; these resulted in lower transfer performance, providing support for a *coherence* effect.

Other research is generally supportive, although not all outcomes always affirm all principles. For example, appropriate graphics are thought to enhance learning, but Richard (2006) conducted an experiment with three groups ( $N=137$ ) who saw pictures and heard words representing different common nouns. These media were presented (i) simultaneously, (ii) separately, or (iii) separately whilst attention was distracted by an unrelated task. The divided attention condition caused a degradation in performance as expected; however – and contrary to Mayer’s (2001; 2005) *temporal contiguity* principles – she found no notable difference in performance between the first two groups.

In a study of the effectiveness of different multimedia combinations for procedural-based tasks, Bhowmick et al (2007) found that for low difficulty tasks the multimedia mix was not critical but, for complex tasks, a combination of audio, video and synchronized text yielded the best results. The study comprised just 24 college students but the evidence appears to contradict both Mayer’s (2001) and Sweller’s (2005) *redundancy principle* and, although Mayer (2001) does define an *individual differences* principle, this applies to different types of *learner* whereas Bhowmick et al (2007) show that different types of *task* can also be significant. Finally, when investigating the *modality* principle, Kalyuga et al (2000) found that an animated diagram with simultaneous narration was more effective (based on students’ application of learned procedural steps) than the diagram alone for low-prior-knowledge students, but

this effect actually *reversed* for students having greater experience, with the narration starting to have a negative effect.

There is also evidence of some evolution of, and variation in, precise definitions. An example is that of *redundancy*, described above by Mayer (2001) as relating to *animated* graphics, whereas Kalyuga et al (1999) take a broader view that encompasses other forms of visual imagery. Sweller (2005) also takes a broad view of redundancy: "Information is best presented in a single form unless that form is not intelligible in isolation" (p159). More recently, Mayer (in Mayer & Johnson, 2008) revised his own definition to acknowledge the potential benefits of brief synchronised text, something he had previously discouraged when used with audio and imagery.

In addition to some examples of contrary evidence, there has also been criticism of Mayer's research both in terms of the robustness of data (Lehman et al, 2007) and conclusions drawn (Sadoski, 2001). It is clear that many of his experiments involve short duration learning activities (often just a few minutes) in fairly familiar, and usually technical, topic areas (e.g. the operation of a bicycle pump; the formation of lightning) and it is rare for him to measure longer term retention of knowledge through any form of deferred post-testing. Questions might also be raised as to the extent to which psychology students (his usual participants) are naturally inclined towards learning these sorts of topic and to what extent some almost inevitable prior knowledge may impact on the outcomes. For example, it would be quite unusual for a young adult to know nothing at all about how a bicycle pump worked or how lightning was formed and so teaching it again would trigger some recall of prior knowledge; but that is *not* the same as taking someone with a rudimentary knowledge of algebra and then going on to teach them to solve Riccati equations.

There are also practical issues with some research. For example, in experiments investigating what would eventually become the *modality* principle, Mayer and Anderson (1991) showed through the use of problem-solving transfer tests (Experiment 3;  $N=48$  psychology students) that teaching the operation of a bicycle pump was more effective when using animation and narration than when using either of these media alone. However, in reality, no CAL designer would ever contemplate teaching the operation of a mechanical device that has concealed but important internal parts using just narration

with no visual media; therefore, the outcomes were interesting from an academic perspective, but offered little useful guidance for practising designers.

A less than straightforward picture is beginning to emerge and, whilst it is true there is general support for the multimedia design principles from a number of sources and scenarios, there are also some exceptions and inconsistencies that cannot be ignored. In order to understand the reasons for these and for CAL designers to take appropriate action, it is necessary to probe further and, since a common thread in the principles is how particular forms of information are processed by the brain, the following section considers how this occurs.

## ***Cognition and Learning Design***

According to Sweller (2005), “Learning is defined as an alteration in long-term memory. If nothing has been altered in long-term memory, nothing has been learned” (p20). Some may take exception to this very cognitivist perspective and the aim here is not to dismiss claims by supporters of alternative views such as Activity Theory (Engeström, 1987) or Situated Cognition (Lave and Wenger, 1991) that learning also involves important external, contextual and social factors. However, it is difficult to dispute that our capacity to solve equations, ride a bike or engage in a philosophical debate must all require as a starting point, some internalised knowledge residing in individuals’ memories.

Sweller (2005) describes long-term memory as being based on schema construction, where schemata<sup>3</sup> are “...cognitive constructs that allow multiple elements of information to be categorised as a single entity” (p21). This is similar to Tulving’s (1972) description of semantic memory as:

...a mental thesaurus, organised knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them. (p386)

The schemata and mental thesaurus descriptions both relate to an information-processing model of learning and also to Schema Theory, which describes not just the existence of a hierarchical schemata-based model of memory but also the interaction of incoming information with existing knowledge. Benjafield (2006) describes this as occurring through a four-stage process of selection,

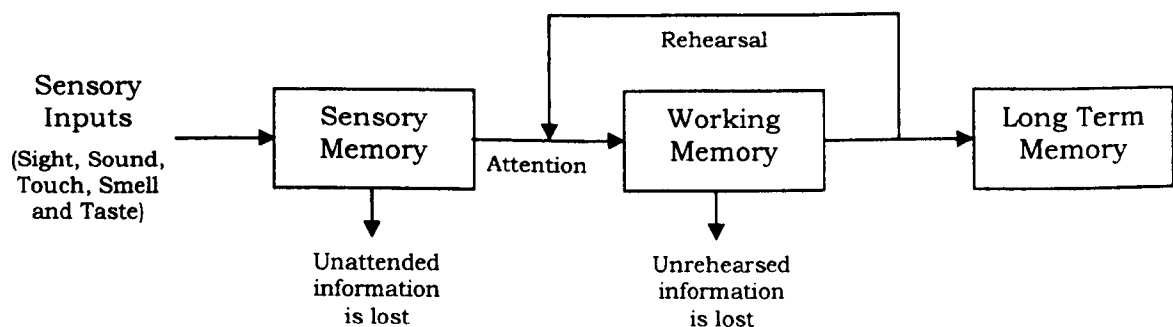
---

<sup>3</sup> Schemata: the plural form of schema.



abstraction, interpretation and the integration of old with new. Schemata construction also underpins advice from Knowles (1973) in his discussion of Andragogy<sup>4</sup> that, when teaching adults, new information should be explicitly linked to existing knowledge whenever possible. Similarly, Rowntree (1990), describing the *tutorial-in-print* prevalent in the early days of the Open University, suggests: “Explain the subject matter in such a way that learners can relate it to what they know already” (p82). Further support comes from other related principles such as Subsumption Theory (Ausubel, 1963) and Elaboration Theory (Reigeluth, 1979), both of which are discussed further in Chapter 3.

The route to long term memory and schemata construction is via working memory (Figure 2.1), the limited capacity of which – cited for many years as  $7 \pm 2$  items (Miller, 1956) – restricts our information processing capability. More recently, Cowan (2001) suggests a lower capacity of  $4 \pm 1$  items and Sweller (2005) that working memory “...can probably process in the sense of combine, contrast or manipulate no more than about 2-4 elements” (p22). Such limitations are not thought to apply to familiar information that is already organised into schemata (van Merriënboer & Ayres, 2005) but they do apply to the *novel, unorganised information* that can often feature in the learning process.



**Figure 2.1. Three stage model of memory**  
(Atkinson and Schiffrin, 1968)

Sweller developed a *cognitive load theory* (1988), based on working memory’s restricted capacity. Sweller et al (1998) define three distinct forms of cognitive load: *intrinsic* (difficulty inherent in each piece of information or task); *germane* (beneficial extra loading which helps schema construction); and *extraneous* (extra loading that serves no useful purpose). Sweller (2005) lists five specific

<sup>4</sup> Andragogy is defined as the science of teaching adults, as distinct from pedagogy which, although widely used in adult education, etymologically should refer just to the teaching of children.

extraneous cognitive load effects he claims designers should take into consideration:

- **Split-attention**, between different but related on-screen artefacts;
- **Modality**, presenting visual and auditory data concurrently;
- **Redundancy**, by using one source of information if possible;
- **Expertise reversal**, presenting information differently for experts and novices;
- **Worked example**, more effective than problem-solving.

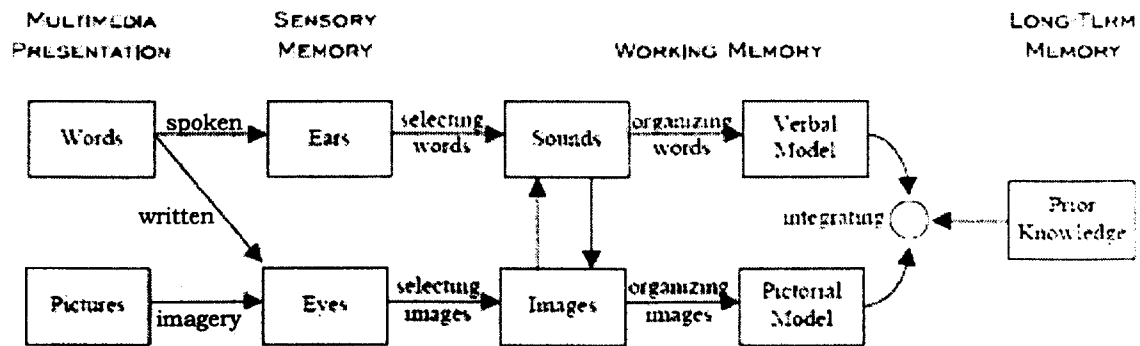
Some duplication/commonality with Mayer's (2001) multimedia design principles will be immediately evident (see also Table 2.1 below) and it is not surprising to discover a similar overlap and occasional collaboration between both men and their various co-authors. Other examples of overlap – in this case in respect of *expertise reversal* – come from Kalyuga et al (2000), Kalyuga (2007) and Seufert et al (2009), all of whom conducted experiments investigating the *modality* principle using narrated versus non-narrated material, finding the effects of audio diminished to the point of reversal in students with greater prior knowledge. However, cognitive load theory is not without its critics and de Jong (2009), whilst in agreement with the underlying principles, notes that the way the theory is constructed makes it difficult to falsify:

...every outcome fits within the theory post-hoc. If learners perform better, a higher cognitive load must have been composed of germane processes; if they perform poorly, a higher cognitive load must have been extraneous. (p21)

## ***Dual coding***

Not only is there a clear overlap between Sweller's (2005) cognitive load theory and Mayer's (2001; 2005) multimedia design principles, but both of these also link closely to Paivio's (1986) *Dual Coding Theory*. The basis of dual coding is that there are two parallel channels within working memory that process verbal and non-verbal data concurrently, as illustrated by Mayer's (2001) model of cognitive multimedia learning in Figure 2.2. Paivio (1986) refers to the dual channels as being functionally independent but with referential interconnections; he also believes each channel operates in different ways:

- Verbal information is represented as *logogens*: word-like codes that are processed sequentially and so are well suited to learning sequences such as the steps of a process or some other chronology.
- Visual information is represented as *imagens* which retain their properties and spatial characteristics and are organised so many parts of a mental image can be processed simultaneously.



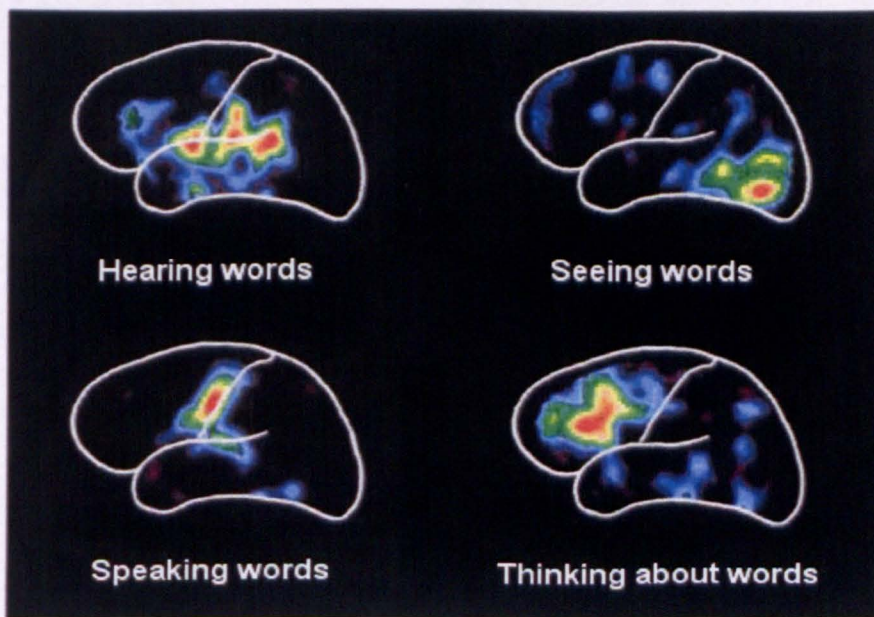
**Figure 2.2. Dual channel cognitive model of multimedia learning**  
(Mayer, 2001, p44)

Mayer holds that dual coding helps us to better understand some of the multimedia design principles. For example, the essence of *modality* is not the number of different media per se, but the channels along which the information is processed. Hence, text plus audio is beneficial (and more so than either of these on their own), as is imagery plus audio; however, text plus imagery is not, because both those media would be competing for limited capacity in the lower channel of Figure 2.2.

Although dual coding is described by Marks (1997) as “...one of the most influential theories of cognition of this century” (p433), it has also been subject to criticism and challenges by psychologists such as Pylyshyn (1981) and Johnson-Laird (1998) who claim that images and verbal information coalesce as a single form of knowledge at deeper levels of processing. Guan (2009) compares dual coding against Baddley’s (2000) model of working memory which comprises a central executive coordinating a phonological loop (processing written or spoken language), a visuo-spatial sketchpad (mental images) and an episodic buffer (integrating information from these sub-systems and long term memory). Where dual coding offers the potential to increase the limited information processing capacity of working memory, Guan suggests that, in Baddley’s model:

...the capacity of the central executive or the episodic buffer is very limited and will not be substantially increased even if the phonological loop and the visuospatial sketchpad are simultaneously used to process information. (p63)

More recent research supports nonunitary models of working memory by showing that maintaining visuospatial information is affected by concurrent spatial tasks but not by concurrent verbal tasks, and vice versa (Robinson & Molina, 2002). Some support for dual coding may also come from brain scans, such as those from Raichle (2007) in Figure 2.3, which suggest that neural activity occurs in different parts of the brain in response to specific stimuli.



**Figure 2.3. PET scans of neural activity in response to different stimuli**  
(Raichle, 2007)

Three main theories or sets of principles have been discussed thus far. These appear to complement each other as Table 2.1 demonstrates and it is not surprising to find that Mayer, Sweller, Paivio and their colleagues have often collaborated and cross-referred to each others' work. However, from an evidential perspective, the very nature of multimedia means that numerous design permutations are possible and it is not always easy to ensure that like is being compared with like, or to say with absolute certainty that X has caused Y to occur. Isolating anomalies and inconsistencies is often made more difficult because much of the research addresses more than one effect/principle, often using different combinations of media and sometimes using competing

terminology. Further investigation is therefore reported below which tries to provide more dependable guidance for practitioners and researchers.

**Table 2.1. Complementary aspects of three cognitive theories**  
 (1Mayer, 2001; 2Sweller, 2005; 3Paivio, 1986)

Multimedia Design <sup>1</sup>	Cognitive Load <sup>2</sup>	Dual Coding <sup>3</sup>
<b>Multimedia Principle</b> Students learn better from words and imagery than from words alone		
<b>Temporal Contiguity Principle</b> Words and corresponding imagery should be presented simultaneously rather than successively		
<b>Spatial Contiguity Principle</b> Students learn better when words and corresponding imagery are adjacent on the page or screen	<b>Split-Attention Effect</b> Multiple sources of essential visual information should be physically integrated	
<b>Modality Principle</b> Students learn better from imagery and narration than from imagery and on-screen text	<b>Modality Effect</b> Presenting information in both visual and auditory modes is superior to presenting purely visually	<b>Dual Coding</b> Working memory is able to process verbal and non-verbal information concurrently
<b>Redundancy Principle</b> Students learn better from imagery and narration than from imagery, narration, and on-screen text	<b>Redundancy Effect</b> Presenting unnecessary information or the same information in different modes should be avoided	
<b>Coherence Principle</b> Extraneous words, imagery, and sounds should be excluded rather than included		

## Mix of Media

This section considers the features and benefits of different media, beginning with an examination of predominantly text-based designs before moving towards combinations of richer media and their effects.

### Text

Despite advances in educational and communication technology, text – in some form or other – remains the central plank of most teaching and learning activities. For e-learning, Weller (2002) contends:

Text forms the glue that binds together any of the other elements of an online course. In many ways, text is better suited to the aims of education than speech. Whereas listening is often a passive activity, reading requires participation and effort. (p131)

So the key issue is not whether to use text, but *how* to use it, and this raises some immediate design issues.

### **Reading on-screen text**

Liu (2005) questioned 113 university students and ‘professionals’ (e.g. teachers and scientists) about their reading habits, finding that 89.4% preferred to read printed documents and only 2.7% preferred electronic. Liu (2006) also found that 84.2% of 133 respondents (from library and information science ( $n=42$ ), business administration ( $n=33$ ), computer science ( $n=35$ ) and social sciences ( $n=23$ )) use electronic sources ‘all the time’ or ‘most of the time’; however, he also found that only 33.8% ‘always’ or ‘frequently’ read online, with 80.5% ‘always’ or ‘frequently’ printing out electronic documents. Pomales-García and Liu (2006a) questioned engineering undergraduate students about their preferences for web-based CAL, finding that “Ninety-four percent of the sample reported that they preferred to read text from paper as opposed to reading text from a computer screen” (p166).

Woody et al (2010) surveyed 91 psychology students on e-book<sup>5</sup> usage, finding no differences between e-book and textbook groups in terms of the achievement of learning outcomes. Previous e-book experience had no strong effect on current preferences, although, “e-book users still preferred print texts for learning” (p.947). In a study into student book-buying patterns, Shepperd et al (2008) discovered that 90% of students who were given the option of purchasing an e-book or a more expensive textbook did not purchase the e-book, despite easy access and an in-class demonstration of the e-book.

These findings back up considerable anecdotal evidence regarding the low popularity of on-screen text. It is also often reported that on-screen reading is slower, although Nielsen (2009) experimented by reading alternate chapters of the same book in print and on a *Kindle* e-book device<sup>6</sup>. He found less than 0.5% difference in his reading speed between the two methods, although he does acknowledge that “one person reading one book is not a proper

---

<sup>5</sup> An e-book is an electronic version of a book, often in PDF, ePub or some similar file format. The e-book can be a page-for-page clone of a printed book or it may contain enhanced features such as hyperlinks, embedded media clips or interactive activities.

<sup>6</sup> e-book readers are handheld devices that display electronic versions of books and journals; imagine something similar to just the screen portion of a laptop computer, but approximately A5 in size.

measurement study” (p1). However, this does raise issues about whether *all* screen-based text is problematic or perhaps just that which relates to *fixed* computer screens (usually on a desk, usually vertical, usually at arm’s length). Mangen (2008) believes all reading is multi-sensory and, in particular, “haptic<sup>7</sup> perception is of vital importance to reading, and should be duly acknowledged” (p405). She accepts that e-book readers and the PC-plus-mouse do offer a haptic experience but that it is not the same as the, “...tactilely richer experience when flipping through the pages of a print book” (p405).

### **Determining the right amount of text**

As there is clear evidence that on-screen text has real or perceived shortcomings, Nielsen (2000) advises “...write no more than 50 percent of the text you would have used to cover the same material in a print publication” (p101). However, he does not substantiate that advice, nor state what is an acceptable norm for a ‘print publication’. Rowntree (1990) also discusses how much to write, noting that, whilst people might read a novel at around 200-300 words per minute, learners engaged in study will be much slower, at around 50-100 words per minute, based on which he suggests aiming for 3000-6000 (printed) words per hour of study (p83). This is useful advice if one’s currency is time but it does nothing to help designers gauge how many words are appropriate to explain topic X to audience Y. In fact, it probably exemplifies one difference between academia (where study is measured in hours, days, weeks, etc, with content added or removed accordingly) and commercial e-learning, where duration is not pre-determined but driven by the learning need.

Apart from user preferences, another reason for reducing text is the nature of online reading. Morkes and Nielsen (1997) found that, of the 51 experienced web users they studied in usability laboratories, only 16% read web pages word-by-word whilst 79% scan most screens. Nielsen (2008) also reports on typical web page content and duration of page visits, concluding that the average user reads around 28% of text on any page. These are useful markers but it is important to note that Nielsen’s research relates to general, and often casual, web use rather than online learning; however, it is clear from the evidence cited above (e.g. Liu, 2005) that many people regard and approach online reading differently to print.

---

<sup>7</sup> Haptic relates to the sense of touch; in this case, holding and using a physical book.

In his PhD research into long online text passages, Parsons (2001) bemoans the absence of substantiated advice on appropriate wordage. One relevant investigation from Mayer et al (1996), examined the most effective way to teach 56 psychology students about lightning formation. In their first experiment, three groups were taught 8 'idea units' using different combinations of a 600 word main text and a summary comprising 5 illustrations with captions (totalling 48-words). Table 2.2 shows that students who received just the captioned illustrations performed best in tests of recall.

**Table 2.2. Recall of knowledge about the formation of lightning**  
(Mayer et al, 1996, p68)

Treatment	Recall
600 word passage	2.8
600 words plus 5 illustrations with 48-word captions	3.7
5 illustrations with 48-word captions	5.6

In a further experiment (N=39), they dispensed with the main (unillustrated) text and considered just the effects of different length captions to accompany the five illustrations from the original summary. The most successful in terms of recall and problem-solving transfer was the short version from Experiment 1, with longer captions yielding negative effects as shown in Table 2.3.

**Table 2.3. Second lightning experiment with different length captions**  
(Mayer et al, 1996, p71)

Treatment	Recall	Transfer
5 illustrations with 600-word captions	2.7	3.0
5 illustrations with 100-word captions	5.2	3.4
5 illustrations with 48-word captions	6.1	4.5

Overall, Mayer et al conclude:

There is a clear pattern in which the more words added to the core verbal explanation, the more poorly the student does in producing the core explanative idea units. These results are consistent with the idea that the additional words overload verbal working memory, drawing limited attentional and comprehension resources away from the core verbal explanation. (p71)

This suggests that there could be some sort of normal distribution around the optimum volume of text for learning effectiveness. However, Mayer et al do not



pursue the matter, nor do they explain how they arrived at their original 600 words, or whether it just 'seemed right'. Another issue, even with fairly succinct writing, is whether or not to scroll text below the base of the screen. Dyson and Kipping (1997) compared the same web content laid out as a single scrolling column against a three-column, non-scrolling format, split across a sequence of screens. They found no difference in reading speed between either format, but users preferred using the non-scrolling screen sequence and achieved better comprehension by this method. More recently, Nielsen (2010) conducted further research into screen layout using eye-tracking data from 21 users and found that they spent 80.3% of their time on a web page viewing 'above the fold' (i.e. immediately visible without the need to scroll). He concluded that, even though users were comfortable with scrolling,

The implications are clear: the material that's the most important for the users' goals ... should be above the fold. Users do look below the fold, but not nearly as much as they look above the fold.  
(p1)

Notwithstanding its ubiquity and efficacy, text needs to be used with care in terms of its volume and presentation and – CAL being regarded as a visual medium – the inclusion of appropriate supporting imagery. Designers must take cognisance of user attitudes and expectations about reading from screen and, whereas learners presume that a textbook or academic journal will be quite expansive, almost the opposite is true of CAL, where verbosity can detract from learning, meaning that succinctness is often a prerequisite. Proposition 1 addresses this aspect of CAL design and, though focused on wordage, it does assume the inclusion of graphics which serve some useful pedagogic purpose, because several screens of purely text do not represent an ideal on-screen learning experience:

### ***Proposition 1***

**Pl<sub>a</sub>.** Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will learn less than those receiving a version with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.

**Pl<sub>b</sub>.** Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will rate this as being less interesting and enjoyable than one with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.

## ***Static imagery***<sup>8</sup>

Bertin (1983) describes two types of imagery: paintings, photographs, drawings, etc, which he refers to as *polysemic*, or capable of subjective, ambiguous or multiple interpretations; and *monosemic* graphics which have unambiguous and unique meaning due to diagrammatic conventions. These conventions – such as the size, placing and style of objects – make it possible to express a great deal of information in ways that are quicker to scan and incur less cognitive load than text. Waller (1981) terms this *visual argument* and it accords with Paivio's (1986) description of *imagens* retaining their properties and spatial characteristics in our memories as discussed earlier.

A different perspective comes from Tversky et al (2002) as they categorise images into those which depict visuospatial entities such as maps, molecular models or technical drawings, and those such as graphs or organisational charts, which represent things that are *not* visuospatial. As an example of the latter, Winn et al (1991) asked graduates to solve relationship problems using either family tree diagrams or written statements. They found that diagrams were more effective for helping people draw inferences about relationships, and those using them solved the problems much more quickly because the information 'search' time was much reduced. One common design technique is to use imagery to convey the 'big picture', followed by a more detailed textual explanation. According to Kulhavy et al's *conjoint retention hypothesis* (1994), the graphics studied first are later activated in working memory as the text is read, leading to better recall and comprehension.

Carlson et al (2003), Griffin and Robinson (2000) and Vekiri (2002) all found learning benefits in using monosemic diagrams depicting physical or conceptual relationships; however, if these were not presented in accordance with Mayer's (2001) principles of spatial contiguity, temporal contiguity and redundancy, they did not improve learning. Similarly, a sequence of illustrations can be beneficial; Zacks & Tversky (2003) found that this enhanced the learning and recall of mechanical assembly procedures. However, Kehoe et al (2009) found that, whilst graphics helped adults follow a sequence

---

<sup>8</sup> 'Static Imagery' is used here as a general descriptor for all non-motion-video and non-animated graphics, photos, illustrations, diagrams, charts, etc.

of instructions relating to computing skills, they had a neutral or negative effect on recall of the whole procedure. They conclude that the graphics may have led to superficial processing, such that:

...the participants may have more effortlessly executed the step-by-step instructions and consequently failed to store them in a strongly connected form. Rather, they were stored as weakly connected sets of stimulus-response associations. Subsequently, when recall of the entire sequence was required, the participants were less able to quickly retrieve the successive steps. (p282)

Imagery improved learning for all student profiles for Lewalter (2003) when, in an experiment with three groups of 20 undergraduates, she compared the effectiveness of text-only versus text & graphics and text & animation for teaching optical phenomena in astrophysics; the text (approx 2100 words) was identical for all three groups. Text-only generated the lowest comprehension and problem solving performance but there was little difference between the static and animated graphic versions. She surmises that, "Arrows and series of frames, which are quite conventional symbols for motion, may be sufficient for the learners to acquire factual knowledge in this case" (p187).

However, one must question whether Lewalter's (2003) zeal for experimental equity and rigour has actually introduced a falseness into her designs that would not be contemplated in the 'real world'. She uses the same 2100 words for all three test groups whereas, in reality, a CAL designer would reduce the amount of text as richer visual media are introduced (indeed it would be one of the reasons for using them); and what text remained would also be edited to accommodate and capitalise on the visual artefacts. Therefore Lewalter's outcomes, whilst academically interesting and undeniably robust, are of only limited use to practising designers.

## **Animation**

Animated graphics may appear to be an engaging and powerful method of explanation – particularly of physical operations (e.g. the internal operation of a car engine or a human organ) or less concrete phenomena such as the air pressures and currents in a weather system – and yet the evidence is far from conclusive. Indeed, one other example of a 'concealed' physical operation is the flushing and refilling of a toilet cistern, but Hegarty et al (2002) found that

learners gained no benefits from the use of animation. According to Kim et al (2007):

Programmers and educators may think animations are appealing and beneficial, but actual learners, even young ones, know better. (p268)

Dooley et al (2000) investigated the use of animation on a biochemistry undergraduate course and concluded that it *was* beneficial for those students with lower grade point averages (GPA), whereas it made no difference for those with a higher GPA. Schnotz and Rasch (2005) investigated the effects of animations – which they described as having two functions: *enabling* (enlarging the set of possible cognitive processes) and *facilitating* (triggering schemata that make cognitive processing easier) – on learners with high and low learning prerequisites, concluding that:

Individuals with high learning prerequisites seem to benefit primarily from the enabling function, whereas individuals with low learning prerequisites seem to be affected primarily by the facilitating function of animations. (p56)

McGregor et al (2004) researched over an extended period to gauge the opinions of 59 students on the use of animation (in conjunction with other methods such as lecture, PowerPoint and real farm equipment) on an agricultural power systems course. Using a four-point scale, the following represent ‘strongly agree’ or ‘agree’ ratings:

- The animations helped me to pay attention to the material (96.6%);
- Animations increased my understanding of the topic (98.3%);
- I used a mental picture of the animations while taking a test (91.5%);
- The animations used in class were distracting (3.4%).

From this, they concluded that “...students overwhelmingly support the use of animation and believe that it enhances their learning experience” (p2). We should not ignore the feel-good factor of animation but it is clear from previous evidence that it does not benefit all situations. The previous example comes from agriculture and, although McGregor does not elucidate, one can well imagine that there will have been sequences, processes or dynamic flows for which animation could have added pedagogic value.

Animation was found by Marbach-Ad et al (2008) to have a beneficial effect on the teaching of molecular genetics (e.g. the structure of DNA). A group of 61 students receiving 20 hours of tuition that included computer animations performed better in a variety of tests than a second group ( $n=71$ ) using illustrations and a control group ( $n=116$ ) receiving traditional lectures.

The research findings for animation are more varied than for any other medium. Ainsworth and van Labeke (2004) suggest this may be attributable to different methods and measures being used to investigate their effect, and that learners' individual differences may affect differing learning performances. Lowe (2003) believes that some learners may tend to process animations only superficially because they do not regard them as cognitively challenging, and Tversky et al (2002) suggest that one possible weakness is:

...animations are fleeting, they disappear, and when they can be reinspected they usually are reinspected in motion, where it may be difficult to perceive all the minute changes simultaneously.  
(p256)

Similar concerns were expressed by Ayres and Paas (2007) but can (and should) be easily overcome through the provision of some sort of play/pause button or slider control. In fact, user controls such as these are exactly the solution described in Betrancourt's (2005) *interactivity principle*. Mayer and Chandler (2001) looked at the effects of including interactivity to control the pacing of a narrated animation describing lightning formation, concluding that:

Clearly, there is more to multimedia learning than simply receiving information that is delivered by a computer. Overall, we found that incorporating a modest amount of interactivity can promote deeper learning from a multimedia explanation if it is done in a theory-based way. (p396)

## **Audio**

In a practitioner survey by Calandra et al (2008), 22 of 23 instructional designers questioned claimed to use audio to varying degrees when designing multimedia lessons. Apart from 'special cases' such as music or language tuition, audio in e-learning usually equates to narration which, according to Aragon (2003), can help to portray the emotions and friendliness of the instructor to the students. Bishop et al (2008) investigated other forms of audio, including music and sound effects, noting that little other research had

been conducted into non-speech. They analysed four commercial ‘award winning instructional units’ covering personal finance, English literature and grammar, finding that narration was most prevalent (2292 instances), followed by effects (619) and music (115). All four units were aimed at schools and no compelling case was made for wider use in adult education or training, but it is interesting to note the main uses to which narration was put:

- Direct learners’ attention (25.3% of instances);
- Tie into previous knowledge (19.9%);
- Hold attention (13.3%);
- Focus attention (13.1%); and
- Organise information (9.0%).

Evidence *against* the use of sounds and music comes from Moreno and Mayer (2000) in experiments to teach the operation of a car braking system. In their second experiment, 75 psychology students were exposed to sound used in one of four different ways and the results in Table 2.4 show that including non-narration audio had a deleterious effect on learning, particularly in a problem-solving transfer test; this accords with Mayer’s (2001) *coherence principle*.

**Table 2.4. Effects of different audio types on learning**  
(Moreno & Mayer, 2000, p123)

	Retention Test*	Transfer Test*
Narration	3.95	5.55
Narration plus sounds	3.12	3.06
Narration plus music	3.11	3.33
Narration plus sounds and music	2.45	3.40

\*Mean scores

This suggests that narration is more effective when presented without other forms of audio, a position that is widely supported by anecdotal evidence from the practitioner community. There is some evidence to the contrary, such as from Mammarella et al (2007), who reported that a ‘Vivaldi effect’ improved cognitive performance when classical music was played, compared with two other groups who heard white noise or no music. However, the learning comprised simple recall tasks and the sample of 24 ‘older adults’ was aged between 73 and 86, thus limiting the generalisability of these outcomes.

There is also evidence that narration is even more effective as a learning aid when used in conjunction with other media types rather than the sole source (e.g. an audio book). Guan (2009) showed that audio alone is around half as effective as on-screen text in experiments with 165 university students using a three-sentence definition of cognitive psychology (mean learning efficiency for text = 4.22; audio = 2.27) and a longer text on the structure of human memory (text = 5.97; audio = 3.10). It is possible that the nature of the material affected the results, the former being fairly abstract and the latter benefitting from visualisation, and so both would benefit from re-reading. In contrast, Mayer et al (2003) experimented with the use of an on-screen 'expert' who responded to preset questions (asked by learners by means of a mouse-click) on the subject of an illustrated electric motor. A group ( $n=26$ ) receiving audio feedback performed better in a problem solving transfer test ( $M=8.43$ ) than a group receiving text responses ( $M=6.54$ ;  $n=26$ ).

Winn (1993) noted that, "Human speech is the most powerful and expressive medium the designer has available for use in instructional messages" (p117) and effective use of narration is described in the *redundancy* and *modality* principles (Mayer, 2001, 2005). Barron and Calandra (2003) analysed nine examples of commercial, audio-enhanced, e-learning courseware. These included minimal music and/or sound effects and, where used, they were short and relevant to the course content, in accordance with the *coherence* principle; they also found that visual and audio content was generally well-synchronised as per the *temporal contiguity* principle. However, most of the programs incorporated full-text combined with full verbatim audio; this not only contradicts the *redundancy* principle but, because most adults read much more quickly than voiceover artists speak, there is also the inevitable frustration of having to wait for several seconds at the end of each screen for the narration to 'catch up' with the text they have already read (Alessi & Trollip, 2001).

## **Video**

Video can mean many things, with genres including documentary, drama, demonstration/illustration (e.g. a procedure), discussions and 'talking heads'. It is also possible within these genres to apply different creative and design styles; for example, Arguel and Jamet (2009) experimented with the insertion of static images amongst motion video of first aid treatments. They found that the

hybrid version was more effective than either the video or the images alone and, contrary to their hypothesis, that displaying fewer images resulted in 13% higher mean test scores than if twice as many (relevant) images were each shown for a shorter duration.

Al-Seghayer (2001) conducted experiments on vocabulary acquisition with 30 English-as-a-Second-Language students who used a multimedia program containing 21 unfamiliar words, presented under three conditions: printed text definition alone, printed text coupled with still pictures, and printed text coupled with video clips. The recall test results were 53%, 67% and 87% respectively, leading him to report that this supports Mayer's (2001) theory of multimedia, and to conclude:

The variety of modality cues can reinforce each other and are linked together in meaningful ways to provide an in-depth experience. Among the possible factors that may explain these results were the following: video better helps learners build a mental image, curiosity increases concentration, and video's combination of modalities (dynamic image and sound) facilitate recall. (p225)

In contrast to this single experiment, McLaughlin et al (2007) conducted a meta-analysis of video-rich e-learning and found mixed results. They then conducted three experiments of their own to examine possible relationships between the effectiveness of video, the specific learning task and the characteristics of the learner. They found audio more effective than video to teach the loading of a medication organiser to both younger (18-30) and older (65-75) adults. For teaching the calibration of a glucometer, performance with video was 11% more accurate than audio for younger adults but marginally worse for older; and assembling *Lego* objects – highly visuospatial tasks – video led to greater accuracy (in many cases more than 100% better than audio) for both age groups. Based on these outcomes, they conclude 'which medium is better?' is the wrong question to ask. "A more appropriate question to ask is under what condition is one medium more appropriate than another?" (p400).

Donkor (2010) compared two experimental groups of vocational students ( $n=34$  and 35) who received print-based or video-based instruction, finding that the video group were rated more highly on a practical exercise (a mean of 35.2 versus 31.8). He concluded that "...video-based instructional materials are superior to the print-based instructional materials" (p107). Whilst this is



factually correct, the topic being taught was mixing and applying mortar to a wall so, from a design perspective, the notion of trying to teach this heavily psychomotor task via print alone is questionable and so must be the value of this research to practitioners.

In an investigation of the effectiveness of an online masters module in statistics, DeVaney (2009) compared a version that contained 17 video tutorials versus one without. He found “no statistically significant differences in academic performance” (p606) although the tutorial students “did have a positive attitude toward the tutorials and would recommend them for other classes” (p606). Laurillard (2002) echoes this attitudinal benefit of video, describing it as a *sit-forward* medium because of its inherent ability to encourage students to engage.

Zhang et al (2006) believe that *interactivity* is the key to video’s effectiveness, in the sense of students being able to control playback. They conducted experiments using 138 undergraduate students, divided in four groups and taught equivalent content by: (1) traditional lecture; (2) text/graphics PowerPoint; (3) non-interactive video; and (4) interactive video which allowed “proactive and random access to video content based on queries or search targets” (p15). They found little difference between methods (2) and (3), although both were more effective than (1), but (4) showed marked improvements in performance and satisfaction over the other methods. They also noted that:

...a number of students in the non-interactive group commented on the difficulty of efficiently skipping or browsing for a specific portion of the video. (p24)

This is an interesting use of the term *interactive* because it relates more to its utility than pedagogic value. In fact Laurillard (2002) takes issue with claims that video (as opposed to broadcast TV) is interactive: “Nothing in the video changes when a student rewinds it, just as nothing in a book changes when you turn the page” (p103). She concedes that video may be active but *not* interactive. Curiously, a study by Hasler et al (2007) found that a group of learners given an interactive button to stop/start the animations learnt better than learners without such a control, even though the former group were observed to rarely use the facility.

## **Media richness**

A common assumption in e-learning is that the 'richness' of a medium refers to its inherent dynamic properties and capacity to stimulate so, for example, video would tend to be regarded as *richer* than audio; and audio would be *richer* than text. It is also widely assumed that richer media are somehow better media for teaching and learning, a view echoed by Liu et al (2009), who claim: "The most media-rich presentation interface (text–audio–video presentation) always generates higher levels of [perceived usefulness] and concentration..." (p605).

