

# Novel Multimedia Instruction for Teaching Anatomy Online

**A thesis submitted in fulfillment of the requirements for the degree of  
Master of Philosophy**

**By**

**Natasha Sutevski**

**School of Medicine**

**Discipline of Anatomy and Histology**

**The University of Sydney**

**2017**

## **Statement of Originality**

*This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.*

*I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.*

A handwritten signature in black ink, appearing to read 'Sutevski', with a large, stylized initial 'S'.

**Natasha Sutevski**

### **Dedication and Acknowledgements**

It's funny how winding routes and explorations can lead us to unexpected places. I feel very fortunate to have had the opportunity to work under the guidance of Associate Professor Kevin Keay and Deborah Bryce on this thesis. During this adventure I encountered a number of bumps and hiccups along the way but was always encouraged and supported. Reflecting on the mentorship I have received astonishes me and I am grateful for the warmth, kindness, humour, patience, charm and grace...not to mention time!

I am grateful to my colleagues Zayn and Clive, who graciously allowed me to share a space with them in the media lab and were always happy to help me with my queries, or take some time out to chat about movies, music and more. Similarly, connecting with fellow students and researchers from labs near and far has been a delightful exercise in nerdy-bonding.

I am also thankful for the assistance from a range of staff and teachers in the Discipline of Anatomy and Histology who helped me with administrative support or words of encouragement along the way.

Thank you to my friends and flatmates who never failed to check in and make me chuckle and to my adventure co-pilot (read: partner) Kalila, who is literally superwoman. Her support along this final leg of the thesis is ineffable.

Finally, thank you to my parents, and family who are always cheering me on no matter what...and my nephews Jake and Liam who never cease to tickle my ~~humerus~~ funny-bone.

## Table of Contents

<b>Statement of Originality .....</b>	<b>ii</b>
<b>Dedication and Acknowledgements.....</b>	<b>iii</b>
<b>List of Figures.....</b>	<b>vi</b>
<b>List of Tables .....</b>	<b>viii</b>
<b>Introduction.....</b>	<b>1</b>
Background of the problem .....	1
Problem Statement .....	7
Purpose and significance of study .....	12
Research Questions and Hypotheses.....	13
Assumptions and limitations of the study .....	15
<i>Philosophical assumptions.</i> .....	15
<i>Limitations.</i> .....	16
<i>Delimitations.</i> .....	16
Operational definitions used in the study.....	18
<b>Theoretical orientation for the study .....</b>	<b>19</b>
Dual Coding theory .....	19
Memory.....	21
Cognitive load theory .....	23
Cognitive Theory of Multimedia Learning.....	27
<b>Experiment 1 .....</b>	<b>34</b>
Methods.....	34
<i>Participants.</i> .....	34
<i>Sampling procedure.</i> .....	36
<i>Sample size, power and precision.</i> .....	36
<i>Measures and covariates.</i> .....	37
<i>Experimental design.</i> .....	39
<i>Materials.</i> .....	40
<i>Apparatus.</i> .....	40
<i>Experimental manipulation.</i> .....	42
<i>Procedure.</i> .....	43
<i>Analysis.</i> .....	44
Results .....	45
<i>Participants.</i> .....	45
<i>Pretest comparisons.</i> .....	50
<i>Posttest comparisons.</i> .....	55
<i>Improvement profiles.</i> .....	63
<b>Experiment 2.....</b>	<b>70</b>
Methods.....	70
<i>Participants.</i> .....	70
<i>Sampling procedure.</i> .....	72
<i>Sample size, power and precision.</i> .....	72
<i>Measures and covariates.</i> .....	72
<i>Experimental design.</i> .....	73
<i>Materials.</i> .....	73

<i>Apparatus</i> .....	73
<i>Experimental manipulation</i> .....	73
<i>Procedure</i> .....	73
<i>Analysis</i> .....	74
Results .....	74
<i>Participants</i> .....	74
<i>Pretest comparisons</i> .....	79
<i>Posttest comparisons</i> .....	84
<i>Improvement profiles</i> .....	92
<b>Discussion</b> .....	<b>100</b>
Conclusions .....	100
Theoretical Implications .....	101
Practical implications.....	106
Future directions .....	107
References .....	<b>110</b>
<b>Appendix A. Pretest Questionnaire and Quiz</b> .....	<b>123</b>
<b>Appendix B. Post test Questionnaire and Quiz</b> .....	<b>128</b>
<b>Appendix C. Instructions for Experiment 1</b> .....	<b>132</b>
<b>Appendix D. Instructions for Experiment 2</b> .....	<b>133</b>
<b>Appendix E. Control Video Stills</b> .....	<b>134</b>
<b>Appendix F. Experiment Video Stills</b> .....	<b>135</b>
<b>Appendix G. Mnemonic exploration</b> .....	<b>138</b>

