Discussion and Recommendations 5

5.1 Introduction

This final chapter is divided into 6sections. The focus of each section is presented in Table 5.1.

Section	Focus
5.2	A summary of the research
5.3 – 5.6	A reflection of the lessons learned from the study under three headings: • Reflecting on the methodology. • Summary of the findings and reflections on how this study supports or contradicts existing empirical research, a substantive reflection. • Reflection on the contribution of this study to the field of cognitive style, cognitive load and multimedia learning, a scientific reflection.
5.7	Recommendations for instructional design practice and further research.
5.8	Conclusion

Table 5.1: Overview of Chapter 5

5.2 Summary of the research

The framework of this study is outlined in Chapter 1. Chapters 1 and 3 provide concise definitions of the constructs investigated in this study.

5.2.1 Purpose of the study

The purpose of this study was to explore the role that cognitive load and cognitive style play in the successful achievement of learning outcomes when narrated animation and static images are used in multimedia learning formats in an authentic learning environment. The study also investigated the relationship between cognitive load, which is influenced by both the nature of the content and the specific design strategies used, and the cognitive style of the individual who uses different multimedia formats.

Cognitive load, a multidimensional construct, was defined as the burden imposed on the cognitive system when performing a particular task, as measured using a subjective rating of mental effort.

A clear distinction was made between cognitive and learning style. The position taken was that these are two independent constructs. The focus of this study was on **cognitive style**, and the widely used definition of cognitive style as 'an individual's preferred and habitual approach to both organising and representing information' (Riding & Rayner, 1998) was applied to this study.

Multimedia learning, also a multidimensional construct, was explained from three perspectives:

- various systems can be used to deliver the instruction, including paper, the computer or a cell phone;
- instruction is presented in one or more modes or presentation formats, for example text, illustrations and animation;
- at least one sensory channel through which information is processed, for example visual, auditory, tactile or any combination of these, is used.

These three perspectives define the domain of multimedia learning.

5.2.2 Background to the study

One of the aims of instructional design practice is to create an environment that facilitates and enables learning. A few key questions, arising out of my working environment, served as the drivers for the first steps of this study. These questions were initially very broad, and focused on issues such as:

- the balance between design and development effort, and between time and learning gain,
- the use of multimedia resources by the learner in the actual learning environment; and
- techniques and methods to improve the instructional design of multimedia instruction.

5.2.3 The literature review

The issues listed in Section 5.2.2 provided the direction for the literature review, which became increasingly narrow in focus. The initial literature review helped build the immediate case for the research questions and rationale of the study. This case was presented in Chapter 1. The aim of the review was to:

- ground the study within an appropriate theoretical framework, which was used to interpret the findings of the study,
- synthesize existing empirical evidence in order to establish the scope of the field, looking both for issues that have been addressed and issues that have been neglected in the literature,
- use this synthesis of existing findings as a benchmark against which to compare the findings of this study.

The outcome of this literature review is presented in Chapter 2. The theoretical frameworks selected for the study included the Cognitive Theory of Multimedia Learning, Cognitive Load Theory (CLT) and Riding's Cognitive Style Model. There were 3 themes in the literature review: cognitive style and multimedia learning, cognitive load and multimedia learning and multimedia learning in the health science education context. The cognitive load research literature directed the instructional design perspective of this study, while the cognitive style research literature directed the focus on the individual learner who engaged with instructional material in a multimedia learning environment.

Each of these themes were explored from various perspectives, including contributions to the relevant theory, measurement of the construct or concept, research initiatives, current trends and directions within the theme, specific research into how the theme is thought to influence learning and achievement and the development of heuristics for instructional design practice.

5.2.4 Finding the research question

One major research question, with five sub-questions, was formulated for the study. These questions have already been presented in both Section 1.6 (page 14) and Section 3.4 (page 127) of this dissertation.

5.2.5 The rationale of the study

The case for the study was built by exploring the theoretical, empirical, methodological, media and contextual dimensions in the cognitive load, cognitive style and multimedia learning literature.

Recent empirical research in the field of cognitive style and multimedia either failed to address the fact that the outcomes of the research might be due to cognitive overload (Ghinea & Chen, 2003) or only hinted that there might be a relationship between cognitive style and cognitive load (Graff, 2003b). The results of Riding, Grimley, Dahraei and Banner's study (2003) seemed to indicate that effective working memory capacity had a major influence on the performance of learners with specific styles. The relationship between these constructs was an avenue of research that did not seem to have been explored in great detail.

As early as 1994 the literature on cognitive load theory stated that there are three factors that contribute to cognitive load: task characteristics, learner characteristics and the interactions between these two (Paas & van Merriënboer, 1994a). One of the learner characteristics listed includes cognitive style. Cognitive load research has explored the influence of prior knowledge and learner experience on cognitive load in considerable depth (Kalyuga, 2006; Kalyuga, Ayres, Chandler & Sweller, 2003; Kalyuga, Chandler & Sweller, 2001), but the field has been strangely silent on the influence of cognitive style.

The content domain of a large majority of the studies included mathematics, science, technical subjects such as electrical circuits, computer applications or statistics. These were conducted primarily within primary, secondary and vocational education contexts. Any one of a number of contexts could have been selected for this study, and many different contexts have been used in cognitive style, cognitive load and multimedia research. The issues of both cognitive style and cognitive load are under-researched in health science education. There is also a lack of research in authentic learning environments and so this study was conducted in an authentic environment, using a topic that is part of the physiology curriculum for second year medical and dental students.

Methodological limitations in previous research included:

- studies that did not measure achievement directly (Riding, Grimley, Dahraei & Banner, 2003),
- small samples of under 100 participants (Ayres, 2006a; Dutke & Rinck, 2006; Mayer, Sobko & Mautone, 2003; Riding & Grimley, 1999: van der Meij, & de Jong, 2006),
- cautions that not all findings from experimental laboratory studies were easily generalisable to the classroom setting (Tabbers, Martens & van Merriënboer, 2004),
- failure to measure the cognitive load of the intervention (Chandler & Sweller, 1991; Ghinea & Chen, 2003; Mayer, Moreno, Boire & Vagge, 1999; Moreno, 2006), and
- giving the participants material to learn that was generally not relevant to their own coursework (Mayer, Fennell, Farmer & Campbell, 2004; Moreno, 2004).

Why consider animation and static images above other media options? The use of static images and text in instructional resources has received considerable attention in the research community since the early 1980s (Carney & Levin, 2002, Mayer, 2003; Mayer & Gallini, 1990; Mayer, Mautone, & Prothero, 2002; McKay, 1999; Moreno & Valdez, 2005; Verdi & Kulhavy, 2002). Improvement in technology has seen the increased use of sound, video, animation and 3D presentation formats in instructional materials. The impact of these newer media formats on cognitive processes are being researched with the same rigour and vigour that have been applied to researching the use of text and images in traditional classroom-based, face-to-face learning environments. My study contributes to the body of research investigating the newer media (animation in particular), with the added dimension of looking at how learners with different cognitive styles use and experience a strategy such as animation.

The empirical, theoretical and media imperatives suggest that there might be a relationship between cognitive load and cognitive style, but the existence and nature of such a relationship has not been explored in any depth.

5.2.6 The research methodology

5.2.6.1 The research approach

This was a quantitative study. An experimental and correlation design was used, in that the study aimed, on the one hand, to determine whether a particular intervention (multimedia learning with animation and images) made a difference for the learning outcomes of a group of participants, and on the other hand, it investigated whether the relationship between two factors (cognitive load and cognitive style) could have impacted the learning outcomes.

5.2.6.2 The research design

The variables under investigation were cognitive style (independent variable), cognitive load (independent variable), presentation format (independent variable) and achievement of learning outcomes (dependent variable).

A between-subjects design was used. In this design, subjects who had different cognitive styles (independent variable) were exposed to a different version (independent variable), and each version had a different cognitive load (independent variable). These three independent variables and their effect on the learning outcomes (dependent variable) were considered in the analysis of the results. The design also had the characteristics of a within-subjects (repeated measures) design. The repeated measure is performance in a knowledge and comprehension test, due to the fact that the same participants completed a pre- and posttest after random assignment to the research interventions.

The participants were randomly assigned to one of two presentation formats of the multimedia program, the narrated animation version (Version 1) or the static images & text version (Version 2). Both groups took Riding's Cognitive Styles Analysis test. Both groups took a pretest to determine prior knowledge at the beginning of the experiment. The pretest tested for recall and comprehension of knowledge. On completion of the instruction both groups took a posttest. The posttest assessed for recall and comprehension of knowledge, and included an additional section that tested for application of knowledge.

5.2.6.3 The research sample

The multimedia program designed to teach this topic is relevant for all health science students who study the topic of the Autonomic Nervous System (the population) for the first time (novice learners). In this study it was used by the students at the University of Pretoria (the target population). The second year medical, dental and physiotherapy students (the sample) were all studying the topic of the multimedia, the Autonomic Nervous System (ANS), at a time that coincided with this research program.

Initial sampling used convenience selection, which was deemed appropriate for this study for the following reasons:

1. The potential participants were available.
2. The group identified for selection had characteristics common to both the wider research population and the target population.
3. The number of students in the group allowed for an adequate sample size.

Once the sample had been identified, participants were randomly assigned to one of the treatment interventions. This is in line with the more rigorous approach required in experimental research.

5.2.6.4 The research data

Two separate datasets were collected for cognitive style. The two datasets were ratios that indicated the position of an individual on each of the two dimensions of Riding's Cognitive Style Model, namely the Wholist-Analytic (WA) and the Verbaliser-Imager (VI) dimension.

The methods of measuring cognitive load have been described in Chapter 2. This study used the self-report rating method. Smith (2007), working with the author of this study, measured the cognitive load using the direct measure dual-task methodology. Some of the data from Smith's study was used to answer the following question

> *What is the correlation between the participant's self-report of cognitive load and the direct measure of the cognitive load of the content?*

Self-report rating of cognitive load required the participant to indicate, on a nine-point scale, the mental effort they invested in understanding the content. Using the mental effort scores from the screens where this self-report was administered, a total mental effort score for the entire program was calculated for each participant. The mean of the individual scores for each intervention provided a total score for the cognitive load of the intervention.

Learner performance was measured using a pre- and posttest design. The computer-based test items were the same for the pretest and posttest, except that the order in which the questions were presented differed for the two tests. A score was calculated by the computer for each participant for both the pre- and posttest. A score of 1 was allocated for each correct answer, giving a total score of 22 for both the pre- and posttest. The posttest also included a pencil and paper test, which tested application of knowledge. There were 2 questions in this section of the posttest. The maximum score for each question was 10.

5.2.6.5 The format of the research intervention

Table 5.2 provides a summary of the structure of the two formats of the program.

	Animation version	Static images & text version
Formats used to present same content	Animation on one screen	Static images and text using six screens
Number of screens in program	19	23
Total number of animations in program	2	3

Table 5.2: A summary of the major design similarities and differences for the programs

The design of the multimedia program was informed by previous cognitive load and multimedia learning research. A detailed summary of the instructional strategies and instructional design was presented in Chapter 3. The actual screens for the intervention are illustrated in Appendices G – K.

5.2.6.6 The research hypotheses

Research hypotheses were formulated, both as null and alternate hypotheses, for each of the research questions. These will be presented in the discussion in Section 5.3.2 of this chapter.

5.2.7 Conducting the study

This study included two pilot studies and a main study. The window of opportunity for conducting both the pilot and main study was very small, due to the fact that the intervention was part of the normal study programme for the participants. Initially only one pilot study was planned but a second pilot study was necessary in order to sort out some technological problems with data collection.

The number of participants in the class and the final sample for the first pilot study and the main study is presented in Table 5.3.

Study	Class / Potential population	Class	Sample	Completed
Pilot 1	Physiotherapists	38	38	35
Pilot 2	Volunteers from the residences on the Faculty of Education campus at the University of Pretoria	250	40	33
Main study	2^{nd} year medical and dental students	262	245	238

Table 5.3: Profile of the research sample

The detail of how each study was conducted was presented in Section 3.18 and 3.19 of Chapter 3.

5.2.8 Analysis of the results

Two hundred and forty five data logs were retrieved for analysis. Once the data had been cleaned and prepared for analysis the final sample included 238 participants.

Data was analysed using the SAS® software system. Descriptive statistics were reported as frequencies, means and standard deviations (M ±SD) or standard error (M ±SE). Analyses of data were performed using regression analysis, the general linear model (GLM), t tests, Chi-square analysis and Pearson's correlation. Confidence intervals of 95% were reported. Differences were considered significant for p values ≤ 0.05. Effect sizes were measured and interpreted using the guidelines provided by Cohen (1988).

Chapter 4 presented the results of the data analysis, using the research questions and hypotheses as a framework for the chapter. These results and the interpretation thereof are summarised and discussed in Section 5.5 of this chapter.

The next three sections discuss and reflect on this study from three perspectives:

- A methodological discussion and reflection, which considers the research methodology and design.
- A substantive discussion and reflection, which compares the findings of this study with other related and similar studies.

- A scientific discussion and reflection, which discusses the contribution of this study to the existing body of knowledge in the three themes or research streams covered in this study. Areas for future research are recommended.

Together, these three perspectives aim to inform the reader about the lessons learnt from this study.

5.3 Methodological reflection

This section is a reflection on the research methodology, considering both the strengths and weaknesses of the final design. Using a framework similar to that of Chapter 3, this section reflects on the considerations and trade-offs that had to be made for each step of the research process. The design and the rationale for each decision made is explained in great detail in Chapter 3. This section reflects on whether the decision made was the correct one and considers the strengths and limitations of each component as it was implemented in the final design of this study.