Other research supports the rich/lean media definition, but not necessarily the effects. Matarazzo and Sellen (2000) and Yeung and Lu (2004) found that lean media (principally text) are as effective as rich media for presenting 'analysable' information such as facts and procedures, whereas richer media such as video are more effective in presenting less-analysable tasks that may contain ambiguity or require subjective interpretation. Sun and Cheng (2007) conducted four experiments using 240 senior high school students to examine the effects of multimedia on learning score and learner satisfaction in two contrasting topic areas: a Chinese poem and number system transformation from a computer concepts course. They found that the media mix in itself did not improve learning performance or satisfaction, but that it depended on the learning situation:

...whether it is learning score as an objective measure or learning satisfaction as a subjective measure, the course unit with high uncertainty and equivocality in content needs high richness media representation. On the other hand, it is ineffective to use high richness media to promote learning performance for the course unit with low uncertainty and equivocality that can be stated clearly in regular text. (p673)

Some of the terminology used here relates to *media richness theory* (Daft & Lengel, 1986) where media richness relates *not* to one medium versus another, but "...the ability of information to change understanding within a time interval" (p560). Media richness theory is rooted in communication studies and information processing rather than education, but it is nonetheless relevant. For Sun and Cheng (2007), *uncertainty* refers to the gap between the amount of information required to perform a task and the amount of information already known; and *equivocality* is ambiguity caused by conflicting interpretations of information or a situation. To paraphrase their findings, they believe rich

media *do* help when the learning task is harder and the explanation ambiguous. However, Otondo et al (2008) suggest a divided effect, with richer media likely to result in greater learner satisfaction but leaner media yielding greater achievement. Robert and Dennis (2005) investigate research into rich media and *Social Presence Theory*, citing Short et al's (1976) assertion that:

...the use of rich media high in social presence induces increased motivation but decreased ability to process, while the use of lean media low in social presence induces decreased motivation but increased ability to process. (p19)

Timmerman and Kruepke (2006) suggest that rich media: "incorporate channels that allow message recipients to utilize a greater number of sensory modalities (i.e. visual, auditory, tactile)" (p78). They investigated the link between media richness and CAL performance, testing three hypotheses through a meta-analysis of 118 previous studies. The hypotheses and results were as follows, with  $r$  denoting effect size and  $k$  the number of relevant studies:

1. Student performance would be greatest when using apparatus, followed by video, audio, text with graphics then text alone. In fact, audio gave the highest performance gains ( $r=0.26$ ;  $k=8$ ), followed by plain text ( $r=0.14$ ;  $k=32$ ), text with graphics ( $r=0.12$ ;  $k=39$ ), video ( $r=0.07$ ;  $k=24$ ) and apparatus ( $r= -0.05$ ;  $k=4$ ).
2. CAL which includes feedback exercises would have larger learning outcome effect sizes than CAL without. In fact, the difference was marginal between feedback examples ( $r=0.13$ ;  $k=60$ ) and non-feedback ( $r=0.14$ ;  $k=25$ ).
3. CAL designed for a specific learning cohort would achieve learning outcomes better than that designed for general publication. This hypothesis was supported by ( $r=0.14$ ;  $k=86$ ) versus ( $r=0.09$ ;  $k=27$ ).

Having failed to support two of their three hypotheses, Timmerman and Kruepke (2006) discuss a number of reasons why almost *any* conclusions might have been questionable. For instance, no account is taken of the compounding effects of variables, so the effect size for text (reported above as 0.14) was actually 0.21 when feedback was provided but only 0.09 when it was not; undergraduates yielded greater performance gains than graduates, and social science topics yielded the greatest performance gains. Timmerman and Kruepke (2006) also note that *repeated* use of CAL:

...yielded higher average effects than when delivered at one point in time. This finding contradicts theoretical arguments that suggest [CAL] performance gains may solely be a function of novelty effects. (p91)

However, there is more to novelty than *amount* of use: originality and scarcity are also significant factors and their statement that "...benefits were greater in studies published from 1985-1994 than 1995-2004" (p90) could well support a novelty effect because, in the decade to 1994, computers and CAL were still relatively scarce.

More recently, Liu et al (2009) used multimedia to teach Microsoft Project to 102 software engineering students, comparing the effects of three different designs (text-audio; audio-video; and text-audio-video). Based on the Technology Acceptance Model (Davis et al, 1989), they found the latter design to be most successful. They concluded: "This study further confirms that course materials that use rich media can promote higher user acceptance through stimulating a higher [perceived usefulness] and concentration" (p606), although it should be noted that this conflicts with Mayer's (2001) *redundancy principle*.

There is considerable anecdotal endorsement for the merits of richer media in teaching and learning, with the multimedia design principles and the media richness discussion above, both supporting this. However, it is seldom tested explicitly; for example, Boster et al (2006) investigated the use of video for education and concluded:

Although much contemporary thinking leads to the expectation that communication technology, such as video streaming, enhances educational performance on the average, a dearth of strong evidence consistent or inconsistent with this claim precludes a thoughtful evaluation of it. (p46)

Danaher et al (2005) investigated the use of rich media in general on behaviour change websites and noted that, "empirical research needs to examine the accuracy of widely held assumptions that video and audio automatically add value to Web-based programs" (p1). Therefore, a second proposition will test the effects of richer media on both students' perceptions and actual performance.

### **Proposition 2**

**P2a.** When questioned prior to study, learners will presume an audio-rich CAL tutorial to be less appealing than a video-rich CAL tutorial but more appealing than a text-with-imagery-based CAL tutorial.

**P2b.** When questioned prior to study, learners will expect to learn more from a video-rich CAL tutorial than from an audio-rich CAL tutorial and to learn least from a text-with-imagery-based CAL tutorial.

**P2c.** When questioned after study, learners will report an audio-rich CAL tutorial to be less interesting and enjoyable than a video-rich CAL tutorial but more interesting and enjoyable than a text-with-imagery-based CAL tutorial.

**P2d.** Learners studying an audio-rich CAL tutorial will learn less than those studying a video-rich CAL tutorial but more than those studying a text-with-imagery-based CAL tutorial.

## **Summary**

The foregoing presents a representative sample of the considerable research into the attributes, use and effectiveness of different media. From this emerges a broad set of ground rules for CAL designers, suggesting that:

- On-screen text should generally be succinct because most students prefer to read longer amounts as print (Liu, 2005, 2006; Mayer et al, 1996; Morkes & Nielsen, 1997; Nielsen, 2000, 2008, 2009; Pomales-García & Liu, 2006a; Shepperd et al, 2008; Woody et al, 2010);
- Diagrammatic graphics often enhance learning (Carlson et al, 2003; Griffin & Robinson, 2000; Lewalter, 2003; Vekiri, 2002; Zacks & Tversky, 2003) although the case for animation is less clear (Ainsworth & van Labeke, 2004; Ayres & Paas, 2007; Dooley et al, 2000; Hegarty et al, 2002; Lowe, 2003; Tversky et al, 2002);
- Commentary usually increases effectiveness when used appropriately with other media (Aragon, 2003; Guan, 2009; Mayer, 2001, 2005; Moreno & Mayer, 2002), but sound effects or music offer no significant benefits (Moreno & Mayer, 2000); and

- Video can also be beneficial (Al-Seghayer, 2001; Pang, 2009), but students must be able to control playback (Laurillard, 2002; Mayer, 2005; Zhang et al, 2006).

However, the preceding discussion also shows that, for almost every theory or guideline, there are inconsistencies and exceptions (Bhowmick et al, 2007; Kehoe et al, 2009; McLaughlin et al, 2007; Richard, 2006; and Seufert et al, 2009). These might be attributable to the subject matter itself, the pedagogic approach, presentation style and tone, technology, environment, learners or combinations thereof and, with all these different theoretical plates spinning and wobbling on their sticks, a confusing picture emerges for researchers and practitioners.

Therefore, whereas this chapter investigated *which* medium or mix of media best enhance the learning experience and the achievement of learning outcomes, the next chapter considers in more detail the nature and effects of *how* learning content is designed and presented to greatest effect.

## Chapter 3

### ***The Nature of Tutorial Content***

Representation involves thinking through the key ideas in the text or lesson and identifying the alternative ways of representing them to students. What analogies, metaphors, examples, demonstrations, simulations and the like can help build a bridge between the teacher's comprehension and that desired for students?

Shulman (1999, p73)

The principles and theories discussed in the previous chapter give reasonable guidance for practitioners to make informed decisions about which medium/media to use when designing computer assisted learning. However, this is only part of the equation; of equal importance is the *content* itself, and designers should appreciate that different components within a section of content can and should serve different purposes in the overall learning process.

An obvious starting point, and one which applies equally to electronic and face-to-face delivery, is the basic notion of making learning activities and material interesting. This immediately presents a design dilemma because 'interesting' is a subjective term and, just as with art, music, literature, cinema, comedy, etc, one person's *brilliant* may well be another person's *awful*. Some investigation into the nature of interest and its potential effects on learners and learning is therefore worthwhile.

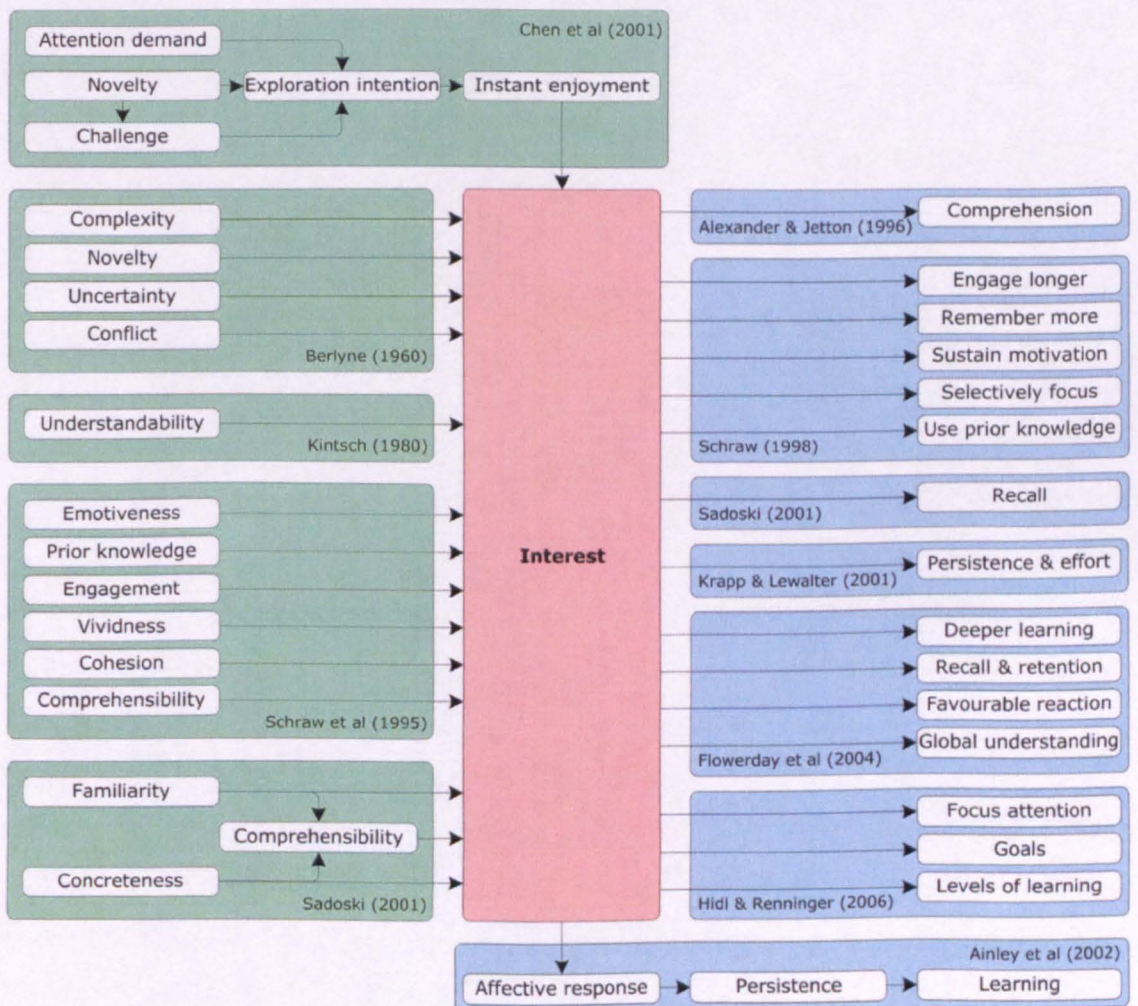
### ***The Nature and Effects of Interest***

Hidi and Baird (1986) caution that "thinking of interest as a general arousal experience is inadequate" (p191). Some things might be regarded as being of 'general interest', we might expect others to be of interest to many people in a particular situation and we might predict others to be regarded as fairly uninteresting by most people. For instance, Silvia (2005) notes that:

...there are different models of what makes art interesting (Berlyne, 1971), what makes text interesting (Schraw & Lehman, 2001), what makes vocations interesting (Savickas & Spokane,

1999) and what makes learning interesting (Hidi, 1990; Krapp, 1999). (p89)

An example of the diversity of perspectives is shown in Figure 3.1, which contains a compilation of some relevant research from a variety of sources. Whilst the diagram shows that some synonymous terminology is used by the different authors, there are also distinct differences in some of the causes investigated and effects found, and whether researchers considered these as discrete or as part of a compounding chain.



**Figure 3.1. Compilation of some causes and effects of interest**

Broader agreement on the nature of interest comes in the form of the following categories and, whilst the exact terms listed vary, they are reasonably synonymous and tend to relate to (i) the *individual learner* and (ii) the *learning scenario*:



- Reader-based and Text-based interest (Hidi & Baird, 1986);
- Personal and Situational interest (Schraw et al, 1995);
- Personal, Situational and Topic interest (Ainley et al, 2002);
- Individual and Situational interest (Hidi & Renninger, 2006).

There is general agreement amongst these researchers that individual interest is personal and enduring, and this relates closely to the notion of *intrinsic motivation*, or that which comes from within due to some inherent appeal, curiosity, drive or beliefs (as opposed to *extrinsic* motivation which represents almost a faux interest driven by some short term goal such as the need to pass a test or gain a reward). Intrinsic motivation can also involve goals but these are set by the individual, often based on *perceived self-efficacy* – an individual’s own assessment of how challenging a task will be – or, according to Bandura (1995), “the belief in one’s capabilities to organize and execute the courses of action required to manage prospective situations” (p2). This belief can of course be heightened by some recent successful performance of a similar task or against a target, referred to by Bandura as ‘mastery experiences’, and teachers and learning designers can support this by devising activities which stretch learners sufficiently without appearing to be either too simplistic or impossible. Pitching things at an appropriate level is also the basis for four design factors described by Malone and Lepper (1987) – challenge, curiosity, control, and fantasy – and the four components of Keller’s (1987) ARCS model of motivational design: attention, relevance, confidence and satisfaction.

All three of these theories can be applied equally to traditional teaching and e-learning and have more recently been adopted to good effect in gaming design, but they can only try to capitalise on individual interest and intrinsic motivation which already exist. In contrast, Ainley et al (2002) considered *situational* interest as an attribute that could be *created* through well designed activities and materials:

Situational interest is elicited by certain aspects of the environment. These include content features [...] and structural features such as the ways in which tasks are organized and presented (p545).

Therefore, in terms of CAL design, it is necessary to consider situational interest in greater detail because multimedia and interactivity offer a greater

range of potential interest attributes than the textbooks that have been the focus of so much research over the past 20 years or so.

### ***Situational interest***

Schraw et al (1995) found six potential sources of situational interest, as depicted in Figure 3.1:

1. Ease of comprehension;
2. Text cohesion (organisation/clarity);
3. Vividness;
4. Engagement;
5. Emotiveness; and
6. Prior knowledge.

They conducted experiments in which 154 undergraduates studied an 800-word *Time* magazine article on Kuwait and then completed a Sources of Interest questionnaire, Perceived Interest Questionnaire and a free recall test. These showed all six factors to be 'significantly related' to situational interest and the first four to successful recall of the text, thus suggesting some correlation between interest and learning. Salomon (1984) hypothesised that some students' interest in electronic media was because they thought it made learning easier and so would require less effort; any resulting lower test scores might therefore be due to expending less effort than to the media per se. This was borne out by Bernard et al (2004) in their meta-analysis of 232 studies into classroom instruction versus distance learning, which concluded:

Interest/satisfaction may not indicate success but the opposite, since students may expend less effort learning, especially when they choose between [distance] and regular courses for convenience purposes. (p415)

Support for an interest-learning correlation also comes from Schiefele (1996), based on experiments with 107 seventeen-year-old students reading text passages on prehistoric peoples. Students indicating low degrees of interest tended to read the text at a verbatim level but those indicating greater interest could provide deeper situational and propositional answers and analyses.

Chen et al (2001) found that situational interest is task-specific. Their interlinked factors shown in Figure 3.1 relate to a non-reading task (viewing video) whereas a second experiment involving a physical task (passing/shooting basketball) showed the same factors but with slightly more complex relationships and different path coefficients. A common factor in both experiments was that “instant enjoyment was a direct source of high situational interest” (p396), contrary to speculation by Reeve (1989) that situational interest *led to* enjoyment.

Pomales-García and Liu (2006a) conducted experiments with 18 engineering undergraduate students who took 12 different web-based CAL modules that included petroleum contamination (duration: 4 minutes, 36 seconds), aquifers (14:55) and air pollution (22:35). They found that:

The study showed no difference in information recall between the different module lengths and formats; however, as module length increased, participants were more likely to not complete the modules. (p163)

Overall, the results show that the seven- and fourteen-minute modules and those containing video are perceived as less difficult than twenty-minute modules or modules containing only audio or text. (p173)

This puts a different perspective on situational interest by suggesting it may relate to more rudimentary learner attitudes: longer programs in this case did not hold interest so well, and less interesting programs (due to length or choice of media) were perceived as harder. However, we also saw above that programs perceived as very interesting may be considered easier, causing learners to invest less effort and learn less deeply (Salomon, 1984). This suggests some tipping point: interest is good for learning but too much can begin to introduce detrimental effects. Sweller (2005) describes a similar characteristic (albeit related to use of media rather than interest per se) in his *expertise reversal effect*, where the use of imagery helped novices but caused those with greater knowledge/experience to perform less well.

## **Concreteness**

One means of satisfying Schraw et al’s (1995) criteria for achieving situational interest (above) is to adopt an appropriate writing style, an example of which is the use of narrative having *concrete* attributes. Concreteness is described by

Sadoski et al (1993) as text that evokes mental images which, in turn, enhance interest. In an early text-based study, Wharton (1980) presented students with one of two versions of a narrative: an original and one that had been rewritten to be more concrete and evocative of imagery; for example:

England's trade was the source of her economic strength, and to undermine it would be to destroy England's power in war.

Became...

England's trade was the *sinew* of her economic strength, and to *shrivel* it would be to *cripple* England's power in war [my italics].

Overall, Wharton changed about one word in eight and participants, given a multiple choice comprehension test, scored more highly and rated the revised passages as more interesting. More recently, Sadoski (2001) reports on experiments using biographies of historic figures to examine the effects of four factors – familiarity, concreteness, comprehensibility and interestingness – on immediate recall, finding, “concreteness was by far the most powerful predictor of comprehensibility and immediate recall [...] these results are highly consistent with Dual Coding Theory” (p268).

Concreteness is thus supported by research evidence but, in practical design terms, there may only be limited opportunities to convert more passive words and phrases into concrete ones without the resultant prose appearing quite contrived or even irritating. An alternative method of engaging learners may be to simply insert some additional interesting material but, as well as adding to length, there is the risk that the interesting material receives more attention from learners than the important material.

## ***Seductive Details***

Seductive details are described by Garner et al (1992) as interesting but unimportant items that are added to learning materials to make them more attractive, but which actually seduce the learner away from the intended learning or outcomes. From an intuitive perspective this seems to make sense, and there can be few practising designers who have not at some point been driven by inherently dull subject matter to seek some injection of interest simply to try and hold the learners' attention.

Harp and Mayer (1997) experimented on the relationship between *emotional* and *cognitive* interest (Kintsch, 1980) and seductive details. According to emotional interest theory, adding seductive details should cause learners to find material more arousing, pay more attention and learn more, but Harp and Mayer's (1997) experiments did not support this. In contrast, according to cognitive interest theory, seductive details should disrupt sense-making and attention, causing learners to comprehend less and rate the material as less interesting than a more succinct version; and this *was* upheld. However, Silvia (2006) points to flaws in Harp and Mayer's (1997) method and conclusions, also noting an absence of corroboration from other researchers.

The majority of research into seductive details has focused on printed text and a frequently cited study by Wade and Adams (1990) reports on experiments with 48 college students who were presented with a 1700-word biographical passage about Horatio Nelson. Mean recall was tested immediately and after one week, and the results – summarised by Renninger et al (1992) in Table 3.1 – show that seductive details were by far the most memorable, with low-interest/high-importance the least.

**Table 3.1. Recall of biographical text based on interest and importance**  
(Wade and Adams, 1990, summarised by Renninger et al, 1992, p266)

Sentence Type	Immediate	Delayed
High Interest, low importance ( <i>seductive details</i> )	65.0%	48.8%
High Interest, high importance	46.0%	32.4%
Low Interest, low importance	34.9%	21.6%
Low Interest, high importance	29.0%	10.5%

These results appear to fit with three hypotheses offered by Harp and Mayer (1998) which state that the inclusion of seductive details can lead to *distraction*, *disruption* and *diversion* in terms of learning important content. Harp and Mayer (1998) conducted experiments, the outcomes of which support the diversion hypothesis, suggesting that learners activate inappropriate schemata based on seductive details rather than the main ideas. However, Goetz and Sadoski (1995) caution that any negative effects from extra seductive content:

...may have been due less to distraction of the reader than to disruption of the text. Perhaps the introduction of an equal amount of uninteresting but peripheral material would have had

the same effect on the recall of the main ideas of the original text.  
(p504)

They also query, in the case of a biographical text, how clear the dividing line is between seductive and non-seductive content. For example, '*during the battle, Nelson's right arm was badly mangled up to the elbow*', was classed as seductive by Wade and Adams (1990, p339) and yet this is factually accurate, it relates to a significant and widely known disfigurement and indicates much about Nelson's fortitude to continue in post. Therefore, whilst it may not relate directly to the learning outcomes, this does not necessarily make it entirely *unimportant*.

There has also been research to show that individual learner attributes can affect the impact of seductive details; for example, working memory capacity (Sanchez & Wiley, 2006) or reading ability (Harp & Mayer, 1997), with the latter reporting that:

...skilled readers possess the metacognitive skills needed to distinguish between portions of a scientific passage that promote cognitive interest (i.e. explanative summaries) and those that promote emotional interest (i.e. seductive text and seductive illustrations). (p100)

Whilst factually accurate, this statement appears to ignore the fact that reading, distinguishing and discarding seductive details must impose some level of extraneous cognitive activity that could impair learning effectiveness. Schraw (1998) investigated seductive details which he categorised as context-independent (generally interesting to most people) or context-dependent (interesting in particular learning situations). He found the latter "took significantly longer to read than main ideas" (p4) but that neither of the seductive detail types had any impact on the recall of main ideas.

Towler et al (2008) examined the effects of seductive details on the learning of IT procedural skills involving Excel data manipulation ( $n=47$  undergraduates) and Word mail merge ( $n=77$ ). They found no effect on knowledge acquisition – contrary to previous research by Mayer et al (2001) – and speculate that this may be due to the difference between learning a stepwise procedure and learning something more systemic; or to differences in the cognitive processes required for recall and recognition tests. In transfer tests they found that seductive details had a *beneficial* effect, which they argued was significant

because, in a training scenario, it is important that students can transfer their learning into a range of work situations and problems.

Thalheimer (2004), in a review of 24 studies dealing with issues relating to the role of seductive details in helping or hindering learning, found 16 demonstrated negative effects on learning. However, he acknowledges the inconclusiveness of this review and questions which aspects of seductive details are beneficial, harmful or neutral to learning, concluding:

It's not any augmentation that will hurt learning. It is only those augmentations that divert the learners' cognitive processing away from the information that should be learned. (p22).

This accords with an earlier study by Goetz & Sadoski (1995) in which they conclude:

In order to provide evidence of a seductive detail effect, it is essential to show that readers are being seduced away from important but uninteresting information *that they otherwise would have learned and remembered*. (p507; italics in original)

### ***Issues with research into seductive details***

The majority of research into a seductive details effect has focused on printed text, and Towler et al (2008) note that:

With one exception (Mayer et al, 2001), no prior studies have examined the effect of seductive details in [technology delivered instruction] environments, despite the increasing popularity of such learning environments and the temptation to build learner interest by augmenting training content (p69).

In addition, the research has often been based on relatively short duration experiments, usually concluding with simple tests of recall conducted within a few minutes of the learning. Furthermore, many of these experiments have been criticised for their lack of a control group although, as Hidi and Baird (1988) lament:

We found it impossible to construct such a control because reducing the base text by removing all of its interesting content would yield merely a generalized summary (p473).

The discussion above may suggest that the problem is simply one of whether 'seductive details' is a single phenomenon which is either present or absent.

However, a distillation of the research points to four significant factors which can affect the nature and amount of potential seduction away from the learning of the main content and which could be addressed by designers. These are:

1. **Importance** and **interestingness** of seductive details relative to main content;
2. **Density** of seductive details relative to main content;
3. **Placement** of seductive details within the overall content; and
4. **Presentation** medium and overall style/tone adopted.

### **Relative importance and interestingness of seductive details**

'Importance' may appear to be quite a subjective and personal characteristic from a learner's perspective but, for designers, content can be assessed much more objectively in terms of whether it relates to the achievement of the learning outcomes. One might question whether all learning – particularly that which is skill- or attitude-based, or associated with higher order cognitive processing – can be adequately defined in the form of outcomes. I would contend that they *can*, provided the outcomes are correctly expressed in terms of measurable verbs such as 'describe', 'interpret', 'demonstrate', 'create', etc. In this way, learners can be assessed to some defined standard of performance whereas, if inappropriate outcome descriptors such as 'understand', 'know' or 'appreciate' are used, this is not directly possible. However, it is possible to unravel the latter through use of the former, such as confirming that a learner knows or can understand X by asking them to *describe, explain or apply* it.

Garner et al (1991) refer to two items of seductive detail added to a biography of Prof Stephen Hawking: one about black holes, the other about a wager Hawking had with a colleague. Both items were regarded as interesting but neither satisfied the learning outcomes; "However, in the case of black holes, the information supported generalizations about Hawking's scientific work on Grand Unification, whereas wager information did not" (p651). This suggests that, although neither were important in terms of the learning outcomes, one was more useful/relevant than the other.

The inherent interestingness of the overall topic will affect the need for and potential effect of any seductive details. Hidi et al (1982) found that texts based mainly on facts, explanations or instructions were generally considered to be



important but uninteresting; anecdotes were interesting but acknowledged as unimportant; and only narrative texts were rated as interesting *and* important. Lehman et al (2007) concludes that “short technical texts ... found negative effects for seductive details, whereas longer texts with a stronger narrative component ... find no effects on learning” (p583). Schraw (1998) also acknowledges that ‘type of text’ may have been a factor in his finding no evidence of a seductive details effect in experiments based on Wade et al’s (1993) Nelson biography.

Garner et al (1991) note that, “One reason why interesting detail is so seductive is that it typically appears in text that is, overall, of very low personal interest to most students” (p657). Harp and Mayer (1997) experimented with teaching 77 psychology students about the formation of lightning where seductive details were included in the form of captioned photos (e.g. an aircraft being struck by lightning, a footballer’s burned and damaged kit) and/or descriptions. As Table 3.2 illustrates, the inclusion of seductive details within a fairly dry scientific explanation inevitably craved the learners’ attention and interfered with schema construction in long term memory, affecting both recall and problem-solving post-test performance.

**Table 3.2. Effect of seductive details on recall and problem-solving scores**  
(Harp and Mayer, 1997, p98)

Content	Recall	Problem solving
Base (non-seductive) description and images	3.8	3.2
Base plus seductive text	2.3	1.6
Base plus seductive images	2.2	2.1
Base plus seductive text and images	0.9	1.2

### **Relative density of seductive details**

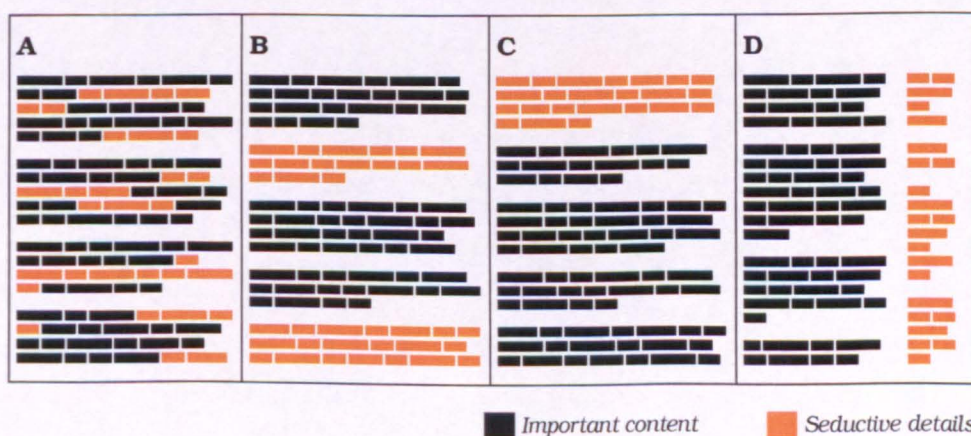
Any seductive effect may depend on the relative proportions of seductive to important content. Garner et al (1989) increased a scientific text on insects by 40% and showed a detrimental effect on learning, but Muller et al (2008) added 50% when experimenting with the multimedia teaching of stellar spectra: fifty five students received a ‘concise’ treatment lasting about 7½ mins, and forty nine received an extra 3¼ mins which included:

...excerpts from an interview with a professional astronomer interspersed throughout the concise treatment. The astronomer discussed exciting topics in the field like the formation of black holes [...] These topics were irrelevant to the learning outcomes. (p214)

Both versions yielded similar post-test results, contrary to the principle of seductive details and also showing that "...in authentic learning settings, interest may mitigate the effects of [Mayer's] coherence principle" (p211). Muller et al do not discuss it, but their findings must also call into question the *redundancy effect* (Mayer, 2001, 2005; Sweller, 2005). However, despite "50%" appearing to represent a lot of seductive details, the validity of Muller et al's findings from these relatively short activities (7½ mins versus 10¾ mins) might be questioned. Also questionable were the negative effects of Harp and Mayer (1997) adding 30% seductive details in the form of text and photos to their exposition of lightning formation; Sadoski (2001) points out that their basic text had 9 main ideas on the causes of lightning, whereas the seductive version added 12 more on the effects of lightning – thus transforming the 'main point' of the lesson.

### Placement of seductive details

The term 'seductive details' is well chosen, because seduction is not typically a solitary or blatant event, but one which unfolds surreptitiously over time. This is illustrated in Figure 3.2(A) where the seductive details are woven into the body of the text – as they were in Wade and Adams' (1990) Nelson narrative – such that it would be difficult to identify and discriminate between seductive and important content.



**Figure 3.2** Examples of seductive details placement

By compartmentalising seductive content (Figure 3.2 B), it can be more obvious that this is non-essential, a good example being an anecdote. Having an unimportant but interesting introduction (C) is not necessarily seductive, but more a well proven technique for gaining the learners' attention' (e.g. Gagné, 1965). Finally, (D) is perhaps an extension of (B) where the designer is flagging that the seductive details are somehow peripheral by their clear spatial separation.

Garner et al (1991) tested for a seductive details effect using a biographic text on Stephen Hawking. Using four different placement methods they concluded that there *was* a detrimental effect but that, "placement of interesting detail in text made no difference whatsoever" (p651). However, these results were based on just four paragraphs of material and small sample sizes ( $n=12$ ) per placement design. Garner & Gillingham (1991) used the same base text with slightly larger samples ( $n=18$ ) and did *not* find a detrimental effect, causing others to speculate that placing the seductive details at the start of the passage had engendered interest but without causing interference (Goetz & Sadoski, 1995; Lehman et al, 2007). In analysing Harp and Mayer's (1998) lightning experiments, Sadoski (2001) noted:

...placing all the additional material about lightning's effects at the end of the passage produced better recall of the foregoing information about lightning's causes, but placing the information about lightning's effects at the beginning of the text produced better recall of that material. (p273)

### **Presentation medium**

Figure 3.2(D) differs from A, B and C in that it represents a design/layout decision rather than a writing/editorial one. Early and pre-multimedia research naturally focused on *textual* content, but CAL offers other media and interactivity options, creating more scope for interest and seduction but also more opportunities to flag less important content. For example: access might be via some form of clickable More button; text may be displayed as *rollover* information, or on tabbed text panels, or there may be hyperlinks to additional data files and online resources (Lynch and Horton, 2001; Nielsen, 2000).

CAL and multimedia also tend to be more graphically rich because, unlike printing, adding colour imagery makes little difference to the cost. This means

graphics can be, and often are, added for 'cosmetic' rather than pedagogic reasons. Levin (1989) refers to graphics which are interesting but have no clear pedagogic value as 'decorative illustrations' but he acknowledges that their affective or motivational value in increasing the learner's interest could be just as important as any pure pedagogic purpose. Sanchez & Wiley (2006) examined the effects of seductive details on high- and low-working-memory-capacity (WMC) learners and found "...the only difference between WMC groups was the amount of irrelevant illustrations viewed, of which high-WMC individuals viewed significantly fewer" (p352).

Evidence suggests that richer media have greater potential for interest, and it may therefore be no coincidence that Wade and Adams' (1990), Garner et al's (1991) and Schraw's (1998) 'poor' learning results were generated when seductive details were added to existing text. In contrast, Levie and Lentz (1982), Gyselinck and Tardieu (1999) and Carney and Levin (2002) all found examples of adding pictures and illustrations to text that were not essential but which did aid learning and, for Muller et al's (2008) 'good' results, the seductive details were spoken by an expert using video – a medium regarded by many as being inherently more stimulating (e.g. Timmerman & Kruepke, 2006).

### ***Rethinking seductive details***

The foregoing discussion shows that effects named, or attributed to, 'seductive details' are many and varied, as is their potential impact on learning; what is very clear is that seductive details should not be considered as being simply present or absent. Consequently, an addendum to the commonly accepted definition is proposed which would constrain matters by focussing on the key deleterious issue, so that:

Interesting but unimportant items that are added to learning materials to make them more attractive, but which actually seduce the learner away from the intended learning or outcomes.

*(Garner et al, 1992)*

Becomes...

Interesting but unimportant items that are added to learning materials to make them more attractive *and which solicit recurrent attention*, thus seducing the learner away from the intended learning or outcomes.

Examples of items that would *not* solicit recurrent attention include: an introductory hook; an anecdote; material which, through placement or other design features, is clearly elective; or a limited amount of content that is interesting but is *not* notably salacious or tangential. Excluding this from the revised definition makes it easier to focus on details that truly do seduce the learner and may therefore impair learning. However, regardless of this redefinition, the balance of evidence for and against a seductive details effect remains inconclusive.

One risk of redefining seductive details is that it might imply to some designers that it is acceptable to add interesting but unimportant content provided it does not appear to detract from the learning. In fact, Mayer's (2001) *coherence principle* remains a good design touchstone because less is often more and, in any event, it is surely incumbent upon designers to make *all* learning content as interesting as they reasonably can.

Notwithstanding the difficulties and reservations with seductive details, discussed above, I contend that there is another discrete category of material that should be considered. Its value lies not in its interestingness or its direct relevance to the learning outcomes, but in acting as a catalyst for the *process of learning*, and it is that which will be considered in some detail for the remainder of this chapter.

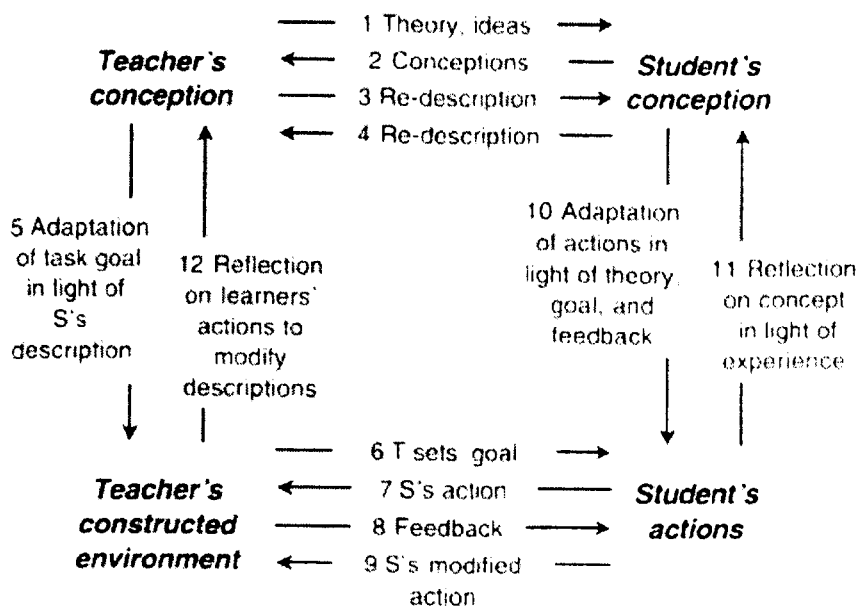
## ***Supporting the Learning Process***

As Shulman's (1999) quote at the start of this chapter reminds us, teachers and presenters routinely use '*analogies, metaphors, examples, demonstrations and the like*' as part of their explanations. Typical use of PowerPoint presentations – described by Tangen et al (2011) as, “text, graphics, images, video and other media [...] displayed on-screen to supplement the verbal narration of the presenter” (p865) – gives an apposite illustration of this. The slideshow often contains fairly succinct on-screen text and diagrams that convey key information which is directly relevant to the purpose, aims or outcomes, with the presenter's commentary providing a supporting narrative which (if done well) glues those facts together, transforming them from a set of statements into a clear, relevant, engaging and effective explanation. It is notable, for example, that when conference presenters make their slides available online for those

who were unable to attend the learning experience, whilst helpful, never seems to quite match that of 'being there'. Shulman (1987) notes that what is important about a didactic approach is both the information presented *and* how effectively this is done; for instance:

The key to distinguishing the knowledge base of teaching lies at the intersection of content and pedagogy, in the capacity of a *teacher to transform the content knowledge he/she possesses* into forms that are pedagogically powerful and yet adaptive to the variations in ability and background presented by the students. (p15; italics in original)

So the role of supportive content can be justified in part by comparison with the everyday process of face-to-face teaching and presenting. Laurillard (2002) describes a Conversational Framework for classroom teaching (Figure 3.3), with 'conversational' being an apposite title, not because the teacher is necessarily conversing with every student, but because of the back-and-forth way in which the teaching and learning interaction unfolds.



