Presentation and Analysis of Empirical Data 4

4.1 Introduction

This chapter presents the analysis of the empirical data for this study, which set out to explore the role cognitive load and cognitive style play in the achievement of learning outcomes, when using animation and static images as 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, using the multimedia formats of animation and static images.

The demographic profile of the sample is described in detail, followed by the results of the analyses undertaken to find answers to the research questions. This analysis and discussion is presented and explored using the research question, the sub-questions and the accompanying hypotheses as a framework. Each major section of the chapter concludes with a summary of the process of analysis and subsequent findings.

I have already stated that the cognitive load and cognitive style research streams are each extensive in their own right. Merging these two fields increased the scope of the research beyond what is normally considered adequate for doctoral work. Since this study follows a relatively new line of research, the analysis undertaken for this study was both broad **and** deep, and extended beyond merely looking for answers that would support or reject the hypotheses. In addition, both these fields are under-represented in the health science education research. My study also addresses this shortcoming. While Chapter 4 presents this broad, deep analysis I will only interpret and discuss the findings related to the hypotheses in Chapter 5, unless the analysis results in unexpected findings, in which case such findings will also be discussed in Chapter 5.

Two hundred and forty five data logs were retrieved by the computer laboratory administrator from the individual computers used for the experiment. Once the data had been cleaned and prepared for analysis the final sample included 238 participants.

Participants were excluded from the final sample if

- they failed to answer any of the questions in the various electronic questionnaires **and** the pre- and posttests, in spite of retrieving a data file for the participant.
- data for **both** the cognitive styles analysis and the self-report of mental effort were incomplete.

4.2 Statistical analysis

The data was analysed by Dr M van der Linde, Department of Statistics at the University of Pretoria, using the SAS® application.[1]

A review of the literature on research methodology indicates that it is generally accepted that researchers want three basic questions answered once the data has been collected and analysed:

- Is this effect real or can it only be attributed to chance?
- If it is real, how large is it? and
- Is it large enough to be useful? (Ellis & Steyn, 2003; Kirk, 2001; Vacha-Haase, 2001; Vaske, 2002; Winkleman, 2001)

These three goals also apply to this study. The first question is answered using null hypothesis significance testing, the result of which determines whether or not the observed effect is due to chance (Onwuegbuzie & Leech, 2004). The question about the size of the effects can be addressed using descriptive statistics, confidence intervals and effect sizes (Kirk, 2001; Onwuegbuzie & Leech, 2004). Determining whether the effect is real and of practical significance does involve an element of subjectivity. It requires a value judgment on the side of the researcher, which is influenced by, amongst others, the value system of the researcher, societal concerns and the assessment of costs and benefits. Kirk is of the opinion that researchers have an obligation to make this kind of judgment and that no-one is in a better position to do so than the person who collected and analysed the data (Kirk, 2001).

In this study descriptive statistics are reported as frequencies, means and standard deviations (M ±SD) or standard error (M ±SE). Inferential analyses of data were performed using regression analysis, the general linear model (GLM), t tests, Chi-square analysis, Pearson's correlation and effect sizes. Confidence intervals of 95% were reported. Differences were considered significant for *p* values ≤ 0.05. This means that the probability to obtain a statistic as or more extreme, if the null hypothesis is true, is smaller than 5%. Of particular interest in inferential analysis is the R^2 value. This value indicates the extent to which the variables that are included in the analysis co-vary (Lowry,

[1] SAS Version 8.2 running under VM/CNS on the University of Pretoria's mainframe system

2007) or to put it in another way, R^2 is a measure of the strength of the relationship between two variables (Garson, 2006). The higher this value the better.

Effect sizes address the issue of statistical significance versus practical significance. The effect size is not dependent on sample size (Vache-Haase, 2001; Winkelman, 2001). The effect size is typically evaluated as small, medium, or large. One commonly used guideline is based on socio-behavioural research, where a small effect size is described as 0.2, a medium effect size as 0.5, and a large effect size as 0.8. An alternative guideline, based on psychotherapy research, describes 0.15 as a small effect size, 0.45 as medium, and large as 0.90 (Winkelman, 2001). The advantage of reporting effect size in an area of knowledge with a growing body of research allows the reader to determine whether there is support for the theory underpinning the research, judge the practical importance of the findings and compare findings across studies despite different analytic procedures (Winkelman, 2001).

I will use the guidelines provided by Cohen (1988) when comparing the means of two samples, where a small effect size is described as 0.2, a medium effect size as 0.5, and a large effect size as 0.8. It is reported as *d*. In a 2 x 2 table the phi (ϕ) coefficient (Ellis & Steyn, 2003), which in SAS is reported together with the Chi-square statistic, is the measure of the effect size and is reported as *w*. The guideline here is that a small effect is described as $w = 0.1$, a medium effect as $w = 0.3$ and a large effect as $w = 0.5$. I will use these guidelines in this study.

4.3 The profile of the participants

One hundred and ninety three (193) participants were registered for the MBChB programme (medical students) and 45 for the BChD programme (dental students).

The demographic issues explored in this study were

- age,
- gender,
- cultural group and
- home language of the participants.

Since prior knowledge (learner experience) influences the self-report rating of mental effort (Ayres, 2006a; Clarke, Ayres & Sweller, 2005) and learning performance (Ginns, 2003; Kalyuga, Chandler & Sweller, 2001) the issue of whether or not the participants had studied the topic previously was explored. Participants were asked to rate their prior knowledge on the topic. These findings are presented and discussed in this section on the profile of the participants.

4.3.1 Age, gender and cultural group

The gender profile of the research sample is presented in Figure 4.1.

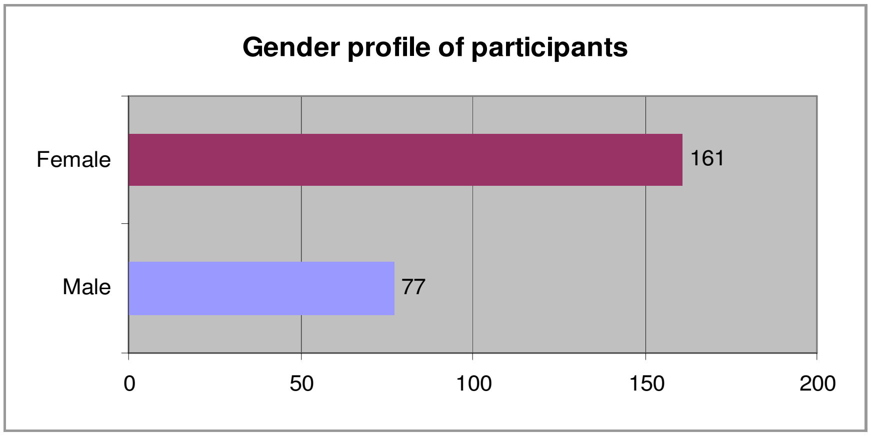

Figure 4.1: Gender profile of the sample

Looking at Figure 4.1 it is evident that the majority of the participants were female. The ratio of female to male students in this class is almost exactly 2:1. The number of female students in this sample does not reflect the gender profile in the South African population, where, using the 2001 Census data, the ratio of females to males for the 15 – 24 year group is almost equal (Statistics South Africa, 2005). The reason for this high proportion of female students reflects current strategy within the National Department of Education in South Africa. Increased enrollment of black and female students in what was historically a white institution for Afrikaans-speaking white South Africans is in line with the requirements for transformation in Higher Education, documented in the White Paper for Higher Education of 1997. One of the aims for higher education is to increase access for black, women, disabled and mature students (Department of Education, 1997).

The age range for the participants was 17 - 37 years, with the majority of the participants in the 19 - 20 year age range. A breakdown of the gender and age distribution is presented in Figure 4.2.

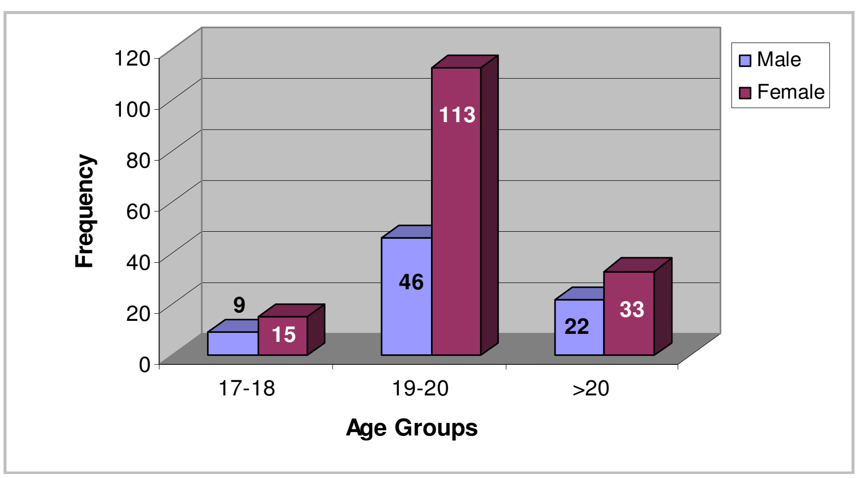

Figure 4.2: Profile of sample by gender and age

The school-leaving age in South Africa is typically between 17 – 18 years of age. The average age for the sample was expected to lie in the 19 - 20 range. Sixty four percent (64%) of the sample are in this expected age range. Figure 4.2 shows a slightly skewed distribution. Fifty five participants (23% of the sample) were 21 years or older. There are at least three factors that contribute to this distribution.

1. The demographic profile of the school leaving population has changed due to the political changes in South Africa since 1994. A larger percentage of school leavers are now in fact older than 18 years.
2. There are generally more applicants for the MBChB programme at the University of Pretoria than there are available places in the programme. Many students who qualify for entry to the MBChB programme, but are not accepted, enrol for a basic science degree (BSc). They complete the BSc degree, a 3–4 year programme, and then go on to enrol for the MBChB programme.
3. Students also transition to the MBChB programme before they complete their BSc degree. This is possible because they are given the opportunity to fill vacancies in the MBChB programme, which arise due to attrition from the MBChB programme.

Participants were randomly allocated to the two versions of the same program (the research interventions. The gender and age distribution of participants across the research interventions, after this random allocation, is displayed in Table 4.1.

	Animation version		**Static images & text version**	
	Male	**Female**	**Male**	**Female**
17-18	1	11	8	4
19-20	21	59	25	54
>20	13	15	9	18
TOTAL	**35**	**85**	**42**	**76**

Table 4.1: Distribution of sample across research intervention, by age and gender

Only one participant was 17 years of age at the time of the research intervention. Ten participants were older than 25 years, and two of these ten were older than 30 years. The oldest participant in the group was 37 years old. Factors that could contribute to this age range have already been discussed earlier in this section. They will not be explored further as they are beyond the scope of this study.

Students at the University of Pretoria are required to indicate their cultural group when enrolling. I collected this data in order to explore the effect of culture on cognitive style. Figure 4.3 displays the profile of the sample by cultural group.

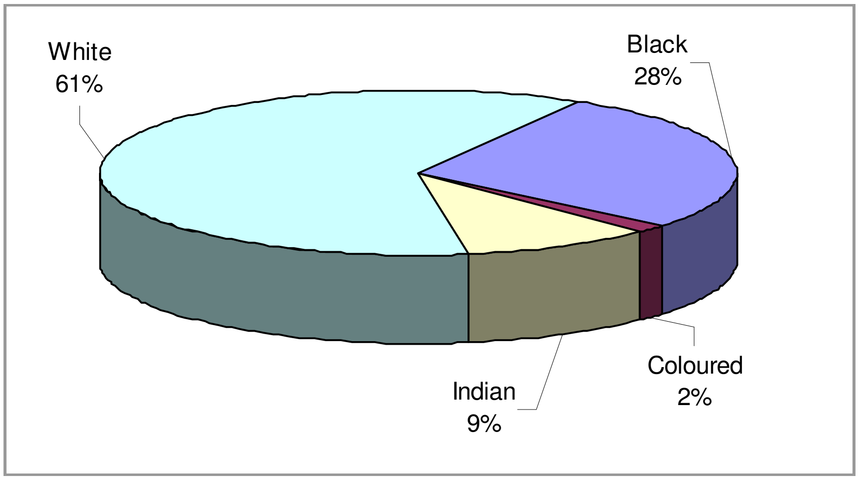

Figure 4.3: Profile of sample by cultural group

The majority of the sample (61%) were White students. Historically, Black and Indian medical students enrolled at the Medical University of South Africa (MEDUNSA) or the University of Durban-Westville respectively. Changes in the structure of South African universities (Department of Education, 1997) and the establishment of a new democracy in 1994 has seen these traditional patterns of enrollment change, with an increase in Black, Coloured and Indian students in universities that previously enrolled only White students. Further discussion of the cultural profile of the sample is beyond the scope of this study.

4.3.2 Home language of the sample

When discussing this study with different physiology lecturers a comment was made by one of them that she experienced that students who were not English first language speakers really battled to understand Physiology (L. Nagel, personal communication, 22 November 2005). Two of the factors Michael (2007) identified that contribute to the difficulty students have in understanding Physiology relate to language. Firstly, the language is mixed: by this the author meant situations where many commonly used words in the discipline take on specific scientific meanings that are different, and often opposite, from their lay meanings. The second factor was that faculty and textbook authors use language imprecisely, where imprecise refers to the fact that there is widespread use of jargon and acronyms, often to the detriment of learning. The University of Pretoria is a multi-cultural university, and tuition in the Faculty of Health Sciences is mostly through the medium of English. There are however 11 official languages in South Africa and the Black students represent a wide spread of cultural and ethnic groups, where mother tongue is not necessarily English.

Participants were asked to indicate whether English was their first, second, third of fourth (and higher) home language. The results are displayed in Figure 4.4.

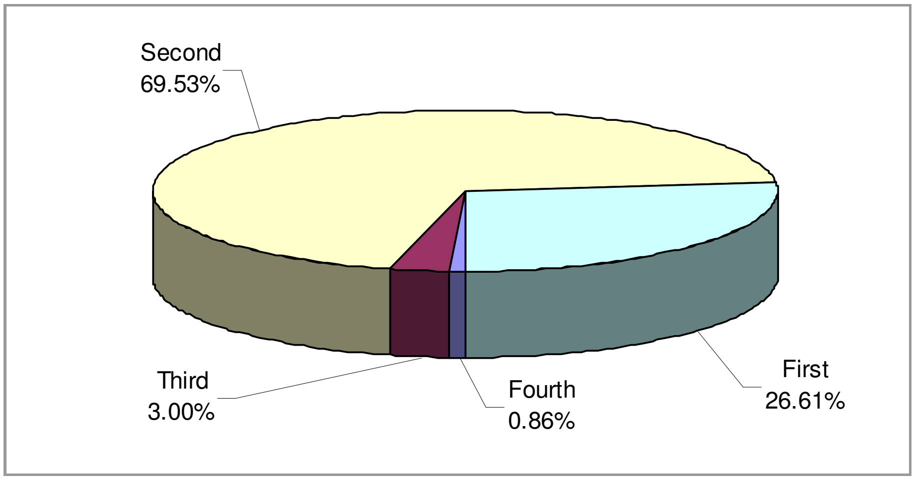

Figure 4.4: Percentage of participants where English is their home language

Only 26.6% of the sample indicated that English was their first home language. The majority of the group (69.5%) use another language as their first language at home. I did not explore which of the other 10 official languages were their first language. The possibility does exist that language plays a role in cognitive load and learning performance and this will be explored later in this chapter.

4.3.3 Prior knowledge of the participants

Prior knowledge was determined in three different ways:

i. Asking participants whether or not they had studied the topic before.

i. Asking participants to rate their knowledge and understanding of the physiology of the Autonomic Nervous System.

ii. Using a pretest to determine knowledge at the level of recall and comprehension.

Data for the first two points listed above were obtained using self-report ratings. The results for these two questions will be presented and discussed in this section. The pretest results will be presented and discussed in Section 4.8 of this chapter. The responses to the question 'Have you studied this topic previously?' are presented in Table 4.2.

	Have you studied this topic previously?	
	Yes	**No**
Animation version	84	35
Static images & text version	83	34
TOTAL	**167**	**69**

Table 4.2: Number of participants who had studied topic previously

Nearly one third of the participant group (29.2%) indicated that they had never studied the topic before. The 70% who answered 'Yes' to this question could have studied aspects of this topic in Biology, a school subject, or while enrolled for a BSc degree in the School of Biological Sciences. It is also possible that they had touched on this topic in Anatomy and/or other units in the Physiology curriculum. I did not interview any of the participants to further explore their responses to this question due to constraints with regard to access to the sample after the session in the computer laboratory.

Participants were asked to rate the level of their knowledge about the topic (The Autonomic Nervous System), illustrated in Figure 4.5. Four options were provided. I think I know and understand

- absolutely nothing about the topic.
- the basic concepts of the topic.
- the concepts beyond a basic understanding.
- the topic at an expert level.

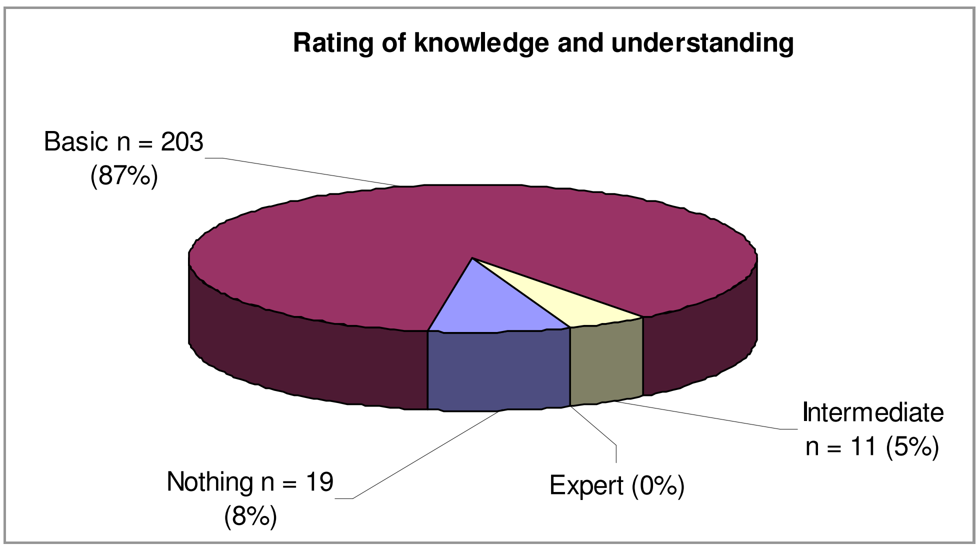

Figure 4.5: Rating of level of knowledge and understanding

From Figure 4.5 we see that only 8% of the sample indicated that they know nothing about the topic. Most of the sample (87%) indicated that they understood the topic at a basic level. None of them rated themselves as experts on the topic.

Cognitive load theory, discussed and illustrated in Chapter 2, proposes that task characteristics and subject characteristics interact to influence the amount of mental effort invested in learning. Asking the participants to rate their knowledge and understanding of the subject takes the simple question 'Have you studied this topic before?' one step further. Perceptions about prior knowledge might influence motivation to learn. Participants who feel that they already know the content might not invest much mental effort in understanding the content. However, by asking them to indicate how much mental effort they actually did invest to understand the content it is possible to start looking at how accurate their perceptions are, particularly if their learning performance is also considered. This rating addresses prior knowledge from a subjective perspective, while the pretest, which is discussed in Section 4.8 of this chapter, addresses prior knowledge from an objective perspective.

4.3.4 In summary: profile of the participants

The profile of the sample used for this study can be summarised as follows: The majority (68%) of the participants are female. The majority (81%) of the participants are medical students. The only other learner group represented in this sample are dental students. The same Physiology course is followed by both these groups at the University of Pretoria, which explains why they are in the same class as the medical students. Due to the fact that this study does not intend to compare cognitive style, cognitive load or learning performance across these two groups no further use will be made of this difference in the learning programme being followed. The majority (64%) of the sample are between 19 – 20 years of age.

The group is culturally diverse and for the purpose of this study four cultural groups were identified: Black, Coloured, Indian and White participants. The majority of the sample are White students (61%) Only 26% of the participants use English as their home language. The majority (70%) of the participants indicated that they had studied this topic previously, but also indicated that they only had a basic understanding of the content.

I now move on the explore the time spent on each version of the program. Time is considered when discussing both cognitive style and cognitive load. The next section analyses the time spent without considering the issues of cognitive load and style.

4.4 Time spent on each version of the intervention

This section looks at the time spent on each version in general, irrespective of cognitive load or style.

4.4.1 Time spent on full program

The animation version had 19 screens. During the testing of the program (before the pilot studies) it was determined that it took users between 30 – 45 minutes to work through the animation version. The participants in this study spent between 15.03 and 76.76 minutes studying the content of the animation version, with a Mean (± SD) of 42.35 (± 11.59) minutes. There were outlier values on both sides of the range. Two participants spent just over 15 minutes study the content (15.03 and 15.5 minutes respectively) and two spent more than 65 minutes studying the content (68.9 and 76.76 minutes respectively). The remaining times were all more than 19 minutes and less than 65 minutes.

The static images & text version had 23 screens, and during testing of the program it took users between 45 – 75 minutes to work through the program. The participants in this study spent between

22.86 and 92.76 minutes studying the content of the static images & text version, with a Mean (± SD) of 45.63 (± 14.01) minutes. There were also two outlier values, 87.45 and 92.76 minutes. The remaining times were all under 83 minutes.

The time spent on the program was divided into three groups. The three 'time' groups were labeled Inadequate (IA), Adequate (AD) and More than Adequate (MA). This categorisation used the time taken during the program testing (before the pilot studies) as a guideline for setting the time parameters for each of the three groups.

The amount of time assigned to each of these three time groups, for each version, is presented in Table 4.3.

	Time - Animation version	**Time - Static images & text version**
Inadequate (IA)	< 30 min	< 40 min
Adequate (AD)	30 – 50 min	40 – 65 min
More than Adequate (MA)	> 50 min	> 65 min

Table 4.3: Time groupings for each version of the program

The frequencies and mean times, for each time group, by version are presented in Table 4.4 (see next page), together with the results of the t test to determine the statistical significance of the means in each category of time.

	Animation			Static images & text			Tests for significance		
	n	**Mean**	**SD**	**n**	**Mean**	**SD**	**t test**	**P value**	**Effect size (d)**
Inadequate	21	25.4429	4.3994	46	33.0065	5.1138	-7.3100	< 0.0001	1.59
Adequate	72	41.6669	6.1519	62	49.9255	6.6207	-10.6849	< 0.0001	1.29
More than Adequate	27	57.3198	5.6094	10	77.1050	9.2193	-9,.8279	< 0.0001	0.65

Table 4.4: Mean time spent on program for each version

With reference to Table 4.4, the results of the t tests for independent samples returned differences between the two versions that were highly statistically significant ($p < 0.0001$). The effect size was also determined for each group and for the categories 'Inadequate' and 'Adequate time' a very large effect size was returned. The effect size for the 'More than Adequate' time category is a medium-to-large effect. These results, indicating both statistical and practical significance, must still be interpreted cautiously as there were more screens in the static images & text version. The results only provide a big picture view of the time spent on the program. They do not indicate exactly where in the program the time was spent. A screen wise comparison, which follows in the next section, will shed more light on this.