5.3.1 The research approach and design

This was a quantitative study, which met the most important characteristics of quantitative research, explained in more detail in Chapter 3, Section 3.5. A quantitative approach aligns this study with the trends of the cognitive load, cognitive style and multimedia research streams, which use predominantly quantitative approaches to research. There are two strengths in working within the boundaries of existing trends:

- New research provides the necessary replication studies for in-depth investigation of specific research questions, which in turn contributes to theory building.
- Researchers are able to make meaningful comparisons between studies.

A potential weakness of always working within a quantitative paradigm is that issues which may arise from within a qualitative paradigm are often neglected. Qualitative designs also need careful planning in order to get to the rich data that is collected during qualitative research. During the analysis of the data collected in my study it became evident that a qualitative approach was needed in order to explore certain issues in depth, but the research design had not planned for a qualitative approach and that avenue of investigation was lost for this study.

The research design implemented was an experimental design using an authentic setting rather than a more controlled laboratory setting. Although the initial sample was a convenience sample, all subsequent allocation to the research intervention was done randomly, thereby ensuring that the design adhered to one of the most important requirements for an experimental design (Creswell, 2005; McMillan & Schumacher, 2001).

Chapter 3 (Section 3.6.1 page 129) made a case for using a between-subjects design, but there were also elements of a within-subjects (repeated measures) design, where the repeated measure included comparing the cognitive load of a specific screen with several other screens in the same version of the program. Nothing in the subsequent implementation of this design indicated that the decisions made were incorrect and the motivation for using this particular experimental design still stands.

The strength of using an authentic context was that the content was relevant to the sample. I do believe that participants are then more willing to invest the mental effort necessary in order to master the new learning. The fact that the content of the research intervention was part of their course also motivated them to use the multimedia program to learn. The participant groups really 'knuckled down' and got on with the task of learning the content after the briefing session. My observation of the groups during the laboratory sessions was that the majority of the participants took the session seriously. There were no disruptive participants and it was never necessary to remind them to stay on task and stay focused. Even after a long session in front of the computer many participants willingly completed the offline posttest that tested application of knowledge. Many of the participants provided substantial answers to the two questions. When the time spent on each version was analysed it was therefore very surprising to see the number of participants who spent less than adequate time, defined specifically for each version, using the multimedia.

The weakness of the authentic setting was the inability to predict all the variables that could influence the study. Even if it is possible to predict these variables, it is not possible to control for every variable in any experimental setting, and especially in an authentic setting. It is therefore possible that variables not identified and controlled for could have influenced the results of the study. For example, due to the fact that the campus did not have an experimental research laboratory for multimedia learning, the study had to be conducted in the laboratories used by the participants every day. Not only was the content authentic, but so was the physical environment. Factors that are difficult, if not impossible, to control in an environment like this include ambient noise, possible distractions from other participants, lighting of the room and the amount of workspace available next to the computer.

I had considered using the laboratory setting where there was more control, even though there were calls in the literature for research to be conducted in more authentic learning environments.

The issues important for the design of this study are listed in Table 5.4. The ease with which each these issues could be included in the design, given the time frames, availability of the sample and ability to randomly assign participants to research intervention, amongst others, are indicated for each of the options that were considered.

Issue	Authentic environment	Laboratory setting
Content relevant to sample	✓	✓
Content from the curriculum of the sample	✓	✓
Groups larger than 30	✓	?
Ability to find enough volunteers	✓	?
Randomisation	✓	✓
Control of variables	?	✓

Table 5.4: Reflecting on the research design

From Table 5.4 it appears that the issues considered were feasible in an authentic setting and so the final decision was made to go ahead with the research design using this authentic setting.

Two alternative approaches to the design of this study that should be considered in any replication studies include the

- point at which cognitive style is measured, and
- work needed to ensure that the cognitive load of the two research interventions differ significantly.

This study measured cognitive style at the start of the laboratory session. This contributed in part to the length of the session. A better design would have been to measure the cognitive style of the participants before the main session, if only to reduce the length of the experimental session. The end-result of this long session was that participants appeared to rush through the two open-ended paper-and-pencil questions that formed part of the posttest, particularly the final question. The low scores obtained by the majority of the participants for this second question could have been due to the fact that they simply did not have enough time to do justice to the question.

Another approach to consider is to use cognitive style as a starting point for randomisation. In other words, take the participants from one style dimension and randomly assign them to one of the two interventions, then take the participants from the second style dimension and do the random assignment. However, since the cognitive style model uses two independent dimensions, and each participant can be placed at some point on the continuum of both dimensions of style, a decision will have to be made about which of the two dimensions to use when grouping the participants.

The original intention was to go into the main study already knowing what the cognitive load of the particular version of the program was. The problems experienced with regard to the measurement of cognitive load in the pilot study have been described in detail in Chapter 3, Section 3.18. As a result the cognitive load was only measured for the first time during the main study. While there **was** a difference in cognitive load between the two versions and it was statistically significant, the effect size was still relatively small ($d = 0.33$). In the planning of future research designs the researcher should consider working on the cognitive load of each version until it is possible to take two interventions with clearly different loads, at a practical level of significance, into a second study.

5.3.2 The research sample

One of the strengths of this study was the large sample used – 238 participants. Once the participants had been randomly allocated to one of two versions they were randomly allocated again to one of two lessons for each version, and the size of each group was still between 50 – 60 participants.

Two of the issues considered in Table 5.4 regarding the planning, the design and sample were the ability to find volunteers and the ability to get research groups that were larger than 30 participants per group.

The participants in this study had an extremely heavy study programme. This in fact applies to all medical students in learning programmes across the globe. Any study that was not immediately relevant to their coursework would have to be conducted outside 'normal' working hours, which in real terms means after hours (after 17:00 during the week or over weekends). This time is usually spent studying or doing clinical work and so the chance of recruiting enough volunteers to ensure that the groups for the interventions were large enough was regarded as a high risk.

Homogeneity of the sample was also important. The cognitive load needed to be measured using learners who were novices with regard to their knowledge and understanding of the topic. Controlling homogeneity of the sample therefore required the sample to be selected from the same year group. Calling for volunteers from across the entire learning program would have interfered with this homogeneity and could have introduced other variables that might have been very difficult to control. The study might have produced an entirely different result.

Two methods for finding and retaining volunteers for a study is to offer them class credit (Dutke & Rinck, 2006; Renkl, Atkinson & Große, 2004; Rieber, Tzeng & Tribble, K. 2004; Wallen, Plass & Brünken, 2005) or pay each participant a small honorarium for their participation (Bodemer, Ploetzner, Feuerlein & Spada, 2004; Dutke & Rinck, 2006; Lee, 2007; Schwan, & Riempp, 2004). There were two limitations to this approach for this study. Firstly I was not in a position to offer extra credit for participation and neither was I able to secure enough funding for this study to pay each participant such an honorarium. Secondly, many learners in South Africa pay for their own tertiary education and therefore do a variety of small jobs (they must be small to fit into their study schedule) for extra money. The end result of paying an honorarium might have resulted in a very skewed sample for the study. The risk of using a skewed sample that was not truly representative of the larger population of medical students was regarded as high and one that could be avoided. This risk provided further justification for using content relevant to the sample in an authentic setting.

One of the limitations of a large sample includes the logistics of getting rich qualitative data to support the quantitative results. Even grouping 238 participants into focus groups of 10 – 15 participants per group would have been an enormous undertaking requiring a great deal of time. This was logistically impossible in this study, given the restrictions placed on my access to the sample. One of the weaknesses identified in the literature review was the small sample size in many of the studies reviewed. I aimed specifically to address this weakness in my study. The trade-off was the opportunity to collect qualitative data within the time frame for the research. The complexity of the design also called for a large sample in order to have large enough groups that could be included in the data analysis.

5.3.3 The research instruments and data

All the instruments, with the exception of the open-ended questions in the posttest, were presented to the participants in electronic format and embedded into the research intervention at the appropriate points. This format worked well and all the data was eventually written out exactly as designed. The instruments appeared easy to use and the basic format was not changed after the pilot study. The problems in the pilot study were due to hardware problems with the computers used during the pilot study and were resolved before the main study took place.

The electronic file produced for each participant was very long - up to 32 pages for each participant. The reason for this was that the same experimental session and sample was used by Smith (2007). The design of the multimedia program, with the instruments embedded in the program, was almost complete when it was decided that Smith would investigate the cognitive load using the direct measurement technique. The mass of data that was generated eventually had to be divided into three different data sets. This proved to be an enormous undertaking, requiring multiple rounds of checking to ensure that no mistakes were made. In hindsight, a better way of getting the data would have been to write it out to several files instead of one large file. Smaller files would have speeded up the process of getting the data cleaned and ready for analysis.

In three of the eleven times participants were required to self-rate the mental effort they invested they were required to consider the content of either two or three screens. This in itself might have imposed unnecessary cognitive load on some participants, even though they had been briefed about the nature of this question. They were not able to go back to previous screens, the reason for this having been explained in Chapter 4. The one change I will recommend in future research is that the self-report rating of cognitive load only require the learner to consider one screen per rating.

Table 5.5 summarises the alignment of the instruments used in this study with the criteria for good instruments. The criteria used are those described by Creswell (2005).

	Self-report rating of Cognitive Load	**Cognitive Styles Analysis**	**Pre- and posttests**
When was the instrument first developed?	1993	1991	Developed specifically for study
Has the instrument been updated?	Some researchers have used a 7-point scale rather than a 9 point scale	Yes	Not applicable
Does the study use the latest version of the instrument?	✓	✓	
Is the instrument widely cited by other authors?	✓	✓	
Are reviews available for the instrument?	✓	✓	
Is there information about the reliability and validity of scores from past uses of the instrument?	✓	✓	
Does the procedure for recording data fit the research questions in this study?	✓	✓	
Does the instrument contain accepted scales of measurement?	✓	✓	

Table 5.5: Alignment of the instruments used in the study with criteria for good instruments

5.3.4 The research intervention

The authentic nature of the study and the requirements from the Department of Physiology (that the content replace one lecture of 40 minutes in duration) determined the scope of the program. An additional period was made available to fit in the Cognitive Styles Analysis, the electronic questionnaires and the posttest. The animation version of the program (19 screens in length) was developed first. Once this version had been developed, the animation (screen 12) was replaced with other screens for the static images & text version. Other design changes were also made in order to make maximum use of the opportunity to explore the influence of different presentation formats on cognitive load and learning performance.

Participants were given unlimited time to complete the program. This aligned with the authentic nature of the study, but made control of some of the variables and the subsequent comparison of the data more complex. An example of this was the fact that participants were allowed to go back to previous screens, but the cognitive load was only measured once. When the time on screen was considered all the entries were used in the time calculation. The total time spent on a screen was then used to see if there was any relationship between time spent on the screen and the self-rating of cognitive load. Participants knew that they could go back to the previous screens. If they had been informed that they only had one chance to study the content they might have either spent more time on a particular screen or invested more metal effort (or both) and subsequently rated the cognitive load differently.

If I had to re-design this study I would make the following changes:

- Allow only forward progression through the program in order to collect the data, after which the participant can be allowed to go back to review selected screens.
- Limit the time to be spent on the program, or at least limit the time on screens where data was collected to facilitate better comparison.
- Consider moving the study from an authentic setting to a more controlled environment in a special experimental laboratory in order to better control the variables.

The actual implementation of the study went surprisingly well considering there were 70 students in the laboratory at one time. I used two laboratories simultaneously and had three research assistants in each laboratory to assist me. The participant groups were extremely co-operative. In spite of this I would recommend that smaller groups work together. If time had allowed I would certainly have considered running 3 – 4 sessions rather than 2 sessions. The session was also long and participant fatigue could have played a role in the outcomes.

5.4 Substantive reflection

This study addressed the following research question:

> ***The research question****: What is the relationship between cognitive style and cognitive load as factors in the achievement of learning outcomes when learning with different multimedia formats of the same content?*

This section will use each of the sub-questions and the appropriate hypotheses as a framework for the discussion and reflection.

5.4.1 The role of cognitive style in an authentic multimedia learning environment

The question:

> What are the cognitive styles of the participants taking part in the study?

5.4.1.1 The cognitive style profile of the participants in the study

The literature review, discussed in Chapter 2, did not provide enough empirical evidence of a specific cognitive profile style for health science education students. There is also limited evidence in the literature that cognitive style influences a person's choice of occupation (Murphy & Casey, 1997).

Fifty-seven percent of (57%) the sample had an Analytic style, 28% an Intermediate style and only 15 % had a Wholistic style.

> *The results from the analysis of the Wholistic-Analytic style dimension **supports** the hypothesis that more than 50% of the sample will have an Analytic or Intermediate style on the Wholistic-Analytic style dimension.*

Forty one percent (41%) of the sample had an Imager style, 31% were Bimodal in style and only 28 % were Verbalisers.