**Figure 3.3. Conversational Framework: activities in the learning process** (Laurillard, 2002, p87)

Considering just the first three steps, categorised by Laurillard as *discursive*, we see an example of the 'sage on the stage' (1) presenting material, (2) gauging the reaction – which may manifest itself not just verbally but perhaps in the form of a dozen blank expressions – and (3) thinking on his or her feet, adapting the explanation as circumstances require. What is notable about activities 1 and 3 is that the primary, factual material cannot change, but the

wrapping material that supports it can. The skilful teacher employs various methods under different circumstances, giving examples, analogies, metaphors, anecdotes, whiteboard sketches, etc, to support, contextualise and reinforce the facts so they may be more easily understood. Mayer (1992) reminds us:

As Resnick (1989) aptly noted, “learning occurs not by recording information but by interpreting it” (p2) ... Accordingly, the teacher becomes a participant with the learner in the process of shared cognition, that is, in the process of constructing meaning in a given situation. (p407)

This insight is important because of the notable difference between face-to-face teaching and tutorial CAL. Most CAL programs simply transmit information and, whilst it is possible to see or hear material again, that is *not* ‘re-description’ (Figure 3.3, Activity 3), it is just the original transmission repeated. Now this is no worse than reading a book, so it is not a poor method per se, but it does lack the sophistication and adaptability of a good teacher.

Pseudo-intelligent programs *are* possible, and Rushby’s (1989) requirement described in Chapter 1 – that computers should “adapt and respond to the learner’s needs, difficulties and progress” (p150) – aspires to this, but such solutions are likely to be complex, costly and therefore rare. The more pragmatic approach is to put maximum effort into designing the initial message to be as effective as possible. Both factual and supporting material should be coherent, relevant, stimulating, engaging, etc, and, as stated above, the essential facts cannot change, but the methods of supporting how they are presented and understood, can. Whilst quick or cheap CAL may just convey information, good CAL does not. As Masie (2009) noted in a presentation to e-learning conference delegates:

Most e-learning is content-rich and context-poor ... Content is the rule, the policy, the theory; context is the story that explains the rule.

Good CAL designers therefore need to use content that can somehow act as a *catalyst for learning* to present sufficient of the ‘the story that explains the rule’. However, they also need to envisage what learners will make of that story, what misconceptions or problems may arise, and how best to overcome them; and they must do all this within the constraints of succinctness that are expected of online learning.

## Catalytic Content

I define catalytic content as content included:

*...for the principal purpose of supporting the process of learning because it somehow introduces, exemplifies, contextualises, substantiates or reinforces that essential primary content which is directly relevant to achieving the learning outcomes.*

Just as a chemical catalyst might be considered to have limited value in its own right, a chemical reaction may occur slowly or not at all without it. Similarly, although catalytic content is not 'critical', without it, primary material would not be as quickly, easily or deeply understood. A very simple example of this comes from an explanation of HTML cell 'padding':

Cell padding is similar to cell spacing except that it determines the amount of space between a cell's content and its border.

The grey text here is catalytic, but the primary content is more easily learned because of its presence; in this case, the catalytic content was *analogous*, helping the learner construct schemata in long term memory by relating unknown information to known.

A different example of catalytic content comes in a video extract from a CAL tutorial on the hydrological cycle and flooding. The first 16 seconds of spoken dialogue below are catalytic because they set the *context* for the primary material which follows:

*Dr Sandy Smith, Open University:* <sup>[1]</sup>"This tiny moorland stream becomes the River Severn. <sup>[2]</sup>Its actual source is a bog a few kilometres uphill from here on the Plynlimon range in mid-Wales. <sup>[3]</sup>Plynlimon is also the site of an important scientific study."

*Jim Hudson, Centre for Ecology & Hydrology:* "We get very heavy rainfall here and our average annual precipitation is about two and a half thousand millimetres..."

In this case, Jim Hudson's statement is afforded relevance and credibility by him being identified by the academic presenter as a scientist (sentence 3) working at a specific location near the source of the River Severn (sentences 1 and 2). In one further illustration below, catalytic content takes the form of an *example*, which supplements a primary explanation of personal taxation:



Personal allowance for 2009-10 is £6,475, with remaining income up to £37,400 taxed at the basic rate of 20%. Any additional income is taxed at the higher rate of 40%.

Nicola earns £36,000. She will pay no tax on the first £6,475 of this, leaving £29,525 which is taxed at 20%, giving her a net income of £6,475 + £23,620 = £30,095.

## ***Analysing and identifying catalytic content***

The three examples above illustrate relatively straightforward distinctions between content that is either primary or catalytic. However, there will be other instances, particularly when there is a greater volume of material, or some combination of richer media, where the distinction may not be so clear. Some ground rules are therefore provided below to assist designers and those who may further research this topic. Because catalytic content, by definition, *cannot* be primary content, delineation of the former is helped here by defining the latter:

1. Primary content relates directly to achievement of the learning outcomes and so its inclusion is essential; if any primary content is removed or omitted from the design, the learner could not be expected to fully achieve the outcomes.
2. Within one activity that a learner could reasonably complete in a single session, information can only be classed as primary once. If it is repeated in the same or another medium or page, it would become catalytic because it is then supporting that initial instance of the message. However, repetition in another activity, chapter, unit, etc, could be categorised as primary once again.
3. There can only be one primary medium at any instance in time. For example, in a CAL tutorial screen which contains voiceover, bullet text and imagery, all of which are relevant and complementary, the primary medium would be that which we could expect most learners to regard as delivering the main message. In this instance, the voiceover would probably solicit greatest attention unless, say, there was a particularly striking image<sup>9</sup>. Once a point has been made, the primary medium may shift, on that or a following screen; however, it

---

<sup>9</sup> It is readily acknowledged that this is a subjective judgement and one which may also vary according to the nature and preferences of different learners.

would be inappropriate to regard the primary medium as switching, almost second by second, as a narrative point unfolds.

4. Whilst there cannot be more than one primary message, there may be some instances where there is none. For example, all the material in a particular screen or narrative might be classed as useful, relevant or interesting but its inclusion serves some other pedagogic purpose, such as a case study or worked example.

Primary content is clearly *essential*, but arguably not *sufficient* when presented in isolation. However, the addition of catalytic content to 'introduce, exemplify, contextualise, substantiate or reinforce' it can lead to a potentially more powerful and effective learning cocktail.

### Potential categories of CAL content

Catalytic content therefore cannot be primary but it must provide one of the supporting functions listed in the original definition above; another essential characteristic, although not explicitly stated in that definition, is *relevance*, in respect of the learner and the existing knowledge to which primary information is being related. Let us return to the short HTML example from above:

Cell padding is similar to cell spacing except that it determines the amount of space between a cell's content and its border.

The analogy works because we can reasonably expect the learner to know about and have some interest in cell spacing, whereas the following analogy is less relevant, probably less familiar and so likely to be less successful:

Cell padding is similar to the safety zone painted on the floor around an industrial machine in that it determines the amount of space between a cell's content and its border.

The examples and discussion so far may imply that content is either primary or catalytic. However, (ignoring text that serves a purely grammatical purpose), there will be other categories, such as the seductive details that were considered earlier.

Figure 3.4 presents seven potential categories of CAL content, based on an extension of Garner et al's (1992) seductive details definition. For pragmatic reasons, interest is restricted to a simple binary characteristic even though it

probably varies along some continuum. In contrast, importance should be unravelled further (i) to ensure that all directly relevant (i.e. critical) material is included and (ii) because content more loosely related to the learning outcomes can be helpful but should be used with due consideration.

1. Interesting but tangential to the topic or aims	7. Catalytic
2. Interesting, relevant, but not matched to learning outcomes	
3. Interesting and directly relevant to learning outcomes	
4. Uninteresting but directly relevant to learning outcomes	
5. Uninteresting, relevant but not matched to learning outcomes	
6. Uninteresting and tangential to the topic or aims	

**Figure 3.4. Potential categories of CAL content**

The categories in Figure 3.4 can therefore be explained thus:

- Category (1) comprises potentially seductive details which, because they may impair learning, should be minimal.
- Category (2) includes devices now excluded from the revised definition of seductive details, such as an anecdote or introductory hook, which can be used sparingly.
- Category (3) is 'ideal' learning material because it is both factually essential *and* interesting.
- Category (4) requires greater learning designer attention because it too is essential, but may *not* be inherently interesting to many learners.  
*Only categories (3) and (4) could be classified as primary content.*
- There should be minimal category (5) and no category (6) content at all.
- Finally category (7), *catalytic content*, is the cement that, when applied to bricks 2-4, can turn them into a sturdy wall.

It may be inferred from the foregoing discussion that catalytic content is somehow less important or valuable than primary content and, although that is *not* the case, they do have different *types* of value. Primary content has a criticality in that its absence would mean some of the learning outcomes could not be achieved; however, catalytic content may have some equivalent value in terms of the effectiveness of the whole learning process. To draw a comparison, if one dines at a Michelin-starred restaurant, the food is undeniably *the* critical

factor, but it would be wrong to claim that this meant the dining room, the service, the wine and the tableware were *not* important; it is the sum of the component parts that results in the success or failure of the entire experience.

When differentiating between primary and catalytic material, provided the learning outcomes are clear, it is reasonably straightforward to analyse text (and voiceover transcript), although the division may not always be quite as neat as in the earlier examples. Take the following example, for which the learning outcomes relate to the differences between evaporation in moorland compared to that in a forested area:

“<sup>[1]</sup>We’ve noted that about fifteen percent of the precipitation in a moorland catchment is evaporated back into the atmosphere, mainly through transpiration <sup>[2]</sup>from the vegetation. Now in the forested catchment it’s <sup>[3]</sup>slightly different. We’re losing about thirty percent of the rainfall over the period, and that’s a combination of transpiration <sup>[4]</sup>through the trees, and also interception <sup>[5]</sup>which is direct drying of the canopy. So overall, we’re getting about double the amount of evaporation from a forested catchment.”

The difference in evaporation between moorland and forest is important although the exact amounts <sup>[1],[3]</sup> are not; the meaning of transpiration and interception are also directly relevant to the outcomes, but were actually explained on the preceding screen, making this repeat occurrence <sup>[2],[4],[5]</sup> catalytic.

### **Other examples of catalytic content**

If catalytic content supports the process of learning, it can apply not just within the body of a teaching narrative but also to pedagogic material that comes before, after and around. This therefore includes content such as introductions, aims/objectives, formative assessments, problems and solutions, worked examples, case studies, sample data, captions, transcripts, summaries, glossary entries, etc. In fact, material of this nature is likely to have less diversionary content (such as seductive details) than the central narrative; even a cursory analysis will show most of it to be catalytic because the reason for its inclusion is usually to support the primary message and the comprehension thereof. It is also worth noting that catalytic content is directed at *pedagogic* support for learning, not other forms of support that are more utilitarian in nature. The latter might include text in ‘header’ and ‘footer’ areas of the screen

such as titles, page numbers and button labels; or 'how to operate this program' information. These tend to be included more to help the learner rather than the learning, and so they would not be classed as catalytic.

The emphasis thus far has been on text, but catalytic content can of course be represented in and by *all* media, including audio and static or moving imagery. For example, a simple and very common use of catalytic material in video is the 'establishing shot', whereby the viewer may see some activity taking place in a room, but they know that room is in a particular office, factory, school, hospital, etc, because they were shown a 2 second clip of the building exterior beforehand. That exterior was not critical to the plot, but it was helpful for making sense of the plot.

Compared to text, analysis of visual imagery is likely to be less straightforward. Relevance to the learning outcomes may be established fairly easily but, if an image is presented in conjunction with text or narration, determining whether the image is primary or catalytic (or peripheral) may be a much more subjective decision. Type of image is pertinent, as it may be argued that one which is monosemic<sup>10</sup> and so follows diagrammatic conventions – such as a map, graph or chart – could be primary content because it may plausibly be interpreted by the learner; the narrative thus provides a supporting description which, albeit important and relevant, may be catalytic. However, a polysemic image such as Van Gogh's *Sunflowers*, may have features or subtle nuances that make it much more open to (mis)interpretation by the learner; hence, in a didactic situation, the narrative discussion or explanation assumes a greater criticality, even though the image is extremely important and would be frequently referenced by the narrative.

The boundary between catalytic and cosmetic imagery is equally fluid and will vary according to circumstances. For example, the portrait of Neil Armstrong on the left in Figure 3.5 might be argued to be cosmetic or seductive in almost any situation. However, the image on the right is (subject to the learning outcomes) more likely to be adjudged as catalytic because it brings with it a context that can reinforce learning: this is not just Neil Armstrong, astronaut; it is Neil Armstrong *walking on the moon*, back in 1969, with all the inherent technological challenges and risks that that entailed.

---

<sup>10</sup> Monosemic and polysemic imagery (Bertin, 1983) were described in Chapter 2.



**Figure 3.5. Cosmetic and catalytic imagery**  
(Google images; copyright unknown)

### ***Theoretical basis for catalytic content***

The descriptions and examples of catalytic content given above have introduced and contextualised the principle and they provide a logical foundation for such a category of material, but much of the foregoing is based more on common sense than on specific theoretical or evidential support. Catalytic content, as defined, is a new concept that has not been subject to previous research as such; however, it is possible to find or infer support from various other pertinent pieces of research and these are discussed in this section.

The role of catalytic content is defined as introducing, exemplifying, contextualising, substantiating and reinforcing primary content, and those characteristics all relate to the general principle of more easily 'making sense' of key information. Content is catalytic because it helps learners relate unknown to known, move from simple to complex, make connections and understand relationships between elements of information: mentally assembling the 1000 jigsaw pieces into one coherent picture. An obvious theoretical starting point is therefore with those principles introduced in Chapter 2 that relate to the construction of schemata in long term memory.

The function of catalytic content accords with the maintenance of a 'mental thesaurus' as described in Tulving's (1972) *semantic memory*. It also relates to *Schema Theory*, which describes not just the existence of a hierarchical schemata-based model of memory but also the interaction of incoming information with existing knowledge. Benjafield (2006) describes this as

occurring through a four-stage process of selection, abstraction, interpretation and the integration of old with new. In a similar vein, Ausubel believes the most important determinant of learning is what the learner already knows. His *Subsumption Theory* (1963) holds that new information is subsumed as it is linked to relevant points in learners' existing cognitive structures, and that this is best achieved by presenting general information first, followed by more specific detail. He supports this classic known-to-unknown method of teaching through the use of *Advance Organisers*:

These organizers are introduced in advance of learning itself, and are also presented at a higher level of abstraction, generality, and inclusiveness; and since the substantive content of a given organizer or series of organizers is selected on the basis of its suitability for explaining, integrating, and interrelating the material they precede, this strategy simultaneously satisfies the substantive as well as the programming criteria for enhancing the organization strength of cognitive structure. (p81)

Ausubel emphasises that advance organisers are *not* the same as overviews and summaries, because these simply emphasise key ideas and are presented at the same level of abstraction as the core material. The 'higher level of abstraction' of advance organisers enable learners to orient themselves to the topic, so they can more easily appreciate the meaning and context of new information through how it links to what they already know. This is similar to Mayer's (2005) *Pretraining Principle*: "people learn more deeply from a multimedia message when they know the names and characteristics of the main concepts" (p170) and also research by Rawson and Kintsch (2002) into the benefits of prior background reading.

Reigeluth's (1979) *Elaboration Theory* offers similar advice: present information in its simplest form first (Reigeluth referred to this as the *Epitome*) and then elaborate, thereby creating meaningful context for the detail. Similarly, in his exposition on *Andragogy*<sup>11</sup>, Knowles (1973) advises that, when teaching adults, new information should be explicitly linked to existing knowledge whenever possible; that linkage is also a component of Gagné's (1965) nine-step model of classroom teaching. Finally, Rowntree (1990), describing the tutorial-in-print prevalent in the early days of the Open University, also advises: "Explain the subject matter in such a way that learners can relate it to what they know already" (p82).

---

<sup>11</sup> Andragogy is defined as the science of teaching adults, as distinct from pedagogy which, though widely misused in adult education, relates to the teaching of children.

Another method of learning support is Bruner's *scaffolding*, although his original research in this field related to children, social interrelationships, and language acquisition in particular. Wood, Bruner and Ross (1976) describe scaffolding as providing learners with assistance or support to perform a task that would ordinarily be beyond their capability. Saye and Brush (2002) propose two forms of scaffolding: soft and hard, with an example of the former being when a teacher circulates the classroom and converses with the students, enquiring about methods and outcomes and providing constructive feedback (thus relating to the discursive activities of Laurillard's (2002) framework in Figure 3.3). Hard scaffolding, in contrast, is not spontaneous but is planned in advance (e.g. models and worksheets) to support students engaging in a difficult activity. This pre-prepared scaffolding is closer to that which would apply to distance learning material and, more recently, Yelland and Masters (2007) describe *technical scaffolding* in which the computer program replaces the teacher, providing support through the tutorial itself or via web links, help pages or other electronic activities and resources.

One concept devised with technology-based learning specifically in mind comes from Bransford and the Cognition & Technology Group at Vanderbilt. *Anchored instruction* (CTGV, 1993) is based on the premise that learning and teaching activities should be designed around an 'anchor' which is typically a relevant case study or worked problem. Based on this, students are then encouraged to explore and to solve complex, realistic problems. In early work by CTGV (1993), video anchors were used, but:

The design of these anchors was quite different from the design of videos that were typically used in education [...] our goal was to create interesting, realistic contexts that encouraged the active construction of knowledge by learners. Our anchors were stories rather than lectures and were designed to be explored by students and teachers. (p52)

### **Coherence and comprehension**

Also relevant is the theory of *text coherence*, described by Ainsworth and Burcham (2007) as, "the extent to which the relationship between the ideas in the text are made explicit" (p288). They investigated the relationship between coherence and another effect: *self-explaining*, whereby students learn more by



generating explanations which state something beyond the information they were given. Ainsworth and Burcham (2007) describe self-explaining thus:

If a text is in someway incomplete [...] then learners generate inferences to compensate for the inadequacy of the text and to fill the gaps in the mental models they are generating. (p287)

They compared the effect on self-explaining of a minimally coherent text about the human circulatory system (1075 words) against a maximally coherent version (3496 words). Coherence was increased in the longer version by:

- Replacing pronouns (e.g. 'they') with nouns ('the valves');
- Adding sentence connectives such as 'therefore', 'however', etc; and
- Adding descriptive elaborations, such as replacing 'the ventricles contract' with 'the ventricles (the lower chambers of the heart) contract'.

The purpose of the latter was to, "link unfamiliar concepts with familiar ones and to provide links with previous information presented in the text" (p291) and so the overlap between text coherence, catalytic content and the other theories noted above is readily apparent. Through experiments with different groups Ainsworth and Burcham (2007) found that self-explaining *does* have a positive effect on learning as expected, but it is, "not sufficient to counteract the negative effects on novices of a lack of text coherence on learning" (p300). In short, the understandability of didactic material is paramount, particularly for those learners with less prior knowledge.

Ainsworth and Burcham's (2007) coherence techniques clearly accord with the principles of catalytic content, although it is noteworthy that they do not appear to have felt constrained by word count, noting, "The second text is noticeably longer because it has provided elaborations that the learner needs to help them make sense of the information" (p289). One might expect that an explanation which is about 3.5 times more expansive than another will, if well-written, almost inevitably have a greater positive effect on learning. Whilst my original definition of catalytic content did not explicitly refer to the volume of material, Mayer's (2001) *coherence principle* should continue to guide designers into applying some degree of moderation over content, most notably in that destined for CAL and other forms of e-learning, where succinctness is an expectation. Catalytic content is therefore not simply a licence to add

unrestricted amounts of explanatory and supporting material in order to get a point across.

A substantial and varied amount of research has also been conducted into the general area of text readability and understanding. Examples include the effects of: syllables per word and words per sentence (Kincaid, 1975); number of passive sentences and average number of verb phrases (Schwarm and Ostendorf, 2005); lexical cohesion (Halliday and Hasan, 1976); and the use of pronouns (Krahmer and Theune, 2002). These are of course relevant to comprehension, but in a way that is subtly different to catalytic content, the essence of which is that one part of the material supports the learning of another (primary) part. The focus therefore is not on *text* comprehension (it is assumed that learners will understand the words themselves), but on *topic* comprehension through a concentration on understanding the context, relevance and relationships between those words. Kintsch (1988) refers to “the local and global meaning and structure of a text” (p163). Catalytic content aims squarely at the global and, whilst the readability research noted above has relevance, this is at more of a local level in terms of good syntactic practice and effective writing rather than techniques specifically to support learning of the topic.

Research into comprehension that is more directly related to catalytic content includes investigations by Goldman (1997) into rhetorical devices that help to organise and signpost content, such as bullets and paragraph indentations or the use of phrases such as ‘in summary’ or ‘the first thing to remember...’. She refers to an earlier study by Goldman and Saul (1990) in which a number of students, upon encountering the enumerator ‘second’ at the start of a sentence, were then seen to check that they knew what had been first. Goldman (1997) also describes the difference between learners simply creating a representation of the text they had read and showing evidence of having integrated it with existing knowledge, thus creating a representation of the meaning and context, the latter referred to by van Dijk and Kintsch (1983) as a *situation model*. McNamara (2002) described the same distinction in terms of two levels of understanding: *textbase*, based purely on information contained in the text, and *situational*, which requires the text to be integrated with real-world knowledge:

A person can memorize Lincoln's Gettysburgh Address without learning the meaning of its contents. At the other extreme, a person can understand the purpose, meaning, and historical consequences of the same address without recalling the exact words (p52).

The literature thus far in this section has been considered in terms of its general endorsement for, or accordance with, catalytic principles. It is also interesting to consider it in the context of the *categories of content* diagram (Figure 3.4) and note that all of these additional principles fulfil a support-for-learning function, which would place them squarely alongside catalytic content rather than in any of the six main categories.

## **Analogy**

One specific catalytic technique described earlier in this section which has received frequent research attention (notably in the areas of science education and psychomotor (e.g. sports) skills) is the use of *analogies*. Rowntree (1990) describes these as, “a potent means of clarifying a new idea” (p144), although he also advises designers, “don’t always be satisfied with the first one that comes to mind” (p148). For example, Gentner and Gentner (1983) describe two common analogies used to explain electrical current: the flow of fluid through pipes (in which electrical resistance would be analogous with pipe diameter); and the flow of crowds through passageways (with resistance analogous with gateways in the passages). They found that learners who had been given the fluid analogy were better at solving battery-related problems, whereas those given the crowd analogy were better at resistor-related problems. Zheng et al (2008) used just the fluid-in-a-pipe analogy in an experiment in which 89 students were taught about voltage, resistance and current flow by one of four combinations of multimedia learning objects (MLOs):

- MLO on water flow (analogy) followed by MLO on electrical circuit
- MLO on electrical circuit with no prior analogy
- Verbal explanation of water flow followed by that of an electrical circuit
- Verbal explanation of electrical circuit only

They found that analogical reasoning significantly improved science learning and that, “multimedia becomes more effective when it is integrated with an

instructional method such as analogy and less so when it is used only as a visual tool” (p474).

In an experiment to teach accounting procedures to 79 undergraduates, Hanson and Phillips (2006) used one of three analogy conditions (none, simple or elaborate) combined with one of two exercise conditions (none or brief), finding that comprehension was greatest when an elaborate analogy was followed by a brief exercise. In a second experiment ( $n=21$ ) using just the three analogy conditions but no exercise, differences in comprehension were not statistically significant but, “positive effects did emerge two weeks later when students learned the allowance method in greater depth” (p2). Also in business studies, Gentner et al (2003) conducted three experiments based on novice undergraduates learning negotiation strategies. The first ( $n=48$ ) showed a benefit for analogical learning using two case studies (compared to a non-case-study control group); the second ( $n=128$ ) showed that comparing two case studies was markedly better than studying them separately; and in the third ( $n=158$ ), performance in a practise face-to-face negotiation exercise was better when the analogical learning had been guided, compared to a group that had simply been instructed to compare two case studies.

McDaniel and Donnelly (1996) experimented with 201 undergraduates who received one of two short didactic texts explaining scientific concepts: one contained a literal description of the concept and the other also included an analogy which related the concept to something that would be familiar to learners. Based on post-tests conducted immediately afterwards, the group who received analogies performed slightly better when answering inferential questions (thus implying a deeper learning/understanding) but were slightly worse on purely factual (recall) questions. However, prior knowledge can impact on the effectiveness of any teaching or learning design approach, and Alexander and Kulikowich (2006) found that, “...the incorporation of analogies may misdirect readers' attention or may increase processing demands, particularly in those cases when readers' [topic] knowledge is low” (p895).

## **Summary**

A number of theories and principles have been discussed above, all of which relate in some way to aiding comprehension and schemata construction by

helping learners relate unfamiliar or complex information to that which they already know and, to some extent, understand. Catalytic content fulfils a similar function but is not so much a specific 'method' as a principle or a perspective on the nature of content which informs a suite of techniques (for example, introductions, analogies, metaphors and worked examples) that can support the learning process. That being the case, catalytic content might appear to offer nothing other than a fancy new label tied to a bag of familiar and well-researched teaching tools but, in fact, its real value is revealed through the key difference between face-to-face and distance learning.

All of the techniques described are routinely applied in a classroom as part of some unfolding process, the human nature of which should determine whether, when and how each is used, and this means that no two instances of any lesson are likely to be exactly the same. Application is usually done intuitively by effective teachers with no time or need for any detailed analysis of the situation. In contrast, didactic material for distance learning, and the methods of presenting and supporting it, must necessarily be determined in advance by designers with the intention of delivering what, for the majority of learners, will be the best *right-first-time* message. The notion of catalytic content makes the analysis of the potential effects of different elements of content – as the examples at the start of this section show – much more explicit and focused, and this should lead to greater learning effectiveness, but within the confines of the brevity that is expected of CAL and other computer based material.

Catalytic content therefore represents a new phenomenon whose principles are supported by related research and whose individual techniques fall into the general category of 'tried and tested'. However, they have not yet collectively been researched from this perspective because investigations in this general domain have tended to focus on relationships between content and the topic, rather than on the role of some content in the *process of learning* the topic. If supported by empirical data, catalytic content may prompt designers to consider whether (or how much) content is included because of its direct relevance to the topic, or for its capacity to support the learning process.

Because of the novelty of this concept, the lack of previous research relating to it, and the potentially confounding effects of different media, investigation hereafter has been restricted to the catalytic properties and effects of *text only*.

The catalytic properties and effects of written wordage within text-with-imagery-based CAL are thus considered in a final proposition.

### **Proposition 3**

**P3a.** Learners studying a text-with-imagery-based CAL tutorial will learn more if there is a higher proportion of catalytic text than those who receive a lower proportion of catalytic text.

**P3b.** When questioned after study, learners will report that a text-with-imagery-based CAL tutorial with a higher proportion of catalytic text is more interesting and enjoyable than one with a lower proportion of catalytic text.

## **Conclusions**

The literature review began in Chapter 2 with a set of multimedia design principles which received broad support from various researchers and research scenarios, but there were also some notable exceptions and inconsistencies which could not be fully explained. In this chapter, the nature and effects of interest were considered and the phenomenon of seductive details was investigated as a potential source of earlier research inconsistencies, but investigations showed a seductive effect to have its own shortcomings. This led to a previously undocumented way of considering learning material: the notion of *catalytic content* which serves to support *the process of learning* by introducing, contextualising, exemplifying, substantiating or reinforcing essential/relevant material.

Three main research propositions have now been put forward, relating to (i) the succinctness and effectiveness of on-screen written text; (ii) user expectations and experiences of different media; and (iii) the effects of different proportions of catalytic content on learning effectiveness and satisfaction. Chapter 4 will next consider the methodological options by which these may be robustly tested.

# **Chapter 4**

## **Research Methodology**

Sometimes policy makers and practitioners are blamed for taking insufficient account of research findings. Alternatively, research is criticized for being insufficiently relevant to the work of those whom it is intended to inform.

Hammersley & Scarth (1993, p216)

Hammersley and Scarth's comments above, although dated, remain pertinent. However, as a former practitioner, I would interpret "taking insufficient account of research findings" differently: it is more the case that most practitioners simply do not have easy access to academic research and, if they did, many would find it written in ways that are not readily understandable or transferrable. Many practising CAL designers operate in the commercial sector and they tend to rely on guidance from others in that arena which is more anecdotal than empirical, or based on one instance, product or project that may not be generalisable. An important personal driver therefore is that, notwithstanding the obvious need for validity, relevance and reliability (Bird and Hammersley, 1996), this research should also be grounded in the real-world, such that practising designers could apply the findings to actual CAL design projects.

Because this research investigates the pedagogic design of Tutorial CAL intended for use by adults, in terms of their attitudes toward, and the effectiveness of, different types and use of media, it seemed clear from the outset that the best approach would involve developing and trialling some actual CAL designs. However, the merits and potential limitations of this needed to be considered from a methodological perspective, together with other quantitative and qualitative options. That discussion forms the starting point for this chapter, leading to the chosen approach, how it was executed, and what practical factors or constraints needed to be taken into account. Methods of data collection are discussed, together with the sample used, the process followed and methods of data analysis.

## Research Approach

There was for many years something of a quantitative–qualitative divide amongst researchers, with Bird (1992) noting:

Until quite recently, researchers have lived in a divided world. One camp looks suspiciously over the fence at the other and shakes its collective head at the follies it sees. (p127)

A good example of this polarisation comes from Kerlinger, who declares, “There’s no such thing as qualitative data. Everything is either 1 or 0” (cited in Miles and Huberman, 1994, p40). Sanders and Liptrot (1993) describe such divisions as an *ideological war zone*, although relations are not quite so strained nowadays, despite the inevitability of many academics and researchers sitting in different philosophical camps. Gummesson (2000) refers to these as *paradigms* which are used to represent people’s “value judgements, norms, standards, frames of reference, perspectives, ideologies, myths, theories, and approved procedures that govern their thinking and action” (p18). Table 4.1 summarises the key features and differences between the positivist and phenomenological paradigms.

Given the opposing nature of many of the factors and terms in Table 4.1, it is not surprising that some researchers holding more polarised views are keen to stand their ground. Howe (1988) refers to this as an *incompatibility thesis*, which holds that qualitative and quantitative research paradigms and methods cannot and should not be mixed. However, there is a growing acceptance that no single method is likely to provide a best fit in terms of research outcomes and validity. Consequently Howe (1988) posits a *compatibility theory* in the belief that:

Although few researchers can be expected to master and pursue both quantitative and qualitative methods, they need at least a rudimentary understanding of what alternative approaches can provide and, accordingly, they should bring a collaborative (rather than paradigm-clique) attitude to research. (p15)

Hussey and Hussey (1997) are amongst many to refer to *triangulation* as the use of different research approaches, methods and techniques in the same study to overcome any potential shortcomings or bias of single methods. According to Patton (2001):



...triangulation strengthens a study by combining methods. This can mean using several kinds of methods or data, including using both quantitative and qualitative approaches. (p247)

**Table 4.1. Features of positivist and phenomenological paradigms**  
(adapted from Mangan et al, 2004, p567)

	<b>Positivist paradigm</b>	<b>Phenomenological paradigm</b>
Beliefs	The world is external and objective Observer is independent Science is value-free	The world is socially constructed and subjective Observer is part of what is observed Science is driven by human interests
Researchers	Look for causality and fundamental laws Reduce phenomena to simple events Formulate hypotheses and then test them	Try to understand what is happening Look at the totality of each situation Develop ideas through induction from data
Methodologies	Cross-sectional studies Experimental studies Models and simulation Surveys Tests	Action research Case studies Ethnography Participative enquiry Hermeneutics
Alternative terms	Quantitative Objectivist Scientific Experimentalist Hypothetico deductive Nomothetic	Qualitative Subjectivist Humanistic/naturalistic Interpretivist/hermeneutic Inductive Ideographic
Data	Numbers, statistics Facts Predictable	Words, imagery, behaviours Meanings, values Dynamic

In a similar vein, Burke Johnson and Onwuegbuzie (2004) define what they term *mixed methods research* as combining, “quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study” (p17). Creswell (2002) notes that, “The situation today is less quantitative *versus* qualitative and more how research practices lie somewhere on a continuum between the two” (p4) and Phillips (2000) also reminds us that, “Neither subjectivity nor objectivity has an exclusive stranglehold on truth” (p123). Finally, Sechrest and Sidani (1995) suggest that, rather than rushing to contrast quantitative and qualitative approaches, it may be helpful to consider their *similarities*, in that both methodologies:

...describe their data, construct explanatory arguments from their data, and speculate about why the outcomes they observed happened as they did. (p78)

Triangulation and mixed methods justify the benefits of a mixed economy but they do not define *what* should be in the mix; this is something that is likely to be driven by both philosophical and situational factors. Table 4.1 includes some but not all possible features which give an overview of the different philosophies; these are now considered in greater detail.

## ***Qualitative research***

Qualitative researchers, as described by Denzin and Lincoln (2000):

...study things in their natural settings, attempting to make sense of or interpret phenomenon in terms of the meanings people bring to them. (p3)

These phenomena are often dynamic and may be related to a particular context or situation and so the outcomes are not necessarily generalisable; there will also almost certainly be subjectivity and so different interpretations and conclusions are possible. Woods (1996) describes most forms of qualitative research as having four main features:

- A focus on natural settings;
- An interest in meanings, perspectives and understandings;
- An emphasis on process; and
- A concern with inductive analysis and grounded theory.

Various research methods are used but many are variations on a general theme of *observation*; this might be through becoming part of the group being studied in their natural setting (ethnography) or by more detached or overt means, including action research, a form of self-reflective enquiry undertaken by practitioners (Kemmis, 1988).

Observation brings the clear advantage that the researcher can experience things as they naturally happen, as opposed to asking people to explain what they *might, would* or *did* do in a particular situation. There are ethical issues with live or recorded observation if it is unknown to the subjects, but these must be weighed against any influence on participants' behaviours (termed *reactivity* by Bird and Hammersley, 1996) when they know they are being

scrutinised. There is also criticism of observational methods by those who object to its subjectivity, such as Ary et al (2006):

Observation in the social sciences is often less objective [than in natural sciences] because it more frequently involves interpretation on the part of the observers. (p17)

But this is exactly what others such as Wiersma and Jurs (2005) see as the method's strength:

...the observer must have the option of interpreting events. Thus, observation extends beyond objective recording of what happens. (p253)

Apart from thinking about the method and the subjects, the potential impact of the *researcher* should also be considered. Delamont (1984) recounts her approach to researching in schools:

When I saw heads [...] the coat was knee-length and very conservative looking, while the dress was mini-length to show the pupils I knew what the fashion was. I would keep the coat on in the head's office, and take it off before I first met the pupils. When observing I tried to dress like the student teachers who were common in all schools. (p25)

Observation (live or recorded) has an obvious role in classroom-focused research (Pennekamp, 2008) but its applicability to CAL is more limited. For studies into program structure, usability and interface design (e.g. Nielsen, 2006; Pomales-García and Liu, 2006b) or readability (Wiley and Rayner, 2000), observation, eye- or mouse-tracking and software logs can all yield valuable data about where users were looking, for how long, and what actions or selections they were making at what times.

However, the emphasis of this research was on how users were *comprehending* the information presented and what they were *learning*; each tutorial was undertaken by solitary learners, mostly in silence, thus making it implausible to accurately gauge the pedagogic effectiveness of the design based on a subjective interpretation of what is being observed. Tracking would make it clear *what* the learner was doing (or where they were looking) at any point in time, but it would not be possible to know *why*, or what sense they were making of what they saw – either as a snapshot or in terms of making connections and constructing a bigger picture. A partial solution to this might be to ask participants to speak their thoughts aloud as they work through the

program, but this would then create an artificial falseness and self-consciousness that is not normally be part of that activity (Holzinger, 2005; Wiersma & Jurs, 2005) as well as adding to cognitive load (Sweller, 1988). It may also be that this 'self-explaining' (Ainsworth & Burcham, 2007) could have a confounding effect by aiding the process of learning. Finally, from an ethical perspective, because I was a higher grade than most participants, I was keen to be able to make a point of saying that nothing they did would be recorded.

Other qualitative methods include various forms of interview with individuals or groups. These are more bounded than observation because they can be steered by both the content and style of questions; there are also opportunities for researchers to ask supplementary questions to probe for specific information (Patton, 2001) and to set context, or ask speculative or hypothetical questions. However, compared to the more quantitative *questionnaire*, subjects can answer 'in their own words' and can volunteer information that may not have been expected, as opposed to trying to pick the best option. According to Collard (2009), "Participants discuss and interpret the world they are experiencing and in doing so, the interview becomes more than a data collecting exercise" (p32).

Depending on the topic, there is a possibility of the interviewer influencing behaviour and responses, particularly if he or she is in any position of authority or perceived influence, causing interviewees to give what they regard as acceptable or expected answers (Robson, 1993; Foster & Parker, 1995). It is also possible for the researcher's own opinions or bias to affect the way in which questions are posed and the interpretation of responses. As it is difficult for most solo researchers to conduct an effective interview *and* take comprehensive notes at the same time, the interview method can also be time-consuming, given the need to conduct the interview, transcribe a recording and then analyse and write it up.

One particular form of interview is the focus group, a collective activity which introduces group dynamics into the process. Collard (2009) believes this helps provide insight into, "...the assumptions, beliefs, ethical perspectives and knowledge of varying cultural agents" (p34) but also cautions against pressure on the researcher to generate conformity and consensus, or to focus more on confirming hypotheses than on exploring interaction between participants.

## **Quantitative research**

Quantitative research appears at first sight to be everything qualitative research is not, with Punch (1998) defining it simply as, "...empirical research where the data is in the form of numbers" (p4). Descriptions such as this encapsulate the objectivity, cause-and-effect, predictability, facts, structure and control that underpin a quantitative approach. Burns (2000) describes the four most important characteristics of this approach as "control, operational definition, replication and hypothesis testing" (p5).

Whereas qualitative research is inductive, embracing discovery and subjectivity, quantitative research is deductive and requires hypotheses or propositions as a starting point. Shields (1998) considers the working hypothesis to formulate, "...a belief about the direction of enquiry but not necessarily its ultimate destination" (p211). This is an important reminder that, despite the objectivity of the quantitative approach, if there is *absolute* certainty of the outcomes at the start, then the value of the research must be questionable.

Quantitative methods are based around the need to collect accurate and coherent data in an objective manner. One means of achieving this is the questionnaire – conducted remotely or in person – where every participant is asked the same questions in the same way and given the same options from which to choose (Scott & Morrison, 2006). This is the nature of a census and of much market research, because questionnaires are cheaper and easier to administer, and provide greater assurance of confidentiality, than interviews (Leary, 1995). They also lead to easy formulation of statistical data, trend analysis and predictions. It is difficult to contest purely factual data such as age, gender, income, etc, (notwithstanding the possibility that some participants may not report accurately) but questions based on opinions can be less reliable, particularly in topic areas where respondents might, "...desire to be seen in a good light" (Robson, 1993, p125) or where it is possible to influence the outcome through prejudicial wording of the question (e.g. "do you agree that we pay too much in taxes?").