The range in time spent studying the content in both versions is wide. Of concern is the number of participants spending less than 30 minutes studying the animation version (17.5 % of the sample) and less than 40 minute for static images & text version (39% of the sample). Due to the time pressures for completing the data collection and the fact that the participants were not available after the laboratory session it was not possible to interview the participants to determine the reasons why some of the participants spent so little time studying the content. No assumptions can be made about these reasons.

4.4.2 Time spent on individual screens

This section looks briefly at the time spent on specific screens across the two versions of the program. The screens of interest are screen 12 in the animation version and screens 13-16 in the static images & text version. Screen 12 in the animation version presented the content using an animation that was 1 minute and 45 seconds in duration. Screens 13-16 in the static images & text version replaced the animation (screen 12). These four screens will be treated as one unit for the purpose of this analysis and discussion.

Since the participants were allowed to go back to previous screens the total time each participant spent on each of these screen units was calculated.

The participants spent between 2 and 23.85 minutes studying the content of screen 12 in the animation version, with a Mean ($\pm$ SD) of 5.4328 ($\pm$ 3.2353) minutes. Four of the participants spent more than 11 minutes on this screen. The participants accessed this screen between 1 and 6 times.

The participants spent between 2.466 and 16.283 minutes studying the content of screens 13-16 in the static images & text version, with a Mean ($\pm$ SD) of 7.5558 ($\pm$ 2.9871) minutes. There were no outlier values.

Using the t test for independent samples, the analysis indicated that the difference between the mean time spent on screen 12 and screens 13-16 was statistically significant, $t = -5.26$, $df = 235$, $p < 0.0001$. The effect size was $d = 0.710$, indicating practical significance. This finding was expected as there were more screens to access in the static images & text version (screens 13-16). In fact I expected the time to be even longer for screens 13-16.

Sixty six participants never accessed screen 16 in the static images & text version. Access to this screen was from screen 14. Participants were required to click on the small magnifying glass, illustrated in Figure 4.6 using the red circle and arrow.

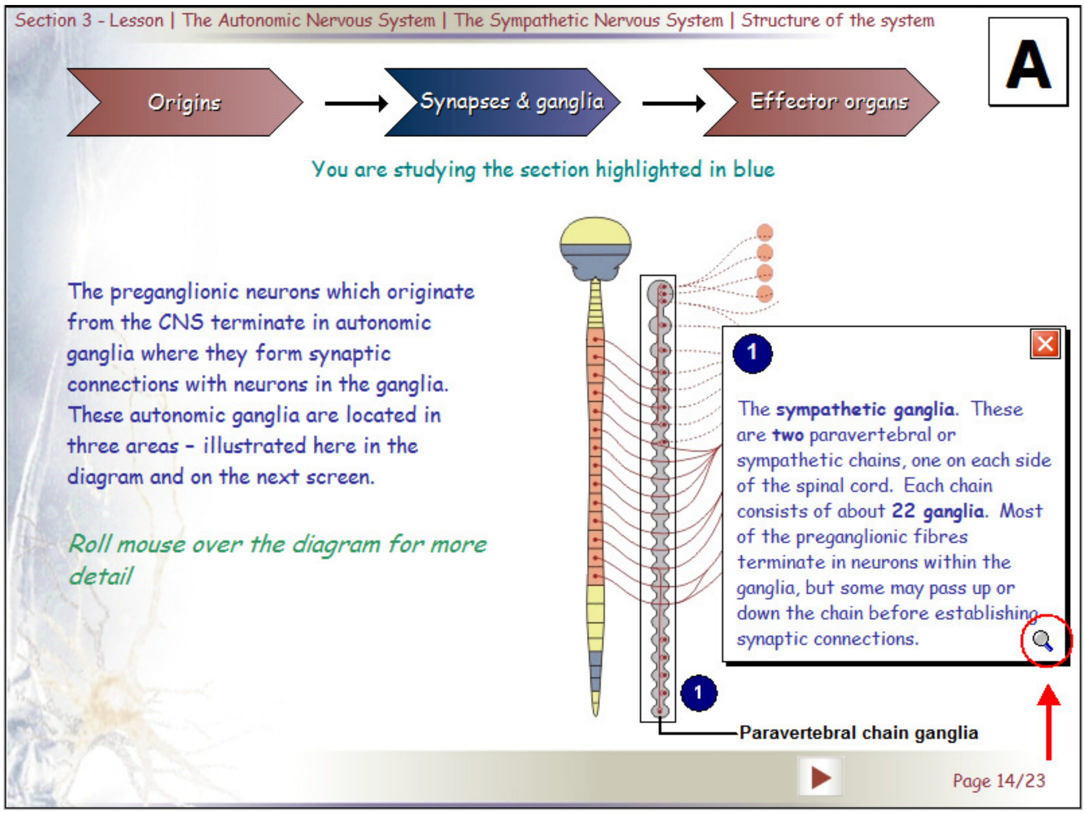

Figure 4.6: Access to screen 16 in the static images & text version

There were no instructions directing the participant to specifically click on this magnifying glass. This turned out to be a design flaw that resulted in 50% of the participants who used this version missing screen 16. The content for screen 16 can be viewed in Appendix I.

The number of times these screens were accessed was determined separately for each screen.

The frequencies for the number of times the screens were accessed are displayed in Table 4.5.

	Screens				
	Vers 1 (n=120)	**Version 2 (n = 118)**			
	12	**13**	**14**	**15**	**16**
No of times accessed	**N**	**N**	**N**	**N**	**N**
1	59	87	40	64	40
2	31	20	39	31	8
3	16	8	24	13	4
4	10	2	9	8	0
5	2	1	1	1	0
6	2	0	1	0	0
More than 6	0	0	4	1	0

Table 4.5: Frequency of access for screen 12 in the animation version and screen 13-16 in the static images & text version

In the animation version (n = 120) the number of participants who viewed screen 12 once or between 2 – 4 times was almost equal: 59 and 57 respectively. In the static images & text version (n = 118) this pattern was different. More than half of the participants only accessed screen 13 once, most of the participants viewed screen 14 between 2 and 4 times and the number accessing screen 15 either once or between 2–4 times was almost equal. The one advantage of the static images & text version was that the learner had more control over how they viewed the content. For example, if they wanted to view a small section of the animation (screen12) again they either had to run the full animation or use the control bar and scroll through the animation until they found the place they were looking for. I did not explore whether users found this action of 'scrolling' through an animation irritating or not, but as an experienced user myself I find this method cumbersome. In the static images & text version the user can click through the screens very quickly. It is much easier to scan the content and find the place you are looking for because there is no movement on the screen as is the case when viewing an animation. The design of the animation in this study also required the participant to actively start the animation. This and other types of user control and usability and the cognitive load associated with this type of navigation could be investigated further in future studies.

4.4.3 In summary: time spent on the program

This section first considered the time spent on the program in general. The participants spent between 15.03 and 76.76 minutes studying the content of the animation version and between 22.86 and 92.76 minutes studying the content of the static images & text version. The time spent on the program was divided into three groups, which were labeled Inadequate (IA), Adequate (AD) and More than Adequate (MA). The effect size was also determined for each group and for the categories 'Inadequate' and 'Adequate time' a very large effect size was returned. The effect size for the 'More than Adequate' time category was in the medium-to-large range. These results, indicating both statistical and practical significance, must still be interpreted cautiously as there **were** more screens in the static images & text version. The results only provide a big picture view of the time spent on the program. Of concern was the number of participants spending less than 30 min studying the animation version (17.5 % of the sample) and less than 40 min for static images & text version (39% of the sample).

Screen-wise comparisons, where the screens of interest were screen 12 in the animation version and screens13-16 in the static images & text version, were conducted. The analysis indicated that the difference between the mean time spent on screen 12 and screens 13-16 was statistically significant. The effect size was $d = 0.710$, indicating practical significance. This finding was expected as there were more screens to access in the static images & text version.

I now move on the explore the role cognitive style plays in an authentic multimedia learning environment.

4.5 Exploring the role cognitive style plays in an authentic multimedia learning environment

This section discusses the cognitive styles of the participants and explores the use of the multimedia within the context of their particular styles. The analysis included looking at whether gender and culture are indicators of cognitive style.

The first sub-question asked in this study was 'What are the cognitive styles of the participants taking part in the study?' The two hypotheses, discussed in Chapter 3, are:

Null Hypothesis 1a	Alternate Hypothesis 1a
There will be no difference in the percentage of the sample having either an Analytic / Intermediate or Wholistic style on the WA style dimension.	*More than 50% of the sample will have an Analytic or Intermediate style on the WA style dimension.*

Null Hypothesis 1b

There will be no difference in the percentage of the sample having either a Verbaliser /Bimodal or Imager style on the VI style dimension.

Alternate Hypothesis 1b

More than 50% of the sample will have a Verbaliser or Bimodal style on the VI style dimension.

4.5.1 Cognitive style as measured using Riding's CSA

The cognitive style of each participant was measured using Riding's Cognitive Style Analysis (CSA). The CSA provides a score for each dimension in the cognitive style model. The ratios for each style dimension typically range from 0.4 through to 4.0 with a central value around 1.0 (Riding, 2005a).

In this study the Wholist-Analytic (WA) ratios of 235 participants (3 participants did not do the CSA) ranged from 0.70 to 4.47 (Mean = 1.57, SD = 0.5898). The Verbaliser-Imager (VI) ratios for these 235 participants ranged from 0.39 to 2.14, (Mean = 1.0868, SD = 0.1983). The WA ratios are above the 0.4 described by Riding for the lower end of the WA style dimension (Riding, 2005a) but just over the upper range of 4.0, indicating a shift to the Analytic end of the dimension for the sample in this study. The VI ratios for this study are just above the 0.4 at the one end of the continuum. The upper limit of the VI ratio for this study, 2.14, is well below the upper limit suggested by Riding.

The correlation between the Wholist-Analytic and Verbaliser-Imager style ratios, using Pearson's Correlation Coefficient, was low (r = 0.0372) and not significant (p = 0.5706). This is in line with correlations reported in the literature (Douglas & Riding, 1993, Rezaei & Katz, 2004; Riding & Grimley, 1999; Riding & Rayner, 1998; Riding & Sadler-Smith, 1992; Riding & Staley, 1998). The results indicate the independence of the two style dimensions, which is also described in the cognitive style model.

4.5.1.1 The Wholistic-Analytic dimension

The literature describes both two and three points along the Wholistic-Analytic dimension, used primarily for the purpose of grouping the participants. Using three categories, the profile of the sample is displayed in Table 4.6.

	WA ratio	Male	Female	TOTAL
Wholist	≤ 1.02	11 (14.7%)	25 (15.6%)	**36**
Intermediate	> 1.02 and ≤ 1.35	18 (24.0%)	47 (29.4%)	**65**
Analytic	> 1.35	46 (61.3%)	88 (55.0%)	**134**
TOTAL		**75**	**160**	**235**

Table 4.6: Profile of the sample for the WA dimension of the CSA using three style groups

The percentages in Table 4.6 are column percentages. Looking at the total number of participants in each style group we see that more than half of the sample are Analytic in style: fifty-seven percent (n = 134) of the sample had an Analytic style, 28% (n=65) had an Intermediate style and only 15 % (n=36) had a Wholistic style.

Table 4.6 also provides a gender perspective of the style results. Sixty one percent (61%) of the male participants and 55% of the female participants were Analytic in style. Only 24% of the males were in the Intermediate style range for the WA dimension, while 29% of the females had an Intermediate style.

If only two categories of WA style are used, gender is not included in the analysis, and a ratio of 1.20 is used as the dividing point along the WA continuum, then the data indicates that 70% of the participants were Analytic in the way they process and organise information and 30% were Wholistic.

> *The results from the analysis of the WA style dimension **support** hypothesis 1a that more than 50% of the sample will have an Analytic or Intermediate style on the WA style dimension.*

4.5.1.2 The Wholistic–Analytic dimension: looking at gender and culture

This section explores the WA style dimension and the relationship between WA style, gender and culture in more detail. As discussed in Chapter 2, many studies look at style and gender, but I could find no literature related to cognitive style and culture.

The mean WA style ratios for the sample, displayed according to gender, are presented in Table 4.7.

		Male		**Female**	
	Ratios	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Wholist	≤ 1.02	11	0.8836 (0.1040)	25	0.8476 (0.0916)
Intermediate	>1.02 and ≤1.35	18	1.1878 (0.0944)	47	1.1949 (0.0892)
Analytic	> 1.35	46	1.9565 (0.6184)	88	1.9375 (0.4435)

Table 4.7: WA style ratios according to gender

A Chi-Square analysis, appropriate to test null hypotheses for categorical data, indicated that there was no significant difference between the gender groups for the Wholistic-Analytic style dimension, $\chi^2 = 0.9232$, df = 2, p = 0.6303. The phi coefficient, a measure of the strength of the relationship between gender and WA style is 0.0627. This is a very small effect indicating that in practice there is no relationship between gender and WA style.

The mean WA style ratios, displayed according to culture, are presented in Table 4.8.

	Black		**Coloured**		**Indian**		**White**	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Wholist ≤ 1.02	5	0.8400 (0.1091)	1	1.0000 (n.a.)	2	0.7250 (0.0353)	28	0.8664 (0.0888)
Intermediate >1.02 & ≤1.35	15	1.2153 (0.0983)	1	1.2400 (n.a.)	7	1.1257 (0.0810)	42	1.1950 (0.0859)
Analytic >1.35	44	2.0107 (0.6193)	2	1.8050 (0.3182)	14	2.1186 (0.6027)	74	1.8751 (0.4058)

Table 4.8: WA style ratios according to culture

These four categories for culture were collapsed into two groups for further analysis: White participants and Non-white participants (included Black, Coloured and Indian participants).

A stepwise regression analysis was performed using the Wholistic-Analytic cognitive style as the dependent variable. This approach was followed in order to determine which of the factors influence cognitive style. The following variables, and all their interactions, were included in the regression analysis:

- Version of the multimedia
- Age
- Gender
- Cultural group
- Pretest scores

The result of the regression was *F*(2, 231) = 6.10, p = .0026, R-square = 0.0502 and C(p) = 0.9483. The value for C(p) (known as Mallow's coefficient) provides an indication of the completeness of the model. The value of this coefficient should be a close as possible to the number of parameters in the regression. The variables retained in the stepwise regression were culture and the interaction between version and gender.

The stepwise regression equation is presented in Table 4.9.

Variable	Parameter Estimate	Standard Error	Type 11 Sum of squares	F-value	p-value
Intercept	2.1166	0.1628	56.0932	169.00	<.0001
Culture	-0.2521	0.0777	3.4932	10.52	0.0014
Version x Gender	-0.0545	0.03467	0.8221	2.48	0.1169

Table 4.9: Stepwise regression equation for WA style

From the data it can be determined that there was a statistically significant main effect for culture, *F*(2, 231) = 10.52, p = 0.0014, but not for the interaction between version and gender (p = 0.1169).

A confirmatory General linear model (GLM) was then run, based on the results from the stepwise regression. The GLM presents the standard error and p-difference tables, the results of which are not provided in the stepwise regression. The GLM allows comparison of the class variable effects (version, gender and culture) using pair wise least squares means comparisons. The model used for the GLM was: WA cognitive style = version, gender, culture, and the interaction between version and gender.

The GLM analysis returned a statistically significant finding, *F*(3, 230) = 3.17, p = 0.0146, R^2 = 0.052. The culture of the participants accounted for the significant result, *F*(1, 230) = 9.12, p = 0.0028, R^2 = 0.052. Further post hoc comparison, using Fischer's Least Squares Means, indicated that the non-White participants were statistically significantly more Analytic that the White participants. The non-White participants had a Mean (±SE) WA ratio of 1.7268 (0.0639) and for the White participants it was 1.4909 (0.0497).

This effect size (*d*) was calculated using the following formula:

$$d = \frac{\left|\overline{x}_i - \overline{x}_j\right|}{\sqrt{MSE}}$$

where

$\overline{x}_i$ and $\overline{x}_j$ are the group means and MSE is the Mean Standard Error of the ANOVA (Ellis & Steyn, 2003).

The effect size obtained was 0.40. This effect size is bordering on the medium effect size range for this study. This is a visible effect. This conclusion is made cautiously as there are not enough similar comparisons (cognitive style and culture) in the cognitive style literature against which to benchmark this finding.

This concludes the section on the Wholistic-Analytic dimension of cognitive style. The next section considers the Verbaliser-Imager dimension of cognitive style.

4.5.1.3 The Verbaliser-Imager dimension

This section presents the analysis of the data to determine the Verbaliser-Imager profile of the sample used in this study.

The literature describes both two and three points along the Verbaliser-Imager dimension, used primarily for the purpose of grouping the participants. Using three categories the profile of the sample is displayed in Table 4.10.

	VI ratio	**Male**	**Female**	**TOTAL**
Verbaliser	≤ 0.98	19 (25.4%)	45 (28.1%)	**64**
Bimodal	> 0.98 and ≤ 1.09	25 (33.3%)	49 (30.6%)	**74**
Imager	> 1.09	31 (41.3%)	66 (41.3%)	**97**
TOTAL		**75**	**160**	**235**

Table 4.10: Profile of the sample for the Verbaliser-Imager Dimension of the CSA using three style groups

The percentages in Table 4.10 are column percentages, reflecting distribution for gender per style sub-group. The data in this table indicates that the sample appears to lean slightly more to the Imager style. Forty one percent (n=97) of the sample had an Imager style, 31% (n=74) were Bimodal in style and only 28 % (n=64) were Verbaliser in style.

Table 4.10 also provides a gender perspective of the style results. Forty-one point three (41.3%) of both the male and female participants were Imagers. The distribution of the Bimodal style was also very similar for male and female participants (33.3% and 30.6% respectively). Only 25.4% of the males and 28.1% of the females were Verbalisers.

If only two categories of VI style are used, gender is not included in the analysis, and a ratio of 1.035 is used as the dividing point along the VI continuum, the data indicates that 40% of the participants are Verbalisers with regard to their inclination to represent information and 60% are Imagers in style. The results therefore indicate that the Imager style was strongly represented in this sample.

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

4.5.1.4 The Verbaliser-Imager dimension: looking at gender and culture

This section explores the Verbaliser-Imager (VI) style dimension and the relationship between VI style, gender and culture in more detail. The mean VI style ratios for the sample, displayed according to gender, are presented in Table 4.11.

		Male		**Female**	
	Ratios	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Verbaliser	≤ 0.98	19	0.8800 (0.1193)	45	0.9002 (0.0943)
Bimodal	> 0.98 and ≤ 1.09	25	1.0392 (0.0294)	49	1.0445 (0.0288)
Imager	> 1.09	31	1.2842 (0.2448)	66	1.2301 (0.1572)

Table 4.11: VI style ratios according to gender

A Chi-Square analysis, appropriate to test null hypotheses for categorical data, indicated that there was no significant difference between the gender groups for the Verbaliser-Imager style dimension, $\chi^2 = 0.2652$, df = 2, p = 0.8758. The phi coefficient, a measure of the strength of the relationship between gender and VI style is 0.0336. This is a very small effect size indicating that in practice there is no relationship between gender and VI style.

The mean VI style ratios, displayed according to culture, are presented in Table 4.12.

	Black		**Coloured**		**Indian**		**White**	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Verbaliser (≤ 0.98)	20	0.8920 (0.1322)	1	0.8200 (n.a)	8	0.9100 (0.0431)	35	0.8940 (0.0941)
Bimodal > 0.98 & ≤1.09	22	1.0545 (0.0292)	1	1.0500 (n.a.)	4	1.0500 (0.0264)	47	1.0359 (0.0276)
Imager (> 1.09)	22	1.2990 (0.2290)	2	1.1750 (0.0210)	11	1.1854 (0.0784)	62	1.2424 (0.1894)

Table 4.12: VI style ratios according to culture

These four categories for culture were collapsed into two groups for further analysis: White participants and Non-white participants (included Black, Coloured and Indian participants).

A stepwise regression analysis was performed using the Verbaliser-Imager cognitive style as the dependent variable. This approach was followed in order to determine which of the factors influence cognitive style. The following variables, and all their interactions, were included in the regression analysis:

- Version of the multimedia
- Age
- Gender
- Cultural group
- Pretest scores

The result of the regression was $F(1, 232) = 5.69$, p = 0.0179, R-square = 0.02 and C(p) = -4.46. Only one variable was retained in the stepwise regression and that was the pretest score. The stepwise regression equation is presented in Table 4.13.

Variable	**Parameter Estimate**	**Standard Error**	**Type 11 Sum of squares**	**F-value**	**p-value**
Intercept	0.9629	0.0533	12.5958	325.71	<.0001
Pretest	0.0106	0.0044	0.2200	5.69	0.0179

Table 4.13: Stepwise regression equation for VI style

No further analysis was conducted for the VI style dimension. The data does not indicate that there is any relationship between gender or culture and the VI dimension of style.

4.5.1.5 In summary: cognitive style profile of the participants in this study

Section 4.5.1 reported on the cognitive style profile of the sample used in this study. The relationship between cognitive style, gender and culture for both the WA and VI style dimension was explored. The profile of the sample in this study was that of a group that was predominantly Analytic (69% of the sample) and Imager in style (60% of the sample).

Regarding style and gender, the results of the analyses indicate that there was no statistically significant difference between the gender groups for both the Wholistic-Analytic style dimension, χ^2 = 0.0232, df = 2, p = 0.6303, and the Verbaliser-Imager style dimension, χ^2 = 0.2652, df = 2, p = 0.8758. The phi coefficient of 0.0627 for the WA style and 0.0336 for the VI style are very small effect sizes and suggest that in practice there is no relationship between gender and cognitive style.

Regarding the WA cognitive style dimension and culture, a stepwise regression determined that there was a statistically significant main effect for culture, $F(2, 231) = 10.52$, p = 0.0014. This was followed up with confirmatory general linear modeling (GLM), which returned a statistically significant finding, $F(3, 230) = 3.17$, p = 0.0146, R^2 = 0.052. The culture of the participants accounted for the significant result, $F(1, 230) = 9.12$, p = 0.0028, R^2 = 0.052. Further post hoc comparison, using Fischer's Least Squares Means, indicated that the non-White participants were significantly more Analytic than the White participants. This effect size (*d*) obtained was 0.40. This effect size is in the small to medium range, but is a visible effect, although the interpretation is made with caution.

A stepwise regression with the VI style as the dependent variable did not return a statistically significant finding for any of the demographic variables entered into the equation.