### List of Figures

<i>Figure 1.</i> Google search trends for the term Massive Open Online Course.....	2
<i>Figure 2.</i> MOOCs and Open Education Timeline .....	4
Figure 3. The cognitive theory of multimedia learning.....	29
Figure 4. The computer lab where the experiment was conducted.. .....	41
Figure 5. Area of study in which students completed their previous degree.....	46
Figure 6. Students' self-rated level of exposure to learning anatomy prior to entry in the SMP by group. ....	49
Figure 7. Pretest self-ratings of both groups .....	51
Figure 8. Pretest Quiz Scores for both groups.....	54
<i>Figure 9.</i> Frequency of responses across scales for self-rated difficulty when learning the muscles of the back (pretest) and self-rated ease when learning the muscles of the back (posttest). ....	57
<i>Figure 10.</i> Gain scores for the control and experiment groups.....	59
<i>Figure 11.</i> Pretest and posttest quiz scores for both groups. ....	61
<i>Figure 12.</i> Distribution of pretest self-rated questions for the control group. ....	65
<i>Figure 13.</i> Distribution of posttest self-rated questions for the control group. ....	66
Figure 14. Distribution of pretest self-ratings for the experiment group.....	68
<i>Figure 15.</i> Distribution of posttest ratings for the experiment group. ....	69
<i>Figure 16.</i> Area of study in which students completed their previous degree .....	76
<i>Figure 17.</i> Students' self-rated level of exposure to learning anatomy prior to entry in the SMP by group. ....	78
<i>Figure 18.</i> Pretest self-ratings of both groups .....	79
<i>Figure 19.</i> Pretest quiz scores for both groups.....	82

<i>Figure 20.</i> Frequency of responses across scales for self-rated difficulty when learning the muscles of the back (pretest) and self-rated ease when learning the muscles of the back (posttest). .....	86
<i>Figure 21.</i> Gain scores for the control and experiment groups.....	88
<i>Figure 22.</i> Frequency of pretest and posttest scores between groups.....	90
<i>Figure 23.</i> Distribution of pretest self-ratings for the control group.....	95
<i>Figure 24.</i> Distribution of posttest self-ratings for the control group. ....	96
<i>Figure 25.</i> Distribution of pretest self-ratings for the experiment group.....	98
<i>Figure 26.</i> Distribution of posttest self-ratings for the experiment group.....	99

**List of Tables**

Table 1. <i>Principles of the cognitive theory of multimedia learning</i> .....	32
Table 2. <i>Age and sex of participants in each group</i> .....	35
Table 3. <i>A comparison of median pretest self-ratings between groups</i> .....	52
Table 4. <i>Shifts in self-rating scores for the control and experiment groups</i> .....	56
Table 5. <i>Students with a significant RCI compared to those without between groups</i> .....	64
Table 6. <i>Age and sex of participants in each group</i> .....	71
Table 7. <i>A comparison of median pretest self-ratings between groups</i> .....	81
Table 8. <i>Shifts in self-rating scores for the control and experiment groups</i> .....	85
Table 9. <i>Students with a significant RCI compared to those without between groups</i> .....	93



And you who claim to demonstrate by words the shapes of man from every aspect of his membral attitudes, dismiss such an idea, because the more minutely you describe, the more you will confuse the mind of the reader and the more you will lead him away from a knowledge of the thing described. Therefore it is necessary both to illustrate and to describe.

- Leonardo da Vinci c.1510 (From Leonardo on the human body)

### **Background of the problem**

The statement by da Vinci about the importance of words and pictures for learning anatomy captures the problem and primary solution addressed by this thesis. Shuttling forward 506 years, the landscape of higher education is dominated by swift technological advances and unprecedented connectivity. As a result, new platforms and models for education have appeared e.g., Massive Open Online Courses (MOOCs), with some heralding paradigm shifting disruptions and possibly the end of traditional universities as we know them (Zemsky, 2014; Yuan and Powell, 2013). However, after the swell of enthusiasm in 2012, dubbed the year of the MOOC, there was disillusionment as results from initial courses came in: completion rates of less than 4% of those who registered, and rather than expanding education to those without easy access, a large proportion of users were people who had completed a degree in higher education (Zemsky, 2014; Rohs and Ganz, 2015). Although the initial hype was followed by backlash and doubters, Zemsky (2014) highlights New York Times writer Tom Friedman's observation that MOOCs are here to stay and likely to evolve into forms not recognisable to the figures who initially catapulted their popularity. Indeed, Figure 1 from google trends shows the search interest for MOOCs rising and falling in regular patterns since their initial rise.