Another enduring quantitative method is the experiment, an approach that is well suited to measuring the effectiveness of different CAL designs because

learning effectiveness can be evaluated by means of participants' test results. This represents a 'cause and effect' and, according to Cohen et al (2000):

...the fundamental purpose of experimental design is to impose control over conditions that would otherwise cloud the true effects of independent variables upon the dependent variables. (p126)

For instance, an independent (causal) variable could be the pedagogic design with the dependent (effect) variable being the post-test scores. Bird and Hammersley (1996) believe experiments can offer advantages over more naturalistic approaches, in that the researcher should be able to:

...establish a research situation in which it is possible to manipulate the variable or variables that are the focus of the research in order to trace their effects, and to control at least some of the relevant extraneous variables, so as to rule out competing explanations for those effects. (p59)

Competing explanations for effects can be due to *confounding* (Wiersma & Jurs, 2005) where "...two (or more) independent and/or other variables are confounded if their effects cannot be separated" (p94). In the case of CAL research, confounding means that each design variation needs to be introduced incrementally and its effects measured. The implications of this are that either (i) few variations can be tested, or (ii) large numbers of participants will be required, or (iii) sample sizes will be small. Therefore, despite the natural suitability of experiments for CAL research, some practical compromises would be required.

Experiments have been criticised because 'laboratory conditions' may give rise to *reactivity* whereby, according to Bird and Hammersley (1996), "the artificiality of the situation may lead [participants] to respond in a different way to normal" (p59). Wiersma and Jurs (2005) list minimal *artificiality* as one of eight criteria for a well designed experiment, requiring that "...the experiment is conducted in such a way that the results will apply to the real educational world" (p103).

Whilst these criteria are valid for more traditional social and education research, it is not clear to what extent they might also apply to e-learning. To take learners out of a classroom, or to break up an existing group, form an unfamiliar one, introduce unusual conditions or an observer, are all artificial and may lead to out-of-the-ordinary results. However, CAL (as defined above) is

undertaken by learners, alone, at a computer, often in silence and/or wearing headphones. Hence, its natural state could be argued as being quite clinical and experiment-like, regardless of any research conditions. There *would* be artificiality in that participants would be invited to a different location to study a computer-based topic not of their choosing, but it would not be unnatural to use the computer and nobody would be asked to do or express anything embarrassing or controversial; any effects of falseness would be consistent across all subjects.

The discussion thus far has been about methods of data collection and the process by which it occurs, but the *purpose* of the research should not be overlooked. A stated aim in this case was that it should be useful to practitioners and so, whilst not diminishing the importance of theory and rigour, it is essential that there is also an appropriate degree of *pragmatism*.

## ***Pragmatism***

Shields (1998) describes pragmatism as “the philosophy of common sense” (p197) and this might reasonably be defined in the form of the question, ‘will it work?’ or, more precisely, ‘will X result in Y?’ However, pragmatic research demands that such questions be supported by appropriate theory which is itself backed by empirical data. A more useful question is actually, ‘*why* will X result in Y?’ and, once that is understood, many more supplementary (e.g. ‘what if...?’) questions can be addressed.

Empirical data brings *actuality* that theory alone cannot. Dewey (in Eldridge, 1998) argues that the question is not that of theory *versus* practice, but rather of intelligent practice versus uninformed practice; he noted that his own contribution “...had not been to practicalize intelligence but to intellectualize practice” (p5). Hirst (1993) believes theory should guide practice but not completely determine it, because some judgement is also required. This accords with the notion of *praxis* which, according to Carr and Kemmis (1983), is based on elements of *theory* which inform *practice* and are applied with *judgement*; Evans (2007) described *praxis* as ‘action-full-of-thought’ and ‘thought-full-of-action’.

If practice and data are ‘the proof of the pudding’, then theory gives direction to the research and context for the data, hard or soft, that emerges. Phenomenologists may claim this introduces constraints or even bias from the outset but, as Kaplan (1964) points out:

Theory is useful because it guides the collection of data and their subsequent analysis, by showing us beforehand where the data are to be fitted, and what to make of them when we get them ... Without a theory, however provisional or loosely formulated, there is only a miscellany of observations. (p268)

Burke Johnson and Onwuegbuzie (2004) endorse pragmatism as “...a philosophy that can help to build bridges between conflicting philosophies...” (p17) and Shields (1998) also cautions against taking too narrow a perspective, citing one of the fathers of pragmatism, Dewey (1910): “It does not pay to tether one’s thoughts to the post of use with too short a rope” (p139).

### ***Chosen research approach***

A great deal of ‘conventional wisdom’ about educational research and research approaches is influenced by face-to-face teaching and the interaction between students and teacher, and students and students. CAL is notably different, as stated earlier, because of the fixed nature of the program and the absence of those other players and the effects they bring. This is *not* to suggest all face-to-face teaching research is somehow irrelevant, but that some may need to be applied or interpreted in different ways. There will also be additional factors not applicable to classrooms that need to be considered for distance learning or CAL. For example, it is reasonable to consider how at-ease the learner is using a computer, whereas we seldom consider how at-ease they are receiving taught lessons because those are regarded simply as accepted practice.

Whilst some degree of mixed methods adds balance to research which could otherwise be quite polarised, the propositions tested in this research naturally led towards a predominantly quantitative approach. This was based on experiments comprising a pre-test, CAL tutorial and post-test, coupled with user opinion ratings. Five different CAL designs were constructed and compared with each other and there was no ‘control’ group, even though Barnes et al (2005) would hold that this omission renders this a ‘pseudo-experiment’. However, for these experiments, a control group would have



served no useful purpose, because the designs were being compared with each other, not some norm or standard.

## ***Development of Experiments***

It was necessary for bespoke CAL tutorials to be developed that exactly met the research requirements of topic coverage, coherence, length, media mix and appropriate amounts of core and catalytic content. Whilst one suitable design could probably have been located from a commercial or open source provider, it would not have been possible to source all the design variants in this way and be sure that the learning content was both different *and* equivalent in the ways required. Consequently, as an already experienced CAL designer and developer, I constructed bespoke tutorials based on a single learning topic so that, notwithstanding the different media mix, there was a reasonable basis for comparing the designs with each other. Choice of learning topic was a key early decision and the following criteria were applied:

- It should be a topic that most participants would find stimulating, challenging and with which they could comfortably engage. This meant aiming at around A-level or first-year degree and avoiding specialist areas such as complex maths, sciences or languages;
- The topic should be one in which participants would probably have some interest and limited familiarity, but which also allowed scope for sufficient new learning;
- The topic should lend itself to CAL tuition and multiple-choice questioning that could ideally span three different cognitive levels – from *remember (1)* to *apply (3)* – of the six defined by Anderson & Krathwohl (2001). This is typical of most CAL activities, the reasons for which are described in more detail in Chapter 1;
- The presence or absence of audio or video should not be a critical pedagogic factor. For example, if the chosen topic involved spoken French or musical appreciation, it would then be unreasonable to compare the outcomes of a text-based design against one containing audio; and
- Some suitable video source material had to be available because it was not possible to prepare bespoke footage from scratch.

The eventual choice was a tutorial on the hydrological cycle and the effects of land use on flooding, using the River Severn at Shrewsbury as a case study. The source material comprised four videos from the Open University module U216, "Environment: habitat and conservation". These totalled 26:02 minutes' duration and were obtained under the terms of a creative commons licence from the OU channel of Apple's iTunesU website<sup>12</sup> and edited by me according to the design requirements described below.

## ***Design variants***

Five design variants were developed using the University of Nottingham's *Xerte* authoring tool. These contained three supplementary screens – introduction, a formative question (judging when other floods might have occurred based on historic rainfall data) and a conclusion – which were *identical for all five design variants*, plus the 'teaching' screens described below:

1. **Video-rich.** Five of the eight screens contained documentary-style video with soundtrack (totalling 545 seconds) plus synchronised key-point text (367 words).
2. **Audio-rich.** Five of the eight screens contained the commentary from the video (525 seconds once some silences had been edited out) plus 54 synchronised still frame images (from the video) and key-point text (387 words).
3. **Text-with-imagery 1.** Text was based on the transcript of the audio. Minor edits were made to accommodate essential differences in the media mix, the need to present a logical sequence of similar length screens (one extra screen was introduced) plus necessary grammatical changes. Each screen also displayed two relevant still frame images. (709 words; 24.4% catalytic).
4. **Text-with-imagery 2.** Screen and image count as per Text-with-imagery 1, but with the text edited to about 60% of its original volume whilst still covering all learning outcomes. The aim was to be succinct but not Spartan; short sentences were retained in preference to bullet points. (406 words).
5. **Text-with-imagery 3.** This was a synthesis of Text-with-imagery 1 and 2, resulting in exactly the same word count as 1, but with a much

---

<sup>12</sup> <http://deimos.apple.com/WebObjects/Core.woa/Browse/itunes.open.ac.uk>

higher proportion of catalytic content. Careful editing was required to ensure the learning outcomes were adequately addressed and that the narrative did not appear awkward or contrived. (709 words; 39.5% catalytic).

Proposition 1 (effects of different word count) was tested using designs 3 and 4, the latter being an edited-down derivative of the former and Proposition 2 (effects of different media) was tested with designs 1, 2 and 3 as these were equivalent in terms of their use of a common soundtrack or transcript. Proposition 3 (effects of catalytic content) was tested using designs 3 and 5 which had the same word count but different proportions of catalytic wordage.

Development of the tutorials was an exercise in pragmatism and reverse engineering rather than best-practice learning design. Instead of beginning with a learning needs analysis, a set of learning outcomes and then designing activities, content and assessment, the starting point was the most high-value asset: the existing video. This was edited down from the 26 minutes on iTunesU to around 9 minutes of material which presented a fairly coherent 'beginning-middle-end' narrative; the 9 minutes was then divided into five screen-sized clips. Brief synchronised text was also provided because this would reinforce the key learning points using a persistent medium alongside the relatively transient video. A set of learning outcomes was derived from this material, plus the text-based introductory and concluding screens. Finally, the pre- and post-test questions were devised (described in *Method of Data Collection*, below), based solely on the outcomes and the content (i.e. no assumptions were made about prior knowledge). I was conscious of the potential components of a CAL tutorial described earlier in Chapter 1 (*explanations, examples, problems, rich media, interactivity, exercises, assessment and feedback*) and decided that it was not appropriate to try and shoehorn all of these into one 10-15 minute tutorial. In particular, the subject matter/level did not particularly lend itself to problem-solving or exercises and, apart from the one instance of a formative question, assessment was accomplished through the paper-based pre- and post-tests.

Other design variants were natural derivatives of the video version, as described above. The only calculated design decision that was required related to the number of still images that would be synchronised with the audio. There can be a tendency for designers to use too many in an attempt to hold interest, but

Arguel and Jamet's (2009) 'less is more' findings (described in Chapter 2) were taken into account and 54 were eventually used for the 525 seconds that played across 5 screens (averaging one new image about every 10 seconds). Each image had pertinent content (such as a diagram to support the oral description) and was synchronised with some significant point in the commentary, providing a subtle cue for users to 'pay attention now'.

As stated in Chapter 3, the catalytic aspects of this research will confine themselves just to written text. This is partly due to the limited capacity of a lone researcher but also to try and mitigate for the potential confounding effects of the catalytic properties of different media; this would make it difficult to claim with certainty that any particular outcome was due to the nature of the text or some other aspect of the media used. Catalytic wordage was thus incorporated throughout the five main teaching screens in accordance with the analysis method/rules previously described in the previous chapter.

The interface and navigation design followed standard Windows and browser conventions, although requisite on-screen instructions were also provided; example screen shots of different variants are given in Appendix C and a detailed analysis of the five teaching screens is provided in Appendix D.

### ***Learning outcomes***

Common learning outcomes for all five designs were as follows:

1. Describe the hydrological cycle and the different forms of evaporation therein.
2. Describe the main terms and factors associated with fresh water flooding and flood prevention.
3. Explain the effects of different types of vegetation and land use on fresh water flooding and flood prevention.

### ***Sample***

An opportunity sample ( $N=80$ ) was taken from volunteers amongst the Open University's Walton Hall campus staff (Table 4.2). The participant pool was restricted to degree-qualified 'academic-related' staff, aged 21-60 from

specialisms ranging from media developers to senior managers; no reward for participation was offered. All participants used networked computers as part of their normal work and the nature of this work also meant that none suffered from any visual or aural impairment that would influence the outcomes.

**Table 4.2. Experiment sample details**

Design		Video	Audio	Text 1	Text 2	Text 3
Sample size		16	16	16	16	16
Male		11 (69%)	10 (63%)	10 (63%)	11 (69%)	9 (56%)
Female		5 (31%)	6 (37%)	6 (37%)	5 (31%)	7 (44%)
Degree		13 (81%)	11 (69%)	13 (81%)	11 (69%)	13 (81%)
Post-graduate		3 (19%)	5 (31%)	3 (19%)	5 (31%)	3 (19%)
Age	21-30	0 (0%)	3 (19%)	1 (6%)	0 (0%)	2 (12%)
	31-40	5 (31%)	6 (38%)	5 (31%)	7 (44%)	3 (19%)
	41-50	3 (19%)	3 (19%)	6 (38%)	2 (12%)	6 (38%)
	51-60	8 (50%)	4 (24%)	4 (24%)	7 (44%)	5 (31%)

Subjects were allocated to one of five CAL design groups on a random basis. Each CAL tutorial covered the same basic learning content and so each participant could only take part in one learning sequence. This, the need to use as homogenous a sample as possible and some understandable restrictions on people's time during the working day, placed some limits on availability. Therefore, sample sizes for each tutorial were lower than originally intended ( $n=16$ ) but this was considered a worthwhile compromise in the interests of being able to compare five different designs.

Given that the tutorials were browser-based, it was theoretically possible to engage a much larger sample via the internet. However, this would create a number of additional problems in terms of the controlled environment (I could not guarantee the suitability and settings of the equipment used, the timing of and between activities, the possibility of cheating); the tests and opinion ratings would need to be computerised in some form or other and it would be difficult to replicate the nature and immediacy of feedback which was obtained through the face-to-face post-interview.

Whilst the experiments and tutorials did not represent an 'actual' teaching and learning scenario, the content was taken from a Level 2 Open University degree module and was edited to around level 1, so it would be fair to claim that the

sample was representative of reasonably bright and motivated adults, comfortable with basic IT use, who might study such a distance learning course.

## ***Method of Data Collection***

Experimental data were collected in the following sequence. Copies of tests, questionnaires and are provided in Appendix B.

### *Prior to the CAL tutorial*

- Brief personal details;
- Pre-test to gauge entry knowledge;
- Pre-opinion questionnaire;

### *Immediately after the CAL tutorial*

- Post-opinion questionnaire;
- Post-test to assess knowledge gain due to CAL; followed by
- Semi-structured interview.

## ***Participant testing***

CAL tutorials were preceded by a pre-test and followed by a post-test, each comprising 15 one-from-four multiple-choice questions (MCQs). Reservations about MCQs are well documented (e.g. Case & Swanson, 2001; Norman, 1996) and some, such as the ability to simply guess, are difficult to dispute. However, many other common criticisms can be attributed more to poor question design than to flaws in MCQs per se. As Kirkwood and Price (2008) note:

For many educators, course assessment is something of an afterthought; something to be considered after the content and approach of their teaching has been determined. (p5)

MCQs that have been ‘thrown together’ as something of an afterthought are inevitably not as robust as they might be. This means that, not only is the assessment poor but, according to Kirkwood and Price (2008), so may be the learning:

Students are likely to adopt a surface approach to learning if they anticipate a form of assessment that requires little more than knowledge-based factual recall (e.g. a quiz, multiple-choice questions or a short-answer examination). (p8)

This research required measures of *relative* performance between different tutorial designs rather than absolute performance. It was therefore assumed that, if well-considered questions were developed, any potential deficiencies would be common to all participants and so could be disregarded. To try and discourage ‘question-spotting’, participants were told that the pre- and post-tests would be different; in fact only 5 questions had different wording, although the sequencing of questions and answers was changed throughout.

## **Questionnaires**

Participants were given an eight-page pack (Appendix B) which included an introduction, personal details and consent form and the pre- and post-tests. The questionnaire was designed to:

- Record participant opinions – about the topic, on-screen learning, and expectations of different media types – before the pre-test and tutorial;
- Record participant opinions – about the learning experience and the merits of the content – after the tutorial; and
- Create a short break of a few minutes and change of mental focus between the tutorial and the post-test. This also meant that questionnaire feedback would – as I required – relate to the tutorial rather than a summation of views on the tutorial *and* the post-test.

Each pre-opinion question (e.g. “This could be quite hard/easy for me to learn”) had a corresponding post-opinion question (“I found this easier/harder to understand than I expected”). A Semantic Differential scale was used (Osgood et al, 1957) containing bipolar descriptions such as:

*I am not familiar  
with this topic*



*I am very familiar  
with this topic*

Earlier pilot research had used a 5-point Likert scale, but this had shown that most people avoided the extremes, leaving essentially a ‘yes’, ‘no’ and ‘neutral’; therefore a 10-point scale was used on this occasion to encourage reasonable differentiation in responses. The scale was not annotated as I did not want

participants to be drawn to simply 'pick a number' and research by Couper et al (2006) had shown there to be little difference in ratings assigned by users when given numbered and unnumbered scales of different designs.

Care was taken to ask no more questions than necessary and to ensure they were sufficiently specific but did not use research, educational or psychological jargon, nor should they 'lead' the respondents. Piloting had also shown that one specific question which seemed acceptable when asked in the future tense ('*how appealing do you think X will be?*'), but somehow came across as unusual when asked in the past tense ('*how appealing was X?*') for reasons that can only be attributed to a quirk of the English language. An equivalent term was sought no single word seemed to suffice and so it was decided to use a combination of two measures – "I found this less/more *interesting* than expected" and "I found this less/more *enjoyable* than expected" – for the expression of post-opinions of appeal. A precedent for this approach comes from Harp and Mayer (1997). They also used a 10-point scale and sought to ask students two questions in each of the three categories: emotional interest, cognitive interest and importance, but without using jargon. So, for example, they determined *emotional interest* from participants' combined responses on the two scales: 'boring-interesting' and 'tiresome-entertaining'.

## ***Interviews***

Short interviews of between 5 and 10 minutes were conducted with each participant immediately after the post-test. These were deliberately very informal, essentially taking the form of a 'how was it for you?' chat with an opportunity to ask questions and volunteer information. Test scores were revealed and any comments or feedback noted; we then briefly discussed the opinions expressed before and after the tutorial. I was able to ask open questions, encourage the airing of opinions and suggestions, seek more specific clarification and respond to any queries. Apart from the flexibility of this approach, another significant benefit was that these inputs came 'in their own words' rather than expressed as ticks on predetermined rating scales. As Woods (1996) describes it:

The interview, therefore, is not just a device for gathering information. It is a process of constructing reality to which both parties contribute and by which both parties are affected. (p91)



I made few notes during the interview but briefly logged any salient points immediately afterwards. No attempt was made to 'score' interview responses as the analysis was essentially internal, ongoing and subjective on my part, occurring at the same time and in interaction with the collection of quantitative data (Lacey, 1976).

## ***Procedure***

The experiments were conducted between Sep 2009 and Jun 2010. They took place at the Open University's main campus, in a quiet room adjacent to the participants' normal open-plan place of work. The *Xerte* authoring tool generates Flash files and these were run in Microsoft Internet Explorer 7.0 directly from hard disk (to avoid any connectivity issues) on a medium-specification laptop with a 15.4" screen and (for speech versions) headphones. Sample screen shots from the different design variants are given in Appendix C. All written material was presented in an A4 stapled pack (Appendix B).

The same process was followed for all participants, regardless of the CAL design variant. Each was conducted individually and had a typical duration of around 40 minutes. Participants were invited to read the Introduction on page one of the pack, which I then summarised before asking for their verbal consent to continue. Some pre-opinions were then recorded and the pre-test completed, as described above. Written consent, together with some brief personal data, were obtained *after* the pre-test in order to try and create a short space between it and the tutorial. I would ideally have conducted the pre- and post-tests separately to the tutorial but it was not practicable to schedule this amongst mine and the participants' other work commitments.

Participants then moved on to their designated CAL tutorial, which was open at the first screen (general introduction and instructions); I remained nearby but did not intervene and no operational issues arose during any experiment. The tutorial was immediately followed by the completion of post-opinions, which again created a short space between the tutorial and the post-test that followed. The session concluded with a short informal interview, as described above. Finally, an identical post-test was repeated four weeks after the original.

## ***Ethical considerations***

The nature of stand-alone CAL meant that volunteers were carrying out very clearly defined, visible and individual learning activities. Tutorials were not observed or recorded, no data was stored by the program and participants were reassured that the *design* was being assessed, not them. The purpose and format of the research was explained, together with participants' right to withdraw at any time. Assurances of confidentiality and data protection were given and informed consent obtained from each participant. The experiment introduction and wrap-up interview allowed any concerns to be aired and discussed, and at no time was anyone asked to express views about the Open University, individuals, projects, methods, regulations or anything else that might have been considered contentious or open to misinterpretation.

The research was approved and complied with the principles and conditions laid down by the OU Human Participants and Materials Ethics Committee (HPMEC). It also complied with the provisions of the Data Protection Act and was registered with the OU Data Protection Office; two copies of data were maintained, at work and home, both under password protection.

All tutorial source material was publicly available under the terms of a Creative Commons licence<sup>13</sup>. Additional rainfall data used on an introductory screen were publicly available from the Met Office website<sup>14</sup>. The *Xerte* authoring tool was also Creative Commons.

## ***Research Integrity***

The integrity of any research should be gauged, according to Bird and Hammersley (1996), by its *relevance* to its intended audience, and its *validity*. Golafshani (2003) concurs, but adds *reliability*, citing Joppe's (2000) definitions:

[Reliability is] the extent to which results are consistent over time and an accurate representation of the total population under study is referred to as reliability and if the results of a study can be reproduced under a similar methodology, then the research instrument is considered to be reliable. (p598)

---

<sup>13</sup> <http://creativecommons.org/>

<sup>14</sup> <http://www.metoffice.gov.uk/climate/uk/stationdata/shawburydata.txt>

whereas...

Validity determines whether the research truly measures that which it was intended to measure or how truthful the research results are. (p599)

## **Validity**

I believe reliability and relevance have been demonstrated in the descriptions above and elsewhere in this thesis, although it is of course difficult to provide absolute 'proof'. Validity *can* be assessed more objectively and Cook and Campbell (1979) describe four different types that should be considered:

- **Internal validity** requires that a cause-and-effect relationship is due to true covariation between the variables; i.e. the *effect* was due to the *cause* and not some other reason. In this case, the pre-test – tutorial – post-test format of the experiment provided internal validity, although it cannot confirm *what* in the tutorial caused a particular effect to occur.
- **External validity** requires a relationship between two constructs to be established that is generalisable to different populations, situations and measures. As stated earlier, the CAL tutorials were not in actual use, but they were typical of this topic, level, genre and user-base, and so it is reasonable to assume that the results would be generalisable across similar scenarios. However, they may not apply fully to more exceptional situations such as specialist topics (e.g. music, languages, physical skills, very technical subjects) younger children, less able or qualified adults or those with some disabilities or special needs. Such examples are discussed in more detail in Chapter 6.
- **Construct validity** relates to how well the tests, definitions, measures, etc, used for the research fit the theory(s) or construct(s) being investigated. Test scores are objective and pre to post improvements can reasonably be claimed to reflect learning effectiveness. Learner opinions are more subjective and their relationship to the propositions are discussed in Chapter 6.
- **Statistical conclusion validity** refers to the extent to which we can reliably draw conclusions from the statistical evidence, and so an important requirement is to conduct the appropriate form of statistical tests; these tests and the reasons for selecting them are described below.

Scandura and Williams (2000) note that no single research strategy "...can adequately cover all four aspects of validity, and so researchers need to adopt different strategies to maximise the four kinds of validity" (p1252); therefore, the design and conduct of the experiments, mix of data types and sources, and methods of data analysis collectively do their best to mitigate against any such concerns.

## ***Methods of Data Analysis***

Data analyses were conducted using the Statistical Package for Social Science (SPSS), version 18.0. Parametric techniques were envisaged, although there were certain prerequisites that had to be satisfied before proceeding with this approach (Pallant, 2007, p203):

- Data should be normally distributed;
- Data should exhibit homogeneity of variance;
- Data should be measured on an interval or ratio scale;
- Data should be obtained using a random sample of the population; and
- Measurements should be independent of one another.

Specific data under consideration fell into two categories:

- **Participant pre- and post-tutorial opinions.** These were expressed on identical 10-point semantic differential scales and, according to Boslaugh and Watters (2008), "when a larger number of data points are offered [...] they can be analyzed as interval data" (p203).
- **Pre- and post-test scores relating to the learning outcomes.** Each test comprised 15 multiple-choice questions and data were treated as ratio-level because they indicate a meaningful magnitude of difference using a fixed zero point. (SPSS refers to this as *scale level data*).

For the 3 main propositions, there were 8 sub-propositions and 5 different design groups (Table 4.3). The number of groups per proposition varied between 2 and 5 and, in some cases, separate but related opinions were considered in combination; therefore, a mix of analysis methods was necessary as described below. All of these analyses were preceded by initial checks for outliers, normality, skew and kurtosis. Shapiro-Wilk tests of normality were

used because, according to Greene and D'Oliveira (1999), these are more reliable than Kolmogorov-Smirnov for smaller samples. SPSS also performs Levene's test as part of T-tests and ANOVA to check for homogeneity of variance.

**Table 4.3. Propositions, methods of analysis and design variants**

		CAL Tutorial Design					
Proposition		Video-rich	Audio-rich	Text-with-imagery 1	Text-with-imagery 2	Text-with-imagery 3	N
Word Count	<p><b>P1a.</b> Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will learn less than those receiving a version with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.</p> <p><b>Method:</b> Two by two mixed ANOVA.</p>			✓	✓		32
	<p><b>P1b.</b> Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will rate this as being less interesting and enjoyable than one with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.</p> <p><b>Method:</b> One way between groups MANOVA.</p>			✓	✓		32
Media Mix	<p><b>P2a.</b> When questioned prior to study, learners will presume an audio-rich CAL tutorial to be less appealing than a video-rich CAL tutorial but more appealing than a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way repeated measures ANOVA.</p>	✓	✓	✓	✓	✓	78
	<p><b>P2b.</b> When questioned prior to study, learners will expect to learn more from a video-rich CAL tutorial than from an audio-rich CAL tutorial and to learn least from a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way repeated measures ANOVA.</p>	✓	✓	✓	✓	✓	78
	<p><b>P2c.</b> When questioned after study, learners will report an audio-rich CAL tutorial to be less interesting and enjoyable than a video-rich CAL tutorial but more interesting and enjoyable than a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way between groups MANOVA.</p>	✓	✓	✓			79
	<p><b>P2d.</b> Learners studying an audio-rich CAL tutorial will learn less than those studying a video-rich CAL tutorial but more than those studying a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> Three by two mixed ANOVA.</p>	✓	✓	✓			79
Catalytic Wordage	<p><b>P3a.</b> Learners studying a text-with-imagery-based CAL tutorial will learn more if there is a higher proportion of catalytic text than those who receive a lower proportion of catalytic text.</p> <p><b>Method:</b> One way between groups ANCOVA.</p>			✓		✓	31
	<p><b>P3b.</b> When questioned after study, learners will report that a text-with-imagery-based CAL tutorial with a higher proportion of catalytic text is more interesting and enjoyable than one with a lower proportion of catalytic text.</p> <p><b>Method:</b> One way between groups MANOVA.</p>			✓		✓	31

**P1a** (effect of word count on test scores). An independent samples T-test was conducted initially to check that the difference between the two design groups at pre-test was not significant. That being the case, a two-

by-two mixed analysis of variance (ANOVA) was used to assess the effect of different word counts on test scores, the between subjects variable being design (Text1; Text2) and the within subjects variable was time (pre-test; post-test). The proposition was predicated on there being differences at post-test and this was tested by the interaction effect between the variables.

**P1b** (effect of word count on tutorial appeal). An independent samples T-test was conducted initially to check that the difference between the pre-opinions of the two design groups (Text1; Text2) was not significant. That being so, a one-way between-groups multivariate analysis of variance (MANOVA) was used which allowed separate expressions of interest *and* enjoyment to be considered in combination.

**P2a** (expectations of media appeal). Learner pre-tutorial opinions regarding the predicted *appeal* of three different media types (Video; Audio; Text) were analysed for the whole sample ( $N=78$ ) using a one-way repeated-measures analysis of variance (ANOVA) since this is the most appropriate method of analysis for parametric data from “one group of subjects ... measured on three different questions or items (using the same response scale)” (Pallant, 2007, p251).

**P2b** (expectations of media effectiveness). This proposition was identical to P2a in terms of its parameters and the method of analysis, except that learner opinions of the predicted *effectiveness* of the media were sought.

**P2c** (effect of media on tutorial appeal). A one-way between-groups ANOVA was conducted on *pre-opinions* of appeal expressed by those in the three design groups (Video; Audio; Text1). There being no significant difference between these, post-opinions (expressed in terms of interest and enjoyment) were analysed using a one way between groups MANOVA which allowed these two dependent variables to be considered in combination.

**P2d** (effect of media on tutorial effectiveness). A one-way between-groups ANOVA conducted on pre-test scores from those in three design groups (Video; Audio; Text1). There being no significant difference, a 3-by-2 mixed ANOVA was used to assess the impact of the three designs

on participants' test scores on two occasions (pre-tutorial and post-tutorial).

**P3a** (effect of catalytic content on tutorial effectiveness). An independent samples T-test was conducted to compare pre-test scores for the two groups (Text1; Text3). Because this showed a significant difference, a one way between groups ANCOVA was conducted on the post-test scores, described by Pallant (2007) as involving "one independent, categorical variable (with two or more levels or conditions) [i.e. design], one dependent continuous variable [i.e. post-test], and one or more continuous covariates [i.e. pre-test]" (p294).

**P3b** (effect of catalytic wordage on tutorial appeal). An independent samples T-test was conducted to check that the difference between the pre-opinions of the two design groups (Text1 and Text3) was not significant. That being so, a one-way between-groups MANOVA was used which allowed separate expressions of interest *and* enjoyment to be considered in combination.

## ***Methodological Limitations***

Regardless of the amount of planning and effort invested, all methods will inevitably have limitations resulting from what was done, how it was done and who was involved. One advantage of a quantitative approach (or, for those with more phenomenologist leanings, *disadvantage*) is that there is much less scope for subjective, and possibly inaccurate, interpretation of data.

One potential limitation lies in the relationship between the researcher and the researched. As an Open University manager, I *was* an insider, but not in the traditional research sense whereby I might be critically scrutinising people, methods or operations of which I was normally an integral part. The OU was a convenient home for the research but it could have been conducted at almost any office-based location. Neither the OU, its materials or its staff were being investigated, so there was no risk of me making *taken-for-granted assumptions* (Hockey, 1993) based on any local knowledge or prejudices.

I was conscious of being a higher staff grade than most of the participants and so there was potentially more of a risk of what Foster and Parker (1995) refer to as a 'power imbalance'. This was most likely to manifest itself at three instances: firstly when asking for volunteers, although I made it clear that participation was not obligatory, particularly when their response or body language quickly made it clear that they were busy or not keen. The second might have influenced opinions expressed on questionnaires and the third could have affected the openness of the interviews. However, in both cases, the situation was fairly informal, I was well known to the participants and the discussion related to the CAL designs (the authorship of which was not attributed to me), not the OU, its methods or staff.

Another methodological limitation was the sample size, for the practical reasons discussed earlier (e.g. availability of staff with a suitable profile, each participant could only be taught the topic once and, because of periodic interventions required on my part, no more than two parallel (but staggered) participants could be handled at one time). This resulted in sample sizes for each design variant at or near the minimum acceptable although, for pre-opinions, the whole cohort was used.

One final factor – more of a risk than a limitation – was the tutorial designs themselves; in particular, whether they were as equivalent and effective as they should be and whether the content coverage and ease/difficulty (and that of the pre- and post-tests) was appropriate for the target audience. These and other issues and outcomes will be discussed further in Chapters 6 and 7.

## ***Summary***

A predominantly quantitative, experiment-based approach was the natural method for this research, with data derived from a combination of user opinions, test scores and short interviews. Pragmatism was also an important component as, notwithstanding the need for relevance, validity and reliability, it was important that the research was also seen to have credibility with, and applicability to, practitioners. This was a factor that applied particularly when considering the pedagogic equivalence of five quite different tutorial designs:



- **Video**, comprising documentary-style video with soundtrack and synchronised key-point text.
- **Audio**, comprising the commentary from the video plus synchronised still frame images and key-point text.
- **Text 1**, based on a transcript of the audio (subject to minimal essential editing) and relevant still frame images (709 words; 24% catalytic).
- **Text 2**, using the same screen and image count as Text 1, but with the text edited to about 60% of its original volume (406 words).
- **Text 3**, carefully edited to achieve the same word count as Text 1, but with a higher proportion of catalytic content (709 words; 40% catalytic).

The results and data analysis from experiments are presented in Chapter 5.

# Chapter 5

## Presentation of Results

Statistics are, no doubt, proverbially mendacious. Like Sam Weller's veal pies, one must know who prepared them and be sure of the ingredients.

Lowell (1913, p121)

Eight sub-propositions relating to the design, use and effectiveness of different forms of e-learning content were tested with 80 participants divided into five groups of 16, with each group taking one of 5 differently designed, but pedagogically equivalent CAL tutorials. This chapter presents the results and analysis of the experiments, each of which followed the same format, regardless of which tutorial was given:

- Pre-test
  - 15 one-from-four multiple choice questions (MCQs)*
- Pre-opinion questionnaire
  - 11 opinions, each on a 1-10 semantic differential scale*
- CAL Tutorial
  - Video, audio or one of three text variants, each around 15 minutes*
- Post-opinion questionnaire
  - 7 opinions, each on a 1-10 semantic differential scale*
- Post-test
  - 15 MCQs, slightly different wording and sequence to pre-test*
- Interview
  - Approximately 5 minutes' informal discussion*

An opportunity sample of 80 was taken from volunteers amongst the Open University's Walton Hall campus staff. The participant pool was restricted to degree-qualified 'academic-related' staff, aged 21-60 from specialisms ranging from media developers to senior managers, all of whom used networked computers as part of their normal work.

Sample details are given in Table 5.1 and, of the eighty, 51 (64%) were male; all had a first degree, with 19 (24%) reporting a postgraduate qualification; ages were recorded in 10-year bands ranging from 21-60. Each participant was allocated at random to one of the five CAL tutorial design groups but, because of the small sample sizes per group ( $n=16$ ), it was not practicable to test for homogeneity of participants according to gender, education and age.

**Table 5.1. Experiment sample details**

<i>Design</i>	<i>Video</i>	<i>Audio</i>	<i>Text 1*</i>	<i>Text 2**</i>	<i>Text 3***</i>
Sample size	16	16	16	16	16
Male	11 (69%)	10 (63%)	10 (63%)	11 (69%)	9 (56%)
Female	5 (31%)	6 (37%)	6 (37%)	5 (31%)	7 (44%)
Degree	13 (81%)	11 (69%)	13 (81%)	11 (69%)	13 (81%)
Post-graduate	3 (19%)	5 (31%)	3 (19%)	5 (31%)	3 (19%)
Age 21-30	0 (0%)	3 (19%)	1 (6%)	0 (0%)	2 (12%)
31-40	5 (31%)	6 (38%)	5 (31%)	7 (44%)	3 (19%)
41-50	3 (19%)	3 (19%)	6 (38%)	2 (12%)	6 (38%)
51-60	8 (50%)	4 (24%)	4 (24%)	7 (44%)	5 (31%)

\* Text 1 contained 709 words (24.4% catalytic)

\*\* Text 2 contained 406 words

\*\*\* Text 3 contained 709 words (39.5% catalytic)

## **Results**

All detailed analyses of data were preceded by initial checks for normality, skew and kurtosis to ensure that parametric tests were appropriate, bearing in mind that Pallant (2007) notes that:

In a lot of research (particularly in the social sciences), scores on the dependent variable are not normally distributed. Fortunately, most of the techniques are reasonable 'robust' or tolerant of violations of this assumption. (p204)

### **Initial checks**

Results of checks on all 26 data sets are presented below, followed by a brief discussion and conclusions.

## Pre-test scores

**Video** ( $n=16$ ). The mean (8.38) and 5% trimmed mean (8.47) values were similar and there were no outliers. The Shapiro-Wilk significance (0.05) indicated normality, there was a negative skew (-0.43) indicating values clustered more at the high end and negative kurtosis (-1.34) due to a relatively flat distribution.

**Audio** ( $n=16$ ). The mean (7.94) and 5% trimmed mean (8.04) values were similar and there were no outliers. The Shapiro-Wilk significance (0.48) indicated normality, there was a negligible skew (-0.05) and negative kurtosis (-0.72).

**Text 1** ( $n=16$ ). The mean (8.88) and 5% trimmed mean (8.97) values were similar and there were no outliers. The Shapiro-Wilk significance (0.20) indicated normality, there was a negative skew (-0.46) and negative kurtosis (-0.85).

**Text 2** ( $n=16$ ). The mean (8.50) and 5% trimmed mean (8.56) values were similar and there were no outliers. The Shapiro-Wilk significance (0.31) indicated normality, there was a negative skew (-0.12) and negative kurtosis (-1.01).

**Text 3** ( $n=16$ ). The mean (6.00) and 5% trimmed mean (6.06) values were similar and there were no outliers. The Shapiro-Wilk significance (0.07) indicated normality, there was a negative skew (-0.33) and negative kurtosis (-1.30).

## Post-test scores

**Video** ( $n=16$ ). The mean (14.06) and 5% trimmed mean (14.18) values were similar and so one outlier was retained in the sample. The Shapiro-Wilk significance (0.001) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.1 and are discussed below. There was a negative skew (-1.34), because 8 subjects achieved the maximum score, and positive kurtosis (1.18).