4.5.2 Cognitive style and time spent on the program

One of the early questions asked while planning this study was 'What do learners actually DO with multimedia programs?' One perspective of the answer is to look at the time learners spend studying content in a multimedia program. A refinement of this question is 'Is the time spent on the program related to cognitive style and/or load?' This section looks at the amount of time spent on the program in context of cognitive style.

The distribution of the WA and VI style across the two versions of the program is presented in Table 4.14.

	Wholistic	Analytic	Verbaliser	Imager
Animation version	37	83	50	70
Static images & text version	36	79	44	71
TOTAL	**73**	**162**	**94**	**141**

Table 4.14: Frequency of WA and VI styles for each version of the program

4.5.2.1 The Wholistic – Analytic dimension and time spent on the program

The hypothesis tested was:

Null Hypothesis 1c

There will be no difference between the Analytic and Wholistic learner in the time spent studying the content of the program.

Alternate Hypothesis 1c

The Analytic learner will spend more time studying the content of the program than the Wholistic learner.

The mean time spent on each version, for the WA style is presented in Table 4.15.

	Wholistic		Analytic	
	n	**Mean (SD)**	**n**	**Mean (SD)**
Animation version	37	40.505 (12.812)	83	43.172 (10.980)
Static images & text version	36	40.957 (13.748)	79	47.537 (13.795)

Table 4.15: Comparison of time spent on program for WA style and version

From Table 4.15 we observe that for both versions the Analytic learner spent more time on average studying the content. This provides tentative support for Hypothesis 1c. The Analytic learner will spend time processing the detail of the content, while the Wholistic learner will scan more quickly to get a big picture view.

A series of t tests for independent samples was performed to determine if the difference in the means were significant. Looking at the animation version a t test for significance indicated that the difference between the Wholistic and Analytic style was not statistically significant, $t = 1.17$, $df = 118$, $p = 0.246$. The effect size was $d = 0.20$, which is small and therefore also not practically significant.

The time the Analytic learner used to study the content of the static images & text version was statistically significantly more than the time the Wholistic learner used, t = 2.37, df = 113, p = 0.0192. The effect size was *d* = 0.47. This effect size is close to the medium effect size range and could be regarded as being a visibly longer time for Analytic.

> ***Hypothesis 1c*** *is therefore supported for the static images & text version of the program, but not for the animation version.*

The data in Table 4.15 also allows a comparison of the time each style spent on the different versions. The Wholistic learner spent almost the same amount of time for each version (40.5 minutes on the animation version and 40.9 minutes on the static images & text version) in spite of the fact that that there were more screens in the static images & text version. This difference in time was not statistically significant, t = -0.15, df = 71, p = 0.8847. The effect size was also very small and therefore not of any practical significance (*d* = 0.03).

The Analytic learner spent an average of 43.1 minutes on the animation version and 47.5 minutes on the static images & text version. Although this difference in time is statistically significant, t = -2.23, df = 160, p = 0.0269, this result must be interpreted cautiously as there were more screens in the static images & text version. The effect size obtained was also small (*d* = 0.31). Proportionally the Analytic learner spent slightly more time studying the content of the animation version.

The analysis then drilled down to look at the time spent by the Analytic and Wholistic learner at screen level. The screens of interest were screen 12 in the animation version and screens 13-16 in the static images & text version. The mean times for each screen are summarised in Table 4.16.

	Wholistic		**Analytic**	
	N	**Mean (SD)**	**n**	**Mean (SD)**
Screen 12: Animation version	37	5.4000 (3.4539)	83	5.4474 (3.1547)
Screens 13 – 16: Static images & text version	36	6.7194 (3.4538)	79	7.8211 (2.601)

Table 4.16: Comparison of time spent on selected screens for WA style

Once again Table 4.16 allows for comparison from two different perspectives. Firstly it is possible to look at the version (at screen level) and compare style differences. The second comparison looks at each style individually and compares the time spent across the two versions.

When comparing the mean time between the styles for each screen we see in Table 4.16 that there is virtually no difference between the Wholistic and Analytic participant in the time spent studying the animation. A t test did not return a statistically significant difference in these mean times, t = -0.07, p = 0.9413. The effect size was also very small *d* = 0.02. Looking at Table 4.16 we also see that both the Wholistic and Analytic learner spent more time on average studying the content of screen 13-16 in the static & images version. There is however a larger difference between the Wholistic and Analytic participants in the mean time spent studying the static images & text version, but it was also not statistically significant at an alpha of 0.05, t = -1.89, p = 0.0607. The effect size however does point to a visible effect (*d* = 0.60).

I then considered the time **each style group** spent on the different screens. In considering the Wholistic style, a t test for independent samples did not return a statistically significant value, t = -1.63, df = 71, p = 0.107, when comparing the mean time of screen 12 and screens 13-16. The effect size was *d* = 0.38, which suggests that the difference might be visible. The result for the Analytic style indicated that the difference in the mean time for screen 12 and screens 13-16 was statistically significant, t = -5.21, df = 160, p < 0.0001. The effect size was *d* = 0.69, indicating a visible effect.

At screen level the data suggests that hypothesis 1c is supported for the static images & text version of the program, but not for the animation version.

Using the time groups defined earlier (Inadequate, Adequate and More than Adequate) a 3 x 2 contingency table was used to analyse the data for time and style for the animation version. This analysis is presented in Table 4.17.

Animation version	**Wholistic**		**Analytic**	
	N	**Col %**	**N**	**Col %**
Inadequate time spent	10	27.03	11	13.25
Adequate time spent	19	51.35	53	63.86
More than Adequate time spent	8	21.62	19	22.89

Table 4.17: Comparison of time spent on the animation version for the WA style dimension

A Chi-Square analysis, appropriate to test a null hypothesis for categorical data, indicated that there was no significant difference in the time groups for the two styles, Wholistic and Analytic, $\chi^2 = 3.4597$, df = 2, p = 0.1773. The effect size was *w* = 0.1698. This is a small effect and cannot be regarded as practically significant.

A similar analysis was done for the static images & text version and the data is presented in Table 4.18.

Static images & text version	Wholistic		Analytic	
	N	Col %	N	Col %
Inadequate time spent	22	61.11	23	29.11
Adequate time spent	11	30.56	49	62.03
More than Adequate time spent	3	8.33	7	8.86

Table 4.18: Comparison of time spent on the static images & text version for the WA style dimension

A Chi-Square analysis, appropriate to test a null hypothesis for categorical data, indicated that there was a significant difference in the time groups for the two styles, Wholistic and Analytic, $\chi^2 = 11.1727$, $p = 0.0037$. The phi coefficient, a measure of the strength of the relationship between the time spent on the program and WA style is 0.3117. This effect size indicates that in practice this difference in time spent by the different style groups is visible. Sixty-two percent of the Analytic learners spent adequate time on the program, in contrast to only 30.56% of the Wholistic learners who spent adequate time on the program. Of concern here is the fact that 61% of the Wholistic learners spent inadequate time on the program compared to only 29.11% of the Analytic learners.

I now turn to the analysis of the Verbaliser-Imager dimension and the amount of time spent on the program by this group of participants.

4.5.2.2 The Verbaliser–Imager dimension and time

The hypothesis tested was:

Null Hypothesis 1d

There will be no difference between the Verbaliser and Imager learner in the time spent studying the content of the program.

Alternate Hypothesis 1d

The Verbaliser will spend less time studying the content of the program than the Imager.

The mean time spent on each version, for the VI style is presented in Table 4.19.

	Verbaliser		**Imager**	
	n	**Mean (SD)**	**n**	**Mean (SD)**
Version 1: Animation	50	40.936 (10.517)	70	43.359 (12.270)
Version 2: Static images & text	44	46.183 (12.912)	71	45.04 (14.229)

Table 4.19: Comparison of time spent on program for VI style and version

From Table 4.19 we observe that the participants with an Imager style spent more time on average studying the content of the animation version. Using the t test for independent samples, the analysis indicated that the difference between the styles for the animation version was not statistically significant, t = -1.13, df = 118, p = 0.2605. The effect size was *d* = 0.20, which is a small effect and can be regarded as having no practical significance.

The two style groups spent about the same time on average studying the content of the static images & text version. The difference was not statistically significant, t = 0.42, df = 113, p = 0.6738. The effect size was *d* = 0.08, which is very small and cannot be regarded as practically significant.

> ***Hypothesis 1d**, which states that the Verbaliser will spend less time studying the content of the program than the Imager is therefore **not supported** for two versions of the program.*

The second hypothesis tested was:

> ***Null Hypothesis 1e***
>
> *There will be no difference in the amount of time the Verbaliser learner spends proportionally studying the two versions of the content.*
>
> ***Alternate Hypothesis 1e***
>
> *The Verbaliser will spend proportionally more time studying the content of the static images & text version than the animation version.*

When comparing the time the Verbaliser learner spent by version the data in Table 4.19 indicates that the Verbaliser learner spent more time on average studying the static images & text version (M [±SD] = 46.183 [12.912]) than studying the content in the animation version (M [±SD]) = 40.936 [10.517]). 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, and therefore the t test for independent samples, which indicated that the difference between the time for the different versions was significant [t = -2.08, df = 92, p = 0.0406], might not be that relevant for this comparison. The effect size for the difference in the means was *d* = 0.38, which indicates that this result might be visible.

An analysis was done of the time spent by the Verbaliser learner and Imager learner at screen level. The screens of interest were screen 12 in the animation version and screens 13-16 in the static images & text version. The mean times for each screen are summarised in Table 4.20.

	Verbaliser		**Imager**	
	n	**Mean (SD)**	**n**	**Mean (SD)**
Screen 12 in animation version	50	5.5403 (3.716)	70	5.356 (2.869)
Screens 13 – 16. in Static images & text version	44	7.0394 (2.7435)	71	7.7469 (3.0188)

Table 4.20: Comparison of time spent on selected screens for VI style

Once again Table 4.20 allows comparisons from two different perspectives. Firstly it is possible to look at the version (at screen level) and compare style differences. The second comparison looks at each style individually and compares the time spent across the two versions

When comparing the mean time between the styles for each screen we see in Table 4.20 that there is a small difference between the Verbaliser and Imager participant in the mean time spent studying the animation. A t test did not return a statistically significant difference in these mean times, t = 0.31, p = 0.7597. The effect size was also very small *d* = 0.05. There is however a larger difference between the Verbaliser and Imager participant in the mean time spent studying the static images & text version, but it was still not statistically significant at an alpha of 0.05, t = -1.26, 0.208. The effect size is also small (*d* = 0.23) and not of any practical significance.

I then considered the time **each style group** spent on the different screens. In considering the Verbaliser style, a t test for independent samples returned a statistically significant value, t = -2.20, df = 92, p = 0.0304 when comparing the mean time of screen 12 and screens 13-16. The effect size was *d* = 0.40, which suggests that the difference is visible. The result for the Imager style indicated that the difference in the mean time for screen 12 and screens 13-16 was also statistically significant, t = -4.82, df = 139, p < 0.0001. The effect size was *d* = 0.79, indicating a practically significant difference.

Using the time groups defined earlier (Inadequate, Adequate and More than Adequate), a contingency table was used to analyse the data for time and style for the animation version. This analysis is presented in Table 4.21.

Version 1	**Verbaliser**			**Imager**		
	N	**Row %**	**Col %**	**N**	**Row %**	**Col %**
Inadequate time spent	11	52.38	22	10	47.62	14.29
Adequate time spent	29	40.28	58	43	59.72	61.43
More than Adequate time spent	10	37.04	20	17	62.96	24.29

Table 4.21: Comparison of time spent on the animation version for the VI style dimension

A Chi-Square analysis, appropriate to test a null hypothesis for categorical data, indicated that there was no significant difference in the time groups for the two styles, Verbaliser and Imager, $\chi^2 = 1.2871$, df = 2, p = 0.0.5254. The effect size was $w = 0.1036$, which is small and does not indicate practical significance.

A similar analysis was done for the static images & text version and is presented in Table 4.22.

Version 2	**Verbaliser**			**Imager**		
	N	**Row %**	**Col %**	**N**	**Row %**	**Col %**
Inadequate time spent	17	37.78	38.64	28	62.22	39.44
Adequate time spent	24	40	54.55	36	60	55.70
More than Adequate time spent	3	30	6.82	7	70	9.86

Table 4.22: Comparison of time spent on the static images & text version for the VI style dimension

A Chi-Square analysis, appropriate to test a null hypothesis for categorical data, indicated that there was no significant difference in the time groups for the two styles, Verbaliser and Imager, $\chi^2 = 0.3702$, df = 2, p = 0.8310. The effect size was $w = 0.0567$, which is very small and does not indicate practical significance.

4.5.2.3 In summary: cognitive style and time spent on program

This section considered the time spent on the program from the perspective of the different cognitive style dimensions. Looking at the animation version we note that the Analytic leaner spent more time studying the program than the Wholistic learner but the difference was not statistically significant. The effect size was *d* = 0.20, which is small and therefore also not practically significant. For the static images & text version the time the Analytic learner used to study the content was statistically significantly more than the time the Wholistic learner used. **Hypothesis 1c** was therefore supported for the static images & text version of the program (p = 0.0192), but not for the animation version (p = 0.246).

A comparison was also done for the amount of the time each style group spent on the different versions. The Wholistic learner spent almost the same amount of time for each version in spite of the fact that that there were more screens in the static images & text version. This behaviour would seem to confirm the approach a Wholistic learner takes to learning, namely that of scanning in order to get a big picture. This reduced the time spent on the static images & text version, even though there were more screens in the static images & text version. Proportionally the Wholistic learner spent more time studying the content of the animation version. This is quite possible since the Wholistic learner had to watch the whole animation in order to understand it and could not scan the information. The Analytic learner spent less time studying the animation version than the static images & text version and the difference in time across the two versions was statistically significant (p = 0.0269). The effect size obtained was also small (*d* = 0.31). Proportionally the Analytic learner spent more time studying the content of the animation version.

Comparisons were then carried out at screen level, where once again the screens of interest were screen12 in the animation version and screens 13-16 in the static images & text version. When comparing the mean time between the styles for each screen there was virtually no difference between the Wholistic and Analytic participant in the time spent studying the animation. There was however a larger difference between the Wholistic and Analytic participant in the mean time spent studying the static images & text version, but it was also not statistically significant. The effect size however was in the medium-to-large range indicating practical significance. In considering the time **each style group** spent on the different screens it was found that the time spent by the Wholistic learner was not statistically different when comparing the mean time of screen 12 and screens 13-16. The effect size was *d* = 0.38, which suggests that the difference might be visible. The result for the Analytic style indicated that the difference in the mean time for screen 12 and screens 13-16 was statistically significant, with a large effect size with practical significance.

Sixty-two percent of the Analytic learners spent adequate time on the program, in contrast to only 30.56% of the Wholistic learners. Of concern here is the fact that 61% of the Wholistic learners spent inadequate time on the program compared to only 29.11% of the Analytic learners.

The last section considered the Verbaliser-Image dimension. Participants with an Imager style spent more time on average studying the content of the animation version but this difference was not statistically significant and the effect size was small. While not significant the Verbaliser did spend more time studying the content of the static images & text version, which had proportionally fewer images than the animation version, and the Imager spent more time studying the content of the animation version, which had proportionally more images than the static images & text version.

An analysis was done of the time spent by the Verbaliser learner and Imager learner at screen level. The screens of interest were screen 12 in the animation version and screens 13-16 in the static images & text version. When comparing the mean time between the styles for each screen a small difference was found between the Verbaliser and Imager participant in the mean time spent studying the animation and it was not statistically significant. There was a larger difference between the Verbaliser and Imager participant in the mean time spent studying the static images & text version, but it was still not statistically significant. Effect sizes were small.

In considering the Verbaliser style, this participant spent more time on screens 13-16 than on screen 12 and the difference was statistically significant ($p = 0.0304$). In considering the Imager style, this participant spent more time on screens 13-16 than on screen 12 and the difference was statistically significant ($p < 0.0001$), with a large effect size pointing to practical significance.

The next section considers cognitive style and the use of the multimedia program.

4.5.3 Cognitive style and use of the multimedia program

Section 4.5.2 presented the findings related to the amount of time spent on the program by participants with different styles. The question asked was 'What do learners actually DO with multimedia programs?' This section analyses the data from another perspective, namely navigation patterns. The data collected also included the exact sequence of navigation through the program for each participant. These navigation patterns were diverse. This section will report on the following:

- A qualitative impression of navigation patterns
- The number of participants who went through the program once from start to finish
- Selected screens that were accessed more than twice and the cognitive style profile of the participants

4.5.3.1 A qualitative impression of navigation patterns and observation of participants during the computer session

The participants were allowed to make notes during the learning session. These notes were collected before they started the posttest. These notes were merely scanned in order to get a general impression of note-taking practices in the sample.

Scribble pages, the name given to this page hand-out for each participant, were collected from 244 participants. Fifty-two participants did not make any entries on this page. The pages received from the remaining 192 participants varied from very detailed, including the use of different colours of pen, to almost nothing. Some participants attempted to write down everything for each screen, while others made detailed notes for selected screens. Some participants only made summaries, which were mostly key words and key ideas. Some of these pages were extremely neat and organised and others were almost undecipherable. There was a combination of textual and diagrammatic notes. The diagrams varied from basic drawings of the autonomic nervous system to the use of structures that resembled mind maps, with text and arrows.

A detailed analysis of these notes is beyond the scope of this study. Detailed analysis of these notes is an avenue for further research. Of particular interest would be to determine if the different style groups approached note-taking differently.

4.5.3.2 Participants who accessed the program once

Only 19 of the 238 participants worked through the program once, completing the program in a linear fashion. An analysis of the cognitive styles of these participants is presented in Table 4.23.

		Animation	**Static images & text**
WA style dimension	**Wholistic**	7	3
	Analytic	7	2
VI style dimension	**Verbaliser**	8	2
	Imager	6	3

Table 4.23: Cognitive style of participants who accessed the program once only

The pattern that emerges in Table 4.23 indicates that in each version the number of participants accessing each program only once were about the same for each of the style groups. These low frequencies make analysis difficult and the results must be interpreted cautiously. More participants accessed the animation version more than once.

A Chi-Square analysis indicated that there was no significant difference between the versions of the program for the Wholistic-Analytic style dimension, $\chi^2 = 0.1478$, df = 1, p = 0.7007. The phi coefficient, a measure of the strength of the relationship between version and WA style was -0.0882. This very small effect size indicates that in practice there is no relationship between version, WA style and the single access to the program.

A Chi-Square analysis indicated that there was no significant difference between the versions of the program for the Verbaliser-Imager style dimension, $\chi^2 = 0.4343$, df = 1, p = 0.5099. The phi coefficient, a measure of the strength of the relationship between gender and VI style is 0.1512. This small effect size indicates that in practice there is no relationship between version, VI style and he single access to the program.

4.5.3.3 Cognitive style profile of participants accessing selected screens more than twice

The majority of the participants entered selected screens more than twice. The screens of interest at this point are screens 2, 12 and 19 in the animation version, and screens 2, 13-16 and 23 in the static images & text version.

Screen 2, which was identical in both versions, had two functions. Firstly it provided the user with a big picture view of the entire program. Secondly, once the participant had worked through the program once they could use this screen as a menu to go back to whatever section of the program they wished to review again. This screen is illustrated in Figure 3.5 on page 148 of this thesis and in Appendix H.

I looked at the pattern of access for the animation version first. Looking at the WA style dimension, screen 2 was accessed between 2 and 16 times by the participants with access between 2 and 8 times accounting for the majority of the access. Looking at the VI style dimension screen 2 was also accessed between 2 and 16 times with access between 2 and 7 times accounting for the majority of the access.

I then looked at the pattern of access for screen 2 for the static image & text version. For both the WA and the VI style dimension, screen 2 was accessed between 5 and 23 times by the participants. Access between 5 and 11 times accounted for the majority of the access for both style dimensions.

The pattern that emerged was that screen 2 was used by some participants as a menu. This superficial analysis could not answer the question of whether the Wholistic learner used screen 2 to get the big picture view of the program. More detailed analysis of the path of navigation by each participant after they had accessed screen 2 is beyond the scope of this study, but could provide an avenue for future research.

A series of Chi-square analyses were conducted to determine if there was any relationship between the number of times screen 12 and 19 were accessed and the cognitive styles of the participants. Table 4.24 displays the results for the animation version for the two style dimensions, including the size effect (*w*).

	Wholistic-Analytic Dimension				**Verbaliser-Imager Dimension**			
	χ^2	**df**	**p**	***w***	χ^2	**df**	**p**	***w***
Screen 12	1.3636	3	0.7141	0.21	3.3767	3	0.3371	0.34
Screen 19	4.4211	3	0.2194	0.40	3.7238	3	0.2920	0.35

Table 4.24: Results of Chi-square analyses to determine relationship between style and access in the animation version

Table 4.25 displays the results for static images & text version for the two style dimensions, including the size effect (*w*). Screens 13-16 were analysed as a group and individually, with the exception of screen 16 due to the fact that there were so few entries into screen 16.

	Wholistic-Analytic Dimension				**Verbaliser-Imager Dimension**			
	χ^2	**df**	**p**	**w**	χ^2	**df**	**p**	**w**
Screen 13-16	12.5269	15	0.6388	0.3300	21.7566	15	0.1143	0.4350
Screen 13	0.9167	2	0.6323	0.2887	0.9167	2	0.6323	0.2887
Screen 14	8.2036	6	0.2236	0.4709	4.7591	6	0.5751	0.3586
Screen 15	1.1703	3	0.7601	0.2256	0.6192	3	0.8920	0.1641
Screen 23	3.0866	4	0.5434	0.3446	6.8014	4	0.1468	0.5115

Table 4.25: Results of Chi-square analyses to determine relationship between style and access in the static images & text version

None of the Chi-square analyses returned a significant finding. The values of many of the cells in each of the contingency tables set up for the Chi-square analyses presented in Tables 4.25 and 4.26 were smaller than 5. Chi-square analysis may therefore not be a valid test. In each case the Fischer's Exact Test was applied, but there were no significant findings either. It must be concluded that there seems to be no relationship between cognitive style and the number of times the screens were accessed. Other analyses, beyond the scope of this study might shed more light on the relationship between cognitive style and the pattern of navigation in multimedia learning.

4.6 Exploring the role cognitive load plays in an authentic multimedia learning environment

Four sub-questions were asked about the cognitive load of the two interventions used in this study.

The first sub-question explored was 'How do the participants rate the cognitive load of selected multimedia content?'

The 9-point rating scale developed by Paas (1993) was used in this study. Cronbach's coefficient alpha, using all the recorded measures of cognitive load, was used to determine the internal consistency of the scale. This coefficient was calculated for each version of the research intervention. The results obtained were $\alpha = 0.76$ for the animation version and $\alpha = 0.82$ for static images & text version. These are acceptable values, indicating that this scale was found to be internally consistent for this study.