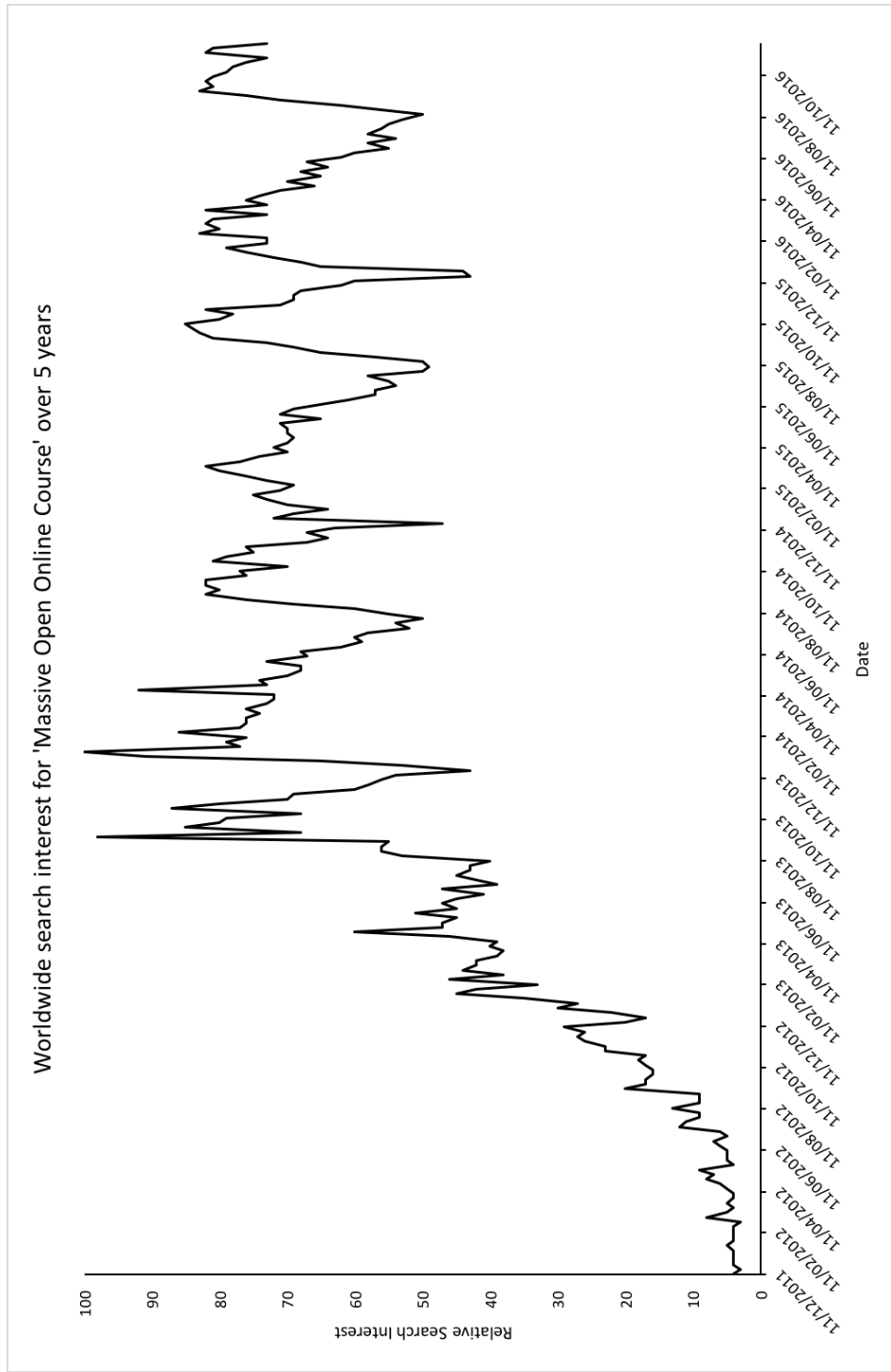


Figure 1. Google search trends for the term Massive Open Online Course

Connectivism is a recent pedagogical model that considers how technology and new sciences (e.g., network theory, chaos theory, complexity theory) redefine learning and knowledge (Siemens, 2004). The model assumes that knowledge is distributed and learning is the process of navigating, growing and pruning connections (Siemens, 2013). Based on these assumptions and several key principles, the first connectivist MOOC (cMOOC) was created. Connectivism and Connective Knowledge (CCK08) was offered both as an open course and in the Certificate of Emerging Technology for learning at the University of Manitoba in 2008. Unlike the mainstream MOOCs (xMOOCs) the course was not limited by strictly defined materials that students must progress through linearly. Rather, students could engage across a range of platforms and extend on, and contribute content (which were aggregated using RSS feeds). This highlights a departure from constructivist, behaviourist, and cognitivist assumptions of learning toward an emphasis on learning as a generative activity coupled with rapidly changing foundations. Figure 2 shows key MOOC and open education developments across 2000-2013. The kernel of mainstream xMOOCs e.g., EdX, Coursera, Udacity, etc began with Stanford's xMOOC on Artificial Intelligence, which in turn was influenced by the unconventional connectivist MOOCs themselves riding on principles of the Open Education movement (Yuan and Powell, 2013).