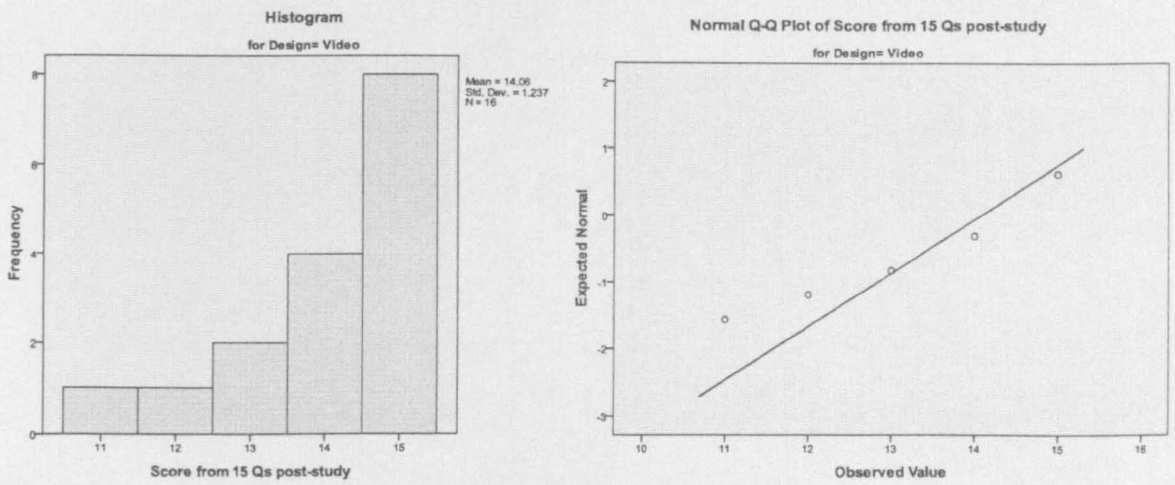


Figure 5.1. Histogram and Q-Q Plot: video post-test scores

**Audio** ( $n=16$ ). The mean (13.44) and 5% trimmed mean (13.54) values were similar and there were no outliers. The Shapiro-Wilk significance (0.05) indicated normality. There was a negative skew (-0.74), because 5 subjects achieved the maximum score, and positive kurtosis (0.31).

**Text 1** ( $n=16$ ). The mean (13.38) and 5% trimmed mean (13.58) values were similar and so two outliers were retained in the sample. The Shapiro-Wilk significance (0.001) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.2 and are discussed below. There was a negative skew (-1.89), because 4 subjects achieved the maximum score, and a large positive kurtosis (3.48) indicating a peaked distribution.

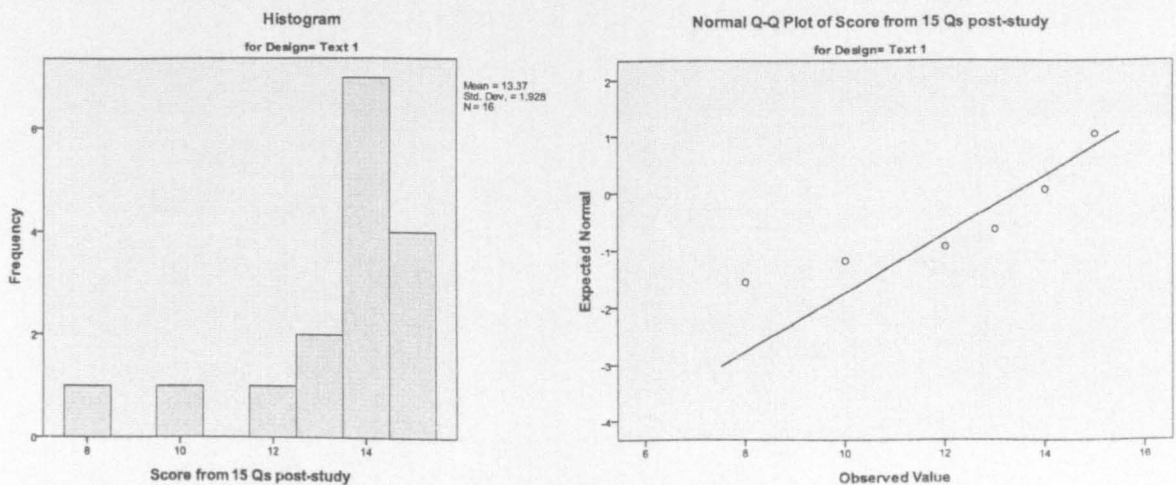
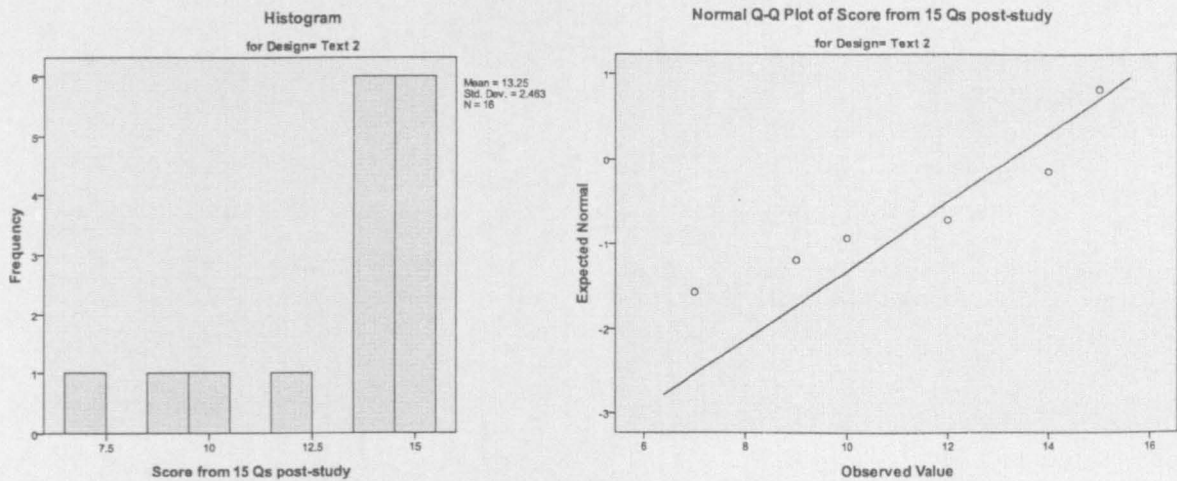


Figure 5.2. Histogram and Q-Q Plot: Text 1 post-test scores

**Text 2** ( $n=16$ ). The mean (13.25) and 5% trimmed mean (13.50) values were similar and so two outliers were retained in the sample. The Shapiro-Wilk significance ( $<0.0005$ ) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.3 and are discussed below. There was a negative skew (-1.64), because 6 subjects achieved the maximum score, and a positive kurtosis (1.77).



**Figure 5.3. Histogram and Q-Q Plot: Text 2 post-test scores**

**Text 3** ( $n=16$ ). The mean (11.38) and 5% trimmed mean (11.42) values were similar and there were no outliers. The Shapiro-Wilk significance (0.74) indicated normality. There was a negative skew (-0.15) and negative kurtosis (-0.85).

### Anticipated appeal

**Video** ( $n=80$ ). The mean (7.09) and 5% trimmed mean (7.18) values were similar and so one outlier was retained. The Shapiro-Wilk significance ( $<0.0005$ ) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.4 and are discussed below. There was a negative skew (-0.94) and positive kurtosis (0.43).

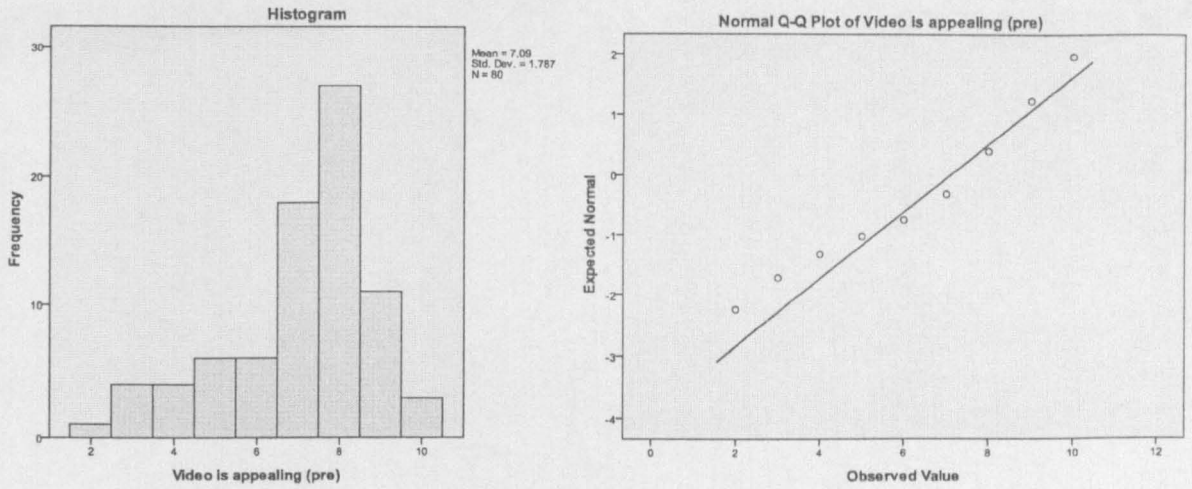


Figure 5.4. Histogram and Q-Q Plot: anticipated appeal of video

**Audio** ( $n=80$ ). The mean (5.75) and 5% trimmed mean (5.79) values were similar and there were no outliers. The Shapiro-Wilk significance (0.003) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.5 and are discussed below. There was a negative skew (-0.34) and negative kurtosis (-0.71).

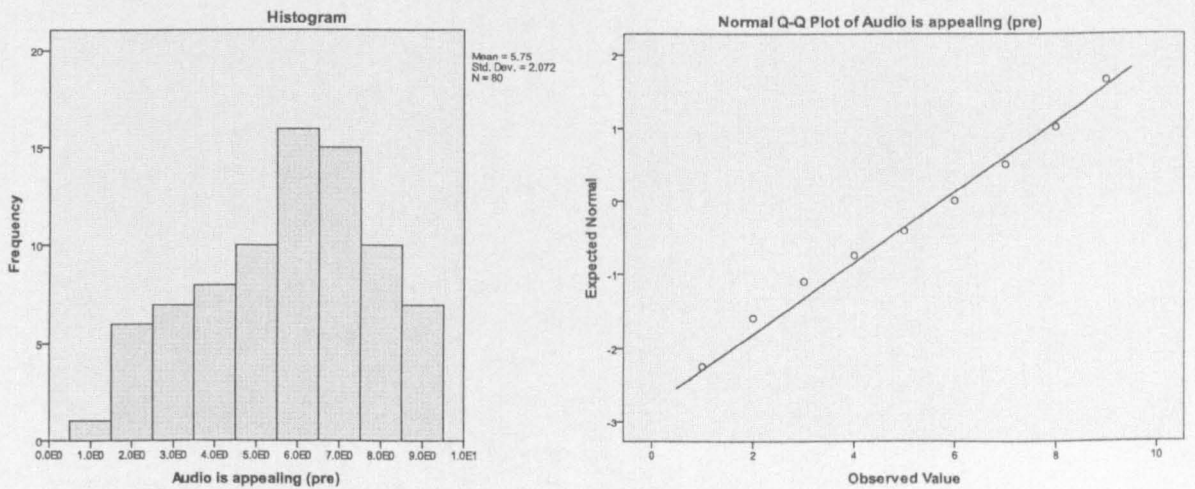


Figure 5.5. Histogram and Q-Q Plot: anticipated appeal of audio

**Text** ( $n=80$ ). The mean (4.43) and 5% trimmed mean (4.36) values were similar and there were no outliers. The Shapiro-Wilk significance (0.005) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.6 and are discussed below. There was a positive skew (0.24) and negative kurtosis (-0.45).

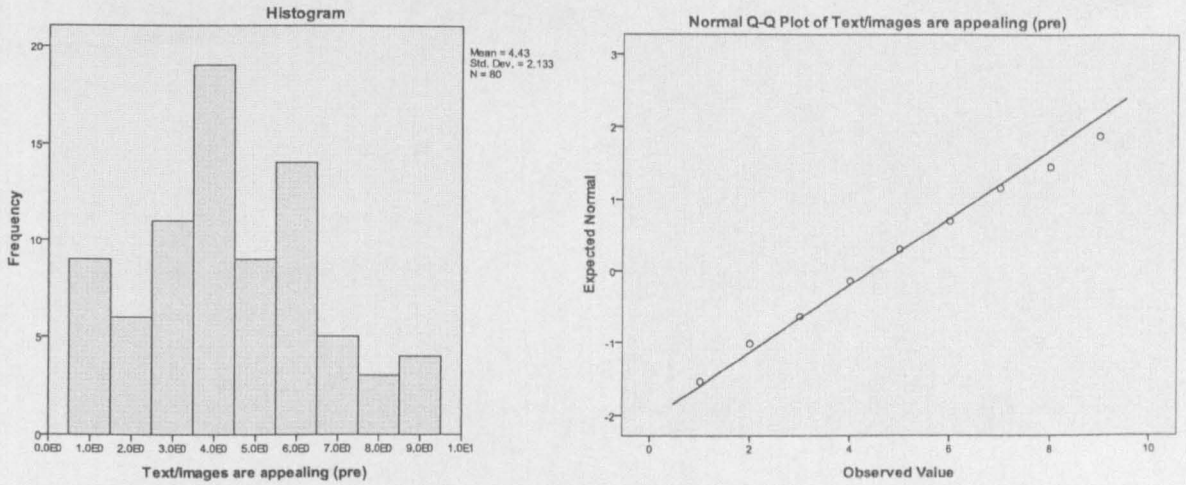


Figure 5.6. Histogram and Q-Q Plot: anticipated appeal of text

### Anticipated effectiveness

**Video** ( $n=80$ ). The mean (6.16) and 5% trimmed mean (6.17) values were similar and there were no outliers. The Shapiro-Wilk significance (0.01) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.7 and are discussed below. There was a negligible skew (0.01) and negative kurtosis (-0.72).

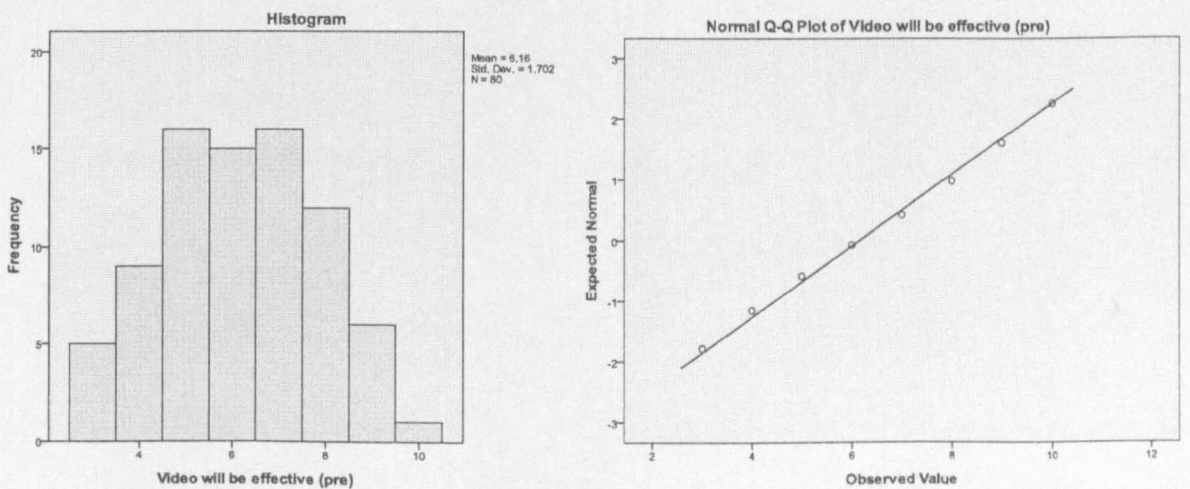
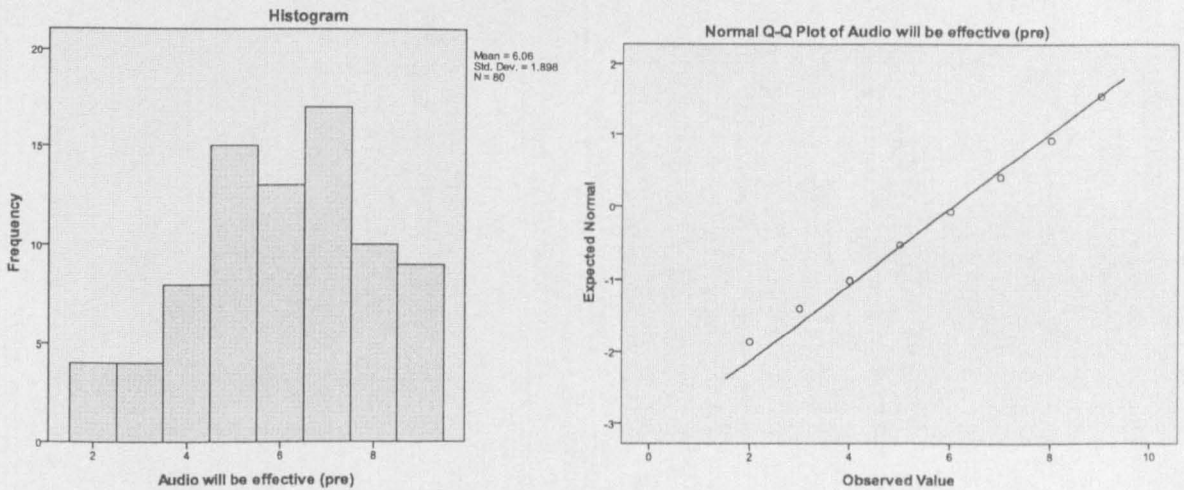


Figure 5.7. Histogram and Q-Q Plot: anticipated effectiveness of video

**Audio** ( $n=80$ ). The mean (6.06) and 5% trimmed mean (6.13) values were similar and there were no outliers. The Shapiro-Wilk significance (0.003) indicated a lack of normality; the normal Q-Q plot and histogram are



shown in Figure 5.8 and are discussed below. There was a negative skew (-0.30) and negative kurtosis (-0.56).



**Figure 5.8. Histogram and Q-Q Plot: anticipated effectiveness of audio**

**Text** ( $n=80$ ). The mean (5.04) and 5% trimmed mean (5.03) values were similar and so one outlier was retained. The Shapiro-Wilk significance (0.06) indicated normality, there was a positive skew (0.15) and negative kurtosis (-0.46).

### Actual interestingness

**Video** ( $n=16$ ). The mean (7.38) and 5% trimmed mean (7.42) values were similar and there were no outliers. The Shapiro-Wilk significance (0.38) indicated normality. There was a negative skew (-0.29) and negative kurtosis (-0.17).

**Audio** ( $n=16$ ). The mean (7.63) and 5% trimmed mean (7.64) values were similar and there were no outliers. The Shapiro-Wilk significance (0.19) indicated normality, there was a negligible skew (-0.06) and negative kurtosis (-1.15).

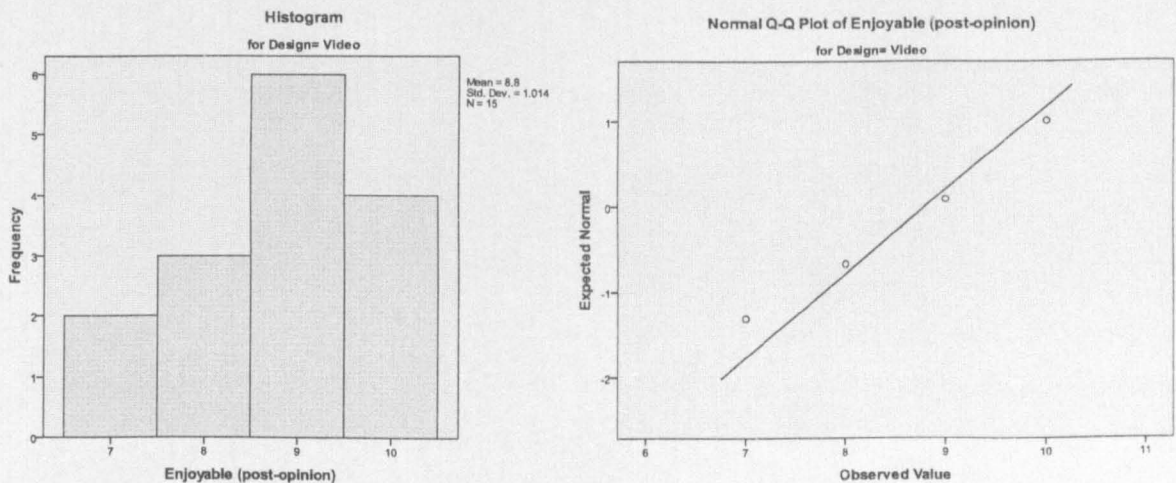
**Text 1** ( $n=16$ ). The mean (6.38) and 5% trimmed mean (6.31) values were similar and there were no outliers. The Shapiro-Wilk significance (0.15) indicated normality, there was a positive skew (0.89) and positive kurtosis (0.60).

**Text 2** ( $n=16$ ). The mean (7.13) and 5% trimmed mean (7.14) values were similar and there were no outliers. The Shapiro-Wilk significance (0.05) indicated normality, there was a negative skew (-0.24) and negative kurtosis (-1.35).

**Text 3** ( $n=15$ ). The difference between the mean (7.06) and 5% trimmed mean (7.24) values meant that one extreme outlier was excluded from the sample. This resulted in improved values of 7.40 and 7.44 respectively, a Shapiro-Wilk significance (0.11) indicating normality, a negative skew (-0.54) and negative kurtosis (-0.43).

### Actual enjoyability

**Video** ( $n=15$ ). The difference between the mean (8.38) and 5% trimmed mean (8.64) values meant that one extreme outlier was excluded from the sample. This resulted in improved values of 8.80 and 8.83 respectively, a Shapiro-Wilk significance (0.04) indicating a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.9 and are discussed below. There was a negative skew (-0.49) and negative kurtosis (-0.60).



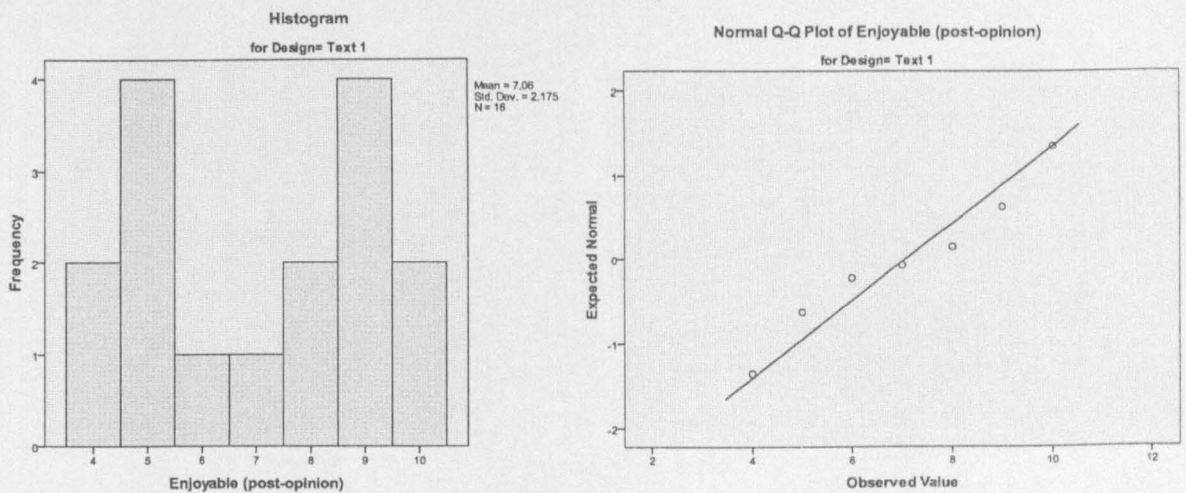
**Figure 5.9. Histogram and Q-Q Plot: actual enjoyability of video**

**Audio** ( $n=16$ ). The mean (8.25) and 5% trimmed mean (8.33) values were similar and so one outlier was retained. The Shapiro-Wilk significance

(0.12) indicated normality, there was a negative skew (-0.81) and positive kurtosis (0.30).

**Text 1** ( $n=16$ ). The mean (7.06) and 5% trimmed mean (7.07) values were similar and there were no outliers. The Shapiro-Wilk significance (0.04) indicated a lack of normality; the normal Q-Q plot and histogram are shown in Figure 5.10 and are discussed below. There was a negative skew (-0.09) and negative kurtosis (-1.93).

**Text 2** ( $n=16$ ). The mean (7.31) and 5% trimmed mean (7.40) values were similar and so one outlier was retained. The Shapiro-Wilk significance (0.16) indicated normality, there was a negative skew (-0.87) and positive kurtosis (1.93).



**Figure 5.10. Histogram and Q-Q Plot: actual enjoyability of Text 1**

**Text 3** ( $n=15$ ). The difference between the mean (7.44) and 5% trimmed mean (7.60) values meant that one extreme outlier was excluded from the sample. This resulted in improved values of 7.80 and 7.78 respectively, a Shapiro-Wilk significance (0.16) indicating normality, a negligible skew (0.06) and positive kurtosis (0.22).

### Assessment of initial checks

Two extreme outliers were removed: #15 from the Video sample and #88 from the Text 3 sample, reducing each of those samples to 15 ( $N=78$ ).

Of the 26 sets of data checked, 16 were shown by Shapiro-Wilk significance to exhibit normality and a further 5 were assessed as acceptable by visual examination of their histograms and Q-Q plots (Figures 5.4, 5.5, 5.6, 5.7 and 5.8 above). This left five sets whose normality was questionable:

- Video post-test scores (Figure 5.1)
- Text 1 post-test scores (Figure 5.2)
- Text 2 post-test scores (Figure 5.3)
- Actual enjoyability of video (Figure 5.9)
- Actual enjoyability of Text 1 (Figure 5.10)

Consideration was given to changing to nonparametric testing for analyses involving these data, but this was rejected for the following reasons:

- Motulsky (2009) advises that, “with small data sets, normality tests have little power” (p349);
- Greene and D'Oliveira (1999) suggest that, “parametric tests are fairly ‘robust’, even when their assumptions are broken” (p99);
- Dubcovsky (2011) notes that normality is, “the least influential assumption on the F test” (p8.1) but that homogeneity of variance (checked by SPSS using Levine’s test as part of a t-test or ANOVA) is much more critical.
- The power of nonparametric tests is low with small data sets (Motulsky, 2009);
- For some of the parametric tests required (e.g. two way ANOVA, MANOVA and ANCOVA) there are no nonparametric equivalents; and
- Most critically, Motulsky (2009) advises that:

When analysing a series of experiments, all should be analysed the same way. Therefore, results from normality tests should not be used to choose a test for each particular experiment. (p179)

Therefore, it was decided to proceed with parametric testing for all samples because 21 of the 26 sets of data exhibited normality.

## **Proposition 1a**

*Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will learn less than those receiving a version with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.*

**Table 5.2. Effect of different word counts on test scores (P1a)**

	n	Pre-Test		Post-Test	
		Mean*	SD	Mean*	SD
<b>Text 1</b> (709 words)	16	8.88	3.90	13.38	1.93
<b>Text 2</b> (406 words)	16	8.50	3.01	13.25	2.46

\*Scores from 15 questions (N=32)

The effect of different amounts of wordage on achieving the learning outcomes was measured by post-test scores for those in the Text 1 (709 words) and Text 2 (406 words) groups. An independent samples T-test was conducted to compare pre-test scores for the two groups, showing no significant difference between these (Table 5.2):  $t(30) = 0.31$ ;  $p = 0.76$  (two-tailed). The magnitude of the differences in the means (mean difference = 0.38; 95%CL: -2.14 to 2.89) was very small (eta squared = 0.003) and so a 2 by 2 mixed ANOVA was conducted which would assess the impact of the two designs on participants' test scores on two occasions (pre-tutorial and post-tutorial). There was a substantial main effect for time (Wilks' Lambda = 0.29;  $F(1,30) = 72.56$ ;  $p < 0.0005$ , partial eta squared = 0.71) with both groups showing a knowledge gain as would be expected. However, the main effect comparing the two designs was not significant ( $F(1, 30) = 0.081$ ;  $p = 0.79$ , partial eta squared = 0.003) and there was no significant interaction between design and time (Wilks' Lambda = 0.99;  $F(1,30) = 0.053$ ;  $p = 0.82$ ; partial eta squared = 0.002), indicating no significant difference in the post-test scores for higher- and lower-wordage CAL tutorials in this experiment and so P1a was not upheld.

## **Proposition 1b**

*Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will rate this as being less interesting and enjoyable than one with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.*

**Table 5.3. Effect of word count on tutorial appeal (P1b)**

	N	Interest		Enjoyment	
		Mean*	SD	Mean*	SD
<b>Text 1</b> (709 words)	16	6.37	1.59	7.06	2.18
<b>Text 2</b> (406 words)	16	7.12	1.46	7.31	1.66

\*Post-opinions, each rated on a 10-point scale (N=32)

Learner post-opinions regarding the *actual appeal* of tutorials having different word counts were sought from those in the Text 1 and Text 2 groups, who were asked to rate their respective tutorial in terms of interest and enjoyment, these having been adjudged to equate to appeal, as discussed in Chapter 4 (p96). An independent samples T-test was conducted to compare pre-tutorial expectations of appeal for the two groups. There was no significant difference between the expectations of students who studied the tutorial using Text 1 ( $M = 5.13$ ,  $SD = 2.03$ ) and those who studied the tutorial using Text 2 ( $M = 4.44$ ,  $SD = 2.42$ );  $t(30) = 0.87$ ;  $p = 0.39$  (two-tailed) and the magnitude of the differences in the means was small (mean difference = 0.69; 95% CL: -0.93 to 2.30; eta squared = 0.025). In terms of post-opinions (Table 5.3), the Text 2 tutorial, with fewer words, appeared to be rated as more interesting and more enjoyable than the Text 1 tutorial. However, a one way between groups MANOVA was used to consider both of these related dependent variables in combination and there was no overall statistical significance ( $F(2, 29) = 1.09$ ;  $p = 0.070$ ; Wilks' Lambda = 0.93; partial eta squared = 0.07) and so P1b was not upheld.

## **Proposition 2a**

*When questioned prior to study, learners will presume an audio-rich CAL tutorial to be less appealing than a video-rich CAL tutorial but more appealing than a text-with-imagery-based CAL tutorial.*

**Table 5.4. Expectations of media appeal (P2a)**

	N	Mean*	SD
A computer-based tutorial containing mostly <b>video</b> with narration is likely to be very appealing	78	7.09	1.80
A computer-based tutorial containing mostly <b>audio</b> and images is likely to be very appealing	78	5.86	1.98
A computer-based tutorial containing mostly <b>text</b> and images is likely to be very appealing	78	4.47	2.12

\*Pre-opinions, each rated on a 10-point scale (N=78)

Learner opinions regarding the expected *appeal* of three different media types were sought from the entire sample ( $N = 78$ ) prior to study (Table 5.4), showing that expectations of video were greater than those for audio and expectations of audio were greater than those for text. A one way repeated measures analysis of variance (ANOVA) was then conducted. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated (approx Chi-Square = 6.49;  $p = 0.039$ ) and so a Greenhouse-Geisser correction was used, showing the effect to be significant ( $F(1.85, 154) = 56.98$ ;  $p < 0.0005$ ; partial eta squared = 0.43). Post hoc pairwise comparisons using the Bonferroni correction gave  $p < 0.0005$  in all cases and so P2a was upheld.

### **Proposition 2b**

*When questioned prior to study, learners will expect to learn more from a video-rich CAL tutorial than from an audio-rich CAL tutorial and to learn least from a text-with-imagery-based CAL tutorial.*

**Table 5.5. Expectations of media effectiveness (P2b)**

	<i>N</i>	<i>Mean*</i>	<i>SD</i>
A computer-based tutorial containing mostly <b>video</b> with narration is likely to be very effective at teaching	78	6.17	1.69
A computer-based tutorial containing mostly <b>audio</b> and images is likely to be very effective at teaching	78	6.13	1.88
A computer-based tutorial containing mostly <b>text</b> and images is likely to be very effective at teaching	78	5.10	2.12

*\*Pre-opinions, each rated on a 10-point scale (N=78)*

Learner opinions regarding the expected *effectiveness* of three different media types were also sought from the entire sample prior to study, showing that expectations of video were marginally greater than those for audio and expectations of audio were greater than those for text (Table 5.5).

A one way repeated measures analysis of variance (ANOVA) was conducted. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated (approx Chi-Square = 12.05;  $p = 0.002$ ) and so a Greenhouse-Geisser correction was used, showing the effect to be significant ( $F(1.74, 154) = 10.96$ ;  $p < 0.0005$ ; partial eta squared = 0.13). Post hoc pairwise comparisons using Bonferroni correction showed the differences between video and text

( $p = 0.002$ ) and audio and text ( $p < 0.0005$ ) to be significant; however, the difference between video and audio was not significant ( $p = 1.00$ ) and so P2b was only partially upheld.

### **Proposition 2c**

*When questioned after study, learners will report an audio-rich CAL tutorial to be less interesting and enjoyable than a video-rich CAL tutorial but more interesting and enjoyable than a text-with-imagery-based CAL tutorial.*

**Table 5.6. Effect of media on tutorial appeal (P2c)**

	<i>n</i>	<i>Interest</i>		<i>Enjoyment</i>	
		<i>Mean*</i>	<i>SD</i>	<i>Mean*</i>	<i>SD</i>
<b>Video</b> with narration	15	7.60	1.45	8.80	1.01
<b>Audio</b> and images	16	7.62	1.71	8.25	1.44
<b>Text</b> and images	16	6.37	1.59	7.06	2.18

*\*Post-opinions, each rated on a 10-point scale (N=47)*

A one way between groups ANOVA was conducted to compare pre-tutorial opinions regarding appeal for the three designs ( $M_{\text{Video}} = 6.40$ ,  $SD = 2.13$ ;  $M_{\text{Audio}} = 6.31$ ,  $SD = 1.92$ ;  $M_{\text{Text}} = 5.13$ ,  $SD = 2.03$ ). It should be noted that this analysis was different to that conducted for P2a ( $N=78$ ) because only 47 opinions were considered in P2c so that any pre-to-post analysis would be valid. The ANOVA showed there to be no significant difference between pre-opinions ( $F(2,44) = 1.95$ ;  $p = 0.15$ ; eta squared = 0.08). Post-opinions were recorded for actual interest and enjoyment (Table 5.6), showing audio to be considered marginally more interesting than video (contrary to the proposition) with both more interesting than text; video was rated as more enjoyable than audio and audio more enjoyable than text. A one way between groups MANOVA was used to consider both of these related dependent variables in combination for the three designs, but there was no overall statistical significance ( $F(4, 86) = 2.57$ ;  $p = 0.044$ ; Wilks' Lambda = 0.80; partial eta squared = 0.11) and so P2c was not upheld.



## **Proposition 2d**

*Learners studying an audio-rich CAL tutorial will learn less than those studying a video-rich CAL tutorial but more than those studying a text-with-imagery-based CAL tutorial.*

**Table 5.7. Effect of media on tutorial effectiveness (P2d)**

	<i>n</i>	<i>Pre-Test</i>		<i>Post-Test</i>	
		<i>Mean*</i>	<i>SD</i>	<i>Mean*</i>	<i>SD</i>
<b>Video</b> with narration	15	8.73	2.74	14.13	1.25
<b>Audio</b> and images	16	7.94	3.62	13.44	1.46
<b>Text</b> and images	16	8.87	3.90	13.38	1.93

*\*Scores from 15 questions (N=32)*

A one way between groups ANOVA was conducted to compare pre-test scores for the three groups (Table 5.7), showing that the differences were not significant ( $F(2,44) = 0.34$ ;  $p = 0.72$ ). The absence of a significant difference allowed a 3 by 2 mixed ANOVA to be used to assess the impact of the three designs on participants' test scores on two occasions (pre-tutorial and post-tutorial). There was a substantial main effect for time (Wilks' Lambda = 0.23;  $F(1,44) = 147.30$ ;  $p < 0.0005$ ; partial eta squared = 0.77) with all groups showing an increase in scores from pre- to post-test as would be expected. However, the main effect comparing the three designs was not significant ( $F(2,44) = 0.423$ ;  $p = 0.66$ ; partial eta squared = 0.019) and there was no significant interaction between design and time (Wilks' Lambda = 0.98;  $F(2,44) = 0.573$ ;  $p = 0.57$ ; partial eta squared = 0.025), suggesting no significant difference in the learning effectiveness of the CAL tutorials using different main media in this experiment and so P2d was not upheld.

## **Proposition 3a**

*Learners studying a text-with-imagery-based CAL tutorial will learn more if there is a higher proportion of catalytic text than those who receive a lower proportion of catalytic text.*

**Table 5.8. Effect of catalytic wordage on tutorial effectiveness (P3a)**

	<i>n</i>	<i>Pre-Test</i>		<i>Post-test</i>		<i>Adjusted Post-test</i>
		<i>Mean*</i>	<i>SD</i>	<i>Mean*</i>	<i>SD</i>	
<b>Text 1</b> (24.4% catalytic)	16	8.88	3.90	13.38	1.93	12.85
<b>Text 3</b> (39.5% catalytic)	15	5.93	2.63	11.40	2.50	11.96

*\*Scores from 15 questions (N=31)*

An independent samples T-test was conducted to compare pre-test scores for the low and high catalytic content groups (Table 5.8) and this showed a significant difference:  $t(29) = 2.45$ ;  $p = 0.021$  (two-tailed). Therefore, a one way between groups ANCOVA was conducted on the post-test scores using pre-test score as the covariate. This showed there to be no significant difference between designs ( $F(1,28) = 1.43$ ;  $p = 0.242$ ; partial eta squared = 0.049). There was a significant relationship between pre- and post-test scores ( $p = 0.001$ ), as indicated by a large partial eta squared value of 0.31. Adjusted mean post-test scores were contrary to the proposition and P3a was not upheld.

### **Proposition 3b**

*When questioned after study, learners will report that a text-with-imagery-based CAL tutorial with a higher proportion of catalytic text is more interesting and enjoyable than one with a lower proportion of catalytic text.*

**Table 5.9. Effect of catalytic wordage on tutorial appeal (P3b)**

	<i>n</i>	<i>Interest</i>		<i>Enjoyment</i>	
		<i>Mean*</i>	<i>SD</i>	<i>Mean*</i>	<i>SD</i>
<b>Text 1</b> (24.4% catalytic)	16	6.37	1.59	7.06	2.18
<b>Text 3</b> (39.5% catalytic)	15	7.40	1.06	7.80	1.08

*\*Post-opinions, each rated on a 10-point scale (N=31)*

An independent samples T-test was conducted to compare pre-tutorial opinions regarding the appeal of a text-with-images tutorial. There was no significant difference between Text 1 ( $M = 5.13$ ,  $SD = 2.03$ ) and Text 3 ( $M = 4.07$ ,  $SD = 2.12$ ):  $t(29) = 1.42$ ;  $p = 0.17$  (two-tailed) and the magnitude of the differences in the means was moderate (mean difference = 1.06; 95%CL: -0.47 to 2.58; eta squared = 0.065). Post-opinions were recorded for actual interest and enjoyment (Table 5.9), appearing to show Text 3, with a greater proportion of catalytic content, to be more interesting and more enjoyable than Text 1.