4.6.1 Self-report of cognitive load

4.6.1.1 The animated version

One hundred and twenty participants (n=120) were given the animation version. The self-reported cognitive load, on the scale of 1 – 9, was measured five times across the 19 screens of content. Table 4.26 summarises the content for the animation version that was studied prior to asking the participants to indicate the amount of mental effort they felt they had to invest in order to understand the content. The actual screens are illustrated in Appendices G - K.

* Self report cognitive load (SRCL)

Screens	SRCL	Question phrased
Screen 5 of 19: Different parts of the Autonomic Nervous System (ANS). Using text as a hyperlink, the user viewed three different images that each illustrated the different parts of the ANS. They could toggle between these different views to compare the difference visually.	SRCL1	Indicate how much mental effort you used to study the information on the screen you have just reviewed.
Screen 12 of 19: Animation of 1 min 45 secs in duration covering the structure and function of the sympathetic nervous system. User could pause, stop and restart the animation and view it as often as needed before moving on.	SRCL2	Indicate how much mental effort you used to study the information in the animation you have just reviewed.
Screen 13 of 19: Two concepts were described, using text and a static image for each. User clicked on a concept and viewed the content in a pop-up.	SRCL3	Indicate how much mental effort you used to study the information on the screen you have just reviewed.
Screen 17 & 18 of 19: The first screen was an interactive screen where users could select an organ and see a visual representation of the effect of the SNS and PNS on the organ. They could toggle between the two and compare the effect. The second screen summarised these effects in a table, using text.	SRCL4	Indicate how much mental effort you used to study the information on the two screens you have just reviewed.
Screen 19 of 19: Animation of 1 min 15 secs in duration explaining the neurotransmitters that function at the synapses in the ANS. User could pause, stop and restart the animation and view it as often as needed before moving on	SRCL5	Indicate how much mental effort you used to study the information on the screen you have just reviewed.

Table 4.26: Overview of content for the animation version on which self-report of cognitive load was based

With the exception of one measurement, the rating was based on the content viewed in the screen that preceded the question. The fourth time the participants had to rate mental effort invested they had to consider two screens. These screens were viewed one after the other and participants could go backwards and forwards between these screens as often as needed before moving on. The participants could **not** go back to view the screen in order to make the decision about how much mental effort they had invested. They only answered the cognitive load question once. If they returned to the screen later, which was possible using the back button in the program, they were not asked to self-rate the mental effort invested a second time.

The rationale for this design, which only measured the self-report of cognitive load once, was to control for the effect of prior knowledge and previous learning on the self-report of cognitive load. The study was designed to measure cognitive load the first time the content was viewed in the program, rather than investigate how cognitive load changed with increasing exposure to the content. However, since the learning environment was an authentic one it was deemed ethically appropriate to allow the learner to view the content as often as they needed prior to taking the posttest, and so they were allowed to go back to content.

Cognitive load was categorised into low (1 - 3.9), medium (4 – 6.9) and high (7 – 9).

The frequencies for the distribution of these cognitive load categories, for each time the cognitive load was measured (SRCL1 to SRCL5) are presented in Table 4.27.

	Self-report Cognitive Load				
	Screen	**n**	**Low**	**Medium**	**High**
SRCL1	5 of 19	117	14	91	12
SRCL2	12 of 19	114	12	71	31
SRCL3	13 of 19	114	36	70	8
SRCL4	17/18 of 19	113	12	77	24
SRCL5	19 of 19	112	14	78	20

Table 4.27: Frequencies of cognitive load reported as low, medium and high for the animation version

Table 4.27 indicates that a higher number of participants rated screens 12, 17 & 18 and 19 as requiring high mental effort. The animation on screen 12 had the highest percentage of high cognitive load ratings (27%), followed by screens 17 & 18 (21%) and screen 19 (17%) respectively. Screen 13 had the highest percentage (31.5%) of low cognitive load ratings compared to the other screens, and the lowest percentage (7%) of high cognitive load ratings.

The means and standard deviations for each instance of cognitive load measurement in the animation version, ranked from high to low, is presented in Table 4.28.

SRCL	Screen	n	Mean (SD)	Minimum	Maximum
SRCL2	12 of 19	114	5.76 (1.59)	1	9
SRCL4	17 & 18 of 19	113	5.56 (1.45)	2	9
SRCL5	19 of 19	112	5.25 (1.63)	1	9
SRCL1	5 of 19	117	4.96 (1.31)	1	8
SRCL3	13 of 19	114	4.22 (1.62)	1	9

Table 4.28: Mean cognitive load ratings for individual measurements of cognitive load in the animated version

Using the same categories of low (1 - 3.9), medium (4 – 6.9) and high (7 – 9) cognitive load, it would appear from the data in Table 4.28 that the mean ratings are all within the range for medium cognitive load, with the mean value for SRCL2 (the animation on screen 12) approaching the higher end of the medium scale. The highest cognitive load was found for screen 12, the animation, and the lowest load for screen 13, a screen that used active text and pop-ups. The pop-ups each displayed one static image and text. The design of screen 13 is illustrated in Appendix H.

This data suggests that animation requires more cognitive resources from working memory than do static images and text, resulting in a higher self-report rating of mental effort invested. The influence of presentation formats on cognitive load will be explored in more detail in the next section.

4.6.1.2 The influence of presentation formats on cognitive load in the animation version

The fourth sub-question asked about cognitive load was ‘To what extent do the presentation formats influence cognitive load?’

A series of univariate analyses for pair-wise dependent samples was conducted. The results of the Sign test (M) indicated whether the difference in the measured cognitive load of the two screens being compared was significant. This test is appropriate for dependent samples where the data has been collected using a rating scale (Garson, 2006). The results from the two-tailed test, together with the associated *p* values, are reported. The Bonferroni correction (McMillan & Schumacher, 2001) was applied to these *p* values.

The results are displayed in Table 4.29. The screen listed in the column marked 1 (one of the presentation formats) had a higher cognitive load than the screen listed in the column marked 2 (the second presentation format).

Screen-wise comparison of cognitive load			
Column 1	**Column 2**	**Sign test statistic (M)**	**p**
Screen	**Screen**		
12	17/18	0.5	> 0.95
12	19	14.5	0.0024*
12	5	-23	0.0008*
12	13	32.5	0.0008*
17/18	19	12.5	0.0176*
17/18	5	-22	0.0008*
17/18	13	-30.5	0.0008*
19	5	-7	> 0.05
19	13	-21	0.0008*
5	13	21.5	0.0008*

* Alpha $p < 0.05$

Table 4.29: A comparison of the cognitive load for selected screen pairs in the animation version

From Table 4.29 we see that there are statistically significant differences for all the comparisons excepting the following screen pairs:

- Screen 12 and 17/18
- Screen 19 and 5

Effect sizes were also calculated using the mean values of the cognitive load for each screen.

The screen pairs that had both statistical and visible to large effects (effect size greater than 0.40) are presented in Table 4.30.

Comparison of Mean Values				Effect Size
Column 1		Column 2		
Screen	Mean (SD)	Screen	Mean (SD)	
12	5.76 (1.59)	5	4.96 (1.31)	0.50
12	5.76 (1.59)	13	4.22 (1.62)	0.95
17/18	5.56 (1.45)	5	4.96 (1.31)	0.41
17/18	5.56 (1.45)	13	4.22 (1.62)	0.82
19	5.25 (1.63)	13	4.22 (1.62)	0.63
5	4.96 (1.31)	13	4.22 (1.62)	0.56

Table 4.30: Effect sizes for comparisons between screen in the animation version

There was a statistically significant difference in the cognitive load of screen 12 and 19 ($p = 0.024$), but the effect size was only ($d = 0.31$). In looking at the design of each of the animations however it is possible that this difference could still be visible. The design issues will be discussed in Chapter 5.

4.6.1.3 The static images & text version

One hundred and eighteen participants (n=118) used the static images & text version to learn the content. The self-reported cognitive load, on the scale of 1 – 9, was measured six times across the 23 screens of content. Table 4.31 summarises the content that was studied prior to asking the participants to indicate the amount of mental effort they felt they had to invest in order to understand the content. The actual screens are illustrated in Appendices G – K.

* Self report cognitive load (SRCL)

Screens	SRCL	Question phrased
Screen 5 of 23: Different parts of the Autonomic Nervous System (ANS). See Table 4.26.	SRCL1	Indicate how much mental effort you used to study the information on the screen you have just reviewed.
Screen 13 of 23: Content about the origins of the preganglionic nerves, presented with text, a static image and a hot spot on this image. When mouse was rolled over the hot spot the pop-up displayed a close up cross-section view of the spinal cord.	SRCL3	Indicate how much mental effort you used to study the information on the screen you have just reviewed.
Screen 14, 15 & 16 of 23: Screen 14 used text and an image to explain content. There were two mouse-overs for the image. Rolling the mouse-over the image displayed a pop-up that provided more information. One of the pop-ups had a link to another screen (screen 16). The screen zoomed in to the larger image and a very short animation, using text labels, explained the path of the preganglionic fibre. Screen 15 explained the content using text and two static images.	SRCL4	Indicate how much mental effort you used to study the information on the three screens covering the synapses, which you have just reviewed.
Screen 17, 18 & 19 of 23: Screen 17 introduced the topic of innervation using text only. Screen 18 was exactly the same as screen 17 in Version 1 (see Table 4.26). The second screen summarised these effects using text links and pop-ups (See Appendix J).	SRCL5	Indicate how much mental effort you used to study the information about the innervation of the target organs.
Screen 20 of 23: Two concepts were described, using text and a static image for each. User clicked on concept and viewed content underneath the link. The user could toggle between the two links and observe the subtle changes in the image.	SRCL6	Indicate how much mental effort you used to study the information on the screen you have just reviewed.
Screen 23 of 23: Animation of 1 min 15 secs. Same as Version 1. See Table 4.26.	SRCL7	Indicate how much mental effort you used to study the information on the screen you have just reviewed.

Table 4.31: Overview of content in the static image & text version on which self-report of cognitive load was based

With the exception of two measurements the rating was based on the content viewed in the screen that preceded the question. The third time the participants rated mental effort invested they had to consider three screens: 14 - 16. These screens, together with screen 13, replaced screen 12 of the animation version. The screens were presented in a linear order, but the participants could go backwards and forwards between these screens as often as needed before moving on. Several participants missed the link to screen 16. The fourth time the participants had to rate cognitive load they also had to consider three screens.

In all instances the participant could not go back to view the screen in order to make the decision about how much mental effort they had invested. They also only answered the question once. If they returned to the screen later, which was possible using the back button in the program, they were not asked to self-rate the mental effort invested a second time. The rationale for this design has already been explained in Section 4.6.1.1.on page 207 of this chapter.

The frequencies for the distribution of the cognitive load is presented in Table 4.32.

	Self-report Cognitive Load				
	Screen	**n**	**Low**	**Medium**	**High**
SRCL1	5 of 23	117	10	85	22
SRCL3	13 of 23	116	22	83	11
SRCL4	14/15/16 of 23	117	12	78	27
SRCL5	17/18/19 of 23	113	7	74	32
SRCL6	20 of 23	115	8	81	26
SRCL7	23 of 23	117	2	80	35

Table 4.32: Frequencies of cognitive load reported as low, medium and high for the static images & text version

Table 4.32 indicates that a higher proportion of participants rated screens 23 and 17 - 19 as requiring high mental effort. The animation on screen 23 had the highest percentage of high cognitive load ratings (30%), followed by screens 17 - 18 (28%) and screens 14 - 16 (23%) and 20 (23%) respectively. Only 1.7% of the participants rated screen 23 (the animation) in the low cognitive load category. Screen 13 had the highest percentage (19%) of low cognitive load ratings when compared to the other screens, and the lowest percentage (9.5%) of high cognitive load ratings.

The means and standard deviations for each instance of cognitive load measured in the static images & text version, ranked from high to low, is presented in Table 4.33.

SRCL	Screen	n	Mean (SD)	Minimum	Maximum
SRCL7	23 of 23	117	6.14 (1.20)	3	9
SRCL5	17/18/19 of 23	113	5.73 (1.43)	1	9
SRCL4	14/15/16 of 23	117	5.50 (1.48)	2	9
SRCL6	20 of 23	115	5.53 (1.39)	1	9
SRCL1	5 of 23	117	5.39 (1.40)	2	9
SRCL3	13 of 23	116	4.80 (1.51)	1	8

Table 4.33: Mean cognitive load ratings for individual measurements of cognitive load for the static images & text version

Using the same categories of low (1 - 3.9), medium (4 – 6.9) and high (7 – 9) cognitive load it would appear from the data in Table 4.33 that the mean ratings of all are within the range for medium cognitive load, with the value for SRCL7 (the animation on screen 23) approaching the higher end of the scale. The highest cognitive load was found for screen 23 (M [±SD] = 6.14 [±1.20]), the animation, and the lowest load for screen 13 (M [±SD] = 4.8 [±01.5]), a screen that used text and a static image with a hot spot. When the hot spot was activated another image was displayed on the screen.

The influence of the presentation formats on cognitive load in the static images & text version will be explored in more detail in the next section.

4.6.1.4 The influence of presentation formats on cognitive load in the static images & text version

The fourth sub-question asked about cognitive load was 'To what extent do the presentation formats influence cognitive load?' This section considers the static images & text version.

A series of Univariate analyses for pair-wise dependent samples was conducted. The results of the Sign test (M) indicated whether the difference in the measured cognitive load of the two screens being compared was significant. This test is appropriate for dependent samples where the data has been collected using a rating scale (Garson, 2006). The results from the two-tailed test, together with the associated *p* values, are reported. The Bonferroni correction (McMillan & Schumacher, 2001) was applied to these p-values.

The results are displayed in Table 4.34. The screen listed in the column marked 1 (one of the presentation formats) had a higher cognitive load than the screen listed in the column marked 2 (the second presentation format.

Screen-wise comparison of cognitive load			
Column 1	**Column 2**	**Sign test statistic (M)**	**p**
Screen	**Screen**		
23	17/18/19	-11.5	0.067
23	20	-20.5	0.001*
23	14/15/16	-15.5	0.002*
23	5	-27.5	0.001*
23	13	-37	0.001*
17/18/19	5	-10.5	0.257
17/18/19	20	6	> 0.95
17/18/19	14/15/16	-10	0.169
17/18/19	13	-22.5	0.001*
20	5	-5	> 0.95
20	14/15/16	2	> 0.95
20	13	-22	0.001*
14/15/16	5	-3	> 0.95
14/15/16	13	-24	0.001*
5	13	16	0.003*

* Alpha $p < 0.05$

Table 4.34: A comparison of the cognitive load for selected screen pairs in the static text & images version

From Table 4.34 we see that there are statistically significant differences for a number of the screen pair comparisons.

Effect sizes were also calculated using the mean values of the cognitive load for each screen. The screen pairs that had both statistical and visible to large effects (effect size close to or greater than 0.40) are presented in Table 4.35.

Comparison of Mean Values				Effect Size
Column 1		Column 2		
Screen	**Mean (SD)**	**Screen**	**Mean (SD)**	
23	6.14 (1.20)	20	5.53 (1.39)	0.43
23	6.14 (1.20)	14-16	5.50 (1.48)	0.43
23	6.14 (1.20)	5	5.39 (1.40)	0.53
23	6.14 (1.20)	13	4.80 (1.51)	0.88
17-19	5.73 (1.43)	13	4.80 (1.51)	0.61
20	5.53 (1.39)	13	4.80 (1.51)	0.48
14-16	5.50 (1.48)	13	4.80 (1.51)	0.46
5	5.39 (1.40)	13	4.80 (1.51)	0.39

* Alpha $p < 0.05$

Table 4.35: Effect sizes for comparisons between screen in the static images & text version

The largest effect size was obtained between screen 23 and 13 ($d = 0.88$). These results once again suggest that animation places a heavier load on the resources needed for cognitive processing than do text and static images & text.

There was no significant difference between the cognitive load of screens 17 – 19 and screens 14 – 16, where the amount of content was approximately the same across the two screen sets. This suggests that amount of content might also influence load. Another method of measuring cognitive load is to use the direct cognitive load measurement technique rather than self-report ratings, as the direct method measures the load at the time it occurs rather than after the event (Brünken, Plass & Leutner, 2003).

In both sets of screens the user was required to interact with images, text and buttons in order to study the content. In both sets of screens there was some form of animation, albeit very simple and very short. Smith (2007) measured the cognitive load of screens 14, 15, 16 and 18 using the direct measurement technique. This data was collected for her study using the same experimental as I did, with the same participants, at exactly the same time. Smith (2007) however did not compare the cognitive load of individual screens within the same version of the program or report on any effect sizes. I am using this data with permission (Appendix C) as part of the exploration into how cognitive load is affected by content and design.

The mean cognitive load values for each of these screens, obtained using the direct measurement technique, is presented in Table 4.36.

Screen	No of Obs	N	Mean (± SD)	Std Error	Minimum	Maximum
14	889	118	5.4563 (± 3.2911)	0.3029	0.2500	10
15	554	118	6.2432 (± 3.5856)	0.3300	0.3333	10
16	256	51	6.6049 (± 2.5330)	0.3547	0.6666	10
18	773	118	5.5594 (± 3.2760)	0.3015	0.6000	10

Table 4.36: Cognitive load using direct measurement technique

General Linear Modeling (GLM) was conducted using repeated measures analyses in a within subjects design. This test is appropriate for continuous data, like the data for the direct measurement of cognitive load. Data from the same sample makes the repeated measures analysis the most appropriate test to use. The analysis returned a statistically significant result, $F\,(2{,}100) = 14.74$, p < 0 0001. Further analysis was done to determine which screen-wise comparisons contributed to this finding. The results are presented in Table 4.37.

Screen	with screen	F-value	P value
14	15	1.76	0.1911
14	16	29.77	< 0.0001 *
15	16	11.60	0.0013 *

* Alpha p < 0,05

Table 4.37: Results of GLM Repeated Measures Analysis for the cognitive load comparisons

The results indicate that there was a statistically significant cognitive load between screen 14 and 16 and between screen 15 and 16. In each case screen 16 had the higher load. Screen 16 had the short non-narrated animation. These results once again provide evidence that animation poses a higher burden on the resources of the working memory than other strategies such as the use of text and static images.

The effect sizes for the pair-wise comparisons of the mean cognitive load of screens 14,15,16 and 18 in the static images & text version are presented in Table 4.38.

Comparison of Mean Values				
Column 1		Column 2		
Screen	**Mean (SD)**	**Screen**	**Mean (SD)**	**Effect Size**
14	5.4563 (3.2911)	15	6.2432 (3.5856)	0.22
14	5.4563 (3.2911)	16	6.6049 (2.533)	0.35
14	5.4563 (3.2911)	18	5.5594 (3.276)	0.03
15	6.2432 (3.5856)	16	6.6049 (2.533)	0.10
15	6.2432 (3.5856)	18	5.5594 (3.276)	0.19
16	6.6049 (2.533)	18	5.5594 (3.276)	0.32

Table 4.38: Effect sizes of comparison of mean cognitive load measured with the direct method

The effect sizes obtained indicate that most of the differences in the set of screens reported on in Table 4.38 are not practically significant. Only two effect sizes are larger than 0.30, but only by a very small margin. The comparison of screen 14 and 16 returned an effect size of 0.35, and that of screen 16 and 18 returned an effect size of 0.32. In spite of the fact that the comparison between screen 15 and 16 returned a statistically significant result ($p = 0.0013$) the effect size obtained was very small ($d = 0.10$) indicating that there is no practical significance in this results. Although these are still small effect sizes it must be pointed out that in both comparisons the screen with the animation (screen 16) had the higher cognitive load.

4.6.1.5 Comparing the animation and static images & text versions

The preceding sections (Section 4.6.1.1 through to Section 4.6.1.4) looked at the cognitive load for each version separately. The study, however, was also designed to explore cognitive load using two different formats and so the analysis of cognitive load must compare the findings across the two versions.

The three hypotheses tested were:

Null Hypothesis 2a

There will be no difference between the animation and static images & text version in the cognitive load of the program as a whole.

Alternate Hypothesis 2a

In considering the program as a whole the animation version will have a higher cognitive load than the static images & text version.

Null Hypothesis 2b

There will be no difference in the cognitive load of the screen using animation and the alternative version using static images & text.

Alternate Hypothesis 2b

The screen using animation will have a higher cognitive load than the alternative version using static images & text.

Null Hypothesis 2c

At screen level there will be no difference in the cognitive load across the versions where the content and presentation format are the same.

Alternate Hypothesis 2c

At screen level, the cognitive load will be the same in each version where the content and presentation format are the same.

This comparison used the total cognitive load values for each participant when calculating the cognitive load of the program as a whole, and the method of calculating cognitive load is provided below.

Self-report of cognitive load (SRCLV) for each participant:

$$SRCLV = AVERAGE(SR_1, SR_2 \ldots SR_n)$$ [2]

Self report of cognitive load for each version:

$$SR1^{3} = AVERAGE\ (SRCLV_1, SRCLV_2, \ldots)\ \text{and}\ SR2 = AVERAGE\ (SRCLV_1, SRCLV_2, \ldots)$$

The total cognitive load for each version was calculated as described.

[2] SR_n refers to the number of screens where cognitive load was measured

[3] SR1: Cognitive load of Version 1, SR2: Cognitive load of Version 2

The Mean (Confidence Limit (CL)) and Standard deviation for the cognitive load of the two versions is presented in Table 4.39.

Version		n	Mean (±CL)	SD	Min	Max
Animation	1	120	5.1501 (± 0.1994)	1.093	1.8	8.6
Static images and text	2	118	5.5147 (±0.2337)	1.034	2	8.3

Table 4.39: Mean cognitive load per version using the subjective rating technique

At an alpha of 0.05, a t test to determine if these two means were statistically significantly different returned a t-statistic of -2.63, p = 0.0091. The results indicate that when considering the cognitive load of the program as a whole, the static images & text version had a statistically significant higher cognitive load than the animation version. The effect size was *d* = 0.33, which is a visible but small effect. Conclusions about practical significance must be interpreted cautiously.

> *The results from the analysis of the total cognitive load of each version* ***do not support*** *hypothesis 2a that proposed that as a whole the animation version would have a higher cognitive load than the static images & text version.*

This was an unexpected finding, following the results presented in the previous three sections.

The next step in the analysis was to consider a series of screen-wise comparisons across the two versions. These sets of screens isolate the instructional strategies and presentation formats: animation versus static images & text, and the use of pop-ups.

The Means, Confidence limits, Standard deviations and Standard errors for the screen-wise comparisons of cognitive load between the two versions is presented in Table 4.40.