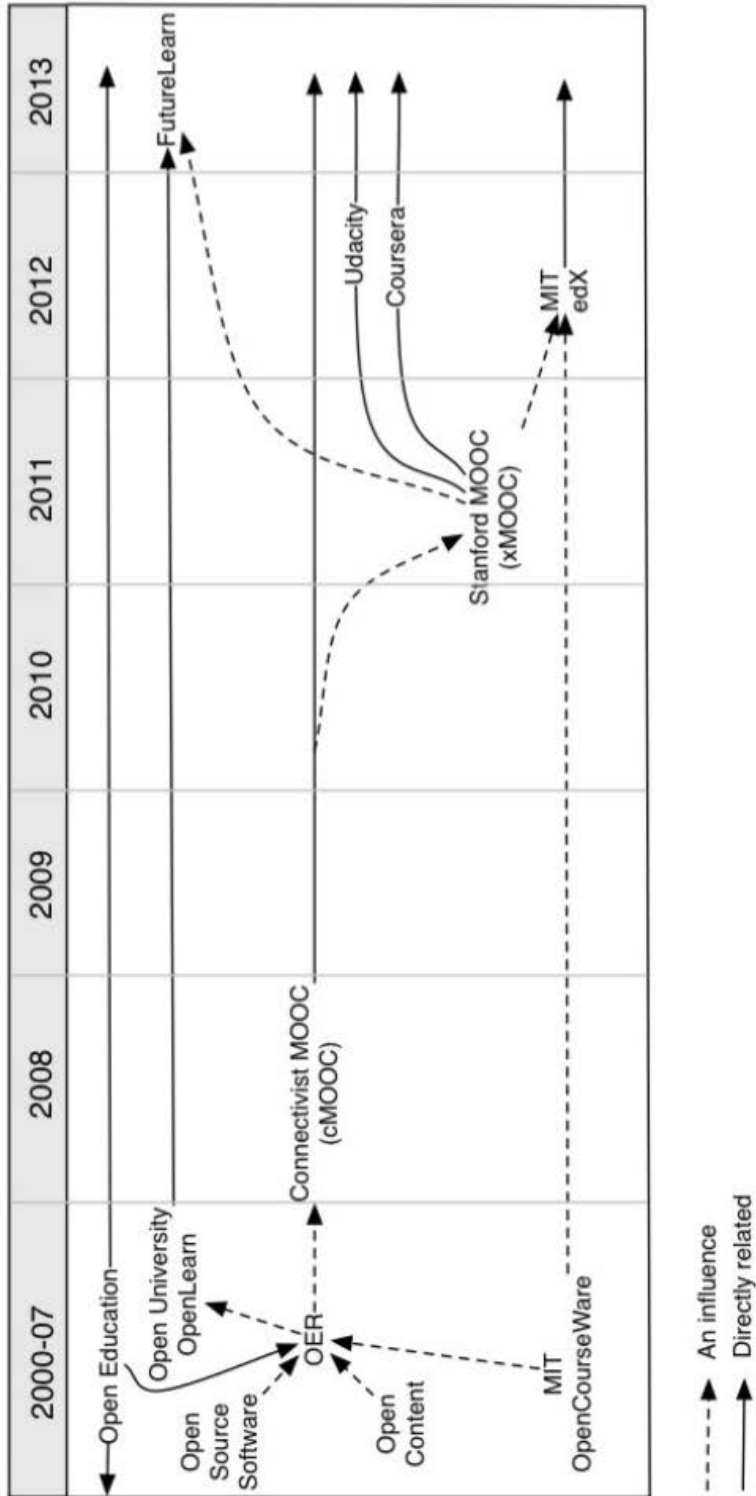


Figure 2. MOOCs and Open Education Timeline from Yuan and Powell, 2013

Further discussion of MOOCs and open education are beyond the scope of this thesis, however, they do bear importance to the topic at hand. For example, the University of Leeds in the UK integrated a MOOC with their anatomy program for 1st year medical students (Swinnerton et al., 2016) and the University of Manchester made use of online discussion boards and chat rooms to develop transferable skills in science students studying anatomy and histology (Choudhury and Gouldsbrough, 2012). In both cases students showed high levels of usage and gave positive feedback about the resources. These results are consistent with reports of students' usage of YouTube as a source of anatomy videos (Jaffar, 2012; Barry et al., 2016). What is salient about these examples is the integration, and in many instances, expectation of online media delivery alongside face-to-face classes (MacLean et al., 2011).

The classical approach to teaching gross anatomy is a combination of didactic lectures, textbook readings and laboratory study that may include dissection or prosected specimens (Sugand et al., 2010; Drake et al., 2009; Trelease, 2016). Unlike subjects that vary in content, human gross anatomy is mostly static (i.e., factual knowledge) and taught across several courses including graduate medicine, undergraduate science and, in limited scope within visual arts. Its fundamental place is in the realm of medical education where, from the early 20<sup>th</sup> century, it was divided into preclinical and clinical anatomy (Drake et al., 2009; Wilhelmsson et al., 2009). Beyond basic anatomy, medical knowledge in general and clinical anatomy are subject to ongoing growth which often leads to curricular changes and the reevaluation of teaching approaches (Johnson et al., 2012; Sugand et al., 2010). More recently, medical courses have moved toward integrative learning approaches which include problem based learning and team based learning (Drake et al., 2009; Pandey and Zimitat, 2007; Bolender et al., 2013; Lazarus et al., 2015). However, evaluation of curricular reformations and new teaching approaches has been difficult due to the lack of uniformity across

institutions (Sugand et al., 2010; Bergman et al., 2011; Trelease, 2016). Although the topics of medical education and clinical anatomy are beyond the scope of this thesis, they reveal the importance of basic anatomy for health professionals and, when, where, and how it is taught.