However, a one way between groups MANOVA was used to consider both of these related dependent variables in combination and there was no overall statistical significance ( $F(2, 28) = 2.17$ ;  $p = 0.133$ ; Wilks' Lambda = 0.89; partial eta squared = 0.13) and so P3b was not upheld.

## ***Interview Outcomes***

Interviews were deliberately quite short and informal, with only noteworthy comments being transcribed afterwards. Such comments fell into three general categories:

**Preferences for other media.** Preferences were expressed for a book (Participant 65, Text 2 group; #76, Text 2; #82, Text 3), audio (#75, Text 2), face-to-face tuition (#6, Video) and a blend of media (#82, Text 3; #91, Text 3). No preferences were expressed for video from those in the non-video groups.

**Use of summaries.** “I liked the summaries at the end of each clip” (Participant 10, Video group). “I found myself not paying full attention to the video but relying more on the text summaries” (#14, Video). “I ended up ignoring the pictures and just listening to the commentary and reading the text” (#22, Audio). “I ended up ignoring the text and just listening to the commentary and watching the pictures” (#24, Audio). “I liked the key points being displayed as the audio played” (#30, Audio).

**Other design suggestions.** “Each clip was quite long; smaller chunks would have been preferable” (Participant 3, Video group). “Formative questions in between screens would have helped” (#5, Video). “A glossary at the end would have been useful” (#66, Text 2). “Greater interactivity would have been good” (#93, Text 3).

## ***Summary of Results***

Of the eight sub-propositions, only one was upheld (P2a, relating to the expected appeal of video-, audio- and text-based tutorials) and one was *partially* upheld (P2b, relating to the expected effectiveness of video-, audio- and text-

based tutorials, two of these three pairwise relationships being significantly different). Comparisons between designs did produce the expected ranking order in three further cases (P1b, P2d and P3b), but the differences were not statistically significant and so support for these propositions cannot be claimed. A summary of the research propositions, designs, sample sizes, methods of analysis and outcomes is given in Table 6.1 at the start of the next chapter, after which possible reasons for, and implications of, these outcomes are discussed.

# **Chapter 6**

## **Discussion of Results and Method**

Research should be more learner focused than technology focused. With little integrated e-learning theory available, it is not surprising that e-learning design appears to have been driven more by advances in technology and 'bells and whistles' than by our long-standing history of cognitive science research and learning theory.

DeRouin et al (2005, p935)

This research investigated principles and guidelines that could support practitioners in the design of stand-alone CAL tutorials, building on evidence and issues arising from existing research literature. Three main propositions were tested; these related to (i) the effects of varying the amount of wordage in a CAL tutorial, (ii) the use of different media, and (iii) the effects of catalytic content, a new phenomenon introduced in Chapter 3.

The primary method of research was a series of experiments in which 80 volunteer OU staff members each took one of five differently designed CAL tutorials addressing the hydrological cycle and the effects of land use on flooding. Each tutorial lasted for approximately 15 minutes (with the introduction, tests, questionnaire and interview extending the overall session to about 40 minutes) and all addressed the same three learning outcomes:

1. Describe the hydrological cycle and the different forms of evaporation therein.
2. Describe the main terms and factors associated with fresh water flooding and flood prevention.
3. Explain the effects of different types of vegetation and land use on fresh water flooding and flood prevention.

The experiment outcomes are summarised in Table 6.1 and this also indicates where data from 2 extreme outliers were excluded from the sample as described in the 'initial checks' section of chapter 5. Of the eight sub-propositions tested, only one was fully upheld (P2a, relating to the expected appeal of video-, audio-

and text-based tutorials) and one was partially upheld (P2b, relating to the expected effectiveness of video-, audio- and text-based tutorials). All outcomes are discussed below, together with an assessment of the validity of the data and the methodology used.

**Table 6.1. Summary of propositions, analysis and outcomes**

Proposition		Video	Audio	Text-1	Text-2	Text-3	N
Word Count	<p><b>P1a.</b> Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will learn less than those receiving a version with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.</p> <p><b>Method:</b> Two by two mixed ANOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P1a was <i>not</i> upheld.</p>			✓	✓		32
	<p><b>P1b.</b> Learners studying a text-with-imagery-based CAL tutorial with a higher number of words will rate this as being less interesting and enjoyable than one with a lower number of words, subject to there being sufficient explanatory material to address the learning outcomes.</p> <p><b>Method:</b> One way between groups MANOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P1b was <i>not</i> upheld.</p>			✓	✓		32
Media Mix	<p><b>P2a.</b> When questioned prior to study, learners will presume an audio-rich CAL tutorial to be less appealing than a video-rich CAL tutorial but more appealing than a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way repeated measures ANOVA.</p> <p><b>Outcome:</b> Differences were as proposed and were statistically significant. P2a was upheld.</p>	✓	✓	✓	✓	✓	78
	<p><b>P2b.</b> When questioned prior to study, learners will expect to learn more from a video-rich CAL tutorial than from an audio-rich CAL tutorial and to learn least from a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way repeated measures ANOVA.</p> <p><b>Outcome:</b> Differences were as proposed but only two of three pairwise comparisons were statistically significant. P2b was <i>partially</i> upheld.</p>	✓	✓	✓	✓	✓	78
	<p><b>P2c.</b> When questioned after study, learners will report an audio-rich CAL tutorial to be less interesting and enjoyable than a video-rich CAL tutorial but more interesting and enjoyable than a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> One way between groups MANOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P2c was <i>not</i> upheld.</p>	✓	✓	✓			79
	<p><b>P2d.</b> Learners studying an audio-rich CAL tutorial will learn less than those studying a video-rich CAL tutorial but more than those studying a text-with-imagery-based CAL tutorial.</p> <p><b>Method:</b> Three by two mixed ANOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P2d was <i>not</i> upheld.</p>	✓	✓	✓			79
Catalytic Wordage	<p><b>P3a.</b> Learners studying a text-with-imagery-based CAL tutorial will learn more if there is a higher proportion of catalytic text than those who receive a lower proportion of catalytic text.</p> <p><b>Method:</b> One way between groups ANCOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P3a was <i>not</i> upheld.</p>			✓		✓	31
	<p><b>P3b.</b> When questioned after study, learners will report that a text-with-imagery-based CAL tutorial with a higher proportion of catalytic text is more interesting and enjoyable than one with a lower proportion of catalytic text.</p> <p><b>Method:</b> One way between groups MANOVA.</p> <p><b>Outcome:</b> Differences not statistically significant; P3b was <i>not</i> upheld.</p>			✓		✓	31

## ***Discussion of Results***

The results given in Table 6.1 are discussed below, organised according to the three main proposition groupings.

### ***Effects of wordage on appeal and effectiveness***

Proposition 1 was predicated on the principle of 'less is more' that forms the basis for Mayer's (2001) third *coherence principle*: "Student learning is improved when unneeded words are removed from a multimedia presentation" (p128). This was discussed in Chapter 2 and is widely supported by other researchers including Guan (2009), Mayer et al (1996) and Mayer, Heiser and Lonn (2001). Results from the Text 1 group of participants ( $n=16$ ) were compared with those from a Text 2 group of the same size. However, whereas the Text 1 design was based on the video/audio narration (709 words), Text 2 used the same screen count and images but the script was edited down to around 60% of its original volume (406 words). This wordage was a pragmatic outcome rather than a target, in that I simply acted as a CAL designer would if instructed to try and make a tutorial 'less wordy', but whilst retaining an appropriate level of explanation. I imposed one further constraint: to retain some form of narrative rather than aiming for a shorter précis or bulleted list, because that would have changed the nature of the shorter tutorial into one in which learners were expected to 'fill in the gaps'.

Proposition 1a related to actual test scores and the results of a 2 by 2 mixed ANOVA conducted on participants' test scores, pre- and post-tutorial, showed the more verbose design to have yielded slightly higher scores. This was contrary to my expectations although it may be accounted for by the fact that, by chance, 3 of this sample of 16 had a geography degree. However, the differences between designs were found not to be significant and so P1a was not upheld.

Proposition 1b considered user post-opinions about how interesting and enjoyable each tutorial had been and, although learners appeared to favour the more succinct design, a one way between groups MANOVA showed the differences not to be significant and so P1b was not upheld.

It was expected that a reduced amount of wordage would be less daunting or onerous for learners and so engender a more favourable disposition to learn, hence my expectation that the more succinct design would be rated more highly for appeal as well as producing better test scores. There is evidence that many learners are simply not comfortable reading large amounts of narrative text from screen (e.g. Liew et al, 2000; Liu, 2005, 2006; Mangan, 2008). Learners are also known to be more inclined to skim-read web pages (Morkes & Nielsen, 1997; Nielsen, 2008) and less inclined to scroll down beyond text that is immediately visible on a screen (Nielsen, 1997, 2010). It is possible that any difference in the number of words across the five screens was too small to make a significant difference to the learners' performance or attitudes.

In terms of just how much text is the right amount, Nielsen's (2000) suggestion of aiming for 50% of what one would write for print is of questionable value, particularly as he does not support this with empirical evidence. Mayer et al (1996) offer some isolated evidence (described in Chapter 2) but, in his PhD research into long online text passages, Parsons (2001) bemoans the absence of substantiated advice on appropriate wordage and so it may be that the good judgement of the CAL designer is the final, albeit potentially inconsistent, arbiter.

### ***Effects of different media on appeal and effectiveness***

There is a great deal of anecdotal support for the notion that using richer media such as video somehow results in 'better' learning. This is backed up by occasional unsubstantiated pieces in trade journals, but relatively little empirical data. Some evidential support for the greater appeal of richer media does manifest itself in terms of improved enjoyment (Maag, 2004), interest (Silvia, 2006) and a more favourable attitude to learning (DeVaney, 2009).

Proposition 2a related to expectations before participants were assigned to a tutorial group, and so data from the whole sample ( $N=78$ ) could be used. They were asked to rate the expected appeal of a video-rich CAL tutorial compared to audio-rich and text-with-imagery. A one way repeated measures ANOVA showed that a video-rich tutorial was perceived as the most appealing, followed by audio and then text, and that these differences were significant; hence P2a was upheld.



Proposition 2c then considered whether these expectations were confirmed by opinions expressed by participants in the Video, Audio and Text 1 groups ( $N=47$ ) *after* they had completed their respective tutorial. Indications were that audio-rich tutorials were favoured over video-rich tutorials, contrary to my expectations. However, a one way between groups MANOVA showed the differences not to be statistically significant and so P2c was not upheld. One possible explanation for the expressed preference for audio relates to the participants, many of whom had some background in the design and development of educational media. Audio with synchronised images and key point text is used quite widely in the commercial training sector but much less so by the Open University and so there may have been a small novelty effect which positively influenced some post-opinions.

It also became clear from the post-experiment interviews that, whilst opinions of different media may hold true for relatively short tutorials, there was less enthusiasm to study a large amount of material using just one medium. Indeed, despite there being a 1-10 scale, 29 of the 78 participants did not rate any of the 3 media options greater than '7' for effectiveness or appeal. When questioned afterwards about what option might have caused them to score an 8, 9 or 10, preferences were mostly expressed for a blend of media and/or interactivity (although 3 stated a preference for a printed book and 2 favoured face-to-face tuition).

Whereas Propositions 2a and 2c tested for expectations and post-opinions relating to predicted and actual *appeal* of different media, Propositions 2b and 2d mirrored these, but for tested for predicted and actual *effectiveness* of a video-rich tutorial compared to audio-rich and text-with-imagery. Evidential support for the greater effectiveness of richer media comes in the form of improved concentration (Liu et al, 2009), recall (Al-Seghayer, 2001; Guan, 2009; Ginns, 2005; Timmerman & Kruepke, 2006) and comprehension (Seufert et al, 2009). In addition, Mayer's (2001) *multimedia* and *modality* principles are also based on benefits offered by appropriate combinations of richer media.

A one way repeated measures ANOVA conducted on data from all 78 participants showed that they did expect video and audio to be more effective than text by a significant amount, but the differences between video and audio were not significant and so P2b was only partially upheld. The participants'

background in educational media may have influenced this and, in subsequent interviews, opinions were expressed that, whilst watching video seemed to be appealing, it may not be as effective in engaging them cognitively as would other forms or combinations of media. The results of such a survey with a less well-informed sample might well be different.

Proposition 2d tested for actual effectiveness of the three different designs, as measured by post-test scores for the 47 participants. The ranking order of the three designs was as proposed; however, a three by two mixed ANOVA showed that the differences were not statistically significant and so P2d was not upheld.

These outcomes tend to accord with previous research findings (e.g. Bader & Strickman-Stein, 2003) that one single medium rarely provides the best solution, particularly for larger amounts of content. It is also known that certain media are better suited to particular topics, scenarios and activities (Bhowmick et al, 2007; Danaher et al, 2005; Liu et al, 2009) and, as Mayer (2001) reminds us: “Learning outcomes depend on the quality of the instructional method rather than on the medium per se” (p71).

### ***Effects of catalytic text on appeal and effectiveness***

Despite general support for the learning benefits of richer media from this and previous research, there are also examples of where media made little difference or did not have the expected effect (e.g. DeVaney, 2009; McLaughlin et al, 2007; Richard, 2006). Researchers have suggested various reasons for this, notably relating to learners’ attributes such as prior knowledge (Lowe, 2004); working memory capacity (Sanchez & Wiley, 2006); visuospatial ability (Mayer and Sims, 1994); age (Mykityshyn et al, 2002); gender (Price, 2006); job status (Lu & Chiou, 2010); and reading ability (Schnotz & Bannert, 2003).

Where attention is turned to the learning material itself, *seductive details* (Garner et al, 1992) have been investigated by some as a potential cause of discrepancies between expectations of particular designs and outcomes, but that theory was shown to be encumbered by its own flaws and inconsistencies (Lehman et al, 2007; Thalheimer, 2004; Towler et al, 2008). However, it became apparent that some content may be included that is relatively neutral in terms of interest or relevance, but whose value lies in acting as a *catalyst* to

*support the process of learning* by somehow introducing, contextualising, exemplifying, substantiating or reinforcing important content.

The original Text 1 tutorial, whose content was based on the transcript of the video and audio variants, contained 709 words, of which 24.4% were judged to be catalytic. A Text 3 variant was then derived which deliberately had the same word count, screen count and imagery, but was edited to achieve a higher proportion of catalytic content (39.5%). This was not a specific target but an outcome of consciously rewriting with catalytic properties in mind. However, it required careful design because of the need to retain sufficient primary learning material to address the learning outcomes whilst increasing the catalytic elements. Apart from achieving the required word count, great care was also taken to create a variant that did not appear contrived and which would present an effective, flowing narrative.

The effect of catalytic content on post-test scores was tested by Proposition 3a and, contrary to my expectations, there was no evidence of a catalytic effect, either in the raw scores or those adjusted for a covariant effect of pre-test scores. The outcomes of Proposition 3b, testing for post-opinions of interest and enjoyment, were closer to those expected, but a one way between groups MANOVA showed the differences between designs not to be significant when both of these related dependent variables were analysed in combination.

Neither P3a nor P3b were upheld, and possible reasons for this, together with a proposal for some more robust and conclusive future research, will be discussed in Chapter 7.

## ***Validity of Data***

Different forms of data validity were discussed in Chapter 4 and are now briefly revisited here in the light of the actual outcomes described above.

### ***Internal validity***

Internal validity requires that any effect or outcome is due to the attributed cause and not some other reason. In this case, the design and conduct of the

experiments appeared to control matters quite well, with pre-test/opinions followed by the tutorial and then post-test/opinions and an interview, all conducted in isolation and within a 45 minute window. Whilst it was likely that participants would have differing levels of prior knowledge, intelligence, interest, learning style/preferences and aptitude, they were assigned to design groups at random and pre-test scores and pre-opinions were checked for significant differences before the method of subsequent analysis was chosen.

One possible threat to internal validity concerns the catalytic effects of content in those three tutorials that were *not* designed to investigate catalytic content. For example, the reason P1 was not upheld may have had little to do with the respective word counts of Text 1 and Text 2, but more to do with their different (by accident) proportions of catalytic text. Whilst there is undoubtedly some truth in this, I would question just how reasonable it is to mitigate against every potential competing effect, particularly whilst maintaining 5 tutorials that were effective and yet did not come across to learners as being contrived. Furthermore, even if the textual content were somehow equalised, it would be reasonable to speculate over the catalytic properties of the images, or the video, or we might move on to consider the readability of the text, or its coherence, or the use of passive sentences. In short, any attempt to compare different multimedia designs as I have done could suffer from this potential problem; it is not practicable to eradicate every competing factor, although this is discussed further in Chapter 7.

One final indeterminate factor that should be considered is whether, within the tutorial itself, participants operated, thought and learned in the way I had expected (and, indeed, whether this mattered). In this respect, contrasting interview comments made it clear that this was not always the case; for example:

*“I found myself not paying full attention to the video but relying more on the text summaries.”*

(Participant 14, Video group)

*“I ended up ignoring the pictures and just listening to the commentary and reading the text.”*

(Participant 22, Audio group)

*"I found I was ignoring the text. I mostly watched the pictures and listened to the commentary."*

(Participant 24, Audio group)

This might suggest that there was unnecessary *redundancy* in some of the tutorial designs, as per Mayer's original (2001) principle. However, he (in Mayer & Johnson, 2008) conducted subsequent experiments that caused him to revise his earlier definition of redundancy in acknowledgement of the potential benefits of using synchronised key-point text, something he had previously discouraged when used with audio and imagery.

The worst-case interpretation of these interview comments is that participants learned more from the key points than the carefully prepared media clips. However, a more positive view is that the key points – as intended – supported and reinforced the clips, thus enhancing the learning. Since there were very few telling or adverse comments and no other obvious methodological flaws, the presumption is that internal validity *was* demonstrated.

### ***Statistical conclusion validity***

Statistical conclusion validity refers to the extent to which we can reliably draw conclusions from the statistical evidence. Initial checks of the data sets showed 5 of 26 *not* to exhibit normality and so a decision had to be made regarding parametric or nonparametric testing. I opted for the former because this offers greater power, because normality (i.e. parametric) was the majority case and because Motulsky (2009) advises that all experiments in a series should be analysed in the same way. However, he also notes that using a parametric test with non-normal data can lead to P values that are inaccurate, cautioning:

A high P value does not prove the null hypothesis. Deciding not to reject the null hypothesis is not the same as believing that the null hypothesis is definitely not true. The absence of evidence is not evidence of absence. (p141)

Pallant (2007) also comments on sample size and significance:

When you have a study where the group size is small (e.g.  $n=20$ ), you need to be aware of the possibility that a non-significant result may be due to insufficient power. (p205)

I am not belatedly offering evidence to invalidate the lack of support for six of the propositions (or to call into question the two that *were* supported) but it is certainly worth future researchers noting that test method and sample size can impact on the shape and robustness of the outcomes. This is something that will be returned to in Chapter 7.

## ***External validity***

External validity relates to the generalisability of outcomes to different populations, situations and measures. As discussed in Chapter 4, although the tutorials were specifically developed by me rather than being taken from an extant teaching and learning scenario, they *were* typical of this topic and genre for post-18 learners studying at introductory degree level and it is reasonable to claim that the results would be generalisable across similar adult learning scenarios. The general design styles were also similar to those used widely in the commercial sector.

Exclusions to these claims of generalisability would be more exceptional situations such as specialist topics (e.g. music, foreign languages, physical skills, very technical, visual or tactile subjects), younger children, less able or less well-qualified adults or those with some disabilities, disaffections, learning difficulties or other specialist needs. IT skills is another pertinent factor but, since CAL is by definition a computer-based medium, these outcomes are no more or less generalisable to computing novices than any other e-learning study or solution. It would be reasonable to claim that the outcomes could apply to those who felt comfortable with basic computing and, as discussed in Chapter 1, this is an increasing sector of the population.

Finally, there is also the matter of professional judgement by designers in the application of specific media to different learning scenarios. For example, although it was shown that video was expected to be effective and appealing for this experimental topic, this does not necessarily mean it would be successful if used to teach calculus, any more than we might expect text to be the best medium to teach plastering.

## ***Evaluation of the Methodology***

Two main methodological factors are considered below: the format of the experiments and the design of the tutorials themselves.

### ***Experiment format***

The inherent nature of stand-alone CAL, plus the format of the experiments meant there was quite tight control over the conduct and consistency of the experiments. There were, however, logistical constraints because of my own availability and the fact that each individual session took around 40 minutes, and no more than two could reasonably be conducted in parallel because of the periodic interventions required on my part. My participants were also employed full time and their availability not only limited the sample size but also the timing and conduct of the experiments, meaning I had little option but to complete each participant in a single session.

Each tutorial was undertaken no more than 5 minutes after completion of the pre-test (a short gap was created by inserting some of the pre-opinion and personal details gathering at this point), and the post-test was similarly completed a few minutes after the tutorial, again punctuated by some opinion gathering. This was not ideal – although it was a flaw that was at least consistent across all designs and participants – but it was enforced by these time/availability constraints. However, the result was high post-test scores, with 24 of the 80 participants scoring a maximum of 15 and mean score across the whole sample of 13.10. I attribute this mainly to closeness of the tutorial and post-test, but there is also likely to be an effect due to the closeness of the pre-test and tutorial, meaning participants were likely to remember pre-test questions and could ‘spot’ for likely post-test questions.

Another factor likely to have influenced these high scores was the topic coverage within the tutorials, since this would have affected the amount of potential learning and the range and difficulty of possible test questions. Tutorial designs will therefore now be considered in general terms and some more specific issues relating to catalytic content will be addressed in Chapter 7.

## ***Tutorial designs***

As discussed in Chapter 4, the choice of topic would always be a compromise, given the mixed background and interests of the sample (despite them being fairly homogenous in terms of occupation and level of education) and the need to pick something that was fairly engaging and familiar whilst not being too well understood or specialised. I was also constrained by needing to use some suitable existing video material, but from which reasonable non-video variants could be adapted, and yet there should be no particular pedagogic advantage or disadvantage when using these different media (e.g. music or spoken language would clearly not work well in a text-only design).

The eventual topic *did* prove acceptable, but it is now clear that it would have benefited from a greater depth of coverage, with many participants commenting afterwards that they were expecting something a little more challenging. The source material came from a second level degree module, but the elements used were probably closer to around A-level. This was due mainly to the content of the video clips and, on reflection, my decision to include video on each of the teaching screens of that design variant imposed an unnecessary restriction; in fact it would have been better to have interspersed a couple of extra text/graphic screens which would have given scope to include some deeper theoretical material.

The choice of material also affected the tutorial in terms of its positioning against Anderson and Krathwohl's (2001) cognitive scale (discussed in Chapter 2). This meant that both the teaching and resulting assessment were rooted around *remember* and *understand* rather than deeper levels such as *apply* or *analyse*. This in turn limited the variety, complexity and total number of questions in each test, as well as the capacity to include very much variety or randomisation. Given that the participants were all degree-qualified, there was less scope for learning improvement than had been envisaged. The mean score across all 80 pre-tests was 7.94 out of 15 (i.e. an entry knowledge of >50%) and there was potential – given the use of multiple choice questions – to achieve this as much by applying general intelligence as any prior knowledge of the subject. This might have been something that should have manifested itself during piloting but unfortunately, due to the shortage of participants, I had chosen to



pilot the *method*, but with different subject matter, in order that I would be able to reuse those people in the main research.

### ***Rigour versus realism***

I was critical in Chapter 2 of other researchers who, in their quest for rigour and equitable designs, had introduced degrees of falseness that practising designers would not contemplate in the 'real world'. Examples included using print to teach students how to apply mortar to a wall (Donkor, 2010), adding graphics and animations to text-based material without a corresponding change or reduction in the amount of text (Lewalter, 2003) and teaching the internal operation of a mechanical pump without the use of imagery (Mayer and Anderson, 1991).

In each of those examples, different design variants were being compared and the falseness was introduced in pursuit of equity. In the case of Lewalter (2003) for example, identical text (approx 2,100 words) was used for groups who received text-only, text & graphics and text & animation versions. In reality (and subject of course to the topic), the design choices would more likely be the 2,100 words of original text alone, but reduced to perhaps around 1,500 words with graphics, or maybe 1,000 with animation: the salient point being that pictures (still or moving) often *can* be worth several hundred words.

Right and wrong here may amount to a matter of perspective. To the purist researcher, Lewalter's (2003) approach is rigorous and allows like to be accurately compared with like. To a practitioner, the experiment is flawed because, although mathematically equal, the three designs do not have any real world equivalence and so cannot usefully be compared. The exact volumes of text are not critical because other media are being introduced which have their own pedagogic features and benefits. What is important is that the sum of the component parts in all designs deliver equivalent (and effective) learning.

This research and its experiments were based on that practical approach and, whilst five tutorial designs were used that offered different visual, aural, cognitive and affective experiences, I am confident they were pedagogically equivalent and all satisfied the fundamental learning design trinity of learning outcomes, matched to activities/content, matched to assessment.

## **Chapter 7**

### **Conclusions and Recommendations**

Remember that things do not often proceed in a neat, linear fashion. Most people experience research as a zig-zag process of continual review and re-adjustment.

McNiff (2002, p12)

The aim of this research was to investigate the pedagogic design of Tutorial CAL intended for use by adults, notably in respect of the anticipated and actual appeal and effectiveness of different types of content. The need for this was reinforced when the literature review began to reveal (i) some inconsistencies in the evidence to support widely accepted multimedia design principles and (ii) a paucity of empirical data to support anecdotal evidence about user expectations of different media. This defined what would become Proposition 2 but, thereafter, the 'zig-zag process' referred to above by McNiff (2002) came to bear.

Firstly, in terms of purely textual content, the period of this research saw the emergence and rapid evolution of a new device: the e-book reader – such as Amazon's *Kindle* or Apple's more versatile *iPad* tablet – and this coincided with a sustained drive by the Open University to move more learning material from printed to online formats. Such material might be viewed on a standard desktop or laptop computer screen, a handheld tablet (typically with a 10" or 7" display) or a mobile phone (typically a 3" or 3½" screen). It was clear that different design considerations would need to be taken into account as evidence showed that not all users were comfortable reading narrative text from screen. However, in terms of the 'right' amount of wordage for e-learning, there was much anecdotal support and general advice but few, if any, evidence-based design guidelines and so this led to what became Proposition 1.

Proposition 3 then emerged after a period of many months contemplating the exact nature, purpose and value of different categories of content. Seductive details had been investigated in some detail as a potential means of explaining the exceptions and inconsistencies that were evident in some of the research into Mayer's (2001, 2005) multimedia design principles; however, seductive

details themselves were shown to be not entirely proven and consistent. The notion of ‘catalytic content’ – that which is included for the purpose of *supporting the process of learning* – emerged about two thirds of the way through the research. As this seemed to represent a credible but previously unexplored theory, Proposition 3 was added to the experiments at this relatively late stage, meaning that some compromises, such as reduced group sizes, were made to accommodate all three research strands.

The description above may appear to represent quite a haphazard approach, although it accords with McNiff’s (2002) quoted observations at the start of this chapter. Further insight comes from Shields (1998) who suggests that such deviations are not unusual and that the main value of a conceptual framework or working research hypothesis – in this case, based on Mayer’s (2001) multimedia design principles – is to formulate, “a belief about the direction of enquiry but not necessarily its ultimate destination” (p211).

The propositions were tested by experiment in which 80 participants each took one of five differently designed CAL tutorials addressing the hydrological cycle and the effects of land use on flooding. The results of these are summarised in Table 6.1 (p120), showing that, of the eight sub-propositions tested, one (P2a) was fully upheld and another (P2b) was partially upheld. Both of these relate to user expectations in terms of the appeal and effectiveness of video-, audio- and text-based CAL tutorials. Six of the eight sub-propositions were *not* upheld but that does not mean that the case for these rests here. Further research is required to confirm or otherwise these results and this is addressed below.

Overall, the experiments ran smoothly: there were 80 willing and interested volunteers and, other than the compressed timescales, there were no notable procedural shortcomings. I designed and produced five CAL tutorials that were pedagogically sound, visually appealing, technically robust and achieved a ‘real world’ equivalence despite their different content, thus satisfying the design requirements without the narrative appearing contrived or laboured. There were also no apparent disconnects between the learning outcomes, activities/content and assessment. The data analysis suggests that the outcomes were valid and, where propositions were upheld, that these were generalisable to similar adult e-learning scenarios. A further positive outcome was the emergence of a theoretical foundation for *catalytic content*, and so it

can reasonably be claimed that the research has usefully contributed to professional knowledge.

Nevertheless, the discussion in Chapter 6 showed that there were some limitations regarding the research design. Even so, recommendations can be made for CAL designers and practitioners as well as suggestions for further research which, in particular, investigates the principle of catalytic content.

## ***Recommendations***

Based on the outcomes previously described, recommendations are set out below for designers, practitioners and future researchers.

### ***Recommendations for CAL designers and practitioners***

Two recommendations are offered to CAL designers and practitioners as a result of propositions upheld by this research:

1. Audio-rich CAL tutorials should be used in preference to text-with-imagery CAL tutorials (where the subject matter, budget and technical restrictions allow) because users expect these to be more appealing and effective, and so will be more predisposed to learn from these.
2. Video-rich CAL tutorials should be used in preference to audio-rich CAL tutorials (where the subject matter, budget and technical restrictions allow) because users expect these to be more appealing and so will be more predisposed to learn from these.

### ***Recommendations for further research***

The following recommendations are offered to future researchers based on the discussion of outcomes and method in Chapter 6:

### **Recommendations pertinent to all propositions**

3. Further research needs to be undertaken into the design and use of different forms of CAL content using larger sample sizes with appropriate homogeneity;
4. The tutorial content, outcomes and assessment should be designed to provide more comprehensive, varied and cognitively challenging topic coverage;
5. The wording of opinion questions needs to be clear and unambiguous;
6. Where a mix of information formats are presented (e.g. video/imagery, audio and key point text), interview questions should ascertain which, and how, learners used these;
7. Greater elapsed time is required between pre-testing, the tutorial and post-testing;
8. Time on task should be recorded and analysed for each participant.

### **Recommendations pertinent to catalytic content**

9. Catalytic content, being a new phenomenon, requires a more precise definition and guidelines for classification that take into account related theory and research findings;
10. The catalytic properties of non-textual content and their possible confounding effects need to be considered and taken into account in future CAL tutorial designs; and
11. Tutorial designs should be subject to inter-rater reliability checks to ensure the categorisation of different types of content is as accurate and consistent as possible.

Catalytic content is a new phenomenon that emerged from this research and, whilst not upheld by experimental outcomes, its description and rationale are plausible. There appears to be evidential support from other complementary theories, and yet it is also distinctive enough to stand apart from these. A description therefore follows which amplifies the recommendations above by discussing and specifying a more rigorous future study into catalytic content, building on the lessons learned from this investigation.

## ***Specification for a Future Study***

It is suggested that some of the reasons for a lack of support for catalytic content and its effect in CAL tutorials relate to general methodological shortcomings described previously. However, others are more specific to the way in which catalytic content was conceived and tested, and so this specification will propose improvements in the following areas:

- A redefinition of what is meant by *catalytic content*;
- The design of experiments, particularly in terms of the tutorial content, assessment and opinion gathering;
- The sample used and procedure followed; and
- The analysis of outcomes.

### ***Redefining catalytic content***

Catalytic content was defined and described in some detail in the latter part of Chapter 3 but it would be prudent to reconsider this, particularly in respect of:

1. **Categorisation.** Revisit the definition and scope of catalytic content. For example, does or should it apply to intro/outtro pages or just the main 'teaching' screens, and how much catalytic content is the 'right' amount? Provide more comprehensive guidelines and criteria for categorisation, supported by some additional examples.
2. **Theory.** Conduct more detailed and wider research to clarify the relationship between catalytic content and other theories such as Ausubel's (1963) Subsumption Theory and Reigeluth's (1979) Elaboration Theory. It should be made clear where and to what extent any aspects of these theories overlap or stand apart from catalytic content and whether they support or conflict with each other.
3. **Non-Textual Catalytic Content.** Begin to consider defining the catalytic properties of non-textual media, particularly where this may be partial or subjective (such as for images) or where a medium is multi-faceted (such as video, which contains visual and aural components, both of which may be catalytic to different extents and at different times). There may also be categories of imagery that should be considered beyond *primary* and *catalytic*, such as *seductive*,

*cosmetic* or others. To address this may seem an ambitious step that should be perhaps deferred until a textual effect is demonstrated; however, it is necessary in order to complete Step 4.

4. **Confounding.** From an experimental research perspective, consider the potential confounding effects of a tutorial containing multiple media with different (and perhaps difficult to quantify) catalytic properties. Apart from clearly delineated cases (e.g. video versus text tutorials), there may be some quite subtle considerations. For example, two designs with the same graphics but different proportions of catalytic text might cause the learner's engagement with those graphics to differ in each case. A further consideration would be whether rewriting for different proportions of catalytic content changes other significant properties, such as the readability, coherence or comprehension, as discussed below.

## ***Experiment design***

### **Design of tutorials**

For the research reported here, the choice of tutorial topic was basically sound, being one that was topical, of general interest and suited to a CAL tutorial approach. No specialist knowledge or aptitude was required but one could reasonably expect that most participants would find some aspects familiar. The choice of tutorial content in future research will *not* be constrained by the need to use suitable existing video as a starting point. This will enable a longer and more challenging topic to be presented and assessed, perhaps increasing from the current 700 words, 5 screens, 10 minutes to something in the region of 1000-1200 words, 8 screens, 15 minutes. This may still be shorter than a typical CAL tutorial, but the combination of extra length and depth – pushing it further up Anderson and Krathwohl's (2001) cognitive taxonomic scale, from just *Remember* and *Understand* to *Apply*, *Analyse* and possibly *Evaluate* – will result in an appropriate difference from the current tutorials.

A paper-based design could be used initially to demonstrate the concept of catalytic content. This need have no images and there would be no need to collect pre- and post-opinions; it would be technically much simpler and quicker to produce and would take any user apprehensions about computing

out of the equation. If successful, a second round of experiments could then be conducted using four different CAL tutorials: high/low catalytic text only and high/low catalytic text sharing common static images. In addition to greater duration and cognitive depth, these should also include greater pedagogic variety and interactivity. For instance, formative questions, problems and feedback could be included, as found in many typical CAL tutorials. However, the basic structure would need to remain sequential so it could be shown that all participants had been presented with the same material in the same order.

### **Writing of content**

To impose too many constraints on the writing and analysis of content could either set an impossible task or result in a suite of tutorials that are so obviously contrived that the learner experience suffers. However, there is good reason to look beyond simply having tutorials of the same length and different proportions of catalytic content. For instance, some consideration should be given as to what else has been affected by the rewriting process that might lead to confounding effects. Amongst the many factors that could be considered, the following seem the most evident:

- **Readability.** For example, Kincaid's (1975) number of syllables per word and words per sentence;
- **Text coherence.** The extent to which explanations are made explicit (Ainsworth & Burcham, 2007);
- **Comprehension devices.** These might include the number of passive sentences and average number of verb phrases (Schwarm & Ostendorf, 2005), lexical cohesion (Halliday & Hasan, 1976) and the use of pronouns (Krahmer & Heune, 2002); and
- **Rhetorical devices.** Appearance and layout can help organise and signpost content, so these might include the use of headings, bullets, paragraph indentations and icons (Goldman, 1997).

There may also be instances where definitions from different sources overlap or conflict with each other. For example, content defined here as *catalytic* may occur at the start of a narrative by way of 'scene-setting', but this might be argued to fall into the category of an *advance organiser* (Ausubel, 1963) or an *epitome* (Reigeluth, 1979) and so warrant some different analytical consideration.



Therefore, as described in Steps 1 and 2 of 'redefinition' (above), some initial work needs to be done to better understand the relationship between these various concepts, where aspects overlap or stand apart, where they support or conflict and to what extent they should be considered in the writing and analysis process. Finally, once this is agreed and implemented, a second and ideally a third reviewer should be used to provide inter-rater reliability checks (Keyton et al, 2004) to assure the most accurate and consistent analysis and categorisation of content.

### **Assessment**

The mean score across all 80 pre-tests was 7.94 out of 15 (i.e. an entry knowledge of >50%) and the mean post-test score was 13.10, with 24 of the 80 participants scoring a maximum of 15. This clearly shows that the questions were insufficiently challenging to discriminate between the true knowledge and understanding of different participants. This was firstly due to the tutorial content itself and the limited scope it allowed for challenging and varied questions; second was the timing of the tests and tutorials and the problems caused by their closeness. Both of these are considered below.

As discussed earlier, the tutorial content and assessment needs to be more cognitively challenging, requiring learners not just to *Remember* and *Understand*, but also to *Apply*, *Analyse* and possibly *Evaluate* (Anderson & Krathwohl, 2001) what they are learning. A new topic and/or treatment and duration will allow for this and may also give scope for some different styles of question and greater variation between the pre- and post-test (whilst still ensuring reasonable equivalence).

Finally, if there is support for the existence of catalytic content then the introduction of a deferred post-test to assess retention of knowledge would be valuable.

### **Learner opinions**

For the initial (paper-based) experiment, it will not be necessary to include opinion gathering as the focus should be investigating support for a catalytic effect on learning. Once this is achieved, some measure of user opinions would

be valuable. The use of 10-point semantic differential scales worked well, but greater care needs to be taken with the wording of 'before and after' questions. For example, the present questionnaire asked:

*I am not likely to find this topic interesting*      □ □ □ □ □ □ □ □ □ □      *I am likely to find this topic interesting*

Followed by...

*I found this less interesting than expected*      □ □ □ □ □ □ □ □ □ □      *I found this more interesting than expected*

The pre-opinion question is clearly worded and should be retained but the post-opinion question adds an unhelpful complication by relating the opinion to prior expectations, whereas a more straightforward form of words could be:

*I did not find this topic interesting*      □ □ □ □ □ □ □ □ □ □      *I found this topic interesting*

Where tutorials present a mix of information formats such as video/imagery, audio *and* key point text, questionnaire or interview questions should ascertain which, if any, of these learners paid most and least attention to.

## **Sample**

Whilst there is no evidential basis for claiming that unsupported propositions were due to sample size, 16 per group (reduced in two instances to 15 because of extreme outliers) is generally regarded as a small and therefore less powerful sample. Interestingly, the two supported propositions both came from data sampled from entire cohort ( $N=78$ ).

Suitable future sample sizes should be determined by *a priori* power analysis, based on the required effect size and significance and the proposed method of analysis. For example, taking a medium partial eta squared of 0.06, an alpha value of 0.05 and a proposed two by two mixed ANOVA (two samples tested pre- and post-tutorial), a sample of 30 per group would be required to achieve an acceptable power value of 0.8 (i.e. an 80% chance of detecting a relationship between the samples).