Strategy	Vers	Screen	n	Subjective rating of cognitive load							
				Mean	**Confidence Limit**	**SD**	**Std Error**	**Otained *t***	**DF**	***p* value**	**Effect size**
Animation versus static image and text	Animation	12	114	5.763	±0.295	1.592	0.149	3.30	230	0.0011	0.40
	Static	13 - 16	118	5.131	±0.240	1.318	0.121				
Same content, same presentation format	Animation	19	112	5.250	±0.306	1.636	0.155	- 4.73	227	< 0.001	0.55
	Static	23	117	6.145	±0.220	1.205	0.111				
	Animation	5	117	4.966	±0.240	1.313	0.121	-2.40	232	0.017	0.33
	Static	5	117	5.393	±0.258	1.408	0.130				
Same content, presentation format different	Animation	13	114	4.228	±0.301	1.624	0.152	-6.56	227	< 0.001	0.81
	Static	20	115	5.539	±0.257	1.391	0.130				
	Animation	17-18	113	5.566	±0.271	1.457	0.137	-0.87	232	0.383	0.12
	Static	17-19	113	5.734	±0.268	1.440	0.135				

Table 4.40: Screen-wise comparison of cognitive load for selected presentation formats

At an alpha of 0.05, a t test to determine if these means were significantly different, returned significant differences for four of the five comparisons. The relevant t-statistics, *p* values and effect sizes are presented in Table 4.40.

The first comparison was between screen 12, the long animation in the animation version, and screens 13-16, which presented the same content using static images & text. The cognitive load for the animation version was significantly higher than the cognitive load for the static images & text version, t = 3.30, df = 230, p = 0.0011. The effect size of 0.40 indicates a visible effect.

> *The results from the analysis therefore **supports** hypothesis 2b which stated that the screen using animation will have a higher cognitive load than the alternative screens using static images & text.*

The next group of comparisons was between screens that had the same content and the same presentation format. The first comparison was between screen 19 from the animation version and screen 23 from the static images & text version. The design of these screens has been described in Sections 4.5.1.1.to 4.5.1.4.

Hypothesis 2c stated that at screen level there will be no difference in the cognitive load across the versions where the content and presentation are the same. The difference for both groups of screen comparisons was significant, as can be seen in Table 4.40, with the difference between screen 19 and screen 23 highly significant (p < 0.001). The effect size was 0.55, also indicating a visible effect. The comparison of screen 5 returned a *p* value of 0.017 and an effect size of 0.33. This small effect size suggests that the difference in cognitive load does not have much practical significance. 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. The reason for this difference will be explored from a cognitive style perspective later in this chapter.

> *The results from the analysis therefore **do not support** hypothesis 2c, which proposed that at screen level there will be no difference in the cognitive load across the versions where the content and presentation format are the same.*

Another reason for the finding might relate to the fact that subjective ratings are always exactly that: subjective. A look at the direct measurement of cognitive load might return different findings.

The mean cognitive load for screens 5, 19 and 23, obtained using the direct measurement technique, and measured by Smith (2007) is presented in Table 4.41.

Screen	Version	N	Mean	SD	Effect size
Animation – Screen 5	1	119	6.1840	3.4439	0.20
Static – Screen 5	2	117	5.4903	3.1933	
Animation- Screen 19	1	119	9.3825	0.7893	0.29
Static – Screen 23	2	117	9.1551	1.3003	

Table 4.41: Cognitive load of selected screens using the direct method of measurement

Both comparisons returned small effect sizes. The effect size of 0.20 for the comparison of the means for screen 5 can be considered not to have any practical significance, and suggests that hypothesis 2c could not be rejected. The effect size for the comparison of the means for screens 19 and 23 is slightly larger but still small. The fatigue effect might still be appropriate.

4.6.1.6 In summary

In this section the cognitive load of each version of the program was explored. A series of screen-wise comparisons of cognitive load, both within the same version of the program and across the two different versions of the program was conducted. The method of measurement was the subjective rating, using the 9-point scale of Paas. Significance testing was conducted using t tests. Effect sizes were calculated for all the comparisons. Where unexpected findings were obtained the direct measurement of the cognitive load for the particular screen or screens was considered, using the data collected by Smith (2007) during the same experiment.

When the program was considered as a whole the total cognitive load of the animation version was M (± SD) = 5.1501 (±0.1994) and that of the static images & text version M (± SD) = 5.5147 (±0.2337). This difference was statistically significant ($t = -2.63$, $p = 0.0091$) with an effect size of 0.33, which is a visible effect.

The cognitive load was divided into three categories: low, medium and high. For both versions of the program the total cognitive load of each measurement was always within the medium range, which was set as greater than 4 and smaller than or equal to 6.9. **A consistent finding was that the screens with animation had the highest cognitive load compared to screens with static images & text.**

Comparison of the cognitive load of screens within the animation version indicated that the screen with the animation had a statistically significantly higher load than the three screens with which it was compared. The effect size for two of the four screen-wise comparisons was larger than 0.50, indicating visible effects. Comparison of the cognitive load of screens within the static images & text version indicated that the screen with the animation had a statistically significantly higher load than the four screens with which it was compared. The effect size for four of the five screen-wise comparisons was larger than 0.45, indicating visible effects.

The next section considers the relationship between cognitive load and cognitive style. This relationship is examined without considering the performance of the learner in the posttest.

4.6.2 The relationship between cognitive load and cognitive style

This section looks at the relationship between cognitive load and cognitive style. Five hypotheses were tested:

Null Hypothesis 3a

There will be no difference between the Wholistic and Analytic learner for the cognitive load of the animation version.

Alternate Hypothesis 3a

The cognitive load of the animation version will be lower for the Wholistic learner than for the Analytic learner.

Null Hypothesis 3b

There will be no difference between the Wholistic and Analytic learner for the cognitive load of the static images & text version, used as an alternative for the animation version.

Alternate Hypothesis 3b

The cognitive load of the static images & text version, used as an alternative for the animation version, will be lower for the Analytic learner than for the Wholistic learner.

Null Hypothesis 3c

There will be no difference between the Verbaliser and Imager learner for the cognitive load of the animation version.

Alternate Hypothesis 3c

The cognitive load of the animation version will be lower for the learner with Imager style than for the learner with Verbaliser style.

Null Hypothesis 3d

The amount of time spent on the program will make no difference to the rating of cognitive load by the Analytic learner.

Alternate Hypothesis 3d

Analytic learners who spent inadequate time on the program will rate the cognitive load lower than other Analytic learners.

Null Hypothesis 3e

The amount of time spent on the program will make no difference to the rating of cognitive load by the Verbaliser and Imager learner.

Alternate Hypothesis 3e

Verbalisers and Imagers who spent inadequate time on the program will rate the cognitive load more highly.

The analyses conducted and reported in this section were done primarily using t tests to establish statistical significance and effect sizes for the comparison of means to establish practical significance.

4.6.2.1 Cognitive load and the WA style

The cognitive load of the program for the participants who were found to have Wholistic and Analytic styles is presented in Table 4.42.

	Wholistic style		Analytic style		Effect size
	n	**Mean (SD)**	**n**	**Mean (SD)**	***d***
Cognitive load	73	5.2356 (1.1035)	162	5.3737 (1.1035)	0.13

Table 4.42: Cognitive load for WA style dimension

From Table 4.42 we can see that the difference in these mean values is small. The effect size for the comparison of mean values is also small, $d = 0.13$, indicating that there is no practical significance between the two groups of participants for the WA dimension of style when version was left out of the analysis.

Further analysis was done to explore the relationship between load and WA style. The cognitive load obtained for each of the WA styles was compared for each version. The results of this comparison are presented in Table 4.43.

	Cognitive load for WA style dimension				
Version	**Wholistic style**		**Analytic style**		**Effect size**
	N	**Mean (SE)**	**n**	**Mean (SE)**	***d***
Animation	37	5.0611 (1.0743)	83	5.1892 (1.1063)	0.12
Static images & text	36	5.4102 (1.1197)	79	5.5650 (1.0147)	0.14

Table 4.43: Cognitive load and the WA style dimension by version

Looking at the means in Table 4.43 we see that the differences in the cognitive load between the Wholistic and Analytic style was small for both the animation and the static images & text version.

Hypothesis 3a predicts that the cognitive load of the animation version will be lower for the Wholistic learner than for the Analytic learner. In Table 4.43 we see that the Wholistic learner did report a lower cognitive load than the Analytic learner for the animation version. A t test did not return a statistically significant difference between the two groups, t = -0.58, df = 116, p = 0.5602. The effect size is very small, *d* = 0.12 and does not provide support for practical significance.

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

Hypothesis 3b predicts that the cognitive load of the static images & text version, used as an alternative for the animation version, will be lower for the Analytic learner than for the Wholistic learner. In Table 4.43 we see that the Analytic learner had a higher cognitive load than the Wholistic learner for the static images & text version. The difference between the two groups was not statistically significant, t = -0.73, df = 113, p = 0.44643. The effect size is also very small, *d* = 0.14 and does not provide support for practical significance.

> *The hypothesis that the cognitive load of the static images & text version, used as an alternative for the animation version, will be lower for the Analytic learner than for the Wholistic learner is therefore **not supported**.*

I then looked at this relationship between cognitive load and WA style from another perspective and compared the cognitive load for the Wholistic learner by version and then the cognitive load for the Analytic learner by version. The results are presented in Table 4.44.

Version	**Cognitive load for WA style dimension**				
	Animation		**Static images & text**		**Effect size**
	N	**Mean (SE)**	**n**	**Mean (SE)**	***d***
Wholistic	37	5.0611 (1.0743)	36	5.4102 (1.1197)	0.31
Analytic	83	5.1892 (1.1063)	79	5.5650 (1.0147)	0.34

Table 4.44: Cognitive load of the WA styles for each version

Looking at Table 4.44 we see that the Wholistic learner experienced a higher cognitive load in the static images & text version. A t test did not return a statistically significant difference between the two versions for the Wholistic learner, t = -1.35, df = 70, p = 0.184. The effect size, *d* = 0.31, was in the small-to-medium range.

In Table 4.44 we see that the Analytic learner also experienced a higher cognitive load in the static images & text version. A t test **did** return a statistically significant difference between the two versions for the Analytic learner, t = -2.24, df 159, p = 0.0262. The effect size, *d* = 0.34, was in the small-to-medium range.

Cognitive load was also divided into three categories: low, medium and high. The mean values of the cognitive load for each of these categories for the WA style dimension for is presented in Table 4.45.

	Cognitive load and the WA style dimension				
	Wholistic		**Analytic**		
	N	**M (SD)**	**N**	**M (SD)**	**Effect size**
Low (≤ 3.9)	8	3.0479 (0.6791)	8	3.0125 (0.7568)	0.05
Medium (>3.9 and ≤ 6.9)	60	5.3781 (0.6266)	142	5.3212 (0.7065)	0.08
High (> 6.9)	4	7.4750 (0.4113)	11	7.7667 (0.5254)	0.56

Table 4.45: Cognitive load levels and the WA style dimension

The effect sizes were calculated for the comparison between the Wholistic and Analytic learner at each level. The data in Table 4.45 indicates that the effect sizes for the low and medium cognitive load groups for this style dimension were very small, 0.05 and 0.08 respectively, and therefore do not have practical significance. The effect size for the high cognitive load group (where the cognitive load was > 6.9) was 0.56, which is a medium effect and can be regarded as a visible effect, in spite of the fact that a t test of the means in the high cognitive load group did not return a statistically significant result, t = -1.00, df = 13, p = 0.3373. The Analytic learner reported a higher cognitive load in this group. Since I could find no studies with which to compare these results, and version was not considered in this analysis the recommendation is that this effect be investigated further.

There is the possibility that time would also influence the relationship between cognitive load, cognitive style and performance. Those Analytic learners who spent inadequate time on the program (for reasons that I did not examine) might have not been able to process as elaborately as they would have liked, and would therefore rate the cognitive load as being lower.

Inadequate time was defined on page 181 of this thesis. It was different for each version of the program due to the fact that the two versions were different in length. The same category of time was used in the analyses to test hypothesis 3d.

The cognitive load ratings for the Analytic learner, for each category of time spent is displayed in Table 4.46.

	Cognitive load rating for the Analytic Learner					
	N	**Inadequate**	**N**	**Adequate**	**N**	**More than Adequate**
Animation	11	5.0909 (1.2308)	53	5.2956 (1.0581)	18	4.9361 (1.1859)
Static images & text	23	5.0725 (0.7244)	49	5.7442 (0.9968)	7	5.9286 (1.4715)

Table 4.46: Cognitive load for the Analytic learner grouped by time spent on program

In Table 4.46 we see that the Analytic learner who spent 'Inadequate' time on the program had a lower cognitive load rating than the Analytic learner who spent 'Adequate' time on the program, for both the animation and the static images & text version. An interesting finding is that the Analytic learner who spent 'More than adequate time' studying the animation version reported the lowest cognitive load for all the 'time groups' in the animation version, while for the static images & text version the cognitive load continues to increase as the Analytic learner spent more time on the program. A series of t tests for independent samples were performed to determine whether the cognitive load ratings for the different time groups were statistically significant.

The results of these t tests, and the associated effect sizes are displayed in Table 4.47.

In Table 4.47 we see that none of the comparisons of cognitive load were statistically significant for the animation version. In looking at the static images & text version we see a different picture. The difference in the cognitive load for Analytic learners who spent 'Inadequate' and 'Adequate time' was statistically significant, $t = -2.89$, $df = 70$. $p = 0.0051$, as was the difference for Analytic learners who spent 'Inadequate' and 'More than Adequate time', $t = -2.12$, $df = 28$. $p = 0.0431$. Medium-to-large effect sizes were returned, indicating practical significance.

> *The hypothesis that the Analytic learners who spent inadequate time on the program will rate the cognitive load lower than other Analytic learners was supported for the static images & text version, but not for the animation version.*

.

				Animation Version					
Time	**n**	**Cognitive load**	**Time**	**n**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	11	5.0909 (1.2308)	Adequate	53	5.2956 (1.0581)	62	-0.57	0.5721	0.17
Inadequate	11	5.0909 (1.2308)	More than Adequate	18	4.9361 (1.1859)	27	0.34	0.7392	0.13
Adequate	53	5.2956 (1.0581)	More than Adequate	18	4.9361 (1.1859)	69	1.21	0.2312	0.30
				Static images & text version					
Time	**n**	**Cognitive load**	**Time**	**n**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	23	5.0725 (0.7244)	Adequate	49	5.7442 (0.9968)	70	-2.89	0.0051*	0.67
Inadequate	23	5.0725 (0.7244)	More than Adequate	7	5.9286 (1.4715)	28	-2.12	0.0431*	0.58
Adequate	49	5.7442 (0.9968)	More than Adequate	7	5.9286 (1.4715)	54	-0.43	0.6686	0.13

* Alpha at $p < 0.05$

Table 4.47: T-test for significance for the comparisons of cognitive load per time group for the Analytic learner

4.6.2.2 Cognitive load and VI style

The cognitive load of the program for the participants who were found to have Verbaliser and Imager styles, using Ridings CSA, is presented in Table 4.48.

	Verbaliser style		**Imager style**		**Effect size**
	n	**Mean (SD)**	**n**	**Mean (SD)**	***d***
Cognitive load	94	5.3070 (0.9073)	141	5.3472 (1.1914)	0.03

Table 4.48: Cognitive load and VI style

From Table 4.48 we see that the difference in these mean values is very small. The effect size for the comparison of these mean values is also very small, $d = 0.03$, indicating that there is no practical significance between the two groups of participants for the VI dimension of style, when version is not considered.

Further analysis was done to explore the relationship between load and VI style. The cognitive load obtained for each of the VI styles was compared for each version. The results of this comparison are presented in Table 4.49.

Version	**Cognitive load for VI style dimension**				
	Verbaliser style		**Imager style**		
	n	**Mean (SD)**	**n**	**Mean (SD)**	**Effect size**
Animation	50	5.1847 (0.9663)	70	5.1248 (1.1850)	0.05
Static images & text	44	5.4458 (0.8241)	71	5.5603 (1.1663)	0.10

Table 4.49: Cognitive load and the VI style dimension by version

Table 4.49 compares the cognitive loads for each VI style group by version. The results of the analysis show that for each version, the differences in the cognitive load means for the two VI style groups being compared were small. Looking at the means in Table 4.47 we see that the difference in the cognitive load mean of the VI styles in the animation version was very small, with a very small effect size, $d = 0.05$. The difference in the cognitive load mean of the VI styles for the static images & text version was slightly larger, but the effect size was also very small, $d = 0.10$. This indicates that there is no practical significance in the differences between these two groups, for both versions.

Hypothesis 3c predicted that the cognitive load of the animation version will be lower for the learner with Imager style than for the learner with Verbaliser style. In Table 4.49 we see that the Imager learner did report a lower cognitive load than the Verbaliser learner for the animation version, but a t test did not return a statistically significant difference between the two groups, t = 0.29, df = 116, p = 0.7701.

> *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 is therefore* ***not supported.***

I then looked at this relationship between cognitive load and VI style from another perspective and compared the cognitive load for the Verbaliser learner by version and then the cognitive load for the Imager learner by version. The results are presented in Table 4.50.

Version	**Cognitive load for VI style dimension**				
	Animation		**Static images & text**		**Effect size**
	n	**Mean (SE)**	**n**	**Mean (SE)**	***d***
Verbaliser	50	5.1847 (0.9663)	44	5.4458 (0.8241)	0.27
Imager	70	5.1248 (1.1850)	71	5.5603 (1.1663)	0.37

Table 4.50: Cognitive load of the VI styles for each version

Looking at Table 4.50 we see that both the Verbalisers and Imagers experienced a higher cognitive load in the static images & text version. A t test did not return a statistically significant difference between for the Verbalisers, t = -1.49, df = 92, p = 0.1649. The effect size, *d* = 0.27, was small and suggest that this difference does not have practical significance. The t test did return a statistically significant difference for the Imagers, t = -2.18, df = 137, p = 0.0307. The effect size, *d* = 0.37, in the small to medium range.

The mean values of cognitive load for the three categories of load are presented in Table 4.51 for the VI style dimension.

	Cognitive load and the VI style dimension				
	Verbaliser		Imager		
	n	M(SD)	n	M (SD)	Effect size
Low (≤ 3.9)	4	3.0000 (0.5416)	12	3.0403 (0.7602)	0.05
Medium (>3.9 & ≤ 6.9)	87	5.3310 (0.6589)	115	5.3435 (0.7029)	0.02
High (>6.9)	3	7.6833 (0.1607)	12	7.6903 (0.5613)	0.01

Table 4.51: Cognitive load levels and the VI style dimension

The effect sizes were calculated for the comparison between the Verbaliser and Imager learner at each level. The data in Table 4.51 indicates that the effect sizes for all three groups were very small, and therefore do not have practical significance. No further analysis was done.

It was argued in Chapter 3 that the format of the program should suit both the Verbaliser and Imager. VI style should therefore not influence learning performance, unless the cognitive load is high and inadequate time was spent in the program. Hypothesis 3e predicts that Verbalisers and Imagers who spent inadequate time on the program will rate the cognitive load more highly.

The cognitive load ratings for the Verbaliser and Imager learners, for each category of time spent in the animation version is displayed in Table 4.52.

	Cognitive load rating for the VI style dimension in the Animation version					
	N	Inadequate	N	Adequate	N	More than Adequate
Verbaliser	11	4.4364 (1.2769)	29	5.5322 (0.6888)	10	5.000 (0.8273)
Imager	10	5.22 (0.9211)	42	5.1151 (1.1179)	16	5.0906 (1.52295)

Table 4.52: Cognitive load for the VI style grouped by time spent on the animation version

Verbalisers, who are thought to share some of the characteristics common to Analytic learners, have a pattern of load for the animation version that is similar to the pattern of the Analytic learner for the animation version (see Table 4.46): low where 'Inadequate' time is spent, higher for the level 'Adequate' time and lower again for the 'More than Adequate' time group. The only difference is that the lowest rating is found in the 'Inadequate' time group. The pattern that emerges for the Imager style in the animation version is quite different. The highest cognitive load was reported by the

Imagers who spent ‘Inadequate’ time on the animation version and the cognitive load decreased as more time was spent on the program.

A series of t tests for independent samples was performed to determine whether the cognitive load ratings for the different time groups were statistically significant. The results of these t tests, and the associated effect sizes are displayed in Table 4.53.

Verbalisers: Animation Version									
Time	**n**	**Cognitive load**	**Time**	**n**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	11	4.4364 (1.2769)	Adequate	29	5.5322 (0.6888)	38	-3.51	0.0012	0.86
Inadequate	11	4.4364 (1.2769)	More than Adequate	10	5.000 (0.8273)	19	-1.19	0.2501	0.44
Adequate	29	5.5322 (0.6888)	More than Adequate	10	5.000 (0.8273)	37	2.00	0.0527	0.42
Imagers: Animation Version									
Time	**n**	**Cognitive load**	**Time**	**n**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	10	5.22 (0.9211)	Adequate	42	5.1151 (1.1179)	50	0.27	0.7846	0.09
Inadequate	10	5.22 (0.9211)	More than Adequate	16	5.0906 (1.52295)	24	0.24	0.8120	0.08
Adequate	42	5.1151 (1.1179)	More than Adequate	16	5.0906 (1.52295)	56	0.07	0.9468	0.02

Table 4.53: T-test for significance for the comparisons of cognitive load per time group for the Verbaliser learner

In Table 4.53, looking at the results for the Verbalisers, we see that two of the three comparisons of cognitive load were statistically significant for the animation version, and medium to large effect sizes were returned for all three comparisons (d = 0.85, d = 0.44 and d = 0.42 respectively), even though the comparison between cognitive load for the 'Adequate' and 'More than Adequate' time group was not statistically significant (although it closely approached statistical significance). These results have both statistical and one has practical significance.

In looking at the Imagers we see a different picture. None of the comparisons were statistically significant and effect sizes were all very small.

> *Hypothesis 3e is therefore not supported in the animation version of the program, although the findings were in the projected direction for the Imager who spent inadequate time on the program.*

The cognitive load ratings for the Verbaliser and Imager learners, for each category of time spent in the static images & text version is displayed in Table 4.54.

	Cognitive load rating for the VI style dimension in the Static images & text version					
	n	**Inadequate**	**n**	**Adequate**	**n**	**More than Adequate**
Verbaliser	17	5.3235 (0.64140	24	5.5326 (0.9756)	3	5.4444 (0.3849)
Imager	28	5.1714 (1.1367)	36	5.7681 (1.0715)	7	6.0476 (1.2009)

Table 4.54: Cognitive load for the VI style grouped by time spent on the static images & text version

The lowest cognitive load was reported for the Verbalisers who spent 'Inadequate' time on the program. This pattern is similar to that of the Verbaliser who used the animation version (see Table 4.52). The pattern that emerges for the Imager style in the static images & text version is the **opposite** of the pattern found for the Imagers who used the animation version. In the static images & text version the **lowest** cognitive load was reported by those who spent 'Inadequate' time on the program and the cognitive load **increased** as more time was spent on the program (see Table 4.52).