Anatomy is one of the oldest branches of medicine (Persaud, 1984) characterised by its high volume of content (approximately 7500 terms see: *Terminologia Anatomica*). Yet over the last 3 decades there has been a decline in teaching time for anatomy by up to 55%, while student-teacher ratios have seen a steep rise (Bergman et al., 2011; Drake et al., 2009; Topping, 2013). Additionally, the place of dissection has become a topic of debate with some results indicating a learning advantage when learning anatomy via dissection as opposed to projections (Bergman et al., 2011). However, the results are not straightforward and it is difficult to objectively evaluate dissection as the optimal method of learning anatomy (Sugand et al., 2010). Practical considerations such as running costs, time and health hazards may also contribute to ongoing debates between ‘traditionalists’ who favour dissection and ‘modernists’ who favour new teaching modalities (Johnson et al., 2012). Despite these debates, it is acknowledged that a single teaching modality is unlikely to meet all learning requirements and that mixed methods result in better learning outcomes (Bergman et al., 2011; Sugand et al., 2010; Estai and Bunt, 2016).

The methods for teaching human anatomy remained relatively unchanged until computer innovations (Trelease, 2016). Advances in information management and analysis with computers have led to several innovations in anatomy - what Robert Trelease (2002) terms anatomical informatics. Many subfields emerge which highlight the intrinsically visual nature of the subject such as: imaging, image processing, virtualisation, virtual reality, modelling and simulation, structural databases, networked information, and computational anatomy (e.g., mathematical models of bodily structures). In his historical overview of technological

innovations and their influence on anatomical sciences education, Trelease (2016) frames the various advances through Rogers' diffusion of innovation theory; this frame is also useful for analysing ideas and ways of thinking. Not surprisingly, the internet and world wide web are the technologies with the broadest and deepest influence across education and society. The most relevant development for the purpose of this thesis is the distribution of high resolution videos via the web.

To summarise the broader context, the diffusion of the internet and world wide web established a new social and educational landscape. Based on this new and shifting landscape an alternative approach to learning termed connectivism was proposed and from it the first MOOC was born. Soon after, course platforms were created based on conventional learning approaches.

In a narrower context, anatomical sciences education, which has a historically prominent place in medicine, has undergone a steep decline in teaching hours as curricula move to integrated formats rather than preclinical and clinical blocks. With a high volume of content and as an inherently visual subject, debates about the effectiveness of old and new teaching modalities persist. Meanwhile, the adoption and development of computers and technology facilitated the growth of anatomical informatics and, most pertinently, the distribution of videos via the world-wide web.