> *The results from the analysis of the Verbaliser-Imager style dimension do **not** support the hypothesis that more than 50% of the sample will have a Verbaliser or Bimodal style in the Verbaliser-Imager style dimension.*

The sample in this study appeared to lean slightly more to the Imager style. Although I conducted extensive searches across several databases, including MEDLINE, I found no literature that reports a Wholistic-Analytic style profile, using Riding's CSA, for health science students, a finding supported by Cook (2005). Luk (1998), using the Field Independence/Field Dependence category of cognitive style, investigated the relationship between cognitive style and academic achievement in a group of nursing students who were following a distance education programme. Two studies were reported. In the first study 53% of the sample were Field Independent (a style label that can be compared with the Analytic style), while in the second study 65% of the sample were Field Independent. Chapman and Calhoun (2006), in a recent study that validated three learning styles, concluded that the medical students were found to be significantly more Field-Independent, favouring a more Analytic approach to learning. It would appear that medical and dental students appear to be more analytic in style, but there is not enough evidence in the literature to confirm that the results of my study are aligned with a particular cognitive style profile for medical and dental students. Such profiles have not been adequately researched and described in the health science education literature. There are **learning style profiles** reported in the literature (Chapman & Calhoun, 2006; Cook & Smith, 2006; Martin, Stark & Jolly, 2000; Smits et al., 2004) but this study has taken the position that cognitive style and learning style are two different constructs.

I found very little literature that reports on empirical studies that investigated the visual-verbal dimension of style in health science education. Effken and Doyle (2001) explored the role of the visual-verbal cognitive style with a small sample of undergraduate nursing students. The style was measured using an adaptation of Richardson's Verbaliser-Visualizer Questionnaire (VVQ). Seventy percent of the sample were found to be visualisers. Effken & Doyle did not indicate whether this profile was typical of nursing students in general. The sample (n=18) was in fact so small it is not possible to generalise to a larger population of health science education students. It is therefore not possible to determine whether the findings of this study are aligned with a general Verbaliser-Imager style profile for health science education students in general, and medical and dental students in particular.

The analysis of cognitive style was extended to consider cognitive style **and** gender and cognitive style **and** culture and cognitive style **and** time spent on the lessons for two reasons:

- The findings in the literature, discussed in Chapter 2, are mixed as far as the relationship between cognitive style and gender is concerned.
- Cognitive style profiles have not been adequately researched and described in the health science education literature. Findings related to gender, culture and time used during multimedia learning extend the knowledge about the cognitive style profile of health education students.

5.4.1.2 Cognitive style and gender

The analysis in this study indicated that there was no significant difference between the gender groups for both the Wholistic-Analytic and Verbaliser-Imager style dimensions. The cognitive style for each gender was determined for each of the three categories of the Wholistic-Analytic cognitive style. The effect size obtained for the Wholistic style ($d = 0.35$) was the only effect size that indicated that there might be a difference in cognitive style for gender. This is still a small effect and is not large enough to conclude without a doubt that the difference between male and female participants, who are health science students, for the Wholistic-Analytic style has any practical significance.

The results reported in the preceding paragraph support the argument that there is, in general, no gender effect for cognitive style (Riding et al., 1995). Graff, Davies and McNorton (2004) explored cognitive style and cross cultural differences in Internet use and attitudes to computers. Cognitive style was measured using the CSI of Allinson & Hayes. They found no gender effect and no interaction between gender and cognitive style. Take note that these conclusions from the literature were made based on empirical research that did not include the health science education field.

Further evidence that there is in fact little difference between the genders for most psychological variables is provided by Hyde (2005), who proposed the Gender Similarities hypothesis. The hypothesis states that most psychological gender differences are in the close-to-zero ($d \leq 0.10$) or small ($0.10 > d < 0.35$) range, a few are in the medium ($0.36 > d < 0.65$) range and very few are in the large (*d* = 0.66 – 1.00) or even very large (*d* < 1.00) range. Hyde reviewed 46 meta-analyses of gender difference studies, including meta-analyses that assessed cognitive variables, and found support for this Gender Similarities hypothesis.

5.4.1.3 Cognitive style and culture

In a stepwise regression, using Wholistic-Analytic style as the dependent variable, culture was retained as one of the variables entered into the regression. A confirmatory GLM (General Linear Model) returned a statistically significant finding, with the culture of the participants accounting for the significant result. For the Wholistic-Analytic style dimension, Non-White participants were statistically significantly more Analytic than the White participants.

The relationship between cultural groups and cognitive style is not addressed extensively in the current cognitive style literature. Graff, Davies and McNorton (2004) also compared cognitive style and differences in Internet use and attitudes to computers using nationality and age as independent variables. They compared the Chinese and British cultures. Chinese students were found to have significantly more positive behavioural attitudes towards computers than UK students. Chinese students seemed to use the Internet more, but there were no cultural (nationality) differences between knowledge of the Internet and ease of use. Graff et al. (2004) concluded that the practical implications of their study was that designers needs to pay more attention to these cultural differences in designing web-based instruction.

The learning styles literature has addressed the issue of culture and ethnicity (Cho & Forde, 2001). Culture has also received attention in the instructional design literature (Chen, Mashhadi, Ang & Harkrider, 1999; Hedberg & Brown, 2002). Cho and Forde (2001) found a significant relationship between ethnicity (White versus Non-white) and learning styles, although they did point out that the purpose of their study was not to conclude that a certain ethnic group falls into a certain learning style category, but rather to encourage educators to use multiple teaching and assessment methods to accommodate style differences. They did not report on the probable causes for these differences.

The European Learning Styles Information Network (ELSIN) announced their 2008 conference towards the end of 2007. The theme for this conference is ‘Style and cultural differences - how can organisations, regions and countries take advantage of style differences’(ELSIN 2008, 2008). The papers presented at this conference might shed more light on the relationship between cognitive style and culture.

What are the implications of the finding of my study, namely that Non-White participants were statistically significantly more analytic than the White participants? The effect size was 0.40, indicating practical significance, which cannot be ignored. The implications of this finding are difficult to determine without knowing what assumptions are made within the research population about the styles and characteristics of students from different cultural groups, especially if these are minority groups. Cultural differences need to be explored simultaneously from multiple perspectives (psychological, cognitive, social, affective) in order to really understand the impact of culture on behaviour and learning.

5.4.1.4 Cognitive style and time spent on program

The analysis was done at two levels. I first looked at the time for the program as a whole before considering time at screen level.

Time spent on the program as a whole for the Wholistic-Analytic style dimension

For the Wholistic-Analytic style dimension it was predicted that the Analytic learner would spend more time studying the content of the program than the Wholistic learner. The Analytic learners generally spent more time than the Wholistic learners studying the content of both versions of the program. This difference was significant for the static images & text version, but not for the animation version. I also found that the Wholistic learner spent about the same amount of time on each version. This behaviour would seem to confirm the approach a Wholistic learner takes to learning (Riding & Rayner, 1998), in that the Wholistic learner's tendency to scan in order to get a big picture reduced the time spent on the static images & text version, even though there were more screens in this version. Proportionally the Wholistic learner spent more time studying the content of the animation version. Use of animation would constrain this scanning strategy of the Wholistic learner as he/she would have to watch the whole animation at least once in order to understand it.

> *The hypothesis that the Analytic learner will spend more time studying the content of the program than the Wholistic learner was therefore supported for the static images & text version of the program, but not for the animation version.*

Time spent on selected screens for the Wholistic-Analytic style dimension

At screen level there was virtually no difference between the Wholistic and Analytic learner in the amount of time spent studying the animation. The difference in the time spent on the static images & text version for the comparison of the Wholistic and Analytic style was not statistically significant, but the effect size indicated that the difference **was** practically significant.

The analysis of the time **each style group** spent on the different screens in the two versions indicated that the Analytic learner spent significantly more time on the static images & text screens (screens 13-16) compared to the animation screen (screen 12). The Wholistic learner spent more time on the static images & text version but it was not significantly more than the time spent on the animation screen.

A suggestion for the fact that both style groups spent approximately the same amount of time studying the content of the animation could be viewed from two perspectives: either animation is a strategy that suits both styles in this style dimension equally well **OR** animation as an instructional strategy does not accommodate either style as effectively as does the strategy that used static images & text across several screens. The animation allows for less control by the individual learner. The learner could pause and restart the animation and use the progress bar to 'scroll' backwards and forwards to selected points in the animation. The use of the control features of the animation was not observed or tracked electronically. Each learner would have had to look at the entire animation at least once, from start to finish in order to understand it, irrespective of their particular cognitive style. The Wholistic learner would most likely only scan the content in the static images & text version, while the design would allow the Analytic learner better opportunity to study the detail without the interference determined by the pace of an animation. Since the Analytic learner processes more elaborately than the Wholistic learner they would take more time to work though the content and this was confirmed in this study. The instructional design of the static images & text version better accommodated the Analytic style and this type of learner seemed to make use of the opportunity by spending more time studying the content on these screens.

Time spent on the program as a whole for the Verbaliser-Imager style dimension

For the Verbaliser-Imager dimension is was predicted that the Verbaliser will spend less time studying the content of the program than the Imager. The intervention was visually rich. There were very few screens that did not use an image to illustrate the content. It was therefore expected that it would suit the Imager style, even though the literature indicates that the Verbaliser style can also use and benefit from visually rich content (Riding et al., 2003).

Participants with an Verbaliser style did spent less time studying the animation version than the participants with an Imager style, but the difference was not significant. The two style groups spent about the same time on average studying the content of the static images & text version and the difference was predictably not significant.

> *The results from the analysis therefore do **not** support the hypothesis that the Verbaliser will spend less time studying the content of the program than the Imager.*

The animation included both images and narration and would therefore suit both style groups. The fact that the mean time for the two styles was not significantly different suggests that a narrated animation, as an instructional strategy, accommodates both the Verbaliser and Imager style. The static images & text version also included both images and text and also appeared to accommodate both the Verbaliser and Imager style.

As discussed in Chapter 3 and repeated here in order to avoid a split-attention effect, there were 29 images across 19 screens in the animation version and 39 images across 23 screens in the static images & text version. There were proportionally more images in the animation version. Verbalisers are more likely to select the text above the image when studying content and so it was predicted that the Verbaliser would spend proportionally more time studying the content of the static images & text version than the animation version.

The analysis of the time **each style group** spent on the two versions indicated that the Verbaliser style spent significantly more time on the static image & text version than the animation version. Proportionally however, because of the difference in the number of screens across the two versions, the Verbaliser learner spent more time studying the content in the animation version than the static images & text version.

> *The hypothesis that the Verbaliser will spend proportionally more time studying the content of the static images & text version than the animation version was therefore* ***not*** *supported.*

Each style group seemed to spend more time on the version that suited their style. The static images & text version was thought to better suit the Verbaliser learner. While the difference was not statistically significant, the Verbaliser **did** spend more time studying the content of the static images & text version than he/she spent studying the animation version. The animation version was thought to better suit the Imager learner. While the difference was not statistically significant the Imager did spend more time studying the content of the animation version than he/she spent studying the static images & text version.

Time spent on selected screens for the Verbaliser-Imager style dimension

I then compared the time **each style group** spent on selected screens. There was very little difference in the time spent by the Verbalisers and Imagers on screen 12 in the animation version. It was not statistically significant. This was similar to the finding for the Wholistic-Analytic style dimension in this study. The difference in time spent by the Verbalisers and Imagers on screens 13-16 in the static images & text version was larger but still not statistically significant. The hypothesis that the Verbaliser will spend less time studying the content of the program than the Imager was also not supported at screen level for both the animation and the static images & text versions.

In considering each style separately the analysis indicated that both the Verbaliser and Imager style spent significantly more time studying the content of the static images & text version. A possible explanation for this could be that the user had more control over the content in the static images & text version, and used the time to review the text and/or images carefully.

In conclusion

The literature appears to be silent on the issue of the amount of time learners with different cognitive styles need in order to learn from multimedia instructional materials. Most of the studies reviewed gave the participants in the study a fixed amount of time to work through the instructional intervention (Ford & Chen, 2001; Graff, 2003b, 2005). These studies were carried out in formal experimental settings rather than in more authentic settings. I do not think that the amount of time needed for studying the content will impact on the instructional design as such, other than making a decision about what to include in each program or topic, but knowledge of time needed by learners with different cognitive styles has practical implication in authentic learning environments. When use is made of electronic resources in any learning environment, but particularly in a blended learning environment, it becomes necessary to know how much time to allocate for these learning opportunities.

The fact that Verbalisers spent more time studying the content of the programs supports the idea that Verbalisers are similar to Analytic learners in terms of their processing needs (Riding et al., 2003). It is possible that the analysis would have yielded different results and/or the findings would have been better explained if I had used certain style combinations in the research design and analysis. The style combinations for Riding's model would yield four groups namely Analytic-Verbalisers, Analytic-Imagers, Wholistic-Verbaliser and Wholistic-Imagers (Riding & Rayner, 1998). Researchers suggest that the Analytic-Verbaliser and Wholistic-Imager styles are unitary and the Analytic-Imager and Wholistic-Verbaliser styles are complementary (Evans, 2004; Riding & Rayner, 1998), but also call for further investigation into the preferences of these groups. A unitary style means that the strengths or weaknesses of the two style dimensions reinforce each other, while a complimentary style means that the strengths of the one dimension compensates for the weaknesses of the other dimension. The suggestion is that learners with a unitary style combination are less able to compensate for the weaknesses in the particular profile, while learners with a complementary style combination will use the strengths of one style to compensate for the weaknesses of the other style. John & Boucouvalas (2002c) explored the effects of the cognitive styles (using Riding's model) on user performance with audio. They considered the dimensions separately and then in combination and found different results when the sample was grouped differently. They concluded that it is not always possible to draw conclusions by analysing the results of one dimension in isolation as the influence of the other dimension can also effect the results.