A series of t tests for independent samples was performed to determine whether the cognitive load ratings for the different time groups were statistically significant. The results of these t tests, and the associated effect sizes are displayed in Table 4.55.

Verbalisers: Static images & text version									
Time	**n**	**Cognitive load**	**Time**	**N**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	17	5.3235 (0.6414)	Adequate	24	5.5326 (0.9756)	39	-0.77	0.4448	0.21
Inadequate	17	5.3235 (0.6414)	More than Adequate	3	5.4444 (0.3849)	18	-0.31	0.7584	0.19
Adequate	24	5.5326 (0.9756)	More than Adequate	3	5.4444 (0.3849)	25	0.15	0.8797	0.23
Imagers: Static images & text version									
Time	**n**	**Cognitive load**	**Time**	**n**	**Cognitive load**	**df**	***t* value**	***p* value**	**Effect size**
Inadequate	28	5.1714 (1.1367)	Adequate	36	5.7681 (1.0715)	62	-2.15	0.0353	0.52
Inadequate	28	5.1714 (1.1367)	More than Adequate	7	6.0476 (1.2009)	33	-1.73	0.0936	0.73
Adequate	36	5.7681 (1.0715)	More than Adequate	7	6.0476 (1.2009)	41	-0.60	0.5545	0.23

Table 4.55: T-test for significance for the comparisons of cognitive load per time group for the Imager learner

In Table 4.55, looking at the results for the Verbalisers, we see that none of the comparisons were statistically significant and effect sizes were all very small. For the Imager style the results are different. The comparison between the 'Inadequate' time and the 'Adequate' group was statistically significant, t = -2.15, df = 62, p = 0.0353 and the effect size (*d* = 0.52) was in the medium-to-large range indicating practical significance. The comparison between the 'Inadequate' time and the 'More than Adequate' group was not statistically significant, t = -1.73.15, df = 33, p = 0.0936, but the effect size (*d* = 0.73) was large indicating practical significance.

> *Hypothesis 3e is therefore **not** supported in the static images & text version of the program.*

4.6.2.3 In summary: cognitive load and cognitive style

The analysis considered each style group separately as these styles are independent of each other. Five hypotheses were tested.

A comparison of the cognitive load obtained for each of the WA styles for each version revealed that there was no statistical or practical significance in the difference in the rating of the two style groups in the animation version. It was therefore not possible to reject the null hypothesis that there will be no difference between the Wholistic and Analytic learner for the cognitive load of the animation version. In considering the static images & text version it was also not possible to reject the null hypothesis which stated that there will be no difference between the Wholistic and Analytic learner for the cognitive load of the static images & text version, used as an alternative for the animation version. The expectation was that the cognitive load would be lower for the Analytic learner. In fact the Analytic learner had a higher cognitive load than the Wholistic learner for the static images & text version, but the difference between the two groups was not statistically or practically significant.

The analysis also compared the cognitive load for the Wholistic and Analytic learner by version. The Wholistic learner experienced a higher cognitive load in the static images & text version. The analysis did not return a statistically significant difference between the two groups, but the effect size (*d* = 0.31) suggested that this findings could have practical significance. The Analytic learner also experienced a higher cognitive load in the static images & text version and there was a statistically significant difference between the two groups. The effect size was in the small to medium range which suggests that this finding could be visible.

The cognitive load was divided into three categories: low, medium and high. The mean values of the cognitive load for each of these categories was compared for the two styles in the WA style dimension. The effect sizes for the comparison of cognitive load between the Wholistic and Analytic styles were very small, 0.05 and 0.08 for the low and medium cognitive load groups respectively. The effect size for the comparison of the high cognitive load group (where the cognitive load was > 6.9) was 0.56, which is a medium effect and can be regarded to be visible. This analysis was followed up with a t test to determine statistical significance. The analysis did not return a statistically significant result.

Similar analyses were conducted for the VI style dimension. A comparison of the cognitive load obtained for each of the VI styles for each version revealed that there was no statistical or practical significance in the difference in the rating of the two style groups in the animation version. It was therefore not possible to reject that null hypothesis that there will be no difference between the Verbaliser and Imager learner for the cognitive load of the animation version.

The analysis also compared the cognitive load for the Verbaliser and Imager learner by version. Both the Verbalisers and Imagers experienced a higher cognitive load in the static images & text version. This result was not statistically or practically significant for the Verbalisers. The result was statistically significant for the Imagers and the effect size, which was in the small to medium range, suggests it might be visible.

The cognitive load was divided into three categories: low, medium and high. The mean values of the cognitive load for each of these categories was compared for the two styles in the VI style dimension. The effect sizes for the comparison of cognitive load between the Verbaliser and Imager styles were very small for all three the cognitive load groups respectively. No further analysis was conducted to determine statistical significance.

The final analysis sought to understand whether the time spent on the program made a difference to the cognitive load for the learners with different style. The analysis was considered by version of the program. Only the Analytic style was considered from the WA dimension. An interesting finding was that the Analytic learner who spent 'More than Adequate' time studying the animation version reported the lowest cognitive load for all the 'time groups' in the animation version, while for the static images & text version the cognitive load continued to increase as the Analytic learner spent more time on the program. Following significance testing it was concluded that the hypothesis that the Analytic learners who spent inadequate time on the program will rate the cognitive load lower than other Analytic learners was supported for the static images & text version, but not for the animation version.

Both the Verbaliser and Imager styles were considered when analysing the results for the VI style. In the animation version a pattern similar to the Analytic learner emerged for the Verbaliser learner and two of the three comparisons of cognitive load were statistically significant with effect sizes that indicate practical significance for all three comparisons. In looking at the Imagers we see a very interesting pattern. In the animation version the highest cognitive load was reported by the Imagers who spent 'Inadequate' time on the animation version and the cognitive load decreased as more time was spent on the program, while in the static images and text version the lowest cognitive load was reported by the Imagers who spent 'Inadequate' time on the animation version and the cognitive load increased as more time was spent on the program. The only statistically significant finding was between the cognitive load for the 'Inadequate' and 'Adequate' time group.

The next section will consider the influence of the both the cognitive style and demographic variables on the subjective rating of cognitive load.

4.6.3 Subjective rating of cognitive load and other variables

In the preceding sections the influence of time on the subjective rating of cognitive load for the different style groups was explored. This section considers the influence of several other variables on the subjective rating of cognitive load. A stepwise regression, followed by general linear modeling was used to conduct the analyses. The model for this stepwise regression analysis used cognitive load as the dependent variable.

The following variables, and all their interactions, were included in the regression analysis:

- Version of the program
- Gender
- Culture
- Had they studied the topic previously or not
- Rating of their knowledge
- English as home language
- WA style
- VI style
- Total time spent on the program

The result of the regression was $F(3, 226) = 8.33$, $p < 0.0001$, R-square = 0.0995 and C(p) = -3.5055. The R-square is in the medium effect size range and is a visible effect. The variables retained in the stepwise regression were home language and the interactions between version and total time spent on the program and between culture and the rating of knowledge about the topic.

The stepwise regression equation is presented in Table 4.56.

Variable	Parameter Estimate	Standard Error	Type 11 Sum of squares	F-value	p-value
Intercept	5.0932	0.3634	212.0975	196.46	< 0.0001
Home Language	212.0975	0.1573	3.3433	3.10	0.0798
Version x total time spent on program	0.0060	0.0022	8.0969	7.50	0.0067
Culture x rating of knowledge	-0.2047	0.0660	10.3901	9.62	0.0022

Table 4.56: Stepwise regression equation for self-report rating of cognitive load

From the data it can be determined that there was no statistically significant main effects, although the main effect for home language approached significance, *F*(3, 226) = 3.10, p = 0.0798. There were two statistically significant interactions: between version and the total time spent on the program, *F*(3, 226) = 7.30, p = 0.0067 and between culture and the rating of knowledge about the topic, *F*(3, 226) = 9.62, p = 0.0022.

A confirmatory General linear model (GLM) was then run, based on the results from the stepwise regression. The GLM presents the standard error and p-difference tables, the results of which are not provided in the stepwise regression. The GLM allows comparison of the class variable effects (home language, version, culture and rating of knowledge about the topic) using pair wise least squares means comparisons.

The model used for the GLM was: Self-report rating of cognitive load = version, culture, rating of knowledge about the topic, home language, the interaction between culture and rating of knowledge about the topic, total time spent on the program and the interaction between version and the total time spent on the program.

The GLM analysis returned a statistically significant finding, *F*(7, 224) = 3.67, p = 0.0009, R^2 = 0.1029.

The rating of knowledge about the topic accounted for the significant result, *F*(1, 224) = 3.98, p = 0.0474. Home language, *F*(1, 224) = 2.79, p = 0.0964, and the total time spent on the lesson, *F*(1, 224) = 2.82, p = 0.0943 approached significance. Further post hoc comparison, using Fischer's Least Squares Means was conducted. The effect size was *d* = 0.47, indicating this is a visible effect.

The Least Squares Means for the cognitive load ratings for the two groups of rating of knowledge about the topic are displayed in Table 4.57.

	N	Cognitive load LS Mean	Standard error	*F* statistic	*p* value	Effect size
Know nothing about topic	19	5.7828	0.2442	3.98	0.0474	0.47
Have basic & intermediate knowledge about topic	214	5.2748	0.0858			

Table 4.57: GLM analysis results for cognitive load and rating of knowledge about the topic

The Least Squares Means for the cognitive load ratings for the two different home language groups are displayed in Table 4.58. The effect size is $d = 0.27$, a small effect.

	N	Cognitive load LS Mean	Standard error	*F* statistic	*p* value	Effect Size
English is first home language	62	5.3947	0.1688	2.82	0.0943	0.27
English not first home language	171	5.6629	0.1377			

Table 4.58: GLM analysis results for cognitive load and home language

Although this result was not statistically significant the results support the personal observation referred to on page 177 of this thesis that learners whose first language is not English find Physiology more difficult.

4.6.4 Cognitive load and time spent on the program

In Section 4.5.2 an analysis was conducted on the amount of time the participants spent on the program and the question 'Is the time spent on the program related to cognitive style?' was asked. This section looks at whether the cognitive load of the program influenced the :

- total amount of time spent on the program for each version, and
- number of times a specific screen was accessed.

The cognitive load was measured after the first entry to specific screens, and then not again. The analysis considered the self-reported cognitive load for selected screens where the participant accessed the screen more than twice. The screens reviewed were screens 12 and 19 from the animation version and screens 13-16 and 23 from the static images & text version.

The mean (SD) time spent on the animation version was 42.3496 (11.5879) minutes and for the static images & text version it was 45.6333 (14.0197) minutes. The effect size for the difference between these means is $d = 0.23$, which is small, indicating no practical significance.

4.6.4.1 Cognitive load and time spent on the animation version

This section looks at the cognitive load and time spent on the animation version. In section 4.5.2.1 (page181) the time spent on the program was divided into three groups: Inadequate (IA), Adequate (AD) and More than Adequate (MA). The cognitive load for each of these time groups, by version is displayed in Table 4.59.

	Cognitive load for time spent on program by version				
	Animation Version		**Static images & text version**		
	n	**Mean (± SD)**	**n**	**Mean (± SD)**	**Effect size (d)**
Inadequate	21	4.8095 (1.1653)	46	5.2457 (0.9694)	0.45
Adequate	72	5.2854 (0.9820)	62	5.6575 (1.0197)	0.36
More than Adequate	27	5.0558 (1.2853)	10	5.8667 (1.2368)	0.66

Table 4.59: Cognitive load for time spent on program by version

4.6.4.2 Cognitive load and time spent on the static version

We see in Table 4.59, when comparing versions, that the static images & text version had a higher cognitive load for all three the levels of time. All the mean values of cognitive load displayed in the table are in the medium range. Effect sizes indicate visible effect for all the comparisons of the means. The lower load for the 'Inadequate' time group was unexpected, and raises the question as to whether or not time is an indicator of cognitive load, since there was a lower load for the level of time that was categorised as 'Inadequate' and higher load for the level of time categorised as 'More than Adequate'. The reasons for spending so little time on the program were not followed up with the participants. It is therefore not possible to make any further assumptions about this data.

4.6.4.3 Cognitive load and multiple access to selected screens

This section considers the cognitive load and the number of times the participant accessed a particular screen. The analysis considered the cognitive load of selected screens where the participants entered the screen **more than twice**. Table 4.60 displays the mean cognitive load for screen 12 in the animation version, where it was accessed more than twice.

	Cognitive Load of Screen 12 in Animation version			
No of times entered screen	**N**	**Mean (SD)**	**Minimum**	**Maximum**
3	16	5.3389 (0.7815)	4	6.8
4	10	5.3167 (1.1832)	2.4	6.6
5	2	5.3000 (0.4243)	5	5.6
6	2	5.6000 (0.8485)	5	6.2

Table 4.60: Cognitive of screen 12 where there were multiple entries to the screen

Screen 12 was accessed between 2 and 6 times. The value of n for each of the items in Table 4.60 was small when compared to the value of n of the animation (n = 120). In Table 4.60 we see that the cognitive load of screen 12 where there were 3 and 4 entries were very similar: 5.3389 and 5.3167 respectively. The effect size for the comparison of these two means was very small $d = 0.02$, indicating no practical significance. The comparison of the means where there were 3 and 6 entries returned an effect size of $d = 0.31$, suggesting small to medium effect size. No t tests were conducted to determine statistical significance due to the very small sample size. It is unlikely that the result will be statistically significant.

Table 4.61 displays the mean cognitive load for screen 19 in the animation version, where it was accessed more than twice.

	Cognitive Load of Screen 19 in Animation version			
No of times	**N**	**Mean (SD)**	**Minimum**	**Maximum**
3	21	5.5143 (1.2596)	3.25	8.6
4	5	5.096 (0.2428)	4.8	5.33
5	1	5.2000 (n.a.)	5.2	5.3
7	1	6.0000 (n.a.)	6	6

Table 4.61: Cognitive of screen 19 where there were multiple entries to the screen

Screen 19 was accessed between 2 and 7 times. The value of n for each of the categories in Table 4.61 is small when compared to the value of n of the animation (n = 120). In Table 4.61 we see that the cognitive load of screen 19 differed for each of the entry groups. There does not seem to be a discernable trend, as an average cognitive load of 5.5143 was obtained when there were 3 entries, but the average load was reduced to 5.096 where the number of entries was 4. The average cognitive load where there 3 entries was also higher than the average cognitive load where the number of entries was 5. Only one participant entered screen 19 five times. The one participant who entered screen 19 seven times reported a cognitive load of 6.000. The effect size for the comparison of 3 and 7 entries was $d = 0.39$, but an interpretation of a visible difference is only applicable to the one participant who accessed screen 19 seven times.

The analysis now turns to selected screens in the static images & text version. Table 4.62 displays the mean cognitive load for screen 23 in the static images & text version, where it was accessed more than twice. This screen was exactly the same as screen 19 in the animation version.

	Cognitive Load of Screen 23 in static images & text version			
No of times entered screen	**N**	**Mean (SD)**	**Minimum**	**Maximum**
3	16	5.3389 (0.7815)	4	6.8
4	10	5.3167 (1.1832)	2.4	6.6
5	2	5.3000 (0.4243)	5	5.6
6	2	5.6000 (0.8485)	5	

Table 4.62: Cognitive load of screen 23 where there were multiple entries to the screen

Screen 23 was accessed between 2 and 6 times. The value of n for each of the categories in Table 4.62 is small when compared to the value of n of the animation (n = 120). In Table 4.62 we see that like the cognitive load of screen 19 (Table 4.61) there seems to be no discernable trend. The average cognitive load for the group where screen 23 was entered 5 time is lower than for the group where the screen was entered 3 times. The highest average cognitive load was reported where the screen was entered 6 times, but only 2 participants entered screen 23 six times. The effect size for the comparison of 6 and 7 entries was $d = 0.35$, but an interpretation must be made very cautiously due to the small sample size.

The number of times the screen group 13-16 was analysed per screen, rather than as a group, since the cognitive load was measured after each of these screens. Table 4.63 (on the next page) displays the mean cognitive load for each group of entries for screens 13, 14,15 and 16.

	Mean Cognitive load for number of times screen accessed															
	3		**4**		**5**		**6**		**7**		**8**		**10**		**12**	
	N	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Screen 13	3	5.9250 (1.0196)	2	5.9167 (0.8250)	1	6.6666 (n.a.)										
Screen 14	24	5.9069 (1.2746)	9	5.1815 (0.7678)	1	5.5000 (n.a.)	1	4.6667 (n.a.)	2	4.1500 0.4950	1	6.6666 (n.a.)			1	6.3333 (n.a.)
Screen 15	13	5.4718 (1.4234)	8	4.7666 (0.7972)	1	6.6666 (n.a.)							1	6.3333 (n.a.)		
Screen 16	4	5.0750 (1.2324)														

Table 4.63: Cognitive load of screens13, 14, 15 and 16 where there were multiple entries to each of these screens

The data displayed in Table 4.63 is the first attempt to separate the cognitive load values of this group of screens, which up to this point has always been reported for the group of screens. Looking at Table 4.63 we see that the cognitive load of screen 13 in the medium-high range. We also see that screen 14 was accessed the most by the participants, with one participant accessing it as many as 12 times. The cognitive load, as measured using the subjective rating, varied between 4.1500 and 6.6666 for this group that accessed screen 14 more than twice. In Section 4.6.1.4 in Table 4.36 on page 217 of this chapter the cognitive load for screen 14, using the direct method of measurement was reported. The cognitive load of screen 14 was not higher than that of screen 15 or 16 and yet more participants accessed this screen (screen 14) more than twice than those who accessed screen 15 and 16 more than twice, where the cognitive load, by direct measurement, was higher than that of screen 14.

4.6.4.4 In summary: other variables influencing cognitive load

Section 4.6.3 considered other variables that could have had an influence on cognitive load, including time. Section 4.6.4 considered cognitive load and the amount of time spent on the program in more depth.

The influence of the version of the program, gender, culture, whether or not participants had studied the topic previously, the rating of their knowledge about the topic, English as home language, WA style, VI style and total time spent on the program on the subjective rating of cognitive load was determined using a stepwise regression. The regression analysis was followed up with a confirmatory GLM. The one variable that appeared to have an influence on the cognitive load was the rating of knowledge about the topic. The finding was in the expected direction in that participants who indicated they had basic or intermediate knowledge about the topic had a lower cognitive load than participants who indicated they had knew nothing about the topic. Home language and the total time spent on the lesson approached significance.

The analysis in Section 4.6.4 explored whether the cognitive load of the program influenced the total amount of time spent on the program for each version, and the number of times a specific screen was accessed. A comparison of the mean time spent on each version, using only effect sizes, indicated that the difference in time was not practically significant. Each version was then considered separately, using the three levels of time that had already been defined. The static images & text version had a higher cognitive load for all three the levels of time and effect sizes indicated practical significance ($d > 0.30$) for all the comparisons of the means. The lower load for the ‘Inadequate’ time group and the higher load for the level of time categorised as ‘More than Adequate’ was unexpected.

Screen 12 in the animation version was accessed between 2 and 6 times. The comparison of the mean cognitive load where there were 3 and 6 entries returned an effect size of $d = 0.31$. Screen 19 in the animation version was accessed between 2 and 7 times and the mean cognitive load varied between 5.096 and 6.000. The size of this sub-sample was very small, making analysis complex, if not impossible.

Screen 23 in the static images & text version was accessed between 2 and 6 times. Like the cognitive load of screen 19 (which was the same screen) there seemed to be no discernable trend. The number of times the screen group 13-16 in the static images & text version was analysed per screen, rather than as a group, since the cognitive load was measured after each of these screens. The cognitive load of screen 13 was in the medium-high range. Screen 14 was accessed the most by the participants, with one participant accessing it as many as 12 times. The cognitive load varied between 4.1500 and 6.6666 for this group that accessed screen 14 more than twice.

4.7 The correlation between self-report of cognitive load and the direct measure of cognitive load

The second sub-question asked about cognitive load in this study was 'What is the correlation between the participant's self-report of cognitive load and the direct measure of the cognitive load?'

4.7.1 Comparing correlation of the measurement techniques by version

The hypotheses tested were:

Null Hypothesis 2d

There will be no difference in the cognitive load for each version when two different methods are used to measure cognitive load.

Alternate Hypothesis 2d

The two different methods used to measure cognitive load will return results that are the same for each version.

Null Hypothesis 2e

There will be no correlation between the self-report method and direct measurement method for determining cognitive load.

Alternate Hypothesis 2e

There will be a positive correlation between the self-report method and direct measurement method for determining cognitive load.

The cognitive load was measured by Smith (2007) using the direct method. A total cognitive load for the program was calculated for each participant.

The Mean and standard deviation for the cognitive load for each of the two versions, using the direct measurement, is presented in Table 4.64.

Version	**n**	**Mean (SD)**
Animation	1198	6.6408 (3.3348)
Static images & text	1463	5.6841 (3.4885)

Table 4.64: Mean cognitive load using direct measurement (Smith, 2007)

Note that the frequencies used to obtain the means are considerably higher when compared to the subjective rating technique (see Section 4.6.1.5 of this chapter, page 218 and 219). This is due to the methodology of the direct measurement technique, where cognitive load was measured at multiple points on each screen, rather than just once.

At an alpha of 0.05, an analysis of variance, using the GLM procedure, revealed a significant difference between the two versions, $F(1, 2659) = 52.39$, $p < 0.001$, $R^2 = 0.0193$ (Smith, 2007, page 71). The results indicate that the animation version had a statistically significantly higher cognitive load than the static images & text version. The R^2, an indicator of the effect size is small and of no practical significance. This result must be interpreted with caution, because the large sample size gives a p-value that was statistically highly significant.

With this data available it was possible to compare the cognitive load obtained for each version using two different techniques. This comparison will focus on the total cognitive load values for each version as a whole. I also explored the correlation between the cognitive load values obtained using these two different techniques.

The Mean and Standard deviation (SD) for the different measurement techniques applied to each of the two versions is presented in Table 4.65.

	Direct Measurement		**Subject rating**	
Version	**N**	**Mean (SD)**	**n**	**Mean (SD)**
Animation	1198	6.640 (3.3347)	118	5.150 (1.0936)
Static images & text	1463	5.684 (3.4385)	118	5.514 (1.0349)

Table 4.65: Means for the cognitive load, using different measurement techniques

The cognitive load value obtained for the animation version, using the direct measurement method, was based on 1198 values and the cognitive load value obtained for the static images & text version was based on 1463 values. In contrast the values obtained using the subjective rating method were based on 118 values for each version (one for each participant). To test for statistical significance it was necessary to pair off the 1198/1463 with the 118/118 observations. A mean cognitive load value was calculated for each participant for the direct measurement method. This value was then paired, per participant, with the value obtained using the subjective rating method. The difference between the mean values was determined and a univariate analysis performed on the difference variable. The Sign test (M) indicated that there was a statistically significant difference in the cognitive load values obtained for the animation version ($M = 24$, $p < 0.0001$), with a visible effect size ($d = 0.45$), but not for the static images & text version ($M = -4$, $p = 0.5195$), where the effect size is also small ($d = 0.05$).