### **Problem Statement**

The overlap of technology with anatomical education has produced many novel learning applications. The integration and use of instructional multimedia in anatomy became widespread in the mid 2000's (Trelease, 2016). A notable resource, initially produced on VHS in 1995, is Robert Acland's narrated dissections which have since been digitised and made available as a web archive. A similar resource that also includes the use of cadaveric dissection

is Anatomedia, while another category of resources make use of computer generated anatomical models with varying degrees of interactivity for e.g., Anatomy.tv by Primal pictures, and the Biodigital Human. The distinction between simulated imagery and real world objects provides a useful delimitation for the analysis of instructional multimedia.

When lectures shifted from the use of 35mm slides and overhead transparencies to computer presentations created in Microsoft PowerPoint, the practicality of including rich multimedia became trivial. Additionally, the practice of recording the audio and/or video of lectures and uploading it to learning management systems grew. Initially, many were concerned about the impact on lecture attendance and learning outcomes (Gupta and Saks, 2013; Bacro et al., 2013; Fei et al., 2013). However, research has shown that overall computer based instruction is equal to or often exceeds conventional methods (McNulty et al., 2009). With the recent interest in MOOCs and the advances of portable technology, many courses have adopted blended learning approaches which integrate online course content and face-to-face classes; also referred to as flipped classrooms (Trelease, 2016).

Early studies on multimedia instruction in anatomy tended to report on student satisfaction, usage statistics and comparisons with traditional approaches (Topping, 2013; Nieder and Borges, 2012; McNulty, 2009; Ruiz et al., 2009; MacLean et al., 2011). The widespread adoption of elearning methods has created a variety of innovations, which although are received favourably, do not provide statistically reliable learning efficacy evidence (Trelease, 2016; Vorstenbosch, 2011). Similarly, Regehr (2004) notes in his review of trends in medical education research that there is a lack of studies motivated and informed by useful theories.

In 2010, Richard Mayer published a call to apply principles from the science of learning to medical education. In his paper, he outlined his cognitive theory of multimedia learning



(CTML) which is an evidence based theory of how people learn from words and pictures and includes principles for effective instructional multimedia. The first application of the theory in the field of medical education came in 2011 (Issa et al., 2011); The study used a pretest/posttest control group design and used a full lecture on shock for students in their 3rd year of medical school at a North American university. The lecture was delivered in person with accompanying slides which were used in their traditional format for the control group and in a modified format for the treatment group. Retention and transfer were measured using 10 open ended questions answered prior to the lecture and 1 hour after the lecture. The results showed significant learning improvements for both groups, however the students in the treatment group had significantly greater scores on retention questions. In a follow up study in 2013 (Issa et al., 2013), the experiment was replicated but used 4 time points to assess retention and transfer (pre/ immediate post/1 week post/4 weeks post). The results showed that students in the treatment group significantly outperformed the students in the control group across all post tests. These studies added support to the cognitive theory of multimedia learning and enriched the evidence base outside the confines of the laboratory.

A substantial amount of research conducted by Mayer and colleagues use short (30s to 180s) lessons that explain systems such as engine brakes, the formation of lightning (Mayer et al., 1999) and a bicycle tyre pump (Mayer and Sims, 1994). In these cases, the pictorial component, whether still or dynamic, were illustrations or diagrams. In the context of anatomy, lessons may include illustrations from textbooks and animations, but also photographs and live action video. The likelihood of the latter increases in clinical anatomy or in lessons that correspond to dissection and prosections. The comparison between media types is beyond the scope of this thesis, though it is a topic with renewed discussion in anatomy given changes to curricula (i.e., reduced teaching hours) and innovations such as:

3D printed reproductions of cadaveric specimens (McMenamin et al., 2014) and the use of digital Anatomage tables (Fyfe et al., 2013). To follow Richard Mayer's suggestion, the more relevant question is: "when and how does a particular type of multimedia, and in this case illustrations and animation, affect learning?" (Mayer and Moreno, 2002).