The one implication of the findings related to cognitive style is that a special effort is needed to assist the Analytic learners in seeing a big picture view. In the context of a health science profession, where the trend is towards holistic, comprehensive patient care, this would mean making sure they are able to view and manage their patient holistically, by providing them with enough opportunity to improve the skills needed to see the big picture view. Another use for this finding lies in a question the instructional designer should ask, namely 'Will the planned design disadvantage the learner in any way?'

5.4.2 The role of cognitive load in an authentic multimedia learning environment

The question:

> How do the participants rate the cognitive load of selected multimedia content?

The cognitive load was determined for the program as a whole and for specific screens in each version. In considering the program as a whole it was hypothesised that the animation version would have a higher cognitive load than the static images and text version.

5.4.2.1 Comparison of cognitive load across the two versions

The results indicated that when considering the cognitive load of the program as a whole, the static images & text version had a statistically significantly higher cognitive load than the animation version. The effect size of 0.33 is in the small to medium range, therefore interpretation of practical significance must be made with caution.

> *The hypothesis that the animation version would have a higher cognitive load than the static images & text version was therefore* ***not*** *supported.*

This was an unexpected finding. There are several factors that could have contributed to this finding. Firstly the length of the program could have played a role. Secondly there were considerably more pop-ups and screens with text /image combinations in the static images & text version compared to the animation version. Although a definite effort was made to control for the split attention effect on the screens with text and images, it is possible that there was still some split attention that cumulatively influenced the extraneous cognitive load. The navigation in some of the screens in the static images & text version was more complex than the animation version, which presented the content in a linear format. In the static images & text version the user had to use a menu to access the different screens. The section could be navigated in a non-linear fashion. This could have increased the extraneous cognitive load.

The findings of my study were similar to those of Tabbers et al. (2004). Working in an authentic learning environment, they found that the mental effort ratings in their study were also low (in what I have categorised as the medium range) and were not statistically significant across the different versions of the research intervention they used.

However, when looking at recent literature a different perspective could also explain the findings of my study. Hegarty, Kriz & Cate (2003) suggested that when a learner uses animation he/she is only required to perceive the temporal changes taking place in the animation, while when using static representation the learner must infer these changes. This is assumed to be more difficult and requires greater mental effort. If this is the case, in order to understand the content of the static images and text version the learner would have had to invest more mental effort compared to the animation version. The higher cognitive load found in the static images & text version could therefore be indicative of germane load rather than extraneous load. This load is not detrimental to learning, provided the total load is kept within the limits of working memory. The cognitive load for each version of the program was between 5 and 6, a range I have described as medium load.

Moreno and Marley (2007) investigated verbal and visual abilities and preferences, learning performance and perceptions about learning using three presentation formats. The multimedia program, covering the topic of Astronomy, included a narrated explanation with animated graphics, static graphics or no graphics (narration alone). Participants were given a program rating questionnaire to complete. An exploratory factor analysis of the items included in the program-rating questionnaire revealed an affective factor and a cognitive load (CL) factor. There were no learning or cognitive load differences between the animation and static images group.

The next step in the analysis was to consider a series of screen-wise comparisons across the two versions. These sets of screens isolated the instructional strategies and presentation formats: animation versus static images & text, and the use of pop-ups. It was hypothesised that the screen using animation will have a higher cognitive load than the alternative version using static images & text. It was also hypothesised that at screen level, the cognitive load will be the same in each version where the content and presentation format are the same.

Returning to the discussion of the findings in my study, the cognitive load for screen 12 (animation) was significantly higher than the cognitive load for the alternative static images & text screens (screen 13-16).

> *The hypothesis that the screen using animation will have a higher cognitive load than the alternative version using static images & text **was** supported.*

I was only able to find two references in the literature since 2006 where the cognitive load was measured in the process of comparing animation and static images as presentation formats. Van Oostendorp and Beijersbergen (2007) used two sets of learning material (working of the human heart and the flushing and refilling of a toilet system) to compare the understanding, confidence and mental effort of three instructional conditions (animation, static images and guided animation). They found no significant difference in the amount of mental effort invested in the different instructional conditions. Höffler, Sumfleth and Leutner (2006) investigated the role of spatial ability when learning from an animation or a series of static pictures. The cognitive load was measured but not reported. I contacted Höffler and was able to review sections of his doctoral thesis (Höffler, 2007). I established two things: firstly he adapted the subjective rating scale of Paas & van Merriënboer (1993) for the measurement of cognitive load and secondly, and more importantly, he found that the cognitive load was not statistically significantly different for the two formats of the instructional material. The results of my study therefore contradicts the findings of both studies cited here.

Theoretically, animation appears to be the strategy of choice when explaining processes (in my study the process and path of a stimulus in the autonomic nervous system) but the research comparing animation with static images does not support this assumption (Tversky, Morrison & Bétrancourt, 2002). Higher cognitive load does not imply poorer performance, but there is a risk that a higher load will have a negative impact on learning (cognitive load and performance will be discussed later in this chapter). Paas, van Gerven and Wouters (2007) suggest that in the research reviewed by Tversky et al. (2002) it is possible that the animations were not designed with sensitivity towards the processing limitations of the working memory and could therefore have interfered with the learning process.

Unfortunately the methods used to measure cognitive load do not inform about the relative contributions of each type of cognitive load (Whelan, 2007). If the cognitive load obtained in the animation version of my study was due to germane load then it is unlikely that this higher cognitive load was detrimental to learning, since the total load was in the category I had described as the medium range. Germane load has been shown to facilitate the transfer of learning (Bodemer et al., 2004; Schnotz & Rasch, 2005). However, if the cognitive load was due to extraneous load then methods must be sought to reduce the cognitive load, so that cognitive resources can be freed up for deeper processing.

Paas et al. (2007) discuss some suggestions for designing animation that is sensitive to the limitations of working memory: segmenting animations, cueing learners to specific features during the animation, increasing the level of interactivity with the animation. Paas et al. (2007) compared the instructional efficiency of three instructional strategies that were used **after** learners had studied an animation. They found that interaction that required the learner to either reconstruct or deconstruct the process studied in the animation was superior to the non-interactive strategy that merely required the study of static images extracted from the animation.

5.4.2.2 The relationship between cognitive load and cognitive style

There was no evidence in the literature of attempts to consider the relationship between cognitive load and cognitive style. Using existing empirical evidence about how the design of learning material influences cognitive load **together with** existing evidence about how learners with different cognitive styles are thought to process information and learn, the reasoning and argument behind the hypotheses of the relationship between load and style were presented in detail in Section 3.16.3 in Chapter 3.

Wholistic learners prefer to learn using big picture views, while the Analytic learner prefers step-by-step detail. The animation presented the information in one continuous session, although the user could control the pace of the animation. It was expected that this would suit the learner who had a

Wholistic style. The animations were also very visual and this was expected to suit the learner who had the Imager style. It was therefore hypothesised that the:

- cognitive load of the animation would be lower for the Wholistic learner than for the Analytic learner.
- cognitive load of the text and static image version, used as an alternative for the animation, would be lower for the Analytic learner than for the Wholistic learner.
- cognitive load of the animation would be lower for the learner with an Imager style than for the learner with a Verbaliser style.

Cognitive load and Wholistic-Analytic style

The difference in the cognitive load of the program for the Wholistic and Analytic style was small for both the animation and the static images & text version. The cognitive load of the animation was lower for the Wholistic learner than for the Analytic learner, but the difference was not statistically significant.

> *The hypothesis that the cognitive load of the animation would be lower for the Wholistic learner than for the Analytic learner was therefore* ***not*** *supported.*

It was hypothesised that the cognitive load of the text and static image screens would be lower for the Analytic learner than for the Wholistic learner. It was found however that the Analytic learner reported a higher cognitive load than the Wholistic version, although the difference was not statistically or practically significant.

> *The hypothesis that the cognitive load of the text and static image screens, used as an alternative for the animation, would be lower for the Analytic learner than for the Wholistic learner was therefore* ***not*** *supported.*

I analysed the cognitive load from another perspective and compared the cognitive load for each style by version used. Wholistic learners who used the static images & text version reported a higher cognitive load than the Wholistic learners who used the animation version. The difference was not statistically significant, but the effect size, in the small to medium range, indicated possible practical significance. Analytic learners who used the static images & text version also reported a higher cognitive load than the Analytic learners who used the animation version and the difference was statistically significant, with an effect size in the small to medium range, indicating possible practical significance.

Why would there be a statistically significant difference between the two versions for the Analytic learner, but not for the Wholistic learner? I present two explanations for this finding. The first relates to the relationship between cognitive style and cognitive load and the second relates to the issue of sample size.

The ability of the Wholistic learner to scan rather than study detail would mean that this learner would not be influenced too much by the fact that the static images & text version was slightly longer, while the Analytic learner would pay great attention to the detail of every screen. The static images & text version was also more text rich than the animation version. The Analytic learner, with their elaborate processing style, would experience and report that they invested more mental effort (and therefore a higher cognitive load) in the static images & text version compared to the animation version. In reconsidering the literature that describes how Wholistic and Analytic learners process information (Riding et.al, 2003; Riding & Rayner, 1998) the direction of the results (Analytic learners report a higher cognitive load than Wholistic learners) provides support for the findings of Riding et al. (2003) namely that the Analytic learner processes information elaborately. I suggest that this elaborate processing by the Analytic learner would require more investment of mental effort (and therefore higher cognitive load) than the scanning strategy of the Wholistic learner. I also suggest that this increased mental effort is a manifestation of germane cognitive load, which is beneficial for learning. The results of this deeper processing should be reflected in the learning performance, as measured by the posttest. I will consider this later in this chapter.

Regarding the issue of sample size, it is documented in the research methodology literature that statistical significance tests have a tendency to yield small *p* values as the size of the data set increases in size (Ellis & Steyn, 2003). The two sets of comparisons in this section of the study are a case in point. The value of n for the comparison of the means for the Wholistic learner was 73, while the value of n for the comparison of the means for the Analytic learner was 162, more than double that of the Wholistic group. The absolute difference in the means (mean difference between the animation and static images & text version for the Wholistic learner and mean difference between the animation and static images & text version for the Analytic learner) were approximately the same and yet it was the group with the larger sample size that returned the significant *p* value. In turning to look at the effect size the data indicates that the effect sizes were very similar. Both were in the small–to-medium range, 0.31 and 0.34 for the Wholistic and Analytic group respectively suggesting that this is a visible effect.

The higher cognitive load for the Wholistic learner in the static images & text version is not necessarily related to their processing style but to the fact that they would have needed more time to study the longer version. Time has been regarded as an indicator of cognitive load (Chandler & Sweller, 1991).

Cognitive load and Verbaliser-Imager style

The total cognitive load for the Verbaliser-Imager style dimension, irrespective of version, was almost exactly the same: mean cognitive load of 5.3070 for the Verbaliser style and mean cognitive load of 5.3472 for the Imager style. When the versions were considered separately the mean cognitive loads for each style were also very close to each other. There does therefore not appear to be any relationship between the Verbaliser-Imager style and cognitive load.

A suggestion for this finding relates to the definition of the Verbaliser-Imager style dimension. The Verbal-Imagery style addresses how the individual is inclined to represent information during thinking. This inclination can be towards verbal representation or thinking by means of mental images (Riding & Rayner, 1998). Representing information and processing information (the focus of the Wholistic-Analytic style dimension) are two different cognitive activities. I suggest that it is the processing of information that imposes the load (germane load) rather than the representation of information.

The learner with an Imager style did report a lower cognitive load than the learner with a Verbaliser style for the animation version, but the difference in the cognitive load was not statistically or practically significant.

> *The hypothesis that the cognitive load of the animation version will be lower for the learner with Imager style than for the learner with Verbaliser style was therefore* ***not supported.***

The relationship between cognitive load and Verbaliser-Imager style was also analysed from another perspective. The cognitive load for both the Verbaliser and the Imager learner was compared by version. Both the Verbalisers and Imagers experienced a higher cognitive load in the static images & text version. This is most likely related to the length of the program. In considering the Verbaliser style, the difference in the cognitive load by version was neither statistically or practically significant.

A comparison of the versions for the Imager style did prove to be statistically significant and the effect size also suggested practical significance. The static images & text version had the higher load and this is most likely related to the length of the program.

5.4.2.3 A comparison across the research interventions where content and presentation format were the same

A comparison of the screens in the two versions that had the same content and the same presentation format was conducted. Two sets of screens were compared: screen 19 (animation version) was compared with screen 23 (static images & text version) and screen 5 (animation version) was compared with screen 5 (static images & text version). The findings in this group of comparisons were completely unexpected when viewed from a purely cognitive load perspective.

Screen 23 had a significantly higher cognitive load than screen 19. Screen 5 in the static images & text version also had a higher cognitive load than the same screen in the animation version.

> *The hypothesis that at screen level, the cognitive load will be the same in each version where the content and presentation format are the same was therefore **not** supported.*

It is possible that participant fatigue contributed to the cognitive load difference between screen 19 and 23. Both screens were the last screen in the program and the static images & text version was longer than the animation version. It is unlikely that the finding with regard to Screen 5 can be explained by participant fatigue, as this was the fifth screen in the program for both versions.

Another explanation for these differences could be related to the method of measuring cognitive load. I reviewed the cognitive load obtained by the direct measurement method for these screens (Smith, 2007). There were very small differences in the mean cognitive load of the screens (screen 19 compared to screen 23, screen 5 in each version). The effect sizes for the comparison of the means were calculated and were found to be too small to have any practical significance. If the direct measurement method is more accurate in determining cognitive load then it can be said that the cognitive load of these screens is similar.