> ***Hypothesis 2d*** *that the two methods used to measure cognitive load will return results that are the same for each version was therefore rejected for the animation version but not for the static images & text version.*

4.7.2 Comparing correlation of the measurement techniques without considering version

The Mean (± SD) total cognitive load, without considering version, using the subjective rating method of measurement was 5.3324 (1.0780). The Mean (± SD) total cognitive load, without considering version, using the direct method of measurement was 6.1781 (2.4995). Pearson product moment correlation analysis between the direct method and subjective rating method cognitive load measures was carried out. A very small positive correlation between these two measures, with a small effect, was found, but it was not significant, $r = 0.076$, $p = 0.2462$.

> ***Hypothesis 2e*** *that there will be a positive correlation between the self-report method and direct measurement method for determining cognitive load was therefore not supported.*

4.7.3 In summary

This section compared the direct measurement measure and subjective rating measure of cognitive load obtained for the program as a whole and for each version of the program.

Using the direct method Smith (2007) found a significantly higher cognitive load for the animation version, in contrast to this study where there was a significantly higher cognitive load for the static images & text version.

A comparison of the two different cognitive load measures by version indicated that the difference in the cognitive load was statistically significantly different for the animation version but not for the static images & text version.

The Pearson product moment correlation analysis revealed a very small positive correlation between the two cognitive load measures that was not significant.

The analysis now turns to consider the results of the posttest and the relationship between cognitive load, cognitive style and learning outcomes.

4.8 The interaction between cognitive style, cognitive load and learning performance in an authentic multimedia learning environment

The final sub-question of this study addresses the issue of learner performance: How do participants with different cognitive styles perform when using the same content with different cognitive load? A related question, not articulated in the list of sub-questions, is whether any learning took place at all This section presents the findings related to the research question under the following headings:

- Results of the pre- and posttest, independent of cognitive style and cognitive load
- Cognitive style and learning performance
- The subjective rating of cognitive load and learning performance
- Cognitive style, cognitive load and learning performance

4.8.1 Results of the pre- and posttest, independent of cognitive style and cognitive load

4.8.1.1 Pretest results

The pretest, which tested knowledge relevant to the content of the multimedia, was the third and final method used to determine the prior knowledge of the sample. It was also the only objective method used. There were nine questions in the pretest (see Appendix D). The test assessed prior knowledge at the first two levels of Bloom's taxonomy (Clark, 2007) namely recall and comprehension. The test was marked electronically and the format of the questions included multiple choice, multiple response, drag and drop and short answer questions. A selection of the questions required the participant to view a static image prior to answering the question.

An example of such a question is illustrated in Figure 4.7.

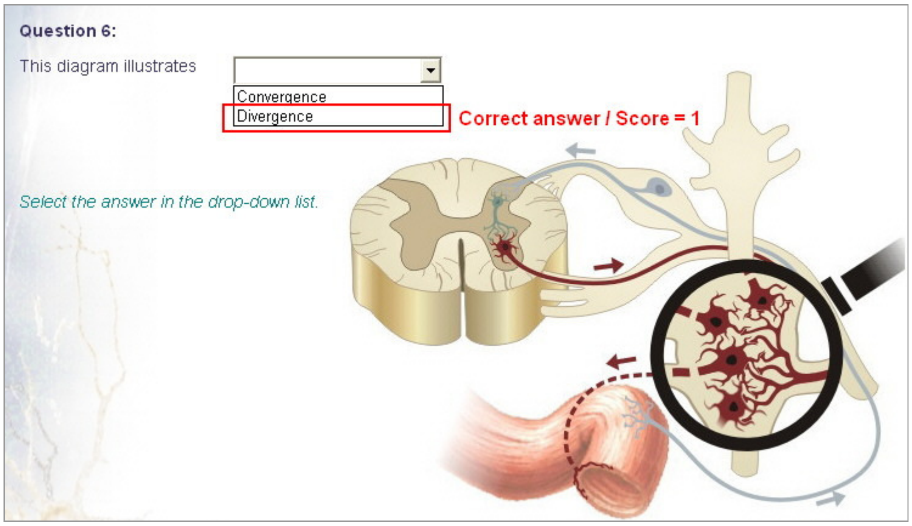

Figure 4.7: Example of question in pretest

The maximum possible score was 22 points. Two-hundred and thirty seven participants completed the pretest. The pretest scores for the sample ranged from 5 – 19, with a Mean (±SD) of 11.7257 (± 2.9308).

Figure 4.8 illustrates the distribution of the pretest results.

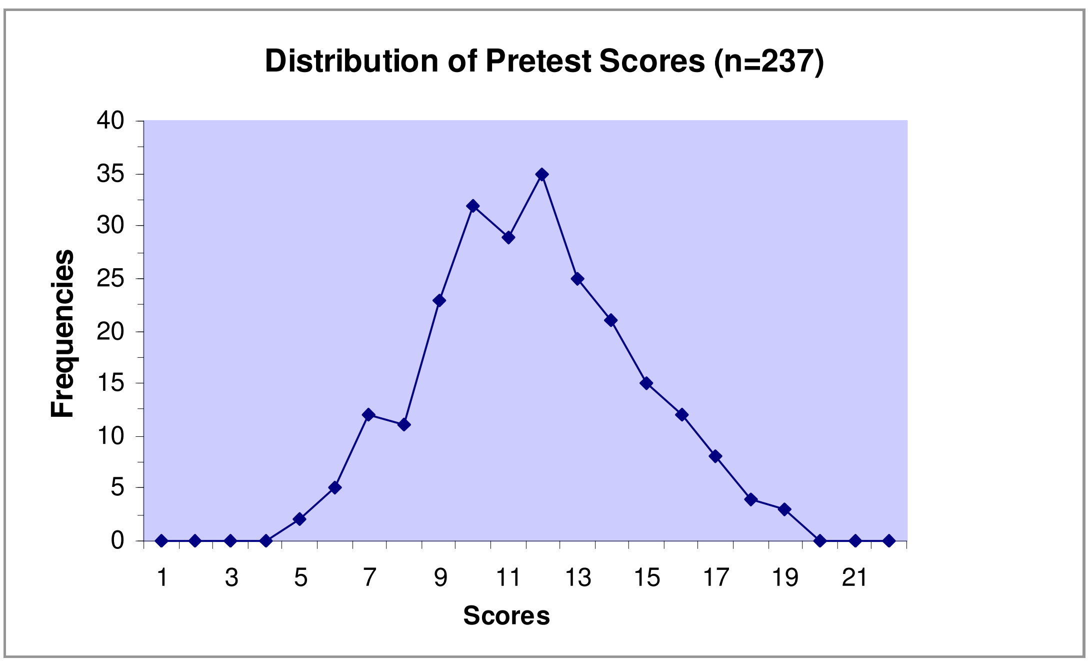

Figure 4.8: Distribution of pretest scores

Figure 4.8 illustrates a relatively symmetric distribution of the pretest scores for this sample. The questions, together with the correct answers (highlighted in bold for the multiple choice questions), the frequency of respondents who answered the question correctly and incorrectly and the percentage of the sample who answered the question correctly are displayed in Table 4.66. The value is highlighted in bold where the percentage of **correct** answers is less than 50% of the sample.

From Table 4.66 we can see that the following questions were answered correctly by the majority of the sample (more than 75%):

- **Question 3.5**: Identifying the target organ in a diagram (86%)
- **Question 8.1**: What effect does increased sympathetic stimulation have on the heart(94%)?
- **Question 8.2**: What effect does increased sympathetic stimulation have on the pupils(90%)?

	Question		Pretest results		
	Question	**Options (Correct answer in bold)**	**Correct (n)**	**Incorrect (n)**	**% Correct**
1	Which area of the brain is most directly involved in the reflex control of the autonomic system?	Cerebellum / Hypothalamus / **Medulla Oblongata** / Cerebral cortex	64	171	**27%**
2.1	Which organs are innervated mainly by the sympathetic nervous system?	Salivary glands / Stomach glands / **Sweat glands** / **Blood vessels of the skin** / Pancreas glands	115	82	**48%**
2.2			169	68	71%
3.1	Drag and drop question - label a diagram with 6 labels	Afferent sympathetic neuron	122	114	51%
3.2		Preganlionic neuron	77	159	**32%**
3.3		Dorsal root ganglion	133	102	56%
3.4		Postganlionic neuron	87	149	**37%**
3.5		Target organ	205	31	86%
3.6		Sympathetic ganglion	131	104	55%
4	Which group of receptors are stimulated when the bladder is full?	Baroreceptors / Stretch receptors / **Volume receptors**	53	184	**22%**
5	Parasympathetic ganglia are located...	in a chain parallel to the spinal cord. / in the dorsal roots of spinal nerves. / **next to or within the organs innervated**. / in the brain.	60	177	**25%**
6	This diagram illustrates	convergence / **divergence**	148	89	62%

	Question		Pretest results		
	Question	**Options (Correct answer in bold)**	**Correct (n)**	**Incorrect (n)**	**% Correct**
7.1	Which neurotransmitters are released by the neurons at the synapses in this diagram?	Ach	166	69	70%
7.2		Ach	153	83	64%
7.3		Ach	72	161	**30%**
7.4		Nor	142	90	60%
7.5		Ach	80	154	**34%**
8.1	Heart	**Increases heat rate** / decreases heart rate	223	12	94%
8.2	Pupils	Constrict / **Dilate**	214	20	90%
8.3	Bladder	**Wall relaxes - bladder fills** / Wall contracts - bladder empties	148	86	62%
8.4	Sweat glands	**None** / Increased secretion / Decreased secretion	137	97	58%
9	This diagram illustrates the	Autonomic Nervous System / Enteric Nervous System / Parasympathetic Nervous System / **Sympathetic Nervous System**	40	193	**17%**

Table 4.66: Pretest questions and answers and the number of participants who answered correctly and incorrectly

Question 3 was a diagram that needed to be labeled. Question 3.5 was the one structure on the diagram that the majority of the sample identified correctly. Their knowledge about the remaining structures was in fact poor, as can be seen in Table 4.66, considering that this content would have been covered in Anatomy as well.

Question 8.1 and 8.2 are common signs of sympathetic stimulation, part of the so-called 'fight or flight' response that is often described in the lay literature. The good performance for these two questions is therefore not surprising.

The questions most poorly answered (less than 30% of the participants answered the question correctly) were:

- **Question 1**: Which area of the brain is most directly involved in the reflex control of the autonomic system?
- **Question 4**: Which group of receptors are stimulated when the bladder is full?
- **Question 5**: Parasympathetic ganglia are located...(had to identify the location)
- **Question 7.3**: Which neurotransmitters are released by the neurons at the synapses in this diagram?
- **Question 9:** This diagram illustrates the… (had to identify the macro-structure of the sympathetic nervous system).

How did prior knowledge, determined by the subjective self-report methods compare with the pretest results? The pretest results were categorised into three performance groups: low (0 – 7), average (8 – 14) and high (15 – 22).

Table 4.67 displays the distribution of the pretest scores for these three performance groupings, further grouped according to their response to the question 'Have you studied this topic previously?'

Have studied content previously	**Pretest results (n=236)**					
	Low (0-7)		**Average (8-14)**		**High (15-22)**	
	n	**%**	**n**	**%**	**n**	**%**
Yes (n=167)	13	7.78%	125	74.85%	29	17.37%
No (n=69)	6	8.70%	51	73.91%	12	17.39%

Table 4.67: Learning performance and self-report of previous exposure to topic

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 in Table 4.67 indicate that this was not the case. The distributions for both groups (those who had not studied the topic previously and those who had) were approximately the same for each performance group. The greater majority of participants also scored in the 'Average' range of performance. These results suggest that when setting up a research design that uses prior knowledge as a variable for creating different groups, actual pretest results are a better criterion to use than self-report measures.

Participants were also asked to rate the level of their knowledge about the topic. Four options were provided: I think I know and understand:

- absolutely nothing about the topic.
- the basic concepts of the topic.
- the concepts beyond a basic level of understanding.
- the topic at an expert level.

None of the participants thought they had expert knowledge about the topic.

Table 4.68 displays the distribution of the pretest scores for the three remaining responses, further grouped according to their response to the request to rate their level of knowledge about the topic.

Rating of knowledge	**Pretest results (n=233)**					
	Low (0-7)		**Average (8-14)**		**High (15-22)**	
	n	**%**	**n**	**%**	**n**	**%**
Nothing (n=19)	3	15.79%	16	84.21%	0	0%
Basic (n=203)	14	6.89%	151	74.39%	38	18.72%
Intermediate (n=11)	0	0%	8	72.73%	3	27.27%

Table 4.68: Learning performance and self-report of level of knowledge of topic

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 (labeled as intermediate in Table 4.63) would fall in the 'High' performance group. The results in Table 4.68 support this expectation. The results also indicate that, once again, the majority of the participants in each of the 'Rating of knowledge' groups scored in the average range of performance.

A Chi-Square analysis indicated that there was no significant difference between the levels of performance in the pretest and the rating of knowledge about the topic, $\chi^2 = 7.0316$, df = 4, p = 0.1432. The phi coefficient, a measure of the strength of the relationship the levels of performance in the pretest and the rating of knowledge about the topic is 0.1737. This is a small effect indicating that in practice there is no relationship between the levels of performance in the pretest and the rating of knowledge about the topic.

Table 4.69 displays the mean scores for the pretest grouped according to their response to the request to rate their level of knowledge about the topic.

Rating of knowledge	**Pretest results (n=233)**					
	Low (0-7)		**Average (8-14)**		**High (15-22)**	
	n	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**
Nothing (n=19)	3	6.3333 (0.5773)	16	10.3750 (1.8211)	0	n.a
Basic (n=203)	14	6.5714 (0.7559)	151	11.3642 (1.7300)	38	16.1578 (1.2197)
Intermediate (n=11)	0	n.a.	8	10.0000 (1.4142)	3	16.3333 (0.5773)

Table 4.69: Pretest results and self-rating of prior knowledge and understanding of topic

The majority of the sample scored in the ‘Average’ range of performance, where the Mean score for the pretest ranged between 10.000 (45%) and 11.3642 (51.6%). Participants who rated their knowledge as ‘knowing and understanding more than the basic concepts’ only managed to obtain an average score of 45% for the pretest. This is a relatively low score, which suggests that self-report ratings of knowledge are not necessarily reliable predictors of prior knowledge.

The final analysis to be discussed with regard to the pretest results is the findings for the pretest by version of content.

The percentages of participants who scored in each of the three levels of performance in the pretest, grouped by the version they subsequently used, are displayed in Table 4.70.

Version	Pretest results (n=237)					
	Low (0-7)		Average (8-14)		High (15-22)	
	n	%	n	%	n	%
Animation	5	26.32%	89	50.57%	25	59.52%
Static images & text	14	73.68%	87	49.43%	17	40.48%

Table 4.70: Pretest by version

A Chi-Square analysis indicated that there was no significant difference between the performance of the two groups of participants in the pretest, $\chi^2 = 5.8056$, df = 2, p = 0.0549. The phi coefficient, a measure of the strength of the relationship is 0.1565. This small effect size indicates that in practice there is also no real relationship between pretest results and grouping by version. The groups could therefore be regarded as equal in terms of prior knowledge, as measured by the pretest.

4.8.1.2 Posttest results

The computer-based test

The posttest was taken as soon as the participants had finished working through the multimedia. It had two sections. A computer-based test and two open-ended questions that were paper-and-pencil based. The computer-based test, which was marked electronically, was identical to the pretest with two exceptions:

- The order in which the questions were asked was changed.
- Question 9 required the user to identify the macro structure in the image. In the pretest the image illustrated the sympathetic nervous system and in the posttest an image of the parasympathetic system was displayed. The wording of the question was exactly the same, and the correct answer was selected from a drop-down list, which was identical in both the pre- and posttest.

The two open-ended questions assessed at the application level of Blooms taxonomy.

The maximum possible score was 22 points for the computer-based test. Each question in the open-ended question counted 10 points. The analysis of these two questions will be considered separately.

Two-hundred and thirty four participants completed the electronic posttest. The total posttest scores for the participants ranged from 8 – 22, with a Mean (± SD) of 17.200 and a SD (± 2.449). Figure 4.9 illustrates the distribution of the posttest results.

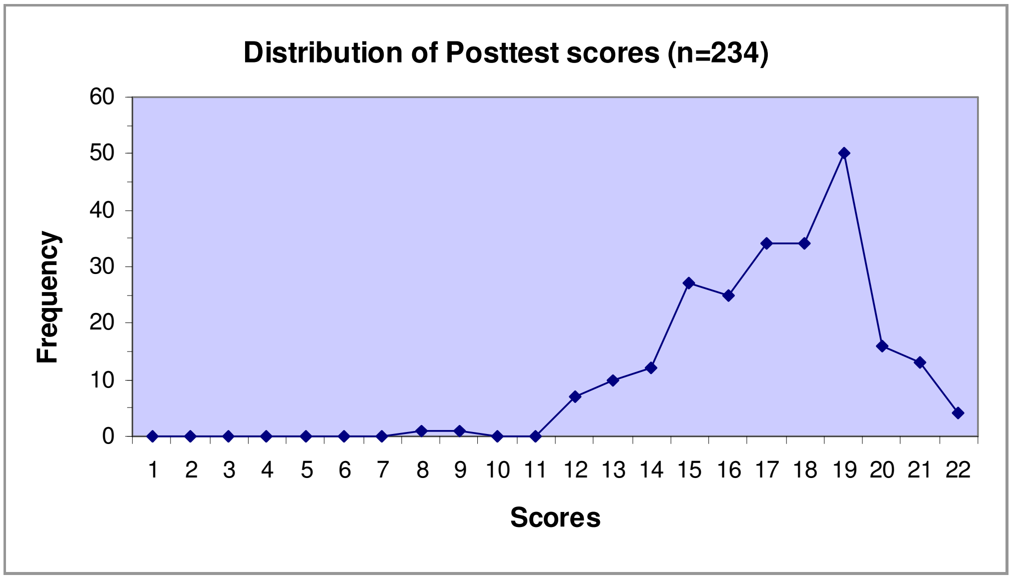

Figure 4.9: Distribution of posttest scores

The distribution of the scores for both the pre- and posttest is displayed in Figure 4.10. This display enables better comparison of the pre- and posttest scores.

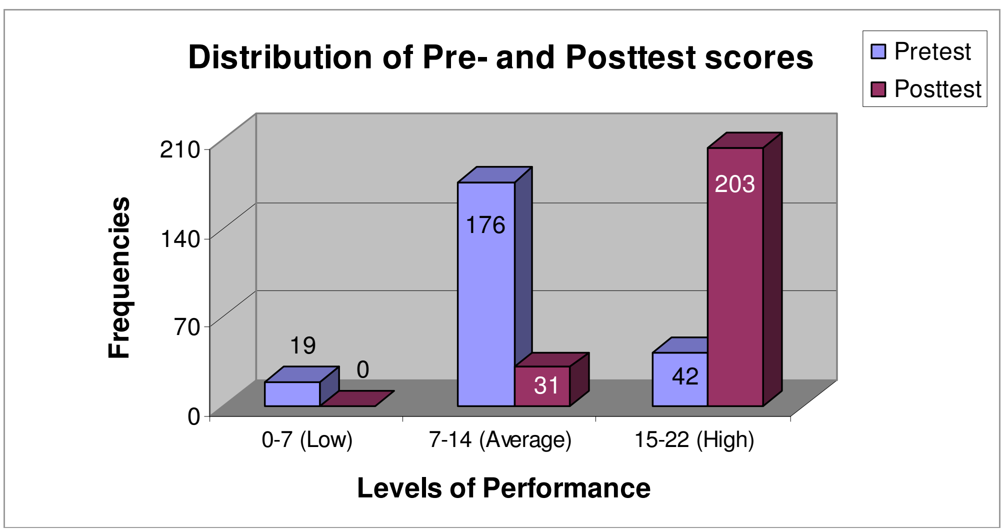

Figure 4.10: Distribution of pre- and posttest scores: Low, Average and High Performance

Figure 4.10 illustrates that in the pretest the scores of the majority of the sample were in the 'Average' level of performance, while in the posttest there was a marked shift in the distribution of the scores, with the majority of the sample scoring in the 'High' level of performance.

The posttest questions, together with the correct answers (highlighted in bold for the multiple choice questions), the frequency of respondents who answered the question correctly and incorrectly and the percentage of the sample who answered the question **incorrectly** is displayed in Table 4.71 (see page 261-262). Where this percentage is less than 50% of the sample the value is highlighted in bold.

From Table 4.71 we can see that the following questions were answered incorrectly by the majority of the sample:

- **Question 1**: Identifying the area of the brain most directly involved in the reflex control of the autonomic system (79.5%).
- **Question 5**: The location of the parasympathetic ganglia.(69.9%)
- **Question 7.3**: The release of the specific neurotransmitters at the synapses between postganglionic neuron and the target organ (61.6%).

Both question 1 and 5 were straightforward multiple choice questions testing knowledge.

- The content for question 1 was presented in screen 15 in the animation version and in screen 21 in the static images & text version.
- The content for question 5 was presented in screen 5 for both versions.
- The content for question 7.3 was presented using a narrated animation. This was screen 19 in the animation version and screen 23 in the static images & text version.

	Question		Posttest results		
	Question	**Options (correct answer in bold type)**	**Correct (n)**	**Incorrect (n)**	**% Incorrect**
1	Which area of the brain is most directly involved in the reflex control of the autonomic system?	Cerebellum / Hypothalamus / **Medulla Oblongata** / Cerebral cortex	48	186	**79.5%**
2.1	Which organs are innervated mainly by the sympathetic nervous system?	Salivary glands / Stomach glands / **Sweat glands** / **Blood vessels of the skin** / Pancreas glands	186	48	20.5%
2.2			212	22	9.4%
3.1	Drag and drop question - label a diagram with 6 labels	Afferent sympathetic neuron	203	30	12.9%
3.2		Preganlionic neuron	204	30	12.8%
3.3		Dorsal root ganglion	215	19	8.1%
3.4		Postganlionic neuron	200	34	14.5%
3.5		Target organ	225	9	3.8%
3.6		Sympathetic ganglion	220	14	6%
4	Which group of receptors are stimulated when the bladder is full?	Baroreceptors / Stretch receptors / **Volume receptors**	199	34	14.5%
5	Parasympathetic ganglia are located...	in a chain parallel to the spinal cord. / in the dorsal roots of spinal nerves. / **next to or within the organs innervated**. / in the brain.	70	163	**69.9%**
6	This diagram illustrates	convergence / **divergence**	202	32	13.7%

	Question		Posttest results		
	Question	**Options (correct answer in bold type)**	**Correct (n)**	**Incorrect (n)**	**% Incorrect**
7.1	Which neurotransmitters are released by the neurons at the synapses in this diagram?	Ach	209	25	10.7%
7.2		Ach	211	23	9.8%
7.3		Ach	89	143	**61.6%**
7.4		Nor	207	26	11.1%
7.5		Ach	138	95	40.7%
8.1	Heart	**Increases heat rate** / decreases heart rate	230	4	1.7%
8.2	Pupils	Constrict / **Dilate**	222	11	4.7%
8.3	Bladder	**Wall relaxes - bladder fills** / Wall contracts - bladder empties	216	18	7.7%
8.4	Sweat glands	**None** / Increased secretion / Decreased secretion	156	78	33.3%
9	This diagram illustrates the	Autonomic Nervous System / Enteric Nervous System / Parasympathetic Nervous System / **Sympathetic Nervous System**	168	71	30.3%

Table 4.71: Posttest questions and answers and the number of participants who answered correctly and incorrectly

The next analysis discussed is the results for the posttest by version of content.