The cognitive theory of multimedia learning is a useful utility as a framework for analysing existing videos. Yue et al (2013) reviewed medical animations publicly available online. Of a total of 860 animations from 20 developers they randomly reviewed a sample of 430 videos to determine if effective multimedia principles were applied. They found that many principles were not applied and that there was an excess of extraneous auditory and visual elements. The scope of analysis was limited to freely available animations and therefore omits several commercial and subscription based services that may be of higher quality. A similar review analysing YouTube videos for learning heart anatomy used a scoring rubric to rate the general quality of videos and the level of anatomical criteria that were conveyed (Raikos et al., 2014). They found that the anatomical criteria were poor and the general quality of the videos were considered borderline. The conclusion was that the absence of content review may contribute to the low quality. In contrast, open educational resources such as mededportal consist of databases of videos, lessons and modules which aim to integrate peer and content review for quality control (Trelease, 2016). Given that the integration of eLearning resources is a priority for many subject domains, the existence of exemplary forms is important for newcomers to the field.

Guiding relevant cognitive processes for learning using multimedia requires consideration of the expertise of the learner and the learning context; individual differences may also contribute to how students use the resources (Nieder et al., 2012). Variations along these dimensions often effect learning outcomes and reveal important boundary conditions.

For instance, as a student moves from novice to expert, the principles that were once beneficial begin to hinder learning. For example, guided instruction is considered essential for novices as they construct new schemas whereas for an expert guided instruction may increase cognitive load due to the redundancy of information and neglect to stimulate germane cognitive load (Kalyuga, 2014).

As an extension to the cognitive theory of multimedia learning, Roxana Moreno's cognitive-affective theory of learning with media (CATLM) incorporates motivation and metacognition as additional variables (Brünken et al., 2010; Moreno, 2005). The CATLM includes the dual channel, limited capacity and active processing assumptions that are the core of Mayer's CTML but extends with 3 more: the affective mediation assumption, the metacognitive mediation assumption and the individual differences assumption (Park et al., 2014).

To summarise, eLearning resources in anatomical education range from open access content, to web based archives and atlases, and animation. Lectures in anatomy, making use of computers and multimedia, became recorded and uploaded to learning management systems for students' review. Studies that compared traditional versus computer resources found that in general there was some advantage of computer based instruction over classical approaches. Eventually with the wider adoption of flipped or blended models of education, the frequency and range of learning innovations increased greatly. However, several critical reviews found that there was a lack of statistically reliable research and a lack of studies informed by or based on strong theory. More recently, experiments were conducted in medical education using a lecture on shock, assessing the effectiveness of the lecture presentation delivered traditionally versus the same lecture which had been adjusted to conform with the principles of the cognitive theory of multimedia learning. The evidence

base for the theory is robust and yet it has not been tested with medical students learning anatomy. In addition to cognitive variables that affect learning, motivation and metacognition have also been implicated as mediators of effective instructional multimedia. Together, learning outcomes from applying the cognitive theory of multimedia to anatomy education, with a consideration of affective factors as potential mediators has not been studied.

### **Purpose and significance of study**

The purpose of this thesis is to apply the principles of effective instructional multimedia, drawn from Richard Mayer's cognitive theory of multimedia learning, to an excerpt about the muscles of the back. Although the theory has accumulated a significant amount of evidence, the subject domain of anatomy does not have any known applications which have been evaluated using a pre-test/post test randomised control design. Additionally, incorporating metacognitive variables for consideration acknowledges the mediating effects on learning as described in Moreno's cognitive affective theory of learning with media. Given the criticism that elearning innovations in anatomical sciences education lack statistically reliable research coupled with the increase in demands and expectations of elearning resources the current study contributes data while drawing on well established theoretical models.

To meet the call to increase statistically reliable research in anatomical sciences education, the current study uses a pretest/post test randomised control group design. This method closely matches the paradigm that Mayer and colleagues have used to establish and test the principles of effective instructional multimedia (Issa et al., 2013; Issa et al., 2011; Mayer et al., 2002). An excerpt about the muscles of the back from a full length lecture was presented without modification to the control group and with modified visuals for the experiment group. In addition to quiz questions, which are used to determine learning outcomes, likert

style questions are used to explore students' self rated levels of confidence and levels of difficulty when learning anatomy and the muscles of the back.

### Research Questions and Hypotheses

The primary research question is:

1. Will an online anatomy lesson consistent with the CTML versus an unmodified 'default' online lesson yield larger learning gains for graduate medical students?	
Part A.	Establishing equivalence
	i - What are the sample characteristics for variables of age, sex, prior degree, and prior learning in anatomy?
	ii - Are the sample demographic variables equivalent between groups prior to the intervention?
	iii - Are pretest quiz results equivalent between groups?
Part B.	Assessing changes
	i - What are students' quiz scores before and after watching an online lesson consistent with CTML?
	ii - What are students' quiz scores before and after watching an online lesson in its default form?
	iii - How do the results of the two lessons compare?