5.4.3 The correlation between self-report and direct measurement as techniques in measuring cognitive load

The question:

> What is the correlation between the participant's self-report of cognitive load and the direct measure of the cognitive load of the content?

I compared two methods of determining cognitive load and investigated the correlation between the cognitive load scores obtained from each technique. The first hypothesis was that the two methods used to measure cognitive load would return results that are the same for each version.

Using the direct method of measurement it was found that the animation version had a significantly higher load than the static images and text version. This contradicts the finding of the subjective rating method where the static images and text version had the significantly higher cognitive load.

The cognitive load measured obtained from the two techniques were compared by version. For the animation version the direct method of measurement returned a significantly higher load than the subjective rating technique, while for the static images & text version the difference between the cognitive load measurements was small and not significant.

> *The hypothesis that the two methods used to measure cognitive load will return results that are the same for each version was supported for the static images & text version, but not for the animation version.*

The next question is 'Which method is the most accurate?' The one advantage of the direct measurement method is that it measures the cognitive load at the time when the load is induced, while the subjective rating is made after the learning event (Brünken, Plass & Leutner, 2003). Paas et al. (2003) state that there is empirical evidence that reliable measures can be obtained with unidimensional scales (like the one used in my study) and that these scales are sensitive to small variations in cognitive load. They are regarded as valid, reliable and unintrusive measures. In the same article they state that the direct method using secondary tasks is also a highly sensitive and reliable measure. Brünken, Plass & Leutner (2003), in their discussion of the different methods used to measure cognitive load, indicate that although the subjective method appears to be reliable, it is still unclear how mental effort relates to actual cognitive load.

Another avenue to explore is to look at what type of cognitive load each method is supposedly measuring. The literature on studies that have compared different methods of measuring cognitive load is sparse. Paas et al. (1994b) compared physiological measures (heart rate variability) and the subjective rating technique and found that the method using heart rate variability was neither reliable or sensitive enough to detect differences in task complexity. The subjective rating scale proved to be a reliable and sensitive measure. Task complexity relates to intrinsic load. Whelan (2006) also compared two approaches to the measurement of cognitive load – self-report questionnaires and dual-task methodology. The different instruments did not produce uniform results. He concluded that each method reflects a different type of cognitive load. The dual task methodology showed its strength in assessing extraneous load, while Paas' instrument was only sensitive at the mean levels of significance to variations in extraneous load and did not show adequate sensitivity to high and low extraneous cognitive load conditions. Whelan (2007) reviewed the literature, looking for alternative approaches to measuring cognitive load. His overview of the efforts in the field of functional magnetic resonance imaging (fMRI) and his subsequent argument for using neuroimaging techniques adds to this new direction for cognitive load measurement research (Tomasi, Chang, Caparelli & Ernst, 2007).

Using the conclusion of Whelan (2006) to explain my findings, I suggest that if the intrinsic load of the material used in my study is regarded as equal for the two versions, there should have been no difference in the cognitive load of the two formats, as measured using the direct method of measurement. This was not the case, suggesting therefore that the extraneous load was different for the two versions and this influenced the outcome of the measurement. If the direct method using dual-task methodology is the more sensitive measure for extraneous load then I suggest that there were extraneous load issues in the animation that outweighed the extraneous load imposed by the navigation through the static images & text version, as discussed in Section 5.4.2.1 of this chapter.

If this is the case, then the design of the animation in my study did not apply enough of the techniques thought to create animations that are sensitive to the limitations of working memory. Research into the efficacy of these techniques was published in the 2007 special issue of the *Applied Cognitive Psychology* journal and include attention cueing (de Koning, Tabbers, Rikers & Paas, 2007) adequate learner control (Hasler, Kersten & Sweller, 2007), segmenting methods (Moreno, 2007) and interactive manipulation of static images following animation (Paas et al., 2007). The animation in my study did allow for learner control in the form of stop, pause and play buttons being available to the participants while watching the animation. The use of these controls was not tracked however.

If the screens presenting content in the static images & text format, used as an alternative for the animation is seen as manipulation of intrinsic load, then I should have found that the static images & text version had the lower cognitive load. This was not so when considering the program as a whole, but it **was** true when I only considered the relevant screens that were of particular interest in the study (screen 12 compared to screens 13-16).

It was also hypothesised that there will be a positive correlation between the self-report method and direct measurement method for determining cognitive load. The correlation between the two methods was found to be positive, but very low ($r = 0.07$) and not statistically significant.

> *The hypothesis that there will be a positive correlation between the self-report method and direct measurement method for determining cognitive load was* ***not*** *supported.*

If the finding of Whelan (2006) is used to explain this result then this low correlation is not surprising as the two methods are not equally sensitive to differences in cognitive load. There is also the possibility that the two methods are not measuring the same type of cognitive load: intrinsic, extraneous and germane load.

5.4.4 Presentation formats and their influence on cognitive load

The question:

> To what extent do the presentation formats influence cognitive load?

This reflection looks at the cognitive loads of the individual screens, both within a particular version and between versions, and considers the design differences between the screens. How did the design of the screens influence cognitive load, and what design strategy might have been responsible for the differences in cognitive load?

5.4.4.1 Comparing screen 12 and screen 19 in the animation version

Both screen 12 and 19 consisted of a narrated animation only. Screen 12 had a statistically significantly higher cognitive load than screen 19, although the effect size was in the small to medium range ($d = 0.31$). This was unexpected, since screen 12 included content that had already been covered in the curriculum, while screen 19 covered new content.

Factors that could have influenced this difference in cognitive load between screen 12 and 19 include the duration of the animation and the transitory nature of the animation. There were a number of different visual views in each animation: parts of the image changed (some of these changes were quite subtle), there were moving elements (excluding text animation), text labels appeared and disappeared, some of the text labels were highlighted and the highlighting technique often moved to another label. Even though narration was added to reduce the processing in visual memory and therefore decrease the extraneous cognitive load, there were still enough visual elements changing to induce extraneous load, over and above the intrinsic load of the content.

The number of changes in the visual views (or scenes) were counted for each of the animations. In screen 12 there were approximately 44 scene changes across the 1 minute 45 seconds. Some of these changes occurred in rapid succession, so close that at times it was difficult to count the changes. In screen 19, on the other hand, there were approximately 30 scene changes across the 1 minute 15 secs. On closer examination of the scene changes in screen 19 it appeared that the time intervals between these scene changes was in most cases longer than those in screen 12. It was at least possible to keep up with the count. While the count is quantitative in nature, the description of the scene changes and reflection on the techniques used is a qualitative observation, but it has provided some valuable insight into the design and subsequent use of animation by learners. Proportionally, the number of scene changes were equal for the two animations, but looking at the design from a cognitive load perspective it seems that these fast scene changes increased the extraneous cognitive load of the animation. The cognitive load experienced by a learner using an animation therefore comes from several sources: the need to search for relevant information within each visual view (extraneous load), the transitory nature of the animation requiring the learner to hold information in memory and process new information (extraneous load) (Ayres & Paas, 2007) and the difficulty of the content (intrinsic load).

However, since the mean cognitive load for screen 12 was within the medium load range (M = 5.76) it can be concluded that not all the design techniques were ineffective. Designers do need to be careful not to reduce the cognitive load too much. The study conducted by Schnotz and Rasch (2005) found that some of the animation used in their study unnecessarily reduced the germane cognitive load associated with deeper more meaningful cognitive processing. In the end, the participants did not invest enough mental effort in order to improve their learning performance sufficiently.

The animation in screen 12 was only 30 seconds longer than the animation in screen 19. Further investigation is necessary to determine the point at which time duration becomes a factor in increased cognitive load of an animation.

5.4.4.2 Comparing screen 5 and screen 13 in the animation version

Screen 5 had the higher mean cognitive load. There were several design issues on screens 5 and 13 that could have influenced the cognitive load. This section briefly discusses each of these issues.

The first issue relates to the images on the screen. Screen 5 and Screen 13 each required the participant to click on a textual hotspot in order to change the image. Participants were required to compare the images presented in each view. Screen 5, illustrated in Figures 5.1 – Figure 5.3, required the learner to toggle between four different images in contrast to the two views of screen 13 (illustrated in Appendix H).

The differences between the images in both screen 5 and 13 were subtle, except for the image illustrated in Figure 5.3. These subtle differences will increase the intrinsic load of the content. When considering the images only, screen 5 could be expected to have a higher cognitive load than screen 13, simply because there were more images to hold in working memory while making the cognitive comparison.