The Mean (SD) posttest score for the participants who studied the animation version was 17.5294 (± 2.2953) and the Mean (SD) posttest score for the participants who studied the static images & text version was 16.8609 (± 2.5646). A *t* test for statistical significance indicated that this difference was statistically significant, $t = 2.10$, df = 232, p = 0.0366. The effect size for the difference between these means was 0.26, indicating an effect in the small to medium range and of no practical significance.

The percentages of participants who scored in each of the three levels of performance in the posttest, grouped by the version they studied, are displayed in Table 4.72.

	Posttest results (n=234)					
	Low (0-7)		Average (8-14)		High (15-22)	
Version	**n**	**%**	**n**	**%**	**n**	**%**
Animation	0	n.a	12	38.71%	107	52.71%
Static images & text	0	n.a	19	61.29%	96	47.29%

Table 4.72: Posttest by version

A Chi-Square analysis indicated that there was no significant difference between the groups reflected in Table 4.72 in the computer-based posttest, $\chi^2 = 2.1098$, df = 1, p = 0.1464. The phi coefficient, a measure of the strength of the relationship is -0.0949. This effect size indicates that in practice there is also no practically significant relationship between the posttest results, when using the three levels of performance, and grouping by version.

The open-ended questions

The participants answered two open-ended questions after completing the computer-based test. These questions tested at the application level of Bloom's taxonomy (Clark, 2007). The questions are presented in Appendix E. The maximum score possible for each question was 10 points. A memorandum was used to guide the marking. These questions were marked by myself.

The scores for the first question (n = 217), independent of version, ranged from 0 – 10, with a Mean (± SD) of 5.7673 (± 1.5921). The scores for the second question (n = 217), independent of version, ranged from 0 – 7, with a very low Mean (± SD) of 1.1613 (± 1.1291).

The analysis was extended to consider the performance for these two questions by version in order to determine whether the version studied had any influence on the results. The mean scores, by version, for Question 1 of the open-ended section of the posttest are displayed in Table 4.73.

	Results for Question 1 of the opened ended posttest question					
	n	**Mean (SD)**	**t**	**df**	**p**	**Effect size**
Animation	109	5.7477 (1.7208)	-0.18	215	0.8561	0.02
Static images & text	108	5.787 (1.4585)				

Table 4.73: Mean score for Question 1 of the open-ended questions for each version of the program

The mean scores for the two groups were almost identical. A t test to determine statistical significance for the difference between the mean scores for Question 1 was done, although it was not expected that there would be a significant difference between the two groups. This was confirmed, t = -0.18, df = 215, p = 0.8561. The very small effect size of d = 0.02 points to no practical significance either. Version of the program did not appear to influence performance in Question 1 in any way.

The mean scores, by version, for Question 2 of the open-ended section of the posttest are displayed in Table 4. 74.

	Results for Question 2 of the opened ended posttest question					
	n	**Mean (SD)**	**t**	**df**	**p**	**Effect size**
Animation	109	1.3119 (1.2149)	1.99	215	0.0481	0.30
Static images & text	108	1.0093 (1.0185)				

Table 4.74: Mean score for Question 2 of the open-ended questions for each version of the program

The participants who studied the animation version had the higher mean for this question. A t test to determine statistical significance between the mean scores for Question 2 was done. The analysis returned a marginally statistically significant result, t = 1.99, df = 215, p = 0.0481. The effect size of d = 0.30, suggests that this difference has no practical significance. These results must be interpreted very cautiously, due to the very low mean scores and the possibility that there were many confounding variables that influenced this outcome.

4.8.1.3 Learning gains

Figure 4.11 displays a detailed comparison of the pre- and posttest results, drilling down to look at the frequencies for each score in the range.

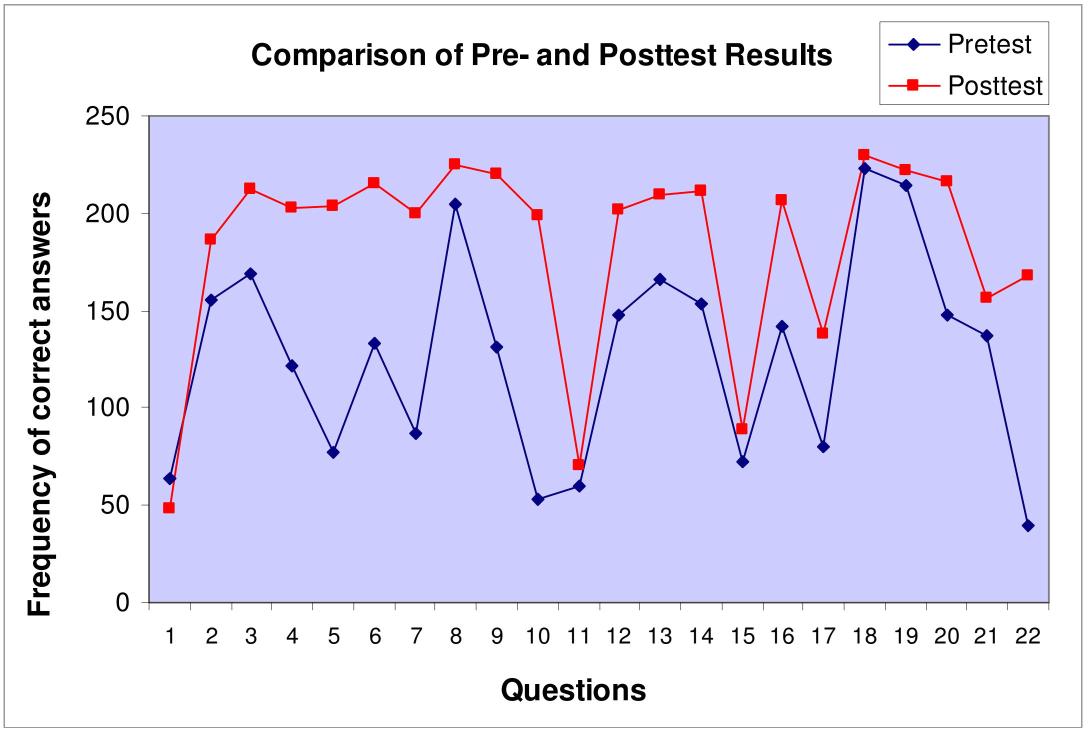

Figure 4.11: Comparison of pre- and posttest scores for each score in the range

In Figure 4.11 we can see that there was an increase in the number of correct answers in all the questions except Question 1, where in fact there were fewer correct answers in the posttest than the pretest. There were relatively small increases in the number of correct answers for Question 5, 7.3 and 8.4. These items have already been identified as problematic in Section 4.8.1.2. It would therefore appear that some learning took place. Was this learning gain significant or not?

The Mean (± SD) for the pretest and posttest, excluding the results of the two open-ended questions, was 11.7257 (2.9308) and 17.2008 (2.4491) respectively. There was a significant increase in the posttest scores. Table 4.75 displays the results of the pre- and posttest by version.

Learning Gain: Knowledge & Recall				
	Version 1: Animation		**Version 2: Static images & text**	
	n	**Mean (SD)**	**n**	**Mean (SD)**
Pretest	119	12.0840 (2.8152)	118	11.3644 (3.0118)
Posttest	119	17.5294 (2.2953)	115	16.8609 (2.5646)

Table 4.75: Mean scores for the pre- and posttests by version

A univariate analysis of the difference between the pre- and posttest results was conducted. The results of the Sign test (M) indicated that the difference was statistically significant, $M = 109$, $p < 0.001$. The learning gain in this study was significant, irrespective of version used.

4.8.2 Cognitive style and learning performance

This section discusses the learning performance of the different style groups in the sample. Since the results of the pretest indicated that there was no difference between the groups only the posttest results will be discussed.

4.8.2.1 Wholistic-Analytic style and learning performance

Using two categories for the WA style dimension it was found that 69% of the sample had an Analytic style and 31% were Wholistic in style.

General Linear Modeling analysis was conducted to determine if cognitive style had any influence on learning performance. The model used for the GLM was : Posttest = Wholistic-Analytic style, pretest and the interaction between WA style and the pretest. The GLM analysis returned a significant finding, $F(3, 227) = 23.74$, $p < 0.0001$, $R^2 = 0.2388$, but it was the pretest that accounted for this significant result, $F(1, 227) = 67.13$, $p<0.0001$. There were no main effects found between WA style and learning performance, $F(1, 227) = 1.99$, $p=0.1592$ and no interaction effects between WA style and the pretest, $F(1, 227) = 1.87$, $p = 0.1732$.

The least squares means for the posttest scores and standard errors for the two styles, the *F* statistic and *p* value are displayed in Table 4.76

	N	LS Mean	Standard error	*F* statistic	*p* value
Wholistic style	73	17.1094	0.2546	1.99	0.1592
Analytic style	162	17.2305	0.1712		

Table 4.76: GLM analysis results for WA style and learning performance

The Wholistic-Analytic style therefore did not appear to have any influence on leaning performance in the posttest.

4.8.2.2 Verbaliser-Imager style and learning performance

Using two categories for the VI style dimension it was found that 40% of the sample had a Verbaliser style and 60 % were Imagers in style.

General Linear Modeling analysis was conducted, with the posttest as the dependent variable and the VI style dimension as the independent class variable. The pretest results were entered as a covariate. The GLM analysis returned a significant finding, $F(3, 227) = 23.58$, $p < 0.0001$, $R^2 = 0.2376$, but it was once again the pretest that accounted for this significant result, $F(1, 227) = 69.69$, $p<0.0001$.

The least squares means for the posttest scores and standard errors for the two styles, the *F* statistic and *p* value are displayed in Table 4.77.

	N	LS Mean	Standard error	*F* statistic	*p* value
Verbaliser style	94	17.3136	0.2271	23.58	0.6012
Imager style	141	17.1603	0.1849		

Table 4.77: GLM analysis results for VI style and learning performance

The Verbaliser-Imager style therefore did not appear to have any influence on leaning performance in the posttest.

In both Tables 4.76 and 4.77, which display the least squares mean scores for the posttest for the WA and VI style groups respectively, it is evident that the style groups performed almost equally well.

For both style groups the results indicate that cognitive style did not have an impact on the learning performance for this group of participants.

4.8.3 The subjective rating of cognitive load and learning performance

This section presents the results of the analyses conducted to determine whether cognitive load influenced learning performance in any way.

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 GLM analysis returned a significant finding, *F*(3, 229) = 24.05, p < 0.0001, R^2 = 0.2395, but it was once again the pretest that accounted for this significant result, *F*(1, 228) = 60.57, p = p<0.0001. There were no main effects found between version and learning performance, *F*(1, 229) = 1.58, p=0.2098 or between cognitive load and learning performance, *F*(1.229) = 1.29, p = 0.2576. Further post hoc comparison, using Fischer's Least Squares Means, demonstrated that the difference in the posttest means for the two versions were not significant when cognitive load was included in the analysis.

The least squares means for the posttest scores, standard errors, *F* statistic and *p* value are displayed in Table 4.78.

	N	LS Mean	Standard error	*F* statistic	*p* value
Animation version	120	17.3714	0.1999	1.58	0.2098
Static images & text version	118	17.0101	0.2025		

Table 4.78: GLM analysis results for version and performance

These results indicated that cognitive load, as measured using the subjective rating technique did not appear to influence the learning performance in any way.

4.8.4 Cognitive style, cognitive load and learning performance

The final sub-question of this study asks 'How is learning performance influenced when content with different cognitive load is studied by learners with different cognitive styles?'

In Chapter 3 it was argued that if cognitive load is high, then cognitive style might influence performance, although this could be moderated by the amount of time spent on the program. The following was predicted:

Null Hypothesis 4a

There will be no difference in the posttest results of the Analytic learner who spent inadequate time on the program and who rated the cognitive load as low and the other Analytic learners.

Alternate Hypothesis 4a

The Analytic learner who spends inadequate time of the program, and who rated the cognitive load as low, will perform more poorly on the posttest.

It was also predicted that the interaction of time and cognitive load will influence the performance of the learners with VI style.

Null Hypothesis 4b

There will be no difference in the posttest results of the Verbaliser and Imager learner who spent inadequate time on the program and who rated the cognitive load as high and the other Verbaliser and Imager learners.

Alternate Hypothesis 4b

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.

These hypotheses are of particular interest since there were proportionally a high number of participants who spent inadequate time on the program (discussed in Section 4.4.1, page 180).

4.8.4.1 Cognitive load, the Analytic style and learning performance

The first analysis considers the Analytic Style. The Mean and standard deviation for the posttest for this style, grouped by time spent on the program (Inadequate, Adequate or More than Adequate) and rating of cognitive load (Low, Medium or High) are displayed in Table 4.79 (next page).

The values of n for each group is very small, since the analysis only used a sub-set of the data. Significance testing is not reliable when the sample is so small. The results displayed here can only point to a trend and no interpretation or conclusions can be made. Hypothesis 4a predicted that the Analytic learner who spends inadequate time on the program, and who rated the cognitive load as low, will perform more poorly on the posttest.

I will consider the animation version first. In Table 4.79 we see that the participant (n = 1) in this group (inadequate time, low cognitive load) in fact had the highest score (posttest score of 21) for the animation version as a whole. The participant (n = 1) who scored the lowest (posttest score of 17) in the animation version spent more than adequate time on the program but rated the program in the high cognitive load range.

	Posttest results for the Analytic Style											
	Animation Version						Static images & text version					
	n	**Inadequate time**	**n**	**Adequate time**	**n**	**More than Adequate time**	**n**	**Inadequate time**	**n**	**Adequate time**	**n**	**More than Adequate time**
Low cognitive load (≤ 3.9)	1	21 (n.a)	2	17.5 (2.1213)	3	19 (2.000)	1	18	1	20	0	n.a
Medium cognitive load (>3.9 and ≤ 6.9)	10	18.4 (1.7127)	48	17.083 (2.4996)	14	17.929 (1.94)	22	16.773 (2.3285)	42	16.952 (2.3783)	3	17.333 (1.5275)
High cognitive load (> 6.9)	0	n.a	3	17.667 (02.0817)	1	17 (n.a)	0	n.a	5	15 (3.000)	2	16.6 (2.1213)

Table 4.79: Posttest results for the Analytic learner, by time and load

Turning to the static images & text version the Analytic learner (n = 1) who spent inadequate time on the program and who rated the cognitive load as low obtained 18 for the posttest. This was not the lowest score, but neither was it the highest. The highest score (posttest score of 20) was for the participant (n = 1) who spent adequate time on the lesson and whose rating for cognitive low was in the low range. The lowest average score was for the participants (n = 5) who spent adequate time on the program, but whose cognitive load rating was in the high range (> 6.9).

There is a trend that suggests that the higher the load the poorer the learning performance.

4.8.4.2 Cognitive load, the Verbaliser-Imager style and learning performance

The Means and standard deviations for the posttest for this style dimension, grouped by time spent on the program (Inadequate, Adequate or More than Adequate) and rating of cognitive load (Low, Medium or High) are displayed for the animation version in Table 4.80 on page 273 and for the static images & text version in Table 4.81 on page 275.

The values of n for each group in the respective versions are very small, since the analyses only used a sub-set of the data. Significance testing is not reliable when the sample is so small. The results displayed here can only point to a trend and no interpretation or conclusions can be made. Hypothesis 4b predicted that 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.

I will consider the animation version first. In Table 4.80 we see that for both the Verbaliser and Imager style there were no participants this group (inadequate time, high cognitive load). In the Verbaliser group the participant (n = 1) with the highest score (posttest score of 21) spent more than adequate time on the program and rated the cognitive load in the low range. This was the highest score for the VI style in the animation version. The participants (n = 28) with the lowest average score (Mean posttest score of 16.857) spent adequate time on the program and rated the cognitive load in the medium range. In the Imager group the participant (n = 1) with the highest score (posttest score of 20) spent inadequate time on the program and rated the cognitive load in the low range. The participants (n = 2) with the lowest average score (Mean posttest score of 15.5) spent more than adequate time on the program and rated the cognitive load in the high range. This was the lowest mean score for the VI style in the animation version.

Once again we see the trend that suggests that the higher the load the poorer the learning performance.

	Posttest results Animation version											
	Verbalisers						Imagers					
	n	**Inadequate time**	**n**	**Adequate time**	**n**	**More than Adequate time**	**n**	**Inadequate time**	**n**	**Adequate time**	**n**	**More than Adequate time**
Low cognitive load (≤ 3.9)	3	19.333 (1.5275)	0	n.a	1	21 (n.a)	1	20 (n.a)	3	18 (1.7321)	3	17 (2.000)
Medium cognitive load (>3.9 and ≤ 6.9)	8	18.125 (2.7999)	28	16.857 (2.8767)	9	18 (1.7321)	9	18.111 (1.2693)	37	17.27 (2.194)	11	18 (1.8974)
High cognitive load (> 6.9)	0	n.a	0	n.a	0	n.a	0	n.a	2	18 (2.8284)	2	15.5 (2.1213)

Table 4.80: Posttest results for the animation version learner for the VI style by time and load

Turning to the static images & text version we see in Table 4.81 we see that for both the Verbaliser and Imager style there were no participants this group (inadequate time, high cognitive load). In the Verbaliser group the participants (n = 2) with the highest mean score (posttest score of 18) spent more than adequate time on the program and rated the cognitive load in the medium range. The participants (n = 22) with the lowest scores (Mean posttest score of 16.455) spent adequate time on the program and rated the cognitive load in the medium range. In the Imager group the participants (n = 3) with the highest mean scores (posttest score of 19) spent inadequate time on the program and rated the cognitive load in the low range. This was the highest score for the VI style in the static images & text version. The participant (n = 1) with the lowest scores (posttest score of 16) spent inadequate time on the program and rated the cognitive load in the high range. This was the lowest mean score for the VI style in the static images & text version.

Once again we see a trend that suggests that the higher the load the poorer the learning performance.

This trend that suggests that the higher the load the poorer the learning performance was confirmed in subsequent analysis. A stepwise regression was performed for the dependent variable posttest outcome. The following variables and their interactions, were included in the regression analysis:

- Version of the multimedia
- Cognitive style (both dimensions)
- Subjective rating of cognitive load
- Total time spent on the lesson

The result of the regression was $F(2, 227) = 4.67$, $p = 0.0103$, R-square = 0.0395, a small effect, and C(p) = 4.0197. The variables retained in the stepwise regression were version and subjective rating of cognitive load.

	Posttest results Static images & text version											
	Verbalisers						Imagers					
	n	Inadequate	n	Adequate	n	More than Adequate	n	Inadequate	n	Adequate	n	More than Adequate
Low (≤ 3.9)	0	n.a	0	n.a	0	n.a	3	19 (1.000)	2	18.5 (2.1213)	0	n.a
Medium (>3.9 and ≤ 6.9)	17	16.765 (2.6582)	22	16.455 (2.5209)	2	18 (2.8284)	24	16.167 (2.9439)	28	17.5 (2.1688)	4	17.5 (3.4157)
High (> 6.9)	0	n.a	0	n.a	0	n.a	1	16 (n.a)	5	16.4 (2.881)	2	16.5 (2.1213)

Table 4.81: Posttest results for the static images & text version learner for the VI style by time and load

The stepwise regression equation is presented in Table 4.82

Variable	Parameter Estimate	Standard Error	Type 11 Sum of squares	F-value	p-value
Intercept	19.7723	0.8709	3021.7804	515.44	< 0.0001
Version	-0.5420	0.3244	16.3605	2.79	0.0962
Cognitive load	-0.3327	0.1494	29.0866	4.96	0.0269

Table 4.82 Stepwise regression equation for posttest results

From the data it can be determined that there was a statistically significant main effect for cognitive load *F*(2, 227) = 4.96, p = 0.0269, and a marginally significant main effect for version, *F*(2, 227) = 2.79, p = 0.0962.

A confirmatory General linear model (GLM) was then run, based on the results from the stepwise regression. The model used for the GLM was: Posttest outcome = version, cognitive load and the interaction between version and cognitive load.

The GLM analysis returned a statistically significant finding, *F*(3, 229) = 3.06, p = 0.0290, R^2 = 0.0386. The subjective rating of cognitive load accounted for the significant result, *F*(1, 229) = 0.03, p = 0.0278. Further post hoc comparison, using Fischer's Least Squares Means, indicated that the posttest score was higher in the animation version, M (± SE) = 17.4599 (±0.2260) compared to the static images & text version, M (± SE) = 16.9292 (±0.2295). The effect size was small, *d* = 0.26, indicating that this might not be a visible effect.

> *Hypothesis 4a that the Analytic learner who spends inadequate time on the program, and who rated the cognitive load as low, will perform more poorly on the posttest is not supported.*

> *Hypothesis 4b that 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 is not supported.*

4.9 Summary of Chapter 4

Chapter presented the results of the data analysis, using the research question and hypotheses as a framework. Both descriptive and inferential statistics were used to conduct the analysis. A detailed demographic profile of the participants was presented at the beginning of the chapter. There were more female participants in the study. The group was made up of predominantly white participants. The age range was between 17 and 37, with the majority of the participants in the 18 – 21 year old range.

The analysis sought to find answers for the following themes.

- The time spent on the program as a whole and on individual screens.
- The cognitive style profile of the participants and the relationship between style and time spent on the program.
- The extent to which cognitive style influences the use of the program. In this section the number of times selected screens were accessed was analysed in depth.
- The cognitive load of each research intervention as a whole.
- The cognitive load of selected screens were determined and compared in order to determine how the presentation format influenced the cognitive load.
- Cognitive load and the relationship with the amount of time spent on the program.
- The relationship between cognitive load and cognitive style.
- The cognitive load obtained using two different measurement techniques was investigated and the correlation between the two methods explored.
- The learning performance of the participants in both a pretest and a posttest and the learning gain.
- The influence of both cognitive load and cognitive style on learning performance.
- The relationship between cognitive load, cognitive style and learning performance.

The results will now be summarised, interpreted and discussed in the Chapter 5, the next and final chapter.

----- o O o -----