Another issue concerns the constituents of the sample. Using Open University staff members was convenient; it ensured a reasonable degree of homogeneity and a familiarity with ICT. There was an issue regarding academic qualifications but this would not have been so critical were it not for the relative ease of the tutorial and tests. There is no reason why further staff members could not be used in future but there is equally no reason why this might not be given to other adults, undergraduate students or even sixth-formers. A slightly broader sample base would result in less homogeneity, making it more representative of the adult learner population, although it may also result in more outliers and skewed data.

### ***Experimental procedure***

The most notable change to the procedure should be simply to insert some extra time between the pre-test and tutorial and the tutorial and post-test. The former period would ideally be a few days so that all thoughts of the pre-test subject matter were overtaken by other events, but the latter period would require further investigation. The real value of 'memory tests' taken immediately after tuition must be questionable (although it is common practice and numerous research papers refer to post-tests with no mention of any delay). However, a decision would be required as to whether to aim for 'recall' immediately, or after a suitable short period (perhaps 30-60 mins, to allow some other activities and information to be processed), or 'retention' after a more prolonged delay of perhaps a few weeks, or perhaps a combination of these. Further research into this would therefore be required.

One other change to procedure would be to record time-on-task for each participant because there may be a relationship between this, the tutorial design and the pre- and post-test scores.

### ***Analysis***

The methods of analysis and statistical tests used were appropriate and could therefore be retained in any future research. Larger sample sizes should lead to data sets that are normally distributed, meaning that parametric methods of data analysis could be used with confidence.

Analysis of data for future experiments that focus only on catalytic content would be more straightforward. There would not be such a variety of propositions, measures and samples to contend with and, with larger sample sizes, there is a greater chance the data will exhibit normality and so any decision about parametric testing should be straightforward.

The net result of all these measures, summarised in the recommendations on pp134-135, should be a more conclusive and reliable set of outcomes that can provide reliable guidelines for CAL designers and form a basis for yet further research into the catalytic effects of content presented by multiple richer media.

## ***Epilogue***

Whilst this slightly zig-zag research journey has focused on the design and use of content and media, it has always done so with learners and learning in mind, and from the perspective of e-learning designers and other practitioners, a position endorsed by Alexander and McKenzie (1998):

The use of a particular information technology did not, in itself, result in improved quality of learning or productivity of learning. Rather, a range of factors ... are necessary for a successful project outcome, the most critical being the design of the students' learning experiences. (p3)

Of course it will always be fun to play with new technological toys as they emerge with what seems to be greater frequency and ever-increasing functionality; indeed, the exploration of new pedagogic opportunities should always be encouraged. However, a sage piece of final advice comes from Richard Mayer (1997), whose work I have not entirely agreed with during this research, but whose book, *Multimedia Learning* (2001), spurred me into action 9 years ago and provided the multimedia design principles so often referred to in this thesis:

Instructional development is too often based on what computers *can* do rather than on a research-based theory of how students learn with technology. (p17)

## References

- Ainley M, Hidi S & Berndorff D (2002) 'Interest, Learning, and the Psychological Processes That Mediate Their Relationship', *Journal of Educational Psychology*, 94(3), pp545-561
- Ainsworth S & Burcham S (2007) 'The impact of text coherence on learning by self-explanation', *Learning and Instruction*, 17(3), pp286-303
- Ainsworth S & van Labeke N (2004) 'Multiple forms of dynamic representation', *Learning and Instruction*, 14(3), pp241-255
- Al-Seghayer K (2001) 'The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study', *Language Learning and Technology*, 5, pp202-232
- Alessi S & Trollip S (2001) *Multimedia for Learning*, Boston, Allyn and Bacon
- Alexander P & Jetton T (1996) 'The role of importance and interest in the processing of text', *Educational Psychology Review*, 8, pp89-121
- Alexander P & Kulikowich J (2006) 'Learning from physics text: A synthesis of recent research', *Journal of Research in Science Teaching*, 31(9), pp895-911
- Alexander S & McKenzie J (1998) *An Evaluation of Information Technology Projects for University Learning (Executive Summary)*, <http://www.dest.gov.au/archive/cutsd/publications/exsummary.html> [accessed 3 Jul 10]
- Ambient Insight (2010) *The Worldwide Market for Self-paced eLearning Products and Services: 2009-2014 Forecast and Analysis*, <http://www.ambientinsight.com/Reports/eLearning.aspx> [accessed 26 Apr 10]
- American Society for Training and Development (2009) *e-Learning Trends 2009 survey*, [http://www.astd.org/LC/2009/0109\\_trends.htm](http://www.astd.org/LC/2009/0109_trends.htm) [accessed 26 Nov 09]
- Anderson T (Ed) (2008) *Theory and Practice of Online Learning, Second edition*, Athabasca University, <http://www.aupress.ca/index.php/books/120146> [Accessed 5 May 09]
- Anderson L & Krathwohl D (Eds) (2001) *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*, New York, Longman
- Aragon S (2003) 'Creating social presence in online environments', in Aragon S (Ed), *Facilitating learning in online environments*, San Francisco, Jossey-Bass
- Arguel A & Jamet E (2009) 'Using video and static pictures to improve learning of procedural contents', *Computers in Human Behavior*, 25, pp354-359
- Ary D, Cheser Jacobs L, Razavieh A & Sorensen C (2006) *Introduction to Research in Education*, Belmont, Thomson Wadsworth
- Atkinson R & Shiffrin R (1968) 'Human memory: A proposed system and its control processes', in: Spence K & Spence J (Eds), *The psychology of learning and motivation, Vol 8*, London, Academic Press
- Ausubel D (1963) *The Psychology of Meaningful Verbal Learning*, New York, Grune & Stratton
- Ayres P & Paas F (2007) 'Making instructional animations more effective: A cognitive load approach', *Applied Cognitive Psychology*, 21, pp695-700

- Ayres P & Sweller J (2005) 'The Split-Attention Principle in Multimedia Learning', in Mayer R (Ed), *The Cambridge book of multimedia learning*, New York, Cambridge University Press
- Baddeley A (2000) 'The episodic buffer: A new component of working memory?' *Trends in Cognitive Science*, 4, pp417-423
- Bader J & Strickman-Stein N (2003) 'Evaluation of New Multimedia Formats for Cancer Communications', *Journal of Medical Internet Research*, 5(3) <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550563/> [accessed 14 Jan 10]
- Bailey G (1993) *Computer-based integrated learning systems*, Washington, International Society for Technology in Education
- Bandura A (1995), *Self-Efficacy in Changing Societies*, New York, Cambridge University Press
- Barker J (1993) 'Training for multimedia', *Inside IT*, 74
- Barnes J, Conrad K, Demont-Heinrich C, Graziano M, Kowalski D, Neufeld J, Zamora J & Palmquist M (2005) *Generalizability and Transferability*, Colorado State University <http://writing.colostate.edu/guides/research/gentrans/> [accessed 11 Dec 09]
- Barron A & Calandra B (2003) 'The use of audio in multimedia learning: Theory and practice', *Journal of Interactive Instruction Development*, 16(1), pp25-34
- Benjafield J (2006) *Cognition* (3<sup>rd</sup> Edition) New York, Oxford University Press
- Bernard R, Abrami P, Lou Y, Borokhovski E, Wade A, Wozney L, Wallet P, Fiset M & Huang B (2004) 'How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature,' *Review of Educational Research*, 74, pp379-439
- Berlyne D (1960) *Conflict, arousal and curiosity*, New York, McGraw-Hill
- Berlyne D (1971) *Aesthetics and psychobiology*, New York, Appleton-Century-Crofts
- Bertin J (1983) *Semiology of Graphics*, Wisconsin, The University of Wisconsin Press
- Betrancourt, M (2005) 'The animation and interactivity principles in multimedia learning', in Mayer R (Ed), *The Cambridge handbook of multimedia learning*, New York, Cambridge University Press
- Bhowmick A, Khasawneh M, Bowling S, Gramopadhye A & Melloy B (2007) 'Evaluation of alternate multimedia for web-based asynchronous learning', *International Journal of Industrial Ergonomics*, 37(7), pp615-629
- Bishop M, Amankwatia T & Cates W (2008) 'Sound's use in instructional software to enhance learning: a theory-to-practice content analysis', *Education Technology Research and Development*, 56, pp467-486
- Bird M (1992) 'Combining quantitative and qualitative research methods: a case study of the implementation of the Open College policy', in Brannen (Ed), *Mixing methods: qualitative and quantitative research*, Aldershot, Avebury/Ashgate Publishing
- Bird M & Hammersley M (1996) *Educational Research in Action: E835 Study Guide*, Milton Keynes, Open University
- Blake C & Scanlon E (2007) 'Reconsidering simulations in science education at a distance: features of effective use', *Journal of Computer Assisted Learning*, 23, pp491-502

- Bloom B (1956) *A Taxonomy of Educational Objectives*, London, Longman
- Boslaugh S & Watters P (2008) *Statistics in a nutshell*, Sebastopol, O'Reilly
- Boster F, Meyer G, Roberto A, Inge C & Strom R (2006) 'Some Effects of Video Streaming on Educational Achievement', *Communication Education*, 55(1), pp46-62
- Brown K (2001) 'Using computers to deliver training: Which employees learn and why?' *Personnel Psychology*, 54, pp271-296
- Burke Johnson R & Onwuegbuzie A (2004) 'Mixed methods research: a research paradigm whose time has come', *Educational Researcher*, 33(7), pp14-26
- Burns R (2000) *Introduction to Research Methods*, London, Sage Publications Ltd
- Calandra B, Barron A & Thompson-Sellers I (2008) 'Audio Use in E-Learning: What, Why, When, and How?', *International Journal on E-Learning*, 7(4), pp589-601
- Carlson R, Chandler P & Sweller J (2003) 'Learning and understanding science instructional material', *Journal of Educational Psychology*, 95, pp629-640
- Carney R & Levin J (2002) 'Pictorial illustrations still improve students' learning from text', *Educational Psychology Review*, 14, pp5-26
- Carr W & Kemmis S (1983) *Becoming critical: Knowing through Action Research*, Geelong, Deakin University Press
- Case S & Swanson D (2001) *Constructing Written Test Questions for the Basic and Clinical Sciences (3rd Edition)*, Philadelphia, National Board of Medical Examiners
- CDW-G (2009) *21st-Century Campus Report: Defining the Vision*, <http://newsroom.cdwg.com/features/feature-11-02-09.html> [accessed 28 Oct 09]
- Chartered Institute of Personnel and Development (2009) *Learning and Development: Annual Survey Report*, London, CIPD
- Chen A, Darst P & Pangrazi R (2001) 'An examination of situational interest and its sources', *British Journal of Educational Psychology*, 71, pp383-400
- Clayton K, Blumberg F & Auld D (2010) 'The relationship between motivation, learning strategies and choice of environment whether traditional or including an online component', *British Journal of Educational Technology*, 41(3), pp349-364
- Cohen L, Manion L & Morrison K (2000) *Research methods in education*, London, Routledge
- Collard J (2009) 'Research methods for leadership in intercultural contexts', in Collard J & Normore A (Eds), *Leadership and Intercultural Dynamics*, Charlotte, Information Age Publishing
- Cook T & Campbell D (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston, Houghton Mifflin
- Couper M, Tourangeau R, Conrad F & Singer E (2006) 'Evaluating the Effectiveness of Visual Analog Scales : A Web Experiment', *Social Science Computer Review*, 24, pp227-245
- Cowan N (2001) 'The magical number 4 in short term memory: A reconsideration of mental storage capacity', *Behavioral and Brain Sciences*, 24, pp87-114
- Creswell J (2002) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd Edn)*, London, Sage Publications Ltd
- Cross, J (2004) *A History of eLearning*, unpublished article for Horizon, 26 Mar 04, <http://internettime.com/Learning/articles/OTH.doc> [Accessed 26 Feb 08]

- CTGV (Cognition & Technology Group at Vanderbilt) (1993) 'Anchored instruction and situated cognition revisited', *Educational Technology*, 33(3), pp52-70
- Daft R & Lengel R (1986) 'Organizational Information Requirements, Media Richness and Structural Design', *Management Science*, 32(5), pp554-571
- Danaher B, Jazdzewski S, McKay H & Hudson C (2005) 'Bandwidth Constraints to Using Video and Other Rich Media in Behavior Change Websites', *Journal of Medical Internet Research*, 7(4) <http://www.jmir.org/2005/4/e49/HTML> [Accessed 21 Jun 10]
- Davis F, Bagozzi R & Warshaw P (1989) 'User acceptance of computer technology: A comparison of two theoretical models', *Management Science*, 35, pp982-1003
- De Jong T (2009) 'Cognitive load theory, educational research, and instructional design: some food for thought', *Instructional Science*, 27 Aug 09 [online: no page numbering]
- DeRouin R, Fritzsche B & Salas E (2005) 'E-Learning in Organisations', *Journal of Management*, 31(6), pp920-940
- DeVaney T (2009) 'Impact of Video Tutorials in an Online Educational Statistics Course', *MERLOT Journal of Online Learning and Teaching*, 5(4), pp 600-608
- Delamont S (1984) 'The old girl network: reflections on the fieldwork at St Lukes', in Burgess R (Ed), *The Research Process in Educational Settings: ten case studies*, Lewes, Falmer Press.
- Denzin N & Lincoln Y (2000) 'Introduction: The discipline and practice of qualitative research', in Denzin N & Lincoln Y (Eds), *Handbook of qualitative research (2nd ed)*, London, Sage Publications Ltd
- Dewey J (1910) *How we think*, New York, Heath & Co
- Donkor F (2010) 'The Comparative Instructional Effectiveness of Print-Based and Video-Based Instructional Materials for Teaching Practical Skills at a Distance', *International Review of Research in Open and Distance Learning*, 11(1), pp 96-116
- Dooley K, Stuessy C, Magill J, & Vasudevan P (2000) 'Cognitive and Affective Outcomes of Animation on Asynchronous Learning of Agricultural Science Concepts', paper presented at the 27th annual National Agricultural Education Research Conference, San Diego
- Dubcovsky J (2011) 'Topic 8: Transformations of the data', class materials for *Design, Analysis, and Interpretation of Experiments*, UC Davis, University of California  
[http://www.plantsciences.ucdavis.edu/agr205/Lectures/2011\\_Lectures/L8\\_Transformation.pdf](http://www.plantsciences.ucdavis.edu/agr205/Lectures/2011_Lectures/L8_Transformation.pdf)  
[Accessed 20 Jun 11]
- Dyson M & Kipping G (1997) 'The legibility of screen formats: Are three columns better than one?' *Computers & Graphics*, 21(6), pp703-712
- Eldridge M (1998) *Transforming Experience: John Dewey's Cultural Instrumentalism*, Vanderbilt University Press.
- Ellis J (1994) *Report on the Introduction of the Morse Trainer at RAF North Luffenham*, RAF Newton, Training, Development & Support Unit
- eNotes (2007) *How products are made > Chalkboard*, <http://www.enotes.com/how-products-encyclopedia/chalkboard> [Accessed 19 Oct 07]
- Engeström Y (1987) *Learning by Expanding: An Activity - Theoretical Approach to Developmental Research*, Helsinki, Orienta-Konsultit



- Ergo (2007) *About Blackboards - Blackboard Technology and Chalkboard History Advances*, [http://www.ergoindemand.com/about\\_chalkboards.htm](http://www.ergoindemand.com/about_chalkboards.htm) [Accessed 19 Oct 07]
- Evans R. (2007) 'Comments on Shulman, Golde, Bueschel, and Garabedian: Existing practice is not the template', *Educational Researcher*, 36, pp553-559
- Fletcher J & Tobias S (2005) 'The Multimedia Principle', in Mayer R (Ed), *The Cambridge book of multimedia learning*, New York, Cambridge University Press
- Flowerday T, Schraw G & Stevens J (2004) 'The role of choice and interest in reader engagement', *The Journal of Experimental Education*, 72(2), pp93-114
- Foster J & Parker I (1995) *Carrying out investigations in psychology: methods and statistics*, Hoboken, Wiley-Blackwell
- Gagné R (1965) *The Conditions of Learning*, New York, Holt, Rinehart and Winston
- Garner R, Alexander P, Gillingham M, Kulikowich J & Brown R (1991) 'Interest and learning from text', *American Educational Research Journal*, 28, pp643-659
- Garner R, Brown R, Sanders S & Menke D (1992) 'Seductive details and learning from text', in Renninger K, Hidi S & Krapp A (Eds), *The Role of Interest in Learning and Development*, Hillsdale, Lawrence Erlbaum Associates
- Garner R, Gillingham M & White C (1989) 'Effects of "seductive details" on macroprocessing and microprocessing in adults and children', *Cognition and Instruction*, 6, pp41-57
- Garner R & Gillingham M (1991) 'Topic knowledge, cognitive interest, and text recall: A microanalysis', *Journal of Experimental Education*, 59, pp310-319
- Gentner D & Gentner D R (1983) 'Flowing waters or teeming crowds: mental models of electricity', in Gentner D & Stevens A (Eds), *Mental Models*, Hillsdale, Lawrence Erlbaum Associates
- Gentner D, Loewenstein J & Thompson L (2003) 'Learning and Transfer: A General Role for Analogical Encoding', *Journal of Educational Psychology*, 95(2), pp393-408
- Ginns P (2005) 'Meta-analysis of the modality effect', *Learning and Instruction*, 15, pp313-331
- Goetz E & Sadoski M (1995) 'The Perils of Seduction: Distracting Details or Incomprehensible Abstractions?', *Reading Research Quarterly*, 30(3), pp500-511
- Golafshani N (2003) 'Understanding Reliability and Validity in Qualitative Research', *The Qualitative Report*, 8 (4), pp597-607
- Goldman S (1997) 'Learning from text: Reflections on the past and suggestions for the future', *Discourse Processes*, 23(3), pp357-398
- Goldman S & Saul E (1990) 'Flexibility in text processing: A strategy competition model', *Learning and Individual Differences*, 2, pp181-219
- Greene J & D'Oliveira M (1999) *Learning to use statistical tests in psychology*, Milton Keynes, Open University Press
- Griffin M & Robinson D (2000) 'Role of mimeticism and spatiality in textual recall', *Contemporary Educational Psychology*, 25, pp 125-149
- Guan Y-H (2009) 'A Study on the Learning Efficiency of Multimedia-Presented Computer-Based Science Information', *Educational Technology & Society*, 12(1), pp62-72

- Gummesson E (2000) *Qualitative Methods in Management Research (2<sup>nd</sup> Ed)*, London, Sage Publications Ltd
- Gyselinck V & Tardieu H (1999) 'The role of illustrations in text comprehension: What, when, for whom, and why?' in: van Oostendorp H & Goldman S (Eds), *The construction of mental representations during reading*, Mahwah, Erlbaum
- Halliday M & Hasan R (1976) *Cohesion in English*, London, Longman Group Ltd
- Hammersley M and Scarth J (1993) *Beware of wise men bearing gifts: A case study in the misuse of educational research*, in: Gomm R and Woods P (Eds), *Educational research in action*, London, Paul Chapman Publishing Ltd
- Hanson E & Phillips F (2006) 'Teaching Financial Accounting with Analogies: Improving Initial Comprehension and Enhancing Subsequent Learning', *Issues in Accounting Education*, 21(1), pp1-14
- Harp S & Mayer R (1997) 'The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest', *Journal of Educational Psychology*, 89(1), pp92-102
- Harp S & Mayer R (1998) 'How seductive details do their damage: A theory of cognitive interest in science learning', *Journal of Educational Psychology*, 90, pp434-441
- Hasler B, Kersten B & Sweller J (2007) 'Learner control, cognitive load and instructional animation', *Applied Cognitive Psychology*, 21, pp713-729
- Hegarty M, Narayanan N & Freitas P (2002) 'Understanding machines from multimedia and hypermedia presentations', in Otero J, Leon J & Graesser A (Eds) *The Psychology of Science Text Comprehension*, Hillsdale, Lawrence Erlbaum
- Hidi S (1990) 'Interest and its contribution as a mental resource for learning', *Review of Educational Research*, 60, pp549-571
- Hidi S, Baird W & Hildyard A (1982) 'That's important but is it interesting? Two factors in text processing', in Flammer A & Kintsch W (Eds), *Discourse processing*, New York, North-Holland
- Hidi S & Baird W (1986) 'Interestingness - a neglected variable in discourse processing', *Cognitive Science*, 10, pp179-194
- Hidi S & Baird W (1988) 'Strategies for increasing text-based interest and students' recall of expository texts', *Reading Research Quarterly*, 23, pp465-483
- Hidi S & Renninger A (2006) 'The Four-Phase Model of Interest Development', *Educational Psychologist*, 41(2), pp111-127
- Hirst P (1974) 'Forms of Knowledge Revisited', in *Knowledge and the Curriculum: a Collection of Philosophical Papers*, London, Routledge
- Hirst P (1993) 'Educational Theory', in Hammersley M (Ed), *Educational Research: Current Issues*, London, Paul Chapman Publishing Ltd
- Hockey J (1993) 'Research Methods: Researching peers and familiar settings', *Research Papers in Education*, 8(2), pp199-225
- Holzinger A (2005) 'Usability engineering methods for software developers', *Communications of the ACM*, 48(1), pp71-74
- Howe K (1988) 'Against the quantitative-qualitative incompatibility thesis, or, Dogmas die hard', *Educational Researcher*, 17, pp10-16

Hussey J & Hussey R (1997) *Business Research: A Practical Guide for Undergraduate and Postgraduate Students*, London, Macmillan Press

Intel (2007), *60 years of the transistor: 1947 – 2007*,  
<http://www.intel.com/technology/timeline.pdf> [Accessed 29 May 10]

iTunesU [Video material from OU course U216: Environment: habitat and conservation] <http://deimos.apple.com/WebObjects/Core.woa/Browse/itunes.open.ac.uk>  
[Accessed 18 Feb 09]

Joint Information Systems Committee (JISC) (2007) *e-Learning pedagogy programme website* [http://www.jisc.ac.uk/whatwedo/programmes/elearning\\_pedagogy.aspx](http://www.jisc.ac.uk/whatwedo/programmes/elearning_pedagogy.aspx)  
[Accessed 4 Oct 07]

Johnson-Laird P (1998) 'Imagery, visualization and thinking', in Hochberg J (Ed), *Perception and Cognition at Century's End: Handbook of Perception and Cognition*, New York, Academic Press

Joppe M (2000) 'The Research Process', in Golafshani N (2003), *Understanding Reliability and Validity in Qualitative Research*, the Qualitative Report, 8(4), Ontario, University of Toronto.

Kalyuga S (2007) 'Expertise reversal effect and its implications for learner-tailored instruction', *Educational Psychological Review*, 19, pp509-539

Kalyuga S, Chandler P & Sweller J (1999) 'Managing split attention and redundancy in multimedia instruction', *Applied Cognitive Psychology*, 13, pp351-371

Kalyuga S, Chandler P & Sweller J (2000) 'Incorporating learner experience into the design of multimedia instruction', *Journal of Educational Psychology*, 92, pp126-136

Kaplan A (1964) *The conduct of enquiry: methodology for behavioural science*, Scranton, Chandler Publishing Co

Kehoe E, Bednall T, Yin L, Olsen K, Pitts C, Henry J & Bailey P (2009) 'Training adult novices to use computers: Effects of different types of illustrations', *Computers in Human Behavior*, 25, pp275-283

Keller J (1987) 'Strategies for stimulating the motivation to learn', *Performance and Instruction*, 26(8), pp1-7

Kemmis S (1988) 'Action Research' in Hammersley, M (Ed), *Educational Research: Current Issues*, 1, London, Paul Chapman/The Open University

Keyton J, King T, Mabachi N, Manning J, Leonard L & Schill D (2004) *Content analysis procedure book*, Lawrence, University of Kansas

Kim S, Yoon M, Whang S, Tversky B & Morrison J (2007) 'The effect of animation on comprehension and interest', *Journal of Computer Assisted Learning*, 23, pp260-270

Kincaid J (1975) Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel

Kintsch W (1980) 'Learning from text, levels of comprehension, or: Why anyone would read a story anyway', *Poetics*, 9, pp87-98

Kintsch W (1988) 'The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model', *Psychological Review*, 95(2), pp163-182

- Kirkwood A & Price L (2008) 'Assessment and student learning: a fundamental relationship and the role of information and communication technologies', *Open Learning*, 23(1), pp5-16
- Knowles M (1973) *The Adult Learner: A Neglected Species*, Houston, Gulf Publishing
- Krahmer E & Theune M (2002) 'Efficient context sensitive generation of referring expressions, in van Deemter K & Kibble R (eds), *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, CSLI Publications
- Krapp A (1999) 'Interest, motivation and learning: An educational psychological perspective', *European Journal of Psychology of Education*, 14, pp23-40
- Krapp A & Lewalter D (2001) 'Development of interests and interest-based motivational orientations: A longitudinal study in vocational school and work settings', in Volet S & Järvelä S (Eds), *Motivation in learning contexts: Theoretical advances and methodological implications*, London, Elsevier
- Kulhavy R, Stock W & Caterino L (1994) 'Reference maps as a framework for remembering text', in Schnotz W & Kulhavy R (Eds), *Comprehension of Graphics*, New York, Elsevier Science
- Lacey C (1976) 'Problems of sociological fieldwork; a review of the methodology of Hightown Grammar', in Shipman M (Ed), *The organisation and impact of social research*, London, Routledge & Kegan Paul
- Larreamendy-Joerns J & Leinhardt G (2006) 'Going the distance with online education', *Review of Educational Research*, 76, pp567-605
- Laurillard D (2002) *Rethinking University Teaching (2<sup>nd</sup> Edition)*, London, Routledge
- Lave J & Wenger E (1991) *Situated Learning. Legitimate peripheral participation*, Cambridge, University of Cambridge Press
- Leary, M (1995) *Introduction to Behavioral Research Methods (2nd Edn)*, Pacific Grove, Brooks/Cole Publishing
- Lehman S, Schraw G, McCrudden & Hartley K (2007) 'Processing and recall of seductive details in scientific text', *Contemporary Educational Psychology*, 32, pp569-587
- Levie W & Lentz R (1982) 'Effects of text illustrations: A review of research', *Educational Communication & Technology Journal*, 30, pp195-232
- Levin J (1989) 'A transfer-appropriate-processing perspective of pictures in prose', in Mandl H & Levin J (Eds), *Knowledge acquisition from text and pictures*, Amsterdam, Elsevier
- Lewalter D (2003) 'Cognitive strategies for learning from static and dynamic visuals', *Learning and Instruction*, 13(2), pp77-189
- Liew C, Foo S & Chennupati K (2000) 'A study of graduate student end-users; use and perception of electronic journals', *Online Information Review*, 24(4), pp302-315
- Liu S-H, Liao H-L & Pratt J (2009) 'Impact of media richness and flow on e-learning technology acceptance', *Computers & Education*, 52, pp599-607
- Liu Z (2005) 'Reading behavior in the digital environment', *Journal of Documentation*, 61(6), pp700-712
- Liu Z (2006) 'Print vs. electronic resources: A study of user perceptions, preferences, and use', *Information Processing and Management*, 42, pp583-592

Low R & Sweller J (2005) 'The Modality Principle in Multimedia Learning', in Mayer R (Ed), *The Cambridge book of multimedia learning*, New York, Cambridge University Press

Lowe R (2003) 'Animation and learning: selective processing of information in dynamic graphics', *Learning and Instruction*, 13, pp157-176

Lowe R (2004) 'Interrogation of a dynamic visualization during learning', *Learning & Instruction*, 14(3), pp257-274

Lowell A (1913) 'The need for a more scientific study of higher education', speech at Mount Holyoke College: The Seventy-Fifth Anniversary.  
<http://www.archive.org/stream/mountholyokecoll00mounuoft/mountholyokecoll00mounuoft.djvu.txt> [accessed 13 Mar 10]

Lu H & Chiou M (2010) 'The impact of individual differences on e-learning system satisfaction: A contingency approach', *British Journal of Educational Technology*, 41(2), pp307-323

Lynch P & Horton S (2001) *Web Style Guide (second edition)*, New Haven, Yale University Press

Maag M (2004) 'The effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self efficacy', *CIN: Computers, Informatics, Nursing*, 22(1), pp26-33

MacLean P & Scott B (2007) 'Learning design: requirements, practice and prospects', *Campus-Wide Information Systems*, 24(3), pp187-198

Malone T & Lepper M (1987) 'Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning' in Snow R & Farr M (Eds) *Aptitude, Learning and Instruction: III. Conative and affective process analyses*, Hillsdale, Erlbaum

Mammarella N, Fairfield B & Cornoldi C (2007) 'Does music enhance cognitive performance in healthy older adults? The Vivaldi effect', *Aging Clinical and Experimental Research*, 19(5), pp394-349

Mangen A (2008) 'Hypertext fiction reading: haptics and immersion', *Journal of Research in Reading*, 31(4), pp404-419

Mangen J, Lalwani C & Gardner B (2004) 'Combining quantitative and qualitative methodologies in logistics research', *International Journal of Physical Distribution & Logistics Management*, 34(7), pp565-578

Marbach-Ad G, Rotbain Y & Stavy R (2008) 'Using Computer Animation and Illustration Activities to Improve High School Students' Achievement in Molecular Genetics', *Journal Of Research In Science Teaching*, 45(3), pp273-292

Marks D (1997) 'Paivio, Allan Urho', in Sheehy N, Chapman A & Conroy W (Eds), *Biographical dictionary of psychology*, New York, Routledge

Marton F & Säljö R (1996) 'Approaches to Learning' in Marton F et al (Eds), *The Experience of Learning*, Edinburgh, Scottish Academic Press

Masie E (2009) 'e-Learning Update: 12 Years Later and 5 Years into the Future', presentation at the *e-Learning 2009 conference* on 11 Nov 2009  
<http://learningwiki.editme.com/Learning2009-Video> [accessed 10 Dec 09]

Matarazzo G & Sellen A (2000) 'The value of video in work at a distance: Addition or distraction', *Behaviour and Information Technology*, 19(5), pp339-348

Mayer R (1992) 'Cognition and instruction: Their historic meeting within educational psychology', *Journal of Educational Psychology*, 84(4), pp405-412

- Mayer R (1997) 'Multimedia learning: Are we asking the right questions?' *Educational Psychologist*, 32, pp1–19
- Mayer R (2001) *Multimedia learning*, New York, Cambridge University Press
- Mayer R (2005) *The Cambridge handbook of multimedia learning*, New York, Cambridge University Press
- Mayer R (2009) *Multimedia learning (second edition)*, New York, Cambridge University Press
- Mayer R & Anderson R (1991) 'Animations need narrations: An experimental test of a dual-coding hypothesis', *Journal of Educational Psychology*, 83, pp484–490
- Mayer R, Bove W, Bryman A, Mars R & Tapangco L (1996) 'When Less Is More: Meaningful Learning From Visual and Verbal Summaries of Science Textbook Lessons', *Journal of Educational Psychology*, 88, pp64-73
- Mayer R & Chandler P (2001) 'When Learning Is Just a Click Away: Does Simple User Interaction Foster Deeper understanding of Multimedia Messages?', *Journal of Educational Psychology*, 93(2), pp390-397
- Mayer R, Dow G & Mayer S (2003) 'Multimedia Learning in an Interactive Self-Explaining Environment: What Works in the Design of Agent-Based Microworlds?', *Journal of Educational Psychology*, 95(4), pp806–813
- Mayer R, Heiser H & Lonn S (2001) 'Cognitive constraints on multimedia learning: When presenting more material results in less understanding', *Journal of Educational Psychology*, 93, pp187–198
- Mayer R & Johnson C (2008) 'Revising the Redundancy Principle in Multimedia Learning', *Journal of Educational Psychology*, 100(2), pp380-386
- Mayer R & Moreno R (2002) 'Animation as an Aid to Multimedia Learning', *Educational Psychology Review*, 14(1), pp87-99
- Mayer R & Sims V (1994) *For whom is a picture worth a thousand words? Extensions of a dual coding theory of multimedia learning*, *Journal of Educational Psychology*, 86(3), pp389–401
- McDaniel M & Donnelly C (1996) 'Learning With Analogy and Elaborative Interrogation', *Journal of Educational Psychology*, 88(3), pp508-519
- McGregor K, Griffith S, Wheat T & Byrd J (2004) *Using Computer-Generated Animation as Additional Visual Elaboration in Undergraduate Courses – Student Perceptions*, Agricultural Services & Development, Tarleton State University, Texas
- McLaughlin A, Rogers W, Sierra E Jr & Fisk A (2007) 'The effects of instructional media: identifying the task demand/media match', *Learning, Media and Technology*, 32(4), pp381-405
- McNamara D (2002) 'Reading Both High-Coherence and Low-Coherence Texts: Effects of Text Sequence and Prior Knowledge', *Canadian Journal of Experimental Psychology*, 55(1), pp51-62
- McNiff, J (2002) *Action research for professional development: Concise advice for new action researchers* [http://www.jeanmcniff.com/Copy booklet for web site.doc](http://www.jeanmcniff.com/Copy%20booklet%20for%20web%20site.doc) [Accessed 22 Mar 06]
- Means B, Toyama Y, Murphy R, Bakia M & Jones K (2009) *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*, U.S. Department of Education, <http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf> [Accessed 30 Jan 10]

- Miles M & Huberman M (1994) *Qualitative Data Analysis*, London, Sage Publications Ltd
- Miller G (1956) 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', *The Psychological Review*, 63, pp81-97  
<http://www.musanim.com/miller1956/> [Accessed 22 Jul 07]
- Mitchell M (2000) 'Presentation on behalf of Wide Learning', *WOLCE 2000 conference*, Birmingham
- Moreno R & Mayer R (1999) 'Cognitive principles of multimedia learning: The role of modality and contiguity', *Journal of Educational Psychology*, 91, pp358-368
- Moreno R & Mayer R (2000) 'A Coherence Effect in Multimedia Learning: The Case for Minimizing Irrelevant Sounds in the Design of Multimedia Instructional Messages', *Journal of Educational Psychology*, 92(1) pp117-125
- Moreno R & Mayer R (2002) 'Verbal Redundancy in Multimedia Learning: When Reading Helps Listening', *Journal of Educational Psychology*, 94(1), pp156-163
- Morkes J & Nielsen J (1997) 'Concise, SCANNABLE, and Objective: How to Write for the Web <http://www.useit.com/papers/webwriting/writing.html> [Accessed 30 Mar 09]
- Motulsky H (2009) *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*, New York, OUP USA
- Muller D, Lee K & Sharma M (2008) 'Coherence or interest: Which is most important in online multimedia learning?' *Australian Journal of Educational Technology*, 24(2), pp211-221
- Mykityshyn A, Fisk A & Rogers W (2002) 'Learning to use a home medical device. Mediating age-related differences with training', *Human Factors*, 44, pp354-364
- Najjar L (1996) 'Multimedia information and learning', *Journal of Educational Multimedia and Hypermedia*, 5, pp129-150
- Nielsen J (1997) Alertbox, *Changes in Web Usability Since 1994*.  
<http://www.useit.com/alertbox/9712a.html> [Accessed 30 Mar 09]
- Nielsen J (2000) *Designing web usability*, Indianapolis, New Riders Publishing
- Nielsen J (2006) Alertbox, *F-Shaped Pattern For Reading Web Content*,  
<http://www.useit.com/alertbox/> [Accessed 19 Apr 2006]
- Nielsen J (2008) Alertbox, *How Little Do Users Read?*  
<http://www.useit.com/alertbox/percent-text-read.html> [Accessed 30 Mar 09]
- Nielsen J (2009) Alertbox: *Kindle 2 Usability Review*  
<http://www.useit.com/alertbox/kindle-usability-review.html> [Accessed 17 Mar 09]
- Nielsen J (2010) Alertbox: *Scrolling and Attention*,  
<http://www.useit.com/alertbox/scrolling-attention.html> [Accessed 22 Mar 10]
- Norman G (1996) 'Multiple choice questions', in Shannon S & Norman G (Eds) *Evaluation Methods: A Resource Handbook (3rd Edition)*, Hamilton, McMaster University
- OECD (2010) *Are the New Millennium Learners Making the Grade?*  
<http://browse.oecdbookshop.org/oecd/pdfs/browseit/9609101E.pdf> [accessed 1 Jun 10]
- Ofcom (2010a) *The Communications Market, 2010: Internet and web-based content*,  
<http://stakeholders.ofcom.org.uk/market-data-research/market-data/communications-market-reports/cmr10/internet-web/> [accessed 23 Aug 10]

- Ofcom (2010b) *Mobile Broadband Research Summary Report*, <http://stakeholders.ofcom.org.uk/binaries/research/consumer-experience/tce-10/mobile-broadband.pdf> [accessed 3 Jan 11]
- Ofcom (2011) *UK fixed broadband speeds*, <http://stakeholders.ofcom.org.uk/market-data-research/telecoms-research/broadband-speeds/speeds-nov-dec-2010/> [accessed 2 May 11]
- Otondo R, Van Scotter J, Allen D & Palvia P (2008) 'The complexity of richness: Media, message and communication outcomes', *Information and Management*, 45(1), pp21–30
- Osgood C, Suci G & Tannenbaum P (1957) *The measurement of meaning*, Urbana, University of Illinois Press
- Paivio A (1986) *Mental Representations: a Dual Coding Approach*, Oxford, Oxford University Press
- Pallant J (2007) *SPSS survival manual (third edition)*, Maidenhead, Open University Press/McGraw-Hill Education
- Pang K (2009) 'Video-Driven Multimedia, Web-Based Training in the Corporate Sector: Pedagogical Equivalence and Component Effectiveness', *International Review of Research in Open and Distance Learning*, 10(3), pp 1-14
- Parsons C (2001) *The efficiency and preference implications of scrolling versus paging when information seeking in long text passages*, PhD thesis, University of Northern Colorado
- Pask G (1960) 'The teaching machine as a control mechanism', *Transactions of the Society of Instrument Technology*
- Patton M (2001) *Qualitative evaluation and research methods (third edition)*, London, Sage Publications Ltd
- Pennekamp M (2008) *A guide to classroom observation and instruction, Teacher Preparation*, Humboldt State University <http://www.humboldt.edu/~thai/observ.html> [Accessed 17 Feb 08]
- Phillips D (2000) *The Expanded Social Scientist's Bestiary*, Lanham, Rowman & Littlefield
- Pomales-García C & Liu Y (2006a) 'Web-Based Distance Learning Technology: The Impacts of Web Module Length and Format', *The American Journal of Distance Education*, 20(3), pp163–179
- Pomales-García C & Liu Y (2006b), 'Web-Based Distance Learning Technology: Interface Design Variables and their Effects', *International Journal of Instructional Technology and Distance Learning*, 3(5), pp27-42
- Prensky M (2001) 'Digital Natives, Digital Immigrants', *On the Horizon*, Vol 9(5)
- Price L (2006) 'Gender differences and similarities in online courses: challenging stereotypical views of women', *Journal of Computer Assisted Learning*, 22, pp349–359
- Punch K (1998) *Introduction to Social Research*, London, Sage Publications Ltd
- Pylyshyn Z (1981) 'The imagery debate: Analogue media versus tacit knowledge', *Psychology Review*, 87, pp16–45
- Raichle M (2007) *PET scans of neural activity*, St Louis, Washington University School of Medicine, [www.alz.org/living\\_with\\_alzheimers\\_your\\_brain.asp](http://www.alz.org/living_with_alzheimers_your_brain.asp) [Accessed 1 Aug 07]