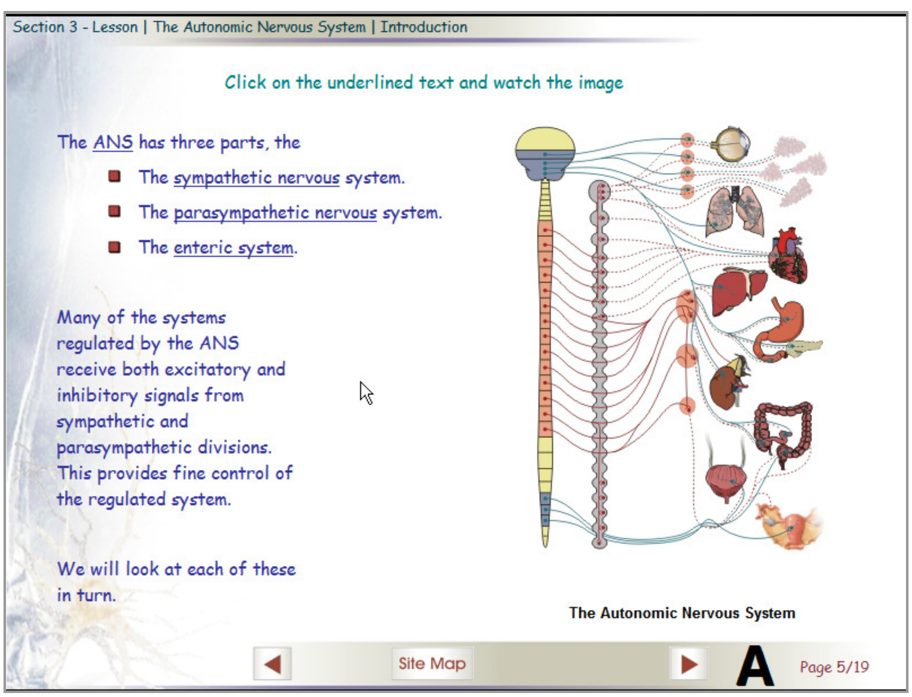

Figure 5.1: Screen 5 – First image

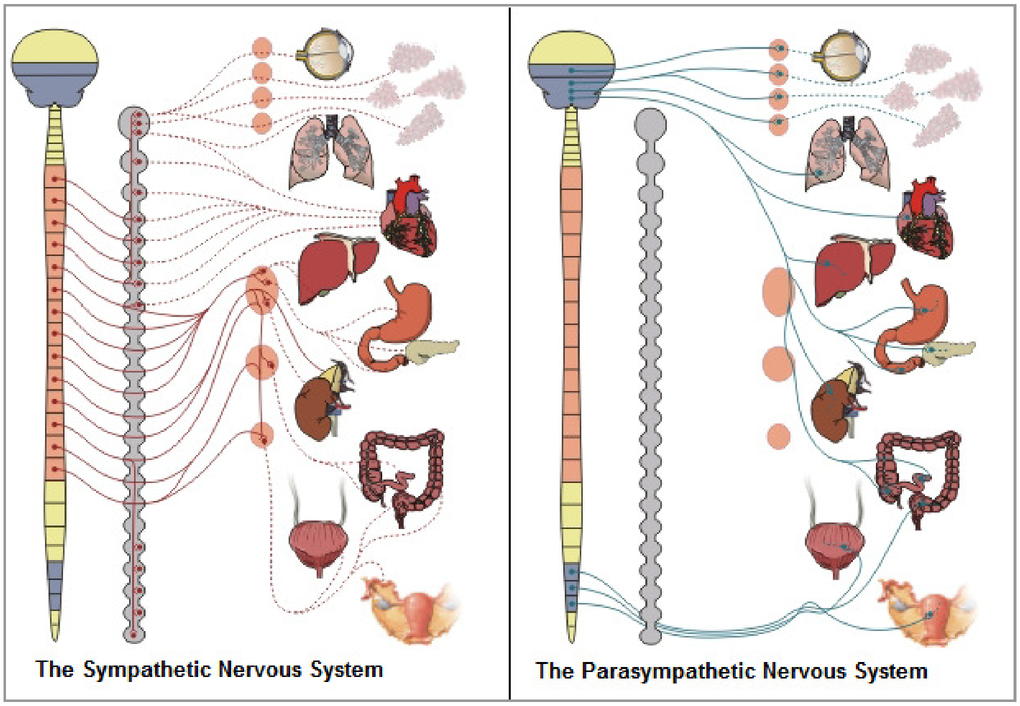

Figure 5.2: Screen 5 – Second and third image

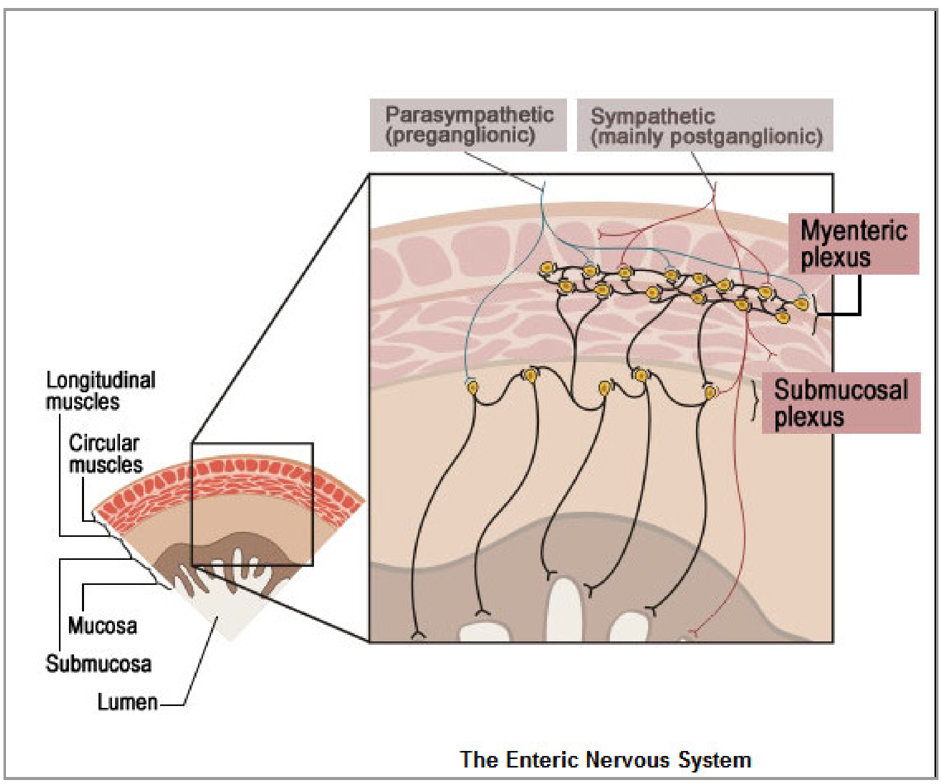

Figure 5.3: Screen 5 – Fourth image

The second issue relates to the design of the image itself. Carney and Levin (2002) reviewed recent research on the use of text and pictures in learning materials. Their review and the ten guidelines they propose for designing text-picture combinations that will facilitate learning led me to the conclusion that at least one of these guidelines was ignored in the images in screen 5, while this was not the case for the image on screen 13. There was poor congruency between the text and images in screen 5, leading to increased cognitive load for screen 5. In fact, the text on screen 5 did not even relate to the images. The images in screen 5, with the exception of the image illustrated in Figure 5.3, were not labeled, while the images in screen 13, although not labeled, had explanatory text that complemented the image.

The third design issue concerns the non-visual content on these screens. Screen 13 included textual information **and** an image, whereas Screen 5 only changed the image. It could be expected that the additional text in the views for screen 13 had the potential to create a split-attention effect, which would increase the extraneous cognitive load of screen 13. However, the contiguity principle was deliberately applied to offset a split-attention effect for screen 13. Since screen 5 had a significantly higher cognitive load than screen 13 ($p = 0.0008$) it suggests that the intrinsic cognitive load contributed more to the total cognitive load than did the extraneous cognitive load embedded in the click actions to view and hide the content. The role that the contiguity effect plays in reducing the potential for a split attention effect on a screen similar to screen 13, where images must be compared, remains an area for further investigation.

5.4.4.3 Comparing screen 14-16 and screens 23 in the static images & text version

The cognitive load of these four screens (13 - 16) was not significantly higher than the cognitive load of the animation of screen 23 ($M = 6.14$). The animation in screen 23, with its 30 ‘scene’ changes (described in Section 5.4.4.1) still resulted in a higher cognitive load than the load experienced across four separate screens of text and images. Chunking of the content, which could be viewed as a manipulation of the intrinsic load, seemed to lower the cognitive load.

5.4.4.4 Comparing screen 20 and screen 13 in the static images & text version

When comparing screen 20 and screen 13 (illustrated in Figure 5.4), where there was a significant difference in the cognitive load (screen 20 had the higher cognitive load) we see another difference in design technique. On entering screen 13 the user can read the text and study the image, which facilitates the understanding of the text. The contiguity principle (Mayer & Moreno, 2003) was applied in this design, in order to keep the extraneous load as low as possible.

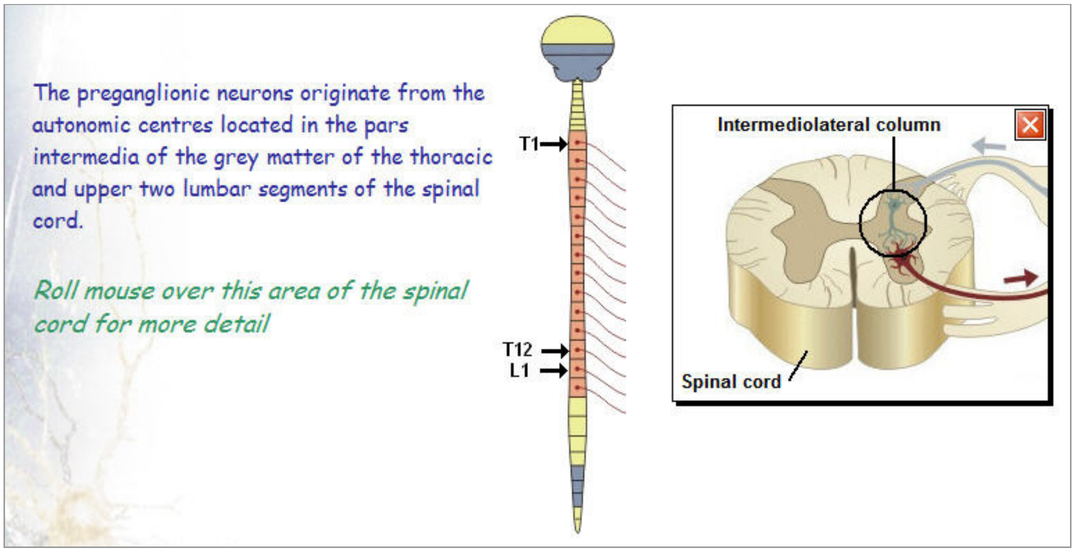

Figure 5.4: Section of screen 13 in Version 2

The initial image is relevant to the last part of the sentence on the screen. Using a mouse-over technique, the user is able to study a second image, which is a cross-section of the spinal cord. The aim of using this image is to provide a better visual in order to understand the text. The image displayed in the mouse-over facilitates an understanding of the first half of the text on the screen. Unlike many pop-ups, which are displayed using the mouse-over technique and which disappear when the user moves the mouse off the hot spot of the image, in this program the user had to click on the icon in the pop-up in order to hide this second image. Although the user must divide his/her attention between the text and the images the load is predicted to be low as there is very little text and the diagrams are not complex. Carney & Levin (2002) suggest that the use of pop-ups is a more effective technique to use when displaying images that need to be used with text. Labeling of the images was kept to a minimum. The user does not have to hold any information in working memory while searching for either the image or the text, due to the application of the contiguity principle. The amount of user control provided also allowed the user to spend as much time as needed to review both the text and the images.

On screen 20 (illustrated in Figure 5.5) the user must compare two concepts. Each concept is explained using both text and a visual image. The user was required to click on the text on the screen (Divergence or Convergence) in order to display the relevant text and image. The aim of this design technique was to facilitate comparison of the images, without the user needing to branch back and forth between screens. The content for the one concept must be held in working memory while the second concept is being studied before any cognitive comparison can take place.

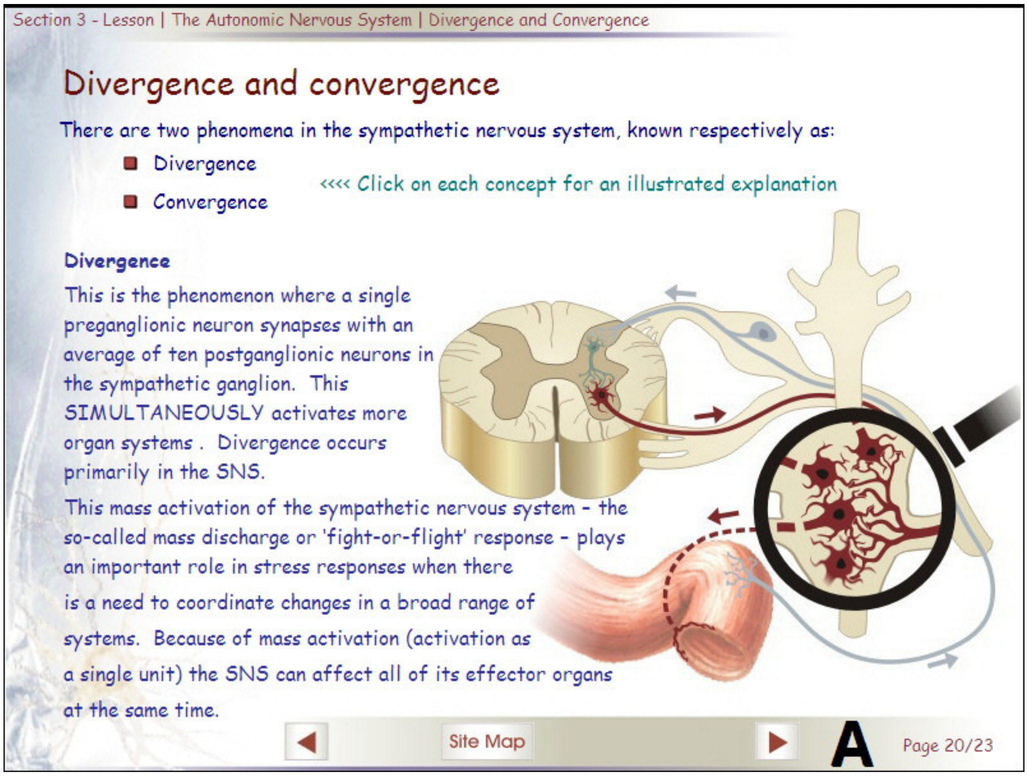

Figure 5.5: Section of screen 20 in Version 2

Considerably less text was displayed for the concept 'Convergence'. A comparison of Figure 5.4 and Figure 5.5 clearly shows that the volume of the content was greater for screen 20 than for screen 13. An alternative design for screen 20 might be to place both images next to each other on the screen, but in order to adhere to the contiguity principle the images would have to be very small. Would reducing the size of the image interfere with learning? Further investigation would need to be done in order answer to this question.

5.4.4.5 Comparing screen 5 and screens 17-19 in the static images & text version

The design of these two screens is discussed since the findings were unexpected. It was expected that the cognitive load of these screens would be significantly different. Screens 17-19 had a higher cognitive load than screen 5 but the difference was neither statistically nor practically significant. Interestingly, the *p* value before the application of the Bonferroni correction was statistically significant. The effect size does not change however, providing support for the value of the effect size in interpreting findings.

There were two differences between these screens: one difference focuses on the content, where screens 17–19 had considerably more content than screen 5. The second difference relates to the design, specifically regarding the user action required to access the content. In screen 5 the user had a one-click sequence to view the relevant content: click on text → view image. There were three new images to view. Each image illustrated the structure of one of the three sub-systems of the ANS. There are subtle differences between the SNS and PNS, but the Enteric system is obviously different. This design was used to allow the user to compare the structure of the sub-systems visually. In screen 18 a similar design approach was used, but the user had at least a two-click sequence in order to view the relevant content: click on text/image → view baseline image → click on one of three buttons to see changes in image. The changes allowed the user to compare the effects of SNS and PNS innervation on the specific organ. The innervation of 5 different organs was presented on this one screen. The comparison was not between the organs, but between the effects of the SNS or PNS on the individual organ. One of the views of screen 18 is illustrated in Figure 5.6.

Although only one image was present on screen 18 at a time (the smaller images functioned as icons), there were a total of 15 different images, with three views for each organ. In both screen 5 and 18 the participant had to hold one or two images in working memory while looking at a second or third image in order to make the comparison. In screen 18 the image of each organ had three different states, which changed dynamically depending on the button clicked to initiate the animation, while in screen 5 there were three **different** images. It may be that the extraneous load imposed by the need to click several times while studying the content imposed a higher load than just clicking on text and being able to view the image without worrying about clicking on another button to change the view.

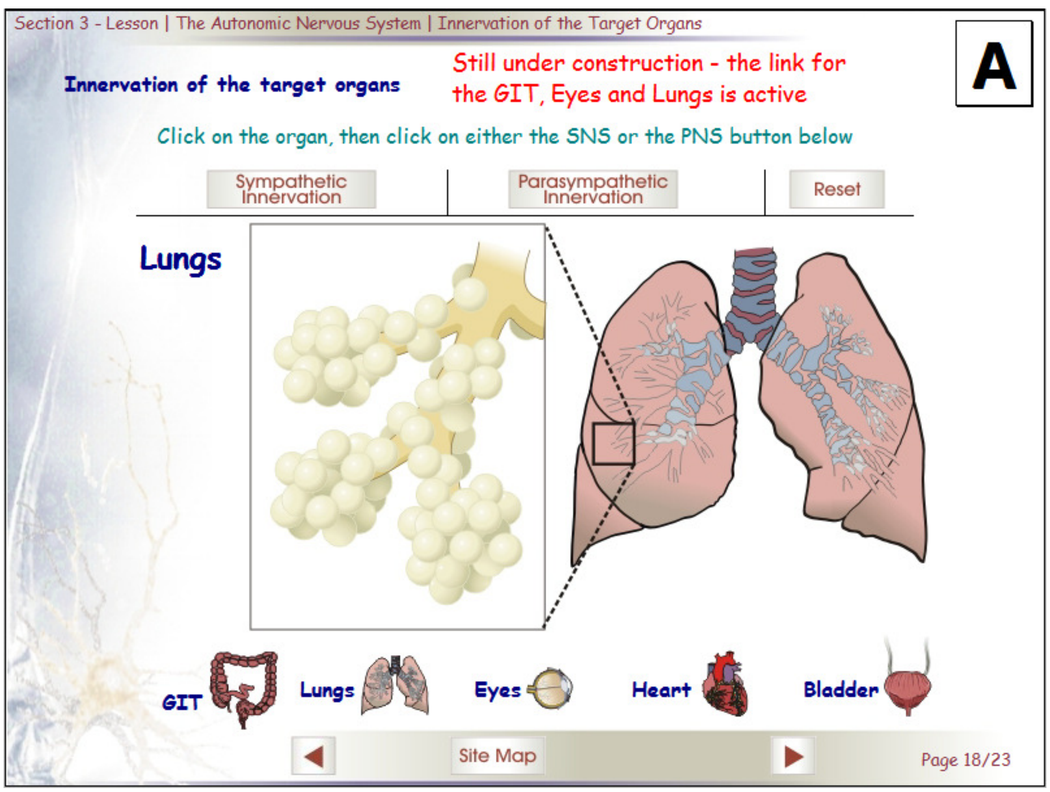

Figure 5.6: View from screen 18 in the static images & text version

In spite of this there was not a significant or practical difference in the cognitive load of these screen groups. Design issues that could account for this include the effective use of learner control, the careful use of text and images in combination with each other and the attention that was paid to the contiguity effect.

5.4.4.6 In summary

A consistent finding was that screens with animation had a higher cognitive load than the screens with static images & text. This finding continues to support the arguments in the literature that animation requires far deeper cognitive processing and requires more cognitive resources from working memory than do static images and text (Chan & Black, 2005; Lowe, 2004).

The presentation formats of the screens that had statistically significant **and** practically significant differences in the cognitive load were discussed. It would appear that the following factors could have played a role in the cognitive load:

- Length of the program and time needed to work through the entire program
- Length of the individual animations
- Number of 'scene' changes in an animation
- Amount or density of the content
- User actions required to navigate through the screens: non-linear navigation appears to impose higher extraneous load.

5.4.5 Cognitive load, cognitive style and learning performance

The question:

> How is learning performance influenced when content with different cognitive load is studied by learners with different cognitive styles?

There was no evidence in the literature of attempts to consider the relationship between cognitive style, cognitive load and learning. The reasoning and argument behind the hypotheses of the relationship between style, load and learning were presented in detail in Section 3.16.2 in Chapter 3.

5.4.5.1 Learning performance in general

Due to the fact that this study was conducted in an authentic learning environment it was important to determine whether there was a learning gain in general, irrespective of cognitive load, cognitive style, time spent on the program and other variables that could influence the outcome. Each participant did a pretest at the start of the experiment and a posttest after studying the content of one of the two versions.

There was a difference in the mean values of the computer-based pre- and posttest and a univariate analysis, which compared the difference between the pretest and posttest results, indicated that this difference was statistically significant. Learning did take place and from an ethical point of view it can be concluded that no participant was disadvantaged by the version they used. Whether this learning was sustained over a period of time was not determined as part of this study.

5.4.5.2 A look at performance in the pretest

Performance in the pretest was categorised into low, average and high. There was no relationship between the pretest results and grouping by version. The groups could therefore be regarded as equal in terms of prior knowledge as measured by the pretest. Two further analyses were conducted using the pretest results.

The first analysis focused on the question: How did prior knowledge, determined by asking whether the participants had studied the topic previously, compare with the pretest results? The expectation was that proportionally more participants who had **not** studied the topic previously would fall in the 'Low' performance group. Similarly, proportionally more participants who **had** studied the topic previously would fall in the 'High' performance group. The results indicated that this was not the case and it was concluded that there is no difference, for all three levels of performance, between the group who answered 'Yes' to this question and the group who answered 'No'.

The second analysis asked: How did prior knowledge, determined by the subjective self-report rating of the knowledge about the topic, compare with the pretest results? The expectation was that proportionally more participants who indicated that they knew nothing about the topic would fall in the 'Low' performance group. Similarly, proportionally more participants who indicated that they had more than a basic understanding of the content would fall in the 'High' performance group. The results supported this expectation. The majority of the participants in each of the 'Rating of knowledge' groups scored in the average range of performance.

The results in this study suggest that self-report ratings of knowledge are not necessarily reliable predictors of prior knowledge. I concluded that actual pretest results are a better criterion to use than self-report measures for determining prior knowledge.

5.4.5.3 Open-ended posttest scores

The posttest included a pencil and paper test at the end of the experiment. Two open-ended questions, testing application of knowledge, were given to the participants (see Appendix E). Many of the participants appeared to put far more effort into Question 1 than Question 2, scoring a relatively good mark for Question 1, but failing (less than 40%) Question 2. The possibility exists that they did not understand the question. In retrospect it was a difficult question, although it had been reviewed by the faculty member responsible for teaching this 2nd year course. It is possible that the sample was truly not able to apply the knowledge they had just learned. None of the participants had had any exposure to the specific clinical situation used in the question. Experience of the condition used in Question 2 would be limited to theoretical knowledge.

Some of the reasons for the low mean for Question 2, not directly related to performance, knowledge and learning, but which can impact this, could include the following:

- Participants were under considerable time pressure at this stage of the study. This test was taken at the end of the session and many of them had a class to attend.
- Fatigue could also have contributed to the low score and they merely scribbled down an answer in order to get it over and done with.
- Lack of interest or commitment to answer to question.

None of these reasons were explored with the sample, as it was not possible to contact them again after the experimental session.

5.4.5.4 Learning performance and cognitive style

General Linear Modeling analysis was conducted to determine if cognitive style had any influence on learning performance. The two style dimensions were considered separately. The pretest was entered as a covariate in this analysis, and in both analyses it accounted for the significant finding. Neither the Wholistic-Analytic or the Verbaliser-Imager style appeared to have any influence on learning performance in the posttest.

The findings of my study support those of two studies and appear to contradict the results of two recent studies.

Calcaterra, Antonietti & Underwood (2005) investigated the effect of the Analytic-Sequential versus Holistic-Intuitive style and hypermedia navigation behaviours on learning outcomes. This study, using undergraduate students, concluded that performance outcomes were related to particular search and navigation patterns and not to the time spent studying the content or the particular cognitive style of the learner. Massa and Mayer (2006) investigated the types of help given to participants categorised as Verbalisers and Visualisers. Using college and non-college students they found that the behaviour of the participants was consistent for their self-reported style (verbalisers tended to reply on textual help and visualisers tend to rely on pictorial help). Verbalisers and visualisers did not differ on the learning test and the researchers concluded that there was no evidence that verbalisers and visualisers should be given different multimedia instructions.

On the other hand Lee (2007) explored the effects of visual metaphor and cognitive style in a hypermedia-based environment and found that the participant’s cognitive styles substantially affected learning performance. The Analytic-Verbalisers performed significantly better than the Wholistic-Imagers and Wholistic-Verbalisers and the largest difference was between the Analytic-Verbalisers and the Wholistic-Imagers. The cognitive style literature (Evans, 2004; Riding et al., 2003; Riding &

Grimley, 1999) describes the Wholistic-Analytic and Verbaliser-Imager style combinations as either unitary or complimentary and in Lee's study we see that the unitary style group outperformed the complementary style group. Grimley (2007), in a study that was very similar to my study, found a main effect for the Wholistic-Analytic style on the overall recall of knowledge. Wholists tended to answer more questions correctly than Analytics. Grimley also found gender effects in this study (all style groupings except Wholistic-Verbalisers showed gender differences) and went into some depth in exploring these gender differences. Grimley (2007) however did not measure the cognitive load of the two interventions (text and pictures and narration and pictures) used in his study.

5.4.5.5 Learning performance and cognitive load

General Linear Modeling analysis was conducted, with the posttest as the dependent variable and the cognitive load and version as the predictor variables. The pretest results were entered as a covariate. The analysis indicated that cognitive load, as measured using the self-reporting technique did not appear to influence the learning performance in this study in any way.

Tabbers et al. (2004) measured mental effort during instruction and testing in an authentic learning environment. The mental effort during instruction was found to be relatively low. Learners who used the visual mode reported higher mental effort than those who used the audio mode, but the difference was not significant. Tabbers at al. (2004) did not report on any analysis of the relationship between mental load **during** instruction and the posttest results. They did report on the analysis of the relationship between the mental effort scores during testing and the subsequent test results. They found that there were significant differences in the mental effort reported for the retention test (learners in the visual condition reported more effort that students in the audio condition) but not for the transfer test. One of the conclusions these researchers discussed was that the authentic environment could have introduced confounding variables that influenced the outcome of the study.

Höffler (2007) investigated learning from instructional animation or a series of static pictures and found that spatial ability accounted for the difference in learning outcome rather than cognitive load, which was not significantly different for the two versions. Höffler, Sumfleth and Leutner (2006) found a strong correlation between spatial ability and learning outcomes in the group that used static pictures, but the correlation between these variables was weak for the group that used animation. Using general linear modeling and controlling for grade-point average they found that the interaction between spatial ability and type of learning material was statistically significant, and this did not change when cognitive load was entered into the linear model.

Another possible explanation for the finding in my study is that since the cognitive load for each version was within an acceptable range it did not influence learning performance negatively. A different finding might have been obtained if the cognitive load for one version was in the low range and the other in the very high range.

5.4.5.6 Cognitive style, cognitive load and learning performance

The relationship and dynamic between style and load appeared to be more complex than initially thought, influenced by different factors that were difficult or even impossible to control in the authentic learning environment. With this in mind, the final set of hypotheses must be viewed as a first attempt to investigate this complex relationship.

We have seen in the preceding sections that neither cognitive style or cognitive load seemed to influence learning performance in this study. Another variable was then introduced into the analysis: the time used by the participants. Early cognitive load studies considered time to be an indicator of cognitive load (Chandler & Sweller, 1991). The results will be discussed briefly before turning to the discussion of the relationship between cognitive style, cognitive load and learning performance.

Two hypotheses were investigated:

- Analytic learners who spend less time on the program will rate the cognitive load more highly.
- Verbalisers and Imagers who spend less time on the program will rate the cognitive load more highly.

The results of a regression analysis used to investigate the hypotheses indicated that cognitive style did not influence the cognitive load. The interaction between version and time spent on the program was retained in the regression, but it only approached significance. A confirmatory GLM (General Linear Model) was conducted and it was established that the only predictor of cognitive load in the model tested was rating of knowledge about the topic. The participants who indicated that they knew nothing about the topic rated the cognitive load significantly higher than the rest of the sample.

> *The hypothesis that Analytic learners who spend less time on the program will rate the cognitive load more highly was therefore **not** supported*

> *The hypothesis that Verbalisers and Imagers who spend less time on the program will rate the cognitive load more highly was therefore **not** supported.*

The final two hypotheses investigated in this study were:

- The Analytic learner who spends inadequate time on the program, and who rated the cognitive load as high, will perform more poorly on the posttest.
- The Verbaliser and Imager learner who spends inadequate time on the program, and who rated the cognitive load as high, will perform more poorly on the posttest.

The analyses looked at the posttest results for the Analytic, Verbaliser and Imager learners, grouped by time spent on the program and rating of cognitive load. Significance testing was difficult as the sample size was small. The data was organised and analysed (using descriptive statistics) in order to determine if a trend could be observed in the data. There was a trend that suggested that the higher the load the poorer the learning performance. There was no clear trend for the amount of time spent on the program.

This trend was confirmed in a subsequent regression analysis where version of the multimedia, cognitive style (both dimensions), subjective rating of cognitive load and total time spent on the lesson were included in the model. The variables retained in the stepwise regression were version and subjective rating of cognitive load. There was a statistically significant main effect for cognitive load and a marginally significant main effect for version. The confirmatory GLM (General linear model) confirmed that the subjective rating of cognitive load accounted for the significant result. The posttest scores were higher in the animation version compared to the static images & text version. This study also determined that the static images & text version had a higher cognitive load than the animation version. I was therefore able to confirm the trend that within an authentic learning environment, where the cognitive load of the entire program was considered, the higher the cognitive load the poorer the learning performance.

The only study found in the literature since February 2006 that is similar to my study (investigates cognitive style and cognitive load in multimedia learning) was that of Grimley (2007). Grimley compared the performance of the cognitive style groups from each dimension, using two different multimedia designs, each with different cognitive loads. The cognitive loads were not specifically measured, but were assumed to be different due to the use of a split-attention effect in the one version. Grimley (2007) argued that, in terms of overall performance, the multimedia material used in his study seemed to suit the Wholistic learner better. Grimley ascribed this to the ability of the Wholistic learner to see the big picture, which required an understanding of both the images and text. Grimley also used gender to explain the differences in the results and this made precise comparison with my study difficult.

The relationship between cognitive load, cognitive style **and** learning performance in an authentic learning environment is less clear than the relationship between cognitive load and style, without the consideration of performance. It is likely that different results will be obtained in a more controlled experimental setting (a dedicated research laboratory) using materials that have larger differences in cognitive load.