- Ravenscroft A (2001) 'Designing E-learning Interactions in the 21st Century: revisiting and rethinking the role of theory', *European Journal of Education*, 36(2), pp133-156
- Rawson K & Kintsch W (2002) 'How does background information improve memory for text content?' *Memory & Cognition*, 30(5), pp768-778
- Reeve J (1989) 'The interest-enjoyment distinction in intrinsic motivation', *Motivation and Emotion*, 13, pp83-103
- Reigeluth C (1979) 'In search of a better way to organize instruction: The elaboration theory', *Journal of Instructional Development*, 2(3), pp8-15
- Renninger K, Hidi S & Krapp A (1992) *The Role of Interest in Learning and Development*, Hillsdale, Lawrence Erlbaum Associates
- Resnick L (1989) 'Introduction', in Resnick L (Ed), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, Hillsdale, Erlbaum
- Richard J (2006) *Remember the Source: Effects of Divided Attention on Source Memory for Modality with Visual and Auditory Stimuli*, Bryn Mawr College, PA. <http://serendip.brynmawr.edu/bb/sourcemon/sourcemon-paper.htm> [Accessed 10 Sep 07]
- Robert L & Dennis A (2005) 'Paradox of Richness: A Cognitive Model of Media Choice', *IEEE Transactions on Professional Communication*, 48(1), pp10-21
- Robinson D & Molina E (2002) *The relative involvement of visual and auditory working memory when studying adjunct displays*, *Contemporary Educational Psychology*, 27(1), pp 118–131
- Robson C (1993) *Real world research: a resource for social scientists and practitioner-researchers*, Oxford, Blackwell
- Rowntree D (1990) *Teaching through self instruction: how to develop open-learning materials*, London, Kogan Page Ltd
- Rushby N (1989) 'Computer-assisted learning', in Eraut M (Ed), *The international encyclopedia of educational technology*, Oxford, Pergamon Press
- Sadoski M (2001) 'Resolving the Effects of Concreteness on Interest, Comprehension, and Learning Important Ideas From Text', *Educational Psychology Review*, 13(3), pp263-281
- Sadoski M, Goetz E & Fritz J (1993) 'Impact of Concreteness on Comprehensibility, Interest, and Memory for Text: Implications for Dual Coding Theory and Text Design', *Journal of Educational Psychology*, 85(2), pp291-304
- Salomon G (1984) 'Television is easy and print is tough: the differential investment in mental effort in learning as a function of perception and attributions', *Journal of Educational psychology*, 76(4), pp647-658
- Sanchez C & Wiley J (2006) 'An examination of the seductive details effect in terms of working memory capacity', *Memory & Cognition*, 34(2), pp344-355
- Sanders S & Liptrot D (1993) *An incomplete guide to basic research methods and data collection for counsellors*, Manchester, PCCS books
- Savickas M & Spokane A (Eds) (1999) *Vocational interests*, Palo Alto, Davies-Black
- Saye J & Brush T (2002) 'Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments', *Educational Technology Research and Development*, 50(3), pp77-96

- Scandura T & Williams E (2000) 'Research methodology in management: current practices, trends and implications for future research', *Academy of Management Journal*, 43(6), pp1248-1264
- Schiefele U (1996) 'Topic interest, text representation, and quality of Experience', *Contemporary Educational Psychology*, 21, pp3-18
- Schnotz W & Bannert M (2003) 'Construction and interference in learning from multiple representation', *Learning and Instruction*, 13(2), pp141-156
- Schnotz W & Rasch T (2005) 'Enabling, Facilitating, and Inhibiting Effects of Animations in Multimedia Learning: Why Reduction of Cognitive Load Can Have Negative Results on Learning', *Educational Technology Research and Development*, 53(3), pp47-58
- Schraw G (1998) 'Processing and Recall Differences Among Seductive Details', *Journal of Educational Psychology*, 90(1), pp3-12
- Schraw G, Bruning R & Svoboda C (1995) 'Sources of situational interest', *Journal of Reading Behavior*, 27(1), pp1-17
- Schraw G & Lehman S (2001) 'Situational interest: A review of the literature and directions for future research', *Educational Psychology Review*, 13, pp23-52
- Schwarm S & Ostendorf M (2005) 'Reading level assessment using support vector machines and statistical language models', in *Proceedings of ACL'05*, pp523-530
- Scott D & Morrison M (2006) *Key Ideas in Educational Research*, London, Continuum Publishing Group
- Scott-Jackson W, Edney T & Rushent C (2007) *Realising value from online learning in management development*, London, Chartered Management Institute
- Sechrest L & Sidana S (1995) 'Quantitative and qualitative methods: Is there an alternative?' *Evaluation and Program Planning*, 18, pp77-87
- Seufert T, Schutze M & Brunken R (2009) 'Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study', *Learning and Instruction*, 19, pp28-42
- Shepherd J, Grace J & Koch E (2008) 'Evaluating the electronic textbook: is it time to dispense with the paper text?' *Teaching of Psychology*, 35, pp2-5
- Shields P (1998) 'Pragmatism as a Philosophy of Science: A Tool for Public Administration', *Research in Public Administration*, 4, pp195-225
- Short J, Williams E & Christie B (1976) *The Social Psychology of Telecommunications*, New York, Wiley
- Shulman L (1987) 'Knowledge and teaching: foundations of the new reform', *Harvard Educational Review*, 57, pp1-22
- Shulman L (1999) 'Knowledge and Teaching: Foundations of the new Reform', in Leach J & Moon R (Eds), *Learners & Pedagogy*, London, Paul Chapman Publishing in association with the Open University
- Silvia P (2005) 'What Is Interesting? Exploring the Appraisal Structure of Interest', *Emotion*, 5(1), pp89-102
- Silvia P (2006) *Exploring the psychology of interest*, New York, Oxford University Press
- Slavin R (1987) 'Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis', *Review of Educational Research*, 57(3), pp293-336

- Sloman M & Rolph J (2003) *E-learning: the learning curve*, London, Chartered Institute of Personnel and Development
- Skinner B & Rogers C (1956) 'Some issues concerning the control of human behavior: A symposium', *Science*, 124, pp1057-66
- Smith S, Salaway G & Caruso J (2009) The ECAR study of undergraduate students and information technology, 2009, EDUCAUSE Center for Applied Research, <http://www.educause.edu/ers0906> [accessed 28 Nov 09]
- Sun P-C & Cheng H (2007) 'The design of instructional multimedia in e-Learning: A Media Richness Theory-based approach', *Computers & Education*, 49(3), pp662-676
- Sweller J (1988) 'Cognitive load during problem solving: Effects on learning', *Cognitive Science*, 12, pp257-285
- Sweller J (2005) 'Implications of cognitive load theory for multimedia learning', in Mayer R (Ed), *The Cambridge book of multimedia learning*, New York, Cambridge University Press
- Sweller J, Van Merriënboer J & Paas F (1998) 'Cognitive architecture and instructional design', *Educational Psychology Review*, 10, pp251-296.
- Tabbers H, Martens R & van Merriënboer J (2004) 'Multimedia instructions and cognitive load theory: Effects of modality and cueing', *British Journal of Educational Psychology*, 71, pp71-91
- Tangen J, Constable M, Durrant E, Teeter C, Beston B & Kim J (2011) 'The role of interest and images in slideware presentations', *Computers & Education*, 56, pp865-872
- Timmerman E & Kruepke K (2006) 'Computer-Assisted Instruction, Media Richness, and College Student Performance', *Communication Education*, 55(1), pp73-104
- Thalheimer W (2004) *Bells, whistles, neon, and purple prose: When interesting words, sounds, and visuals hurt learning and performance - a review of the seductive-augmentation research*, Somerville, Work-Learning Research Inc, [http://www.work-learning.com/seductive\\_augmentations.htm](http://www.work-learning.com/seductive_augmentations.htm) [Accessed 27 Apr 04]
- Towler A, Kraiger K, Sitzmann T, Van Overberghe C, Cruz J, Ronen E & Stewart D (2008) 'The seductive details effect in technology-delivered instruction', *Performance Improvement Quarterly*, 21(2), pp65-86
- Tulving E (1972) 'Episodic and semantic memory', in: Tulving E & Donaldson W (Eds), *Organization of Memory*, New York, Academic Press Inc
- Tversky B, Morrison J & Betrancourt M (2002) 'Animation: can it facilitate?' *International Journal of Human-Computer Studies*, 57, pp247-262
- van Dijk T & Kintsch W (1983) *Strategies of discourse comprehension*, New York, Academic Press
- Van Merriënboer J & Ayres P (2005) 'Research on Cognitive Load Theory and Its Design Implications for E-Learning', *Educational Technology Research and Development*, Vol. 53(3), pp5-13
- Vaughan T (1994) *Multimedia: making it work*, Berkeley, Osborne MacGraw-Hill
- Vekiri I (2002) 'What is the value of graphical displays in learning?' *Educational Psychology Review*, 14, pp261-312
- Wade S & Adams B (1990) 'Effects of importance and interest on recall of biographical text', *Journal of Reading Behavior*, 22, pp331-353

- Wade S, Schraw G, Buxton W & Hayes M (1993) 'Seduction of the strategic reader: Effects of interest on strategies and recall', *Reading Research Quarterly*, 28, pp3-24
- Waller R (1981) 'Understanding network diagrams', paper presented at the *Annual Meeting of the American Educational Research Association*, Los Angeles
- Waschull S (2001) 'The online delivery of psychology courses: Attrition, performance, and evaluation', *Teaching of Psychology*, 28, pp143-146
- Weller M (2002) *Delivering Learning on the Net*, Abingdon, Routledge Falmer
- Wharton W (1980) 'Higher imagery and the readability of college history texts', *Journal of Mental Imagery*, 4, pp129-147
- Wiersma W & Jurs S (2005) *Research Methods in Education*, Boston, Pearson
- Wiley J & Rayner K (2000) 'Effects of titles on the processing of text and lexically ambiguous words: evidence from eye movements', *Memory & Cognition*, 28, pp1011-1021
- Winn J (1993) 'Perception principles', in Fleming M & Levie H (Eds) *Instructional message design: Principles from the behavioral and cognitive sciences*, Englewood Cliffs, NJ, Educational Technology Publications
- Winn W, Li T & Schill D (1991) 'Diagrams as aids to problem solving: Their role in facilitating search and computation', *Educational Technology Research & Development*, 39(1), pp17-29
- Wood D, Bruner J, & Ross G (1976) 'The role of tutoring in problem solving', *Journal of child psychology and psychiatry*, 17, pp89-100
- Woods P (1996) 'Qualitative Research', in: *Educational Research in Action: E835 Study Guide*, Milton Keynes, Open University
- Woody W, Daniel D & Baker C (2010) 'E-books or textbooks: Students prefer textbooks', *Computers & Education*, 55, pp945-948
- Xerte (CAL authoring tool, University of Nottingham)  
<http://www.nottingham.ac.uk/xerte/> [Ver 2.5.4, accessed 19 Jun 09]
- Yelland N & Masters J (2007) 'Rethinking scaffolding in the information age', *Computers and Education*, 48, pp362-382
- Yeung W & Lu M (2004) 'Gaining competitive advantages through a functionality grid for website evaluation', *Journal of Computer Information Systems*, 44(4), pp67-77
- Zacks J & Tversky B. (2003) 'Structuring information interfaces for procedural learning', *Journal of Experimental Psychology*, 92, pp88-100
- Zhang D, Zhou L, Briggs R & Nunamaker J Jr (2006) 'Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness', *Information & Management*, 43, pp15-27
- Zheng R, Yang W, Garcia D & McCadden E (2008) 'Effects of multimedia and schema induced analogical reasoning on science learning', *Journal of Computer Assisted Learning*, 24, pp474-482

## **Appendix A: Glossary of Terms**

Andragogy	The science of teaching adults, as distinct from pedagogy which, although widely referred to in adult education, etymologically relates to the teaching of children
Blog	Web log (akin to a personal or shared online diary)
CAI	Computer Assisted (or Aided) Instruction
CAL	Computer Assisted (or Aided) Learning
CBT	Computer Based Training
CD-ROM	Compact Disc – Read Only Memory
CIPD	Chartered Institute of Personnel & Development
CLT	Cognitive Load Theory
DCT	Dual Coding Theory
DE	Distance Education
DL	Distance Learning
DNA	Deoxyribonucleic Acid
DVD	Digital Versatile Disc
e-Book	Electronic version of a book (often in PDF or e-Pub format)
e-Learning	Learning afforded or supported by computer, either online (possibly involving collaboration with others) or via disc
e-Reader	A device for reading e-Books
FE	Further Education
GPA	Grade Point Average
Haptic	Relating to the sense of touch
HE	Higher Education
HTML	Hypertext Markup Language
ICT	Information and Communication Technology
IT	Information Technology
JISC	Joint Information Steering Committee
Kindle	e-Book reader marketed by Amazon
L&D	Learning and Development
LTM	Long Term Memory
Matching test	Measures recall of knowledge through labelling artefacts on a diagram
MLO	multimedia learning object
OU	Open University (UK)
PBL	Problem Based Learning
PC	Personal Computer
PET	Positron Emission Tomography
RAF	Royal Air Force
RBS	Royal Bank of Scotland
Recall test	Measures recollection of factual information
ROI	Return on Investment
Schemata	Plural form of Schema
SME	Subject Matter Expert
S-R	Stimulus-Response (behaviourism)
STM	Short-Term Memory
TEL	Technology Enhanced (Enabled) Learning
Transfer test	Measures ability to transfer theoretical knowledge into solving written problems
WM	Working Memory
WMC	Working Memory Capacity
WYSIWYG	What-you-see-is-what-you-get

## **Appendix B: Pack for Experiment Participants**

*(Note: actual font size has been reduced to fit thesis page size)*

### **EdD Research - Introduction**

This research into the design of computer-assisted learning (CAL) is being conducted by Jim Ellis (j.r.ellis@open.ac.uk - Ext 58849) as part of the OU Doctorate in Education (EdD) programme.

The research will be conducted mainly through experiments using computer-based CAL tutorials with different media and design characteristics, and assessing the impact on learners and learning. Although the CAL design may vary, the format of each session will be the same for every participant:

- **This Introduction.**
- **Consent.** Participation is voluntary, but the university requires that you indicate your written consent using the attached proforma.
- **Paper-based pre-questions.** Comprising 11 opinion ratings, 15 multiple-choice questions and some brief personal information.
- **CAL session.** The program will be ready to use and its operation is straightforward. I will be nearby but I do not aim to interfere and I do not require you to make any verbal comments unless you encounter a significant problem.
- **Paper-based post-questions.** Comprising 7 reaction ratings followed by 15 *different* multiple-choice questions.
- **Discussion.** A short, informal chat to discuss your experience, opinions, test scores and any issues you wish to raise.
- **Deferred Post-test.** Subject to your availability, it would be helpful to conduct a further post-test two weeks later to indicate how much learning you have retained.

For your assurance...

- This research:
  - Has been approved by, and will comply with the conditions of, the OU Human Participants and Materials Ethics Committee (HPMEC)
  - Will comply with the provisions of the Data Protection Act and has been registered with the OU Data Protection Office
- The experiment should take around 40 mins
- Participation is voluntary and you may withdraw from the research *at any time*, in which case I will delete any individual data I have gathered
- Your scores and any information you volunteer will remain confidential and will be codified in a way which means data cannot be traced back to you personally
- You will not be recorded in any way and the computer will not track any data from your session
- I am happy to discuss the research more fully with you (subject to the confidentiality of others) after the experiment
- Please remember that *it is the CAL design that is being evaluated, not you* – so if you think you're performing poorly, this could be an outcome that indicates an ineffective design approach

Thank you for agreeing to support my research.

*Jim Ellis*

*If you are satisfied, you may tear off and keep this sheet  
Please then continue on the next page*

**Pre-Opinions (1)**

This e-learning tutorial covers the hydrological cycle and factors that may cause or prevent flooding, using the River Severn at Shrewsbury as a case study. The tutorial is based on material available via iTunesU from the OU course U216, "Environment: habitat and conservation".

Please rate yourself below in terms of the following 'starting conditions', and please be honest! For example, it *doesn't matter* if you're not very interested in the topic - but it will help my analysis if you indicate that fact in your rating.

*Please tick **one** box in each case*

I am not familiar with this topic area

I am very familiar with this topic area

This could be quite hard for me to learn

This should be quite easy for me to learn

This topic is not important to me

This topic is important to me

I am not likely to find this topic interesting

I am likely to find this topic interesting

I do not enjoy learning from on-screen material

I enjoy learning from on-screen material

*When you have completed this page, please turn over and continue*

## Pre-test

*There are 15 questions, each of which has one correct answer. Please indicate this by marking the appropriate box in each case.*

1. The hydrological cycle is:
  - The transition of water from ice to liquid to vapour as temperature is increased
  - The transition from reservoir to domestic water, sewage, purification and river
  - The transition of water from vapour to liquid to ice as temperature is decreased
  - The transition from precipitation to rivers, sea, evaporation and clouds
2. Precipitation caught on the surface of vegetation and which evaporates is called:
  - Groundwater
  - Transpiration
  - Interception
  - Run-off
3. Catchment is defined as which of the following:
  - The area through which a river and its tributaries flow
  - The area a river and its tributaries supplies with domestic water
  - The area feeding a river and its tributaries with precipitation
  - The area into which a river and its tributaries could potentially flood
4. Run-off will be increased by:
  - Allowing sheep to graze
  - Planting broad-leafed trees
  - Introducing Argys
  - Reducing the watershed
5. Transpiration is:
  - Snowfall that starts to melt as the temperature rises
  - Precipitation caught on the surface of vegetation which then evaporates
  - Precipitation that falls directly into streams, rivers, lakes and the sea
  - Water absorbed through the roots of a plant which evaporates through the leaves
6. A watershed is defined as which of the following:
  - The boundary between different floodplains
  - The boundary between different catchments
  - The theoretical height above which a river is not expected to flood
  - The theoretical volume of precipitation a river can absorb without flooding
7. Flood Gates are:
  - Used to release water from the floodplain as river levels fall
  - Used to keep water off the floodplain as river levels rise
  - Placed across the river further upstream to limit the rate of flow
  - Placed across tributaries further upstream to limit the rate of flow
8. Planting broad-leafed trees:
  - Compacts the soil making it less permeable
  - Encourages worms and small animals which aerate the soil, increasing run-off
  - Encourages worms and small animals which aerate the soil, increasing absorption
  - Provides shelter for grazing sheep during periods of heavy precipitation



9. Flooding is usually caused by:
- Prolonged heavy rain around the point of the breach
  - Prolonged heavy rain throughout the catchment
  - Prolonged heavy rain around the source of the river and its tributaries
  - Prolonged heavy rain beyond the watershed
10. Water stored underground between soil or rock particles is referred to as:
- Groundwater
  - Transpiration
  - Interception
  - Run-off
11. An Argy:
- Is a natural embankment that prevents minor flooding
  - Is a natural embankment that prevents major flooding
  - Is a man-made embankment that prevents minor flooding
  - Is a man-made embankment that prevents major flooding
12. During and after heavy precipitation, floodplains in rural areas should:
- Be protected through extensive the use of Argys
  - Be allowed to remain flooded and drained only as river levels fall
  - Be protected by increasing the watershed
  - Be allowed to remain flooded for as short a period as possible
13. Run-off is:
- Water running downhill on the surface of non-permeable ground
  - Water flowing downhill under the surface of non-permeable ground
  - Water flowing downhill under the surface of land planted with broad-leaved trees
  - Water running downhill on the surface of ground planted with broad-leaved trees
14. Total evaporation in a forested catchment compared to moorland is typically:
- About twice as high due to interception and transpiration
  - About twice as high due to transpiration and run-off
  - About half the amount due to run-off and groundwater
  - About half the amount due to groundwater and interception
15. A characteristic of groundwater is that it:
- Remains in-situ until it evaporates
  - Erodes banks of rivers and streams, encouraging flooding
  - Evaporates through the surface of foliage
  - Flows only slowly downhill into springs, rivers and the sea

*When you have completed the test, please turn over and continue*

**Background Information and Declaration**

Please tick **one** box in each case

Gender	<input type="checkbox"/> Male	<input type="checkbox"/> Female
Age Group	<input type="checkbox"/> Up to 20	<input type="checkbox"/> 21 - 30
	<input type="checkbox"/> 31 - 40	<input type="checkbox"/> 41 - 50
	<input type="checkbox"/> 51 - 60	<input type="checkbox"/> 61 and over
Education	<input type="checkbox"/> Degree or equivalent	<input type="checkbox"/> Post graduate

- I have received an introduction explaining the background and purpose of the research and the proposed conduct of experiments
- I am a willing volunteer, I have had an opportunity to ask questions and have been satisfied with the answers to these
- I am satisfied with the safeguards in place regarding my personal data, confidentiality and my right to withdraw at any time

Participant Signature \_\_\_\_\_ Date \_\_\_\_\_

**Pre-Opinions (2)**

You will receive *one of three* tutorials – each covering the same learning ground, but designed in different ways. Please rate your expectations as follows...

A computer-based tutorial containing mostly **video with narration** is likely to be:

Not very appealing                      Very appealing

Not very effective at teaching                      Very effective at teaching

A computer-based tutorial containing mostly **audio and images** is likely to be:

Not very appealing                      Very appealing

Not very effective at teaching                      Very effective at teaching

A computer-based tutorial containing mostly **text and images** is likely to be:

Not very appealing                      Very appealing

Not very effective at teaching                      Very effective at teaching

*When you have completed this page, please begin the tutorial **at the computer***

## Post-Opinions

Please rate yourself below in terms of your actual experience.

*Please tick **one** box in each case*

I knew less about this than I originally thought

I knew more about this than I originally thought

I found this harder to understand than expected

I found this easier to understand than expected

None of the content helped me learn

All of the content helped me learn

I found this less interesting than expected

I found this more interesting than expected

Less content would have been beneficial

More content would have been beneficial

None of the content was relevant to the outcomes

All of the content was relevant to the outcomes

I found this less enjoyable than expected

I found this more enjoyable than expected

*When you have completed this page, please turn over and continue*

## Post-test

*There are 15 questions, each of which has one correct answer. Please indicate this by marking the appropriate box in each case.*

1. Interception is:
  - Precipitation caught on the surface of vegetation which then evaporates
  - Water absorbed through the roots of a plant which evaporates through the leaves
  - Precipitation that falls directly into streams, rivers, lakes and the sea
  - Potential flood water that is restrained by dams or floodgates
2. The hydrological cycle is:
  - The transition of water from ice to liquid to vapour as temperature is increased
  - The transition from reservoir to domestic water, sewage, purification and river
  - The transition of water from vapour to liquid to ice as temperature is decreased
  - The transition from precipitation to rivers, sea, evaporation and clouds
3. Catchment is defined as which of the following:
  - The area through which a river and its tributaries flow
  - The area a river and its tributaries supplies with domestic water
  - The area feeding a river and its tributaries with precipitation
  - The area into which a river and its tributaries could potentially flood
4. Groundwater is:
  - Water stored underground between soil or rock particles
  - Water that forms into puddles on the surface of non-permeable ground
  - Any precipitation that has fallen and can potentially evaporate
  - Water remaining on the surface of the floodplain after a flood
5. Planting broad-leafed trees:
  - Compacts the soil making it less permeable
  - Encourages worms and small animals which aerate the soil, increasing run-off
  - Encourages worms and small animals which aerate the soil, increasing absorption
  - Provides shelter for grazing sheep during periods of heavy precipitation
6. Water absorbed through the roots of a plant which then evaporates through the leaves is referred to as:
  - Groundwater
  - Transpiration
  - Interception
  - Run-off
7. A watershed is defined as which of the following:
  - The boundary between different floodplains
  - The boundary between different catchments
  - The theoretical height above which a river is not expected to flood
  - The theoretical volume of precipitation a river can absorb without flooding
8. Run-off will be increased by:
  - Allowing sheep to graze
  - Planting broad-leafed trees
  - Introducing Argys
  - Reducing the watershed

9. Total evaporation in a forested catchment compared to moorland is typically:
- About twice as high due to interception and transpiration
  - About twice as high due to transpiration and run-off
  - About half the amount due to run-off and groundwater
  - About half the amount due to groundwater and interception
10. A man-made embankment that prevents minor flooding is referred to as:
- A Floodbreak
  - An Aquate
  - A Floodgate
  - An Argy
11. A characteristic of groundwater is that it:
- Remains in-situ until it evaporates
  - Erodes banks of rivers and streams, encouraging flooding
  - Flows only slowly downhill into springs, rivers and the sea
  - Evaporates through the surface of foliage
12. During and after heavy precipitation, floodplains in rural areas should:
- Be protected through extensive the use of Argys
  - Be allowed to remain flooded and drained only as river levels fall
  - Be protected by increasing the watershed
  - Be allowed to remain flooded for as short a period as possible
13. Water flowing downhill on the surface of non-permeable ground is referred to as:
- Groundwater
  - Transpiration
  - Interception
  - Run-off
14. Flood Gates are:
- Used to keep water off the floodplain as river levels rise
  - Used to release water from the floodplain as river levels fall
  - Placed across the river further upstream to limit the rate of flow
  - Placed across tributaries further upstream to limit the rate of flow
15. The canopy of a forested area:
- Reduces transpiration
  - Increases run-off
  - Delays rainfall from reaching the ground
  - Delays evaporated moisture from reaching the atmosphere

*End*

## Appendix C: CAL Tutorial Screen Samples

Rivers and flooding  
5. Evaporation

There are some natural features and phenomena which slow down the rate at which heavy precipitation enters the river.

Click the Play button to start the slideshow (110 sec)

Evaporation:

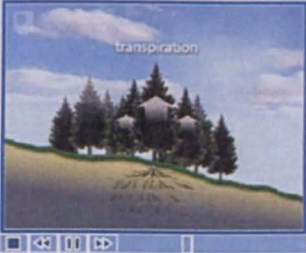
- Water evaporates from the ground

Interception

- Water evaporates from surface of vegetation

Transpiration:

- Water absorbed through roots
- Evaporates through leaves



transpiration


Video / Audio

Rivers and flooding  
6. Hydrological Cycle (2)

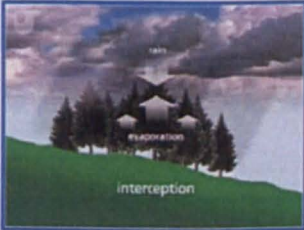
Precipitation which falls onto foliage can evaporate back into the atmosphere as part of a process called interception.

Water can also be transferred through the roots of a plant into its leaves, from which it can evaporate. This process is called transpiration.

Researchers at Plymton, near the source of the Severn, compared a moorland catchment and a forested catchment.



transpiration



rain  
evaporation  
interception

They found that about 15% of the precipitation in a moorland catchment was evaporated back into the atmosphere, mainly through transpiration from the vegetation.

In the forested catchment they found that about 30% was lost through a combination of transpiration through the trees, and interception from the canopy.


So overall, there was about double the amount of evaporation from a forested catchment.

Text (709 words)

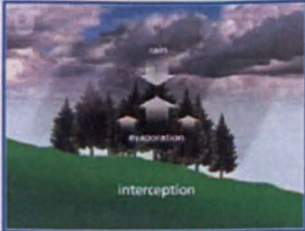
Rivers and flooding  
6. Hydrological Cycle (2)

The forest canopy of mature trees slows precipitation from reaching ground.

Rainfall, etc, also evaporates from the surface of foliage in a process called Interception.



transpiration



rain  
evaporation  
interception

Water is also absorbed through the roots of trees and evaporates through the leaves in a process called Transpiration.

The effects of interception and transpiration mean a forested area generates about twice as much total evaporation as moorland.

Text (406 words)

## Appendix D: Analysis of Text-Based Design Variants

**Table D.1.** Textual analysis of five main 'teach' screens in text-based tutorials

Primary text  
Catalytic text

Text-with-imagery 1	Text-with-imagery 2	Text-with-imagery 3
<p>Shrewsbury's floods did not always coincide with its peak rainfall, because the water level is determined mainly by factors upstream. <b>The catchment is the area of land that feeds rainwater downhill into a particular river.</b> The catchment of the River Severn, including all of its tributaries, is shown here. <b>The higher land that separates the catchments of different rivers is called a watershed.</b> Near the source of the Severn at Plynlimon in mid-Wales, average annual rainfall is about 2,500mm, but Shrewsbury gets only one third of that amount. Flooding is caused by heavy rain in the uplands, plus heavy rainfall across the rest of the lower part of the catchment. If there's a long period of wet weather the ground can't store all of the precipitation, so the excess goes straight into the river. If there's more water than the river channel can hold, it spills out onto the flood plain. In order to alleviate the flooding, we need to either reduce the amount of water getting into the river, or to delay its progress.</p> <p style="text-align: right;">175 total 113 primary 62 catalytic</p>	<p>Shrewsbury's floods did not always coincide with its peak rainfall, because the water level is determined mainly by factors upstream. <b>A river's catchment is the total area feeding that river and its tributaries with water.</b> The watershed is the high ground between catchments. Near the source of the Severn at Plynlimon in mid-Wales, average annual rainfall is about three times that of Shrewsbury. Flooding is caused by heavy rain in the uplands, plus heavy rain lower down the catchment. The ground can't store all the precipitation so the river overfills and spills onto the floodplain.</p> <p style="text-align: right;">95 total 71 primary 24 catalytic</p>	<p>Shrewsbury's floods did not always coincide with periods of peak rainfall because the water level of the Severn in the town is determined mainly by factors upstream. <b>A river's catchment is the total area of land that feeds rainwater downhill into that particular river.</b> The catchment of the River Severn, including all of its tributaries, is shown in this map. <b>The watershed is the high ground between catchments.</b> Near the source of the Severn at Plynlimon in mid-Wales, average annual rainfall is about 2500mm, which is three times that further downstream at Shrewsbury. Flooding is caused by heavy rain in the uplands, plus heavy rain across the rest of the catchment. If this occurs for a prolonged period, the ground can't store all of the precipitation, so the excess flows straight into the river. Then, if there's more water than the river channel can hold, it spills out onto the natural flood plain, as it did 3 times in Autumn 2000.</p> <p style="text-align: right;">161 total 100 primary 61 catalytic</p>
<p>Water flows through the catchment as part of a process called the hydrological cycle. It moves around the earth in a continuous cycle, driven by the sun's energy and by gravity. <b>Precipitation - rain, snow or hail - falls onto the catchment and flows out to sea via rivers and lakes. It then evaporates back into the air and the cycle repeats itself.</b> There are some natural features and phenomena which slow down the rate at which precipitation enters the river. For example, some evaporates directly off the land straight back into the atmosphere. Also, in forested areas, the canopy of mature trees is very efficient at slowing the progress of rainfall as it comes through. (Continued...)</p> <p style="text-align: right;">117 total 79 primary 38 catalytic</p>	<p>In the basic hydrological cycle: <b>Precipitation (rain, hail, snow) falls onto the land; It is carried via rivers and lakes to the sea; It evaporates back into the atmosphere; and falls again as precipitation.</b> There are some natural features and phenomena which slow down the rate at which heavy precipitation enters the river. For example, some evaporates directly off the land back into the atmosphere. (Continued...)</p> <p style="text-align: right;">66 total 46 primary 20 catalytic</p>	<p>To alleviate flooding at Shrewsbury, we must either reduce the total amount of water getting into the river, or to delay its progress. Moisture moves around the earth in a continuous cycle, driven by the sun's energy and by gravity. We can think of this cycle in global terms by including the oceans and polar ice; however, <b>in the basic hydrological cycle, precipitation such as rain, hail and snow, falls onto the land, is carried via rivers and lakes to the sea, evaporates back into the atmosphere, forms clouds and falls again as precipitation.</b> There are some natural features and phenomena which slow down the rate at which heavy precipitation enters rivers and lakes. For example, some which falls onto rural and urban land remains in-situ and is evaporated directly from back into the atmosphere. The forest canopy of mature trees also physically slows precipitation from reaching ground.</p>

Text-with-imagery 1	Text-with-imagery 2	Text-with-imagery 3
<p>Precipitation which falls onto foliage can evaporate back into the atmosphere as part of a process called interception. Water can also be transferred through the roots of a plant into its leaves, from which it can evaporate. This process is called transpiration. Researchers at Plynlimon, near the source of the Severn, compared a moorland catchment and a forested catchment. They found that about 15% of the precipitation in a moorland catchment was evaporated back into the atmosphere, mainly through transpiration from the vegetation. In the forested catchment they found that about 30% was lost through a combination of transpiration through the trees, and interception from the canopy. So overall, there was about double the amount of evaporation from a forested catchment.</p> <p style="text-align: right;">121 total 88 primary 33 catalytic</p>	<p>The forest canopy of mature trees slows precipitation from reaching ground. Rainfall, etc, also evaporates from the surface of foliage in a process called Interception. Water is also absorbed through the roots of trees and evaporates through the leaves in a process called Transpiration. The effects of interception and transpiration mean a forested area generates about twice as much total evaporation as moorland.</p> <p style="text-align: right;">63 total 63 primary 0 catalytic</p>	<p style="text-align: right;">148 total 71 primary 77 catalytic</p> <p>Apart from the physical slowing of precipitation by the forest canopy and evaporation from the ground, rainfall, etc, which falls onto the surface of foliage also evaporates back into the atmosphere in a process called Interception. Water in the ground is also absorbed through the roots, up through the trees and evaporates through the leaves in a process called Transpiration. Researchers at Plynlimon, near the source of the Severn, compared evaporation in a moorland catchment and a forested catchment. They found that about 15% of precipitation over moorland was evaporated, but this rose to about 30% for woodland. Thus, the effects of interception and transpiration mean a forested area generates about twice as much total evaporation as moorland.</p> <p style="text-align: right;">118 total 63 primary 55 catalytic</p>
<p>In the upper part of the Severn catchment, high numbers of grazing sheep compact the soil. This means the ground isn't very permeable so much of the rainwater stays on the surface and runs downhill into the nearest stream. This part of the hydrological cycle is called run-off. Run-off prevents groundwater, which is water that's stored underground in the pores between soil or rock particles, eventually discharging into springs, rivers, or the sea. Groundwater travels only slowly through the water cycle so, if rainwater can be absorbed underground, this will delay it reaching the river and keep river levels down. There's no getting away from the fact that grazing sheep compact the soil. However, if a field has been planted with broadleaf trees, any water here is absorbed straight away. Vegetation under the trees provides the basis of a food chain for worms, voles and other animals. The animals break up and aerate the soil so it absorbs more groundwater.</p> <p style="text-align: right;">160 total 120 primary 40 catalytic</p>	<p>Livestock such as grazing sheep compact the soil, greatly reducing its absorption. This causes Run-off, where most water remains on the surface and runs quickly downhill into streams and lakes. Planting strips of broad leaf trees encourage worms and other small animals which break up the soil, making it more absorbent. Absorbed precipitation is termed Groundwater because it is stored between the soil particles and so travels downhill much more slowly than run-off.</p> <p style="text-align: right;">73 total 73 primary 0 catalytic</p>	<p>The upper part of the Severn catchment comprises large areas of moorland used for farming. Livestock such as grazing sheep compact the soil over a period of time, greatly reducing its absorption. This causes Run-off, where most precipitation remains on the surface and runs quickly downhill into streams and lakes. A local farmer described this as looking like 'silver sheets' on some extremely wet days. Planting strips of broad leaf trees across the slope can break the flow by encouraging vegetation, worms, voles and other small animals which break up and aerate the soil, making it more absorbent. Absorbed precipitation is termed Groundwater because it is stored and flows between the soil or rock particles, eventually discharging into springs, rivers, or the sea. Groundwater therefore moves through the hydrological much more slowly than run-off.</p> <p style="text-align: right;">134 total 83 primary 51 catalytic</p>
<p>We cannot afford to build more dams and reservoirs simply to prevent flooding. In the past,</p>	<p>We cannot afford to build dams and reservoirs simply to prevent flooding. We also know that</p>	<p>Physically controlling the flow of a river is expensive so we cannot afford to build dams and</p>



Text-with-imagery 1	Text-with-imagery 2	Text-with-imagery 3
<p>water was drained off the land as quickly as possible, but doing this in the upper parts of the catchment just causes flooding further downstream. Some areas of flood plain in rural areas can form natural reservoirs and it is now considered better to hold water on the land in these areas for a few days. The natural storage capacity of the flood plain can also be enhanced by a system of man-made embankments called Argys, which help prevent floodwater coming onto the field behind them during smaller floods. During bigger floods, water will overtop the Argys, filling the flood plain. Large floodgates are then used to contain the water, releasing it slowly back into the river as levels fall.</p> <p style="text-align: right;">136 total 136 primary</p>	<p>draining land as quickly as possible in the upper parts of the catchment can just cause flooding downstream. Some areas of flood plain in rural areas can form natural reservoirs and it can help to hold water on the land here for a few days. Flood plain capacity can be enhanced by man-made embankments called Argys. These keep water off the fields during smaller floods. During bigger floods, water overtops the Argys, filling the flood plain. Large floodgates can then contain the water, releasing it slowly back into the river as levels fall.</p> <p style="text-align: right;">109 total 109 primary 0 catalytic</p>	<p>reservoirs simply to prevent flooding. We also know from recent history that draining land as quickly as possible in the upper parts of the catchment can just cause flooding downstream. Some areas of flood plain in rural areas, such as at the confluence of the Severn and Vyrnwy Rivers, can form natural reservoirs and it can help to hold water on the land here for a few days until river levels fall. Flood plain capacity can be enhanced by man-made embankments called Argys. These date back to the 18<sup>th</sup> century and they keep water off the fields during smaller floods. During bigger floods, water overtops the Argys, filling the flood plain. Large floodgates can then be used to contain the water, releasing it slowly back into the river as levels fall.</p> <p style="text-align: right;">148 total 112 primary 36 catalytic</p>

**Table D.2.** Textual analysis of five main 'teach' screens

	Text 1	Text 2	Text 3
Total wordage	709	406	709
Primary words	536 (75.6%)	362 (89.2%)	429 (60.5%)
Catalytic words	173 (24.4%)	44 (10.8%)	280 (39.5%)

**Table D.3.** Textual analysis of all nine tutorial screens

	Text 1	Text 2	Text 3
Total wordage	1036	727	1036
Primary words	536 (51.7%)	356 (49.0%)	429 (41.4%)
Catalytic words	414 (40.0%)	285 (39.2%)	521 (50.3%)
Other *	86 (8.3%)	86 (11.8%)	86 (11.5%)

\* Intro hook; program instructions