5.5 Scientific reflection

This final section of Chapter 5 focuses on four topics:

- Contribution of this study to the scientific body of knowledge in the cognitive style, cognitive load and multimedia learning fields
- Implications of this study for instructional design
- Implications of this study for using multimedia in the learning environment
- Recommendations for future research

5.5.1 Contribution of this study to the body of knowledge

I believe that the most significant contribution of this study lies in the fact that it is one of the first attempts to empirically explore the relationship between cognitive style and cognitive load. The cognitive load theory was discussed in depth and illustrated in Chapter 2. According to cognitive load theory, cognitive style is one of the subject characteristics thought to influence mental effort and performance. Reports of mental effort invested by an individual, together with their subsequent performance, provides in indication of the cognitive load of learning material.

A second unique contribution relates to the context selected for this study. Empirical studies that address either cognitive load or cognitive style have been under-researched in the health science education context. There are a few more recent publications that address the theory of cognitive load (Khalil, Paas, Johnson & Payer, 2005a, 2005b). My study has thoroughly investigated the role of both cognitive load and cognitive style **and** the relationship between these two factors in a health science education context.

A third unique contribution lies in the fact that this study has addressed the call for research in more authentic learning environments. Such environments are complex and the control of variables are difficult. Authentic environments are not usually the first choice for the context of cognitive load and/or cognitive style research. I was able to undertake a rigorous quantitative study in an authentic environment.

A fourth unique contribution is that my study is one of the few studies that compares one of the most widely used techniques to measure cognitive load (subjective rating of mental effort) with a technique that is not widely used, but which is thought to be a more accurate method of determining cognitive load, since it measures cognitive load at the time at which it occurs (direct method using dual-task methodology).

I therefore present the following theses:

1 While health education learners appear to be more Analytic in the way they process information, and while they display more of an Imager style in the way they represent information during thinking, this profile is not yet conclusive for the health education field, due to the lack of empirical studies with which to compare the findings of this study.

2 It cannot be assumed that there are no cultural differences in the way learners process information during multimedia learning, as this study found a significant difference in the Analytic style of learners from different cultural groups.

3 Cognitive style does not appear to make any difference to the amount of time needed to study from animations. There was no significant difference in the amount of time the Wholistic and Analytic learners spent studying the animation, and there was no significant difference in the amount of time Verbaliser and Imager learners spent studying the animation.

Use of animation will most likely constrain the scanning strategy of the Wholistic learner as he/she would have to watch the whole animation at least once in order to understand it.

4 Cognitive style does make a difference to the amount of time needed to study multimedia material that uses static images and text. The Analytic learners spent significantly more time than the Wholistic learner studying the content of program. Since the Analytic learner is not constrained by the temporal nature of the animation he/she can (and therefore does) take all the time he/she needs to process the content. The Verbalisers appeared to need more time than Imagers, but the difference was not significant.

5 The static images & text version of the program as a whole, used in an authentic learning environment, had a significantly higher cognitive load than the animation version. Since the learner must infer the processes explained in the static images & text rather than merely perceiving them in an animation, more mental effort is required to understand and learn the content in the static images & text version. The non-linear nature of the static images & text version and the total length of the program are also thought to have increased the extraneous load of the program as a whole.

6 Animation as an isolated learning event requires more cognitive resources from working memory than do screens presenting the same information using static images and text. This resulted in a higher self-report rating of mental effort invested for the animation screen. This higher cognitive load for the animation was a consistent finding whenever an animation screen was compared with a non-animation screen using static images & text.

7	If the extraneous cognitive load can be controlled and minimised the Analytic learner will still report a higher cognitive load than the Wholistic learner, due to the more elaborate processing strategies used by the Analytic learner, which implies deeper processing and a higher germane cognitive load.
8	Measurements of cognitive load, using the same content, but different measurement techniques were not correlated. The possibility exists that each measurement technique focuses more strongly on a different type of cognitive load.
9	The more mental images the learner is required to compare, which are not presented contiguously, the higher the rating of mental effort.
10	If cognitive load is kept low as possible learning will take place irrespective of the learning strategy used. This study showed that significant learning took place in both versions of the program.
11	There is no conclusive evidence that animation is a better instructional strategy for learning in health science education. The use of animation must be carefully considered in order to justify the cost and time of such development.
12	There is no clear and simple relationship between cognitive load, cognitive style and learning performance in an authentic learning environment in health science education. The relationships are complex and require further investigation.

5.5.2 Implications of this study for instructional design

This study confirms the validity of existing instructional design guidelines for multimedia learning material. These include the practice of controlling for the split attention effect and paying attention to the contiguity principle. I would like to suggest that instructional designers should **first** control for the split-attention effect and **then** consider the contiguity principle. Design strategies that include the use of pop-ups can pay attention to both of these effects. Using a small pop-up that the user can drag and position in such a way that any underlying text or image can still be viewed, rather than placing the information on two separate screens, will address both the split-attention and contiguity effects. The extra click action required to open and close pop-ups does not seem to influence extraneous cognitive load adversely.

My study demonstrated clearly that the learners had significant gains in learning irrespective of the version used. Animation is not necessarily a superior strategy. The development of animation is time-consuming and costly. The use of animation must be carefully considered when there are time and funding constraints. This study did not provide enough convincing evidence that animation was more beneficial in teaching the topic 'The Autonomic Nervous System'.

There is still much to learn about the effective design of animation. When designing animation, attention needs to be paid to the speed at which the views or scenes change. It appeared as if the views in some of the animations used in this program changed too fast. Learners did have control over the pace of the animation, but few learners were observed as actually using the controls. The designer could consider introducing more learner control at specific points in the animation, for example where there is a change in focus of the content. Learners need to be specifically instructed in the use of control features. An alternative is to develop several shorter animations that are viewed in a specific sequence.

5.5.3 Implications of this study for using multimedia in the learning environment

A finding in this study was that an unacceptably high percentage of participants spent inadequate time on the multimedia programs. Some of these participants seemed to merely click through the program, spending less than 1 minute on a screen. The amount of learning that takes place in such a situation must be queried. While this study did not follow these participants up in order to obtain some qualitative data about the reasons for this pattern of use, it is clear that more guidance needs to be provided to learners on how to use multimedia programs when learning. Note-taking while they are using the multimedia program could be encouraged. Learners need to be advised on how much time they should spend on each screen. It was noted that some of the learners clicked very quickly through the entire program before going back to review each screen in more detail. This might be related to cognitive style and a qualitative follow-up of multimedia use might provide a better perspective on this pattern.

5.5.4 Recommendations for future research

I have identified various avenues for future research during the course of this study. The sources of these recommendations include:

- Limitations in the research methodology of my study
- New literature that appears to be aligned with my study
- Findings in my study that suggest other avenues for inquiry

I have divided the recommendations into three areas, although there may be overlap and some of the recommendations can be combined in a single study.

5.5.4.1 Exploring the relationships between cognitive load, cognitive style and multimedia learning

Wholistic learners are unlikely to invoke germane load to the same extent that Analytic learners with their elaborate processing style do. Therefore, if the extraneous load is low the Analytic learner should perform better than the Wholistic learner, since the deeper processing (germane load) should result in better performance. If the extraneous load is high we can expect to see the Analytic learner perform more poorly compared to the Wholistic learner, as their style will still result in some germane load and the total cognitive load might reach a level where it is detrimental to learning. In order to test these predictions I recommend moving the research back into a more controlled experimental setting, using shorter interventions that have large differences in cognitive load.

I would also recommend that future research controls the amount of time the learner is given to study the content, as time seemed to play some role in this study.

Since it appears as if the Wholistic learner will not process as deeply as the Analytic learner future research could investigate and compare strategies that can assist the Wholistic learner with deeper processing.

The relationship between cognitive load, cognitive style and learning performance in multimedia learning needs further investigation in a variety of disciplines, with a variety of different learner groups. Further studies could replicate the current design and/or improve on the design, using the suggestions made in Section 5.3 of this chapter.

5.5.4.2 Instructional design practice

The participants in this study were allowed to make notes while using the multimedia program. These notes were only scanned to get a high level overview of the nature of the note-taking. A detailed analysis of these notes was beyond the scope of this study. Detailed analysis of the note-taking practices of learners using multimedia is an avenue for further research. Of particular interest would be to determine the extent to which different style groups approach note-taking, and how note-taking influences the subjective ratings of mental effort. Electronic and paper-based note-taking could also be compared.

The role that the contiguity effect plays in reducing the potential for a split attention effect, and subsequent cognitive load, on screens where comparison of images is required also needs further investigation. An alternative design could be considered where both images are placed on the same screen, but in order to adhere to the contiguity principle the images would have to be very small. Would this interfere with learning? Would this influence the cognitive load of the instruction? Further investigation would need to be done in order answer to these questions.

One screen in this program provided an overview of the content and it functioned as a menu to facilitate non-linear navigation. More detailed analysis of the path of navigation by each participant after they had accessed this overview/menu screen was beyond the scope of this study, but could provide an avenue for future research that considers how best to design such screens. The navigation path could also provide an indication of which sections of the content were accessed more often, and this analysis could be followed up using both a qualitative and quantitative approach.

5.5.4.3 Cognitive load research

This study only measured the cognitive load the first time the learner accessed the screen. The learners were however allowed to move back and forth through the program. If the cognitive load had been measured with each entry into the program it would have been possible to determine if the rating of mental effort invested decreased as they became more familiar with the content.

Further investigation is necessary to determine the point at which time duration becomes a factor in the cognitive load of an animation.

The one change I will recommend in future research is that the self-report rating of cognitive load only require the learner to consider one screen per rating. The fact that the participant had to keep two screens, and even three, in memory might have inadvertently influenced the cognitive load rating.

5.5.4.4 Cognitive style research

Further investigation is necessary into the cultural differences in cognitive style. This research should focus specifically on environments where many different cultures co-exist and co-learn, rather than looking at differences between cultures separated by distance (Graff, Davies & McNorton, 2004).

The two style dimensions (WA and VI styles) could be combined to reflect four different possible groupings (Wholistic-Verbaliser, Wholistic-Imager, Analytic-Verbaliser and Analytic-Imager) as each learner has a particular style on each of the dimensions. This would allow the investigation of the impact of unitary and complementary cognitive styles on the cognitive load of instructional material.

5.5.4.5 Use of multimedia material for learning

The patterns of navigation in the program were only explored briefly. Reasons for viewing each screen only once, or several times, which could be explored in future research, include attitudes toward the use of a computer to study, personal issues that influence the time participants wanted to spend with the program, fatigue, motivation and interest in topic and general study habits and learning style.

5.6 Conclusion

This study explored the role that cognitive load and cognitive style play in multimedia learning from a number of different perspectives: time spent on the program, comparison of the cognitive load for a particular style across the two versions (illustrated in Figure 5.7 using the yellow arrow), comparison of the cognitive load of a particular version across the two styles (illustrated in Figure 5.7 using the blue arrow), results of learning performance.

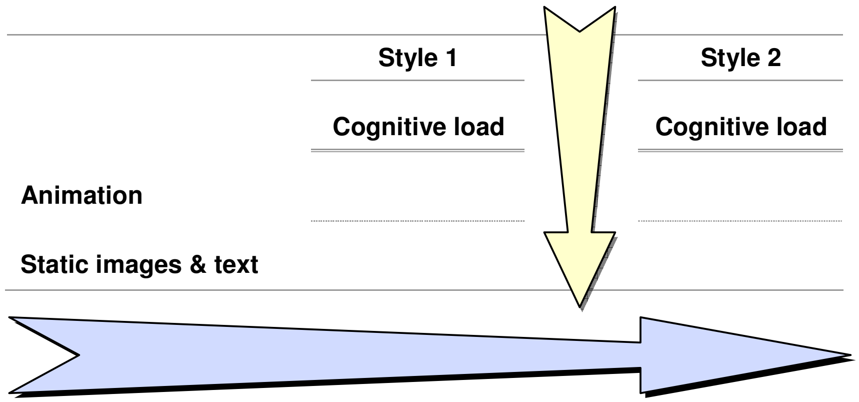

Figure 5.7: Perspectives of the study

The strategies of animation and static images & text were compared within an authentic learning environment. Since the environment was authentic it was important to establish whether learning actually took place.

The cognitive style profile of the research sample leaned towards the Analytic style in the Wholistic-Analytic dimension and toward the Imager style in the Verbaliser-Imager dimension. It has yet to be determined whether this is a typical cognitive style profile for health education learners. The cognitive loads of the respective research interventions were significantly different, yet neither version appeared to have an excessive cognitive load that negatively influenced learning. Significant learning took place for all the participants in this study. Surprisingly, it was found that when the program was considered as a whole, the version that used predominantly animation had the lower cognitive load as measured using the subjective rating technique. When the analysis drilled down to specific screens and compared animation and static images & text, the results consistently showed that animation had a higher cognitive load than static images & text.

This study established that there is once again empirical evidence that cognitive load influences learning performance. It was possible to establish that it seems as if the Analytic cognitive style influences the subjective rating of cognitive load, but further empirical investigation of this relationship is necessary. The proposal is that the Analytic style influences the germane load experienced during learning. Since researchers are currently unable to measure the three different types of load separately, this proposal remains an area for further investigation.

The subjective cognitive load rating of the program was compared with the cognitive load rating measured using the direct measurement method. The direct measurement method (also applied in an authentic learning environment) found that the animation version had the higher cognitive load. The correlation between these two methods of measurement was very low and not significant, thereby confirming a suggestion in recent literature that each method might be measuring different aspects of cognitive load.

A final comment on my experience as a researcher is adequately summed up in the quote by Lloyd Alexander.

> *We learn more by looking for the answer to a question and not finding it than we do from learning the answer itself.'*
>
> *(Lloyd Alexander, n.d).*