Secondary question of the thesis:

2. What are the metacognitive ratings of medical students when learning anatomy and the muscles of the back and do they change depending on the lesson format?	
A	Does perceived confidence/difficulty when learning anatomy and the muscles of

	the back vary depending on the type of multimedia lesson?
B	What are the effect sizes of the changes?
C	(Consistency check) Does self-rated difficulty when learning the muscles of the back (pretest) correlate with self-rated ease when learning the muscles of the back (posttest)

Exploratory data analysis

Post-hoc category membership via Jacobsons RCI:

What are the characteristics of students who made significant learning gains?

### **Assumptions and limitations of the study**

**Philosophical assumptions.** The constructivist view of learning is a dominant paradigm originating with philosopher Giambattista Vico in the 18<sup>th</sup> century. It was extended and developed into several varieties by Jean Piaget, John Dewey, Jerome Bruner, Lev Vygotsky and others. At its core it holds that knowledge is not transmitted from one person to another, but rather is constructed in the mind of the individual and depends on their prior knowledge (Narayan et al., 2013). Further an individual is assumed to be an active sense maker who may elaborate and interpret their experiences (Perkins, 1992). Thus, the philosophical assumptions are a relativist ontology and subjectivist epistemology.

The assumptions carry themselves across to methods of instruction. For example, discovery learning, whereby a student is not explicitly guided by a teacher and left to find and make sense of material on their own, is a (radical) constructivist position. This form is often contrasted with guided instruction which has shown in empirical studies to yield better learning outcomes. The issue, according to Mayer (2009b) is the conflation of constructivism as a prescription for instruction as opposed to a theory of learning; where the former assumes behavioural activity is equal to active learning and the latter assumes that engaging specific cognitive processes is equal to active learning. Constructivism as a theory of learning is the foundation of the cognitive theory of multimedia learning (CTML); where the emphasis is on the cognitive processes that are most important for the effective construction of knowledge: selecting, organizing and integrating information. According to Mayer, it is possible to reconcile information processing theories (such as CLT) and constructivist theories whereas others such as Bednar et.al (1992) contend that the philosophical differences make the combination inconceivable.

A discussion of the philosophical assumptions of a learning theory may seem far removed from its practice, however, its consequences often appear in complex scenarios. For example, a well structured subject domain such as mathematics or certain science subjects are well suited to constructivist approaches such as CTML. However, for ill structured domains such as medicine (diagnosis), literary interpretation or history, certain principles of instruction may in fact hinder learning (Spiro et al., 1992). For instance, introducing a complex topic with a simple example may impede learning via what Rand Spiro terms seductive reductions (Spiro, 2009) or by inducing the Einstellung effect (Schwartz et al., 2009; Bilalic et al., 2008; Sweller et al., 2011). Multimedia use when teaching ill-structured subjects in medicine is beyond the scope of this thesis, however, it does illuminate the context-dependent nature of instruction.

**Limitations.** Although participants were randomly assigned to control and experiment groups, students who volunteered to participate may not accurately represent the population in general. Further, the multimedia content presented a specific area of anatomy (the muscles of the back with some information on innervation) which may not be representative of other areas of study in anatomy.

**Delimitations.** The decision to use multiple choice questions allowed for simple administration and analysis as opposed to free response format (FRF) questions. Additionally, there were only 6 questions used to measure knowledge, which was largely declarative; these questions were defined by the learning objectives from the existing lecture. However, the experiment used a short excerpt from a lecture and applied several principles from the CTML; given the 'steam roller' approach to applying principles and the focused topic of instruction, a large number of questions was not deemed appropriate. Further, since



medical students were participants we did not want to create a lengthy voluntary task given their already heavy workload.

A variable that has been shown to predict performance in anatomy is visual-spatial ability. A mental rotation test (MRT) was not included in the experiments since the focus was on the application and evaluation of the CTML.

Finally, likert items asked students to rate their perceived confidence in understanding anatomy and the muscles of the back and, perceived difficulty when learning anatomy and the muscles of the back. Although these questions do not form a validated scale and depart slightly from common measures such as mental effort, it has been shown that constructs such as self-efficacy (Burgoon, 2008) and perceived difficulty are related to task performance (DeLeeuw et al., 2008).

**Operational definitions used in the study**

Population variables included:

- 1) age – which could be entered as a free text number
- 2) sex – which provided three options male, female, other (although, in hindsight the third option is more suitably coded as ‘x(Indeterminate/Intersex/Unspecified)’ or ‘Unspecified’ in accordance with the Australian Government Guidelines on the Recognition of Sex and Gender (2013).
- 3) Prior area of study – which provided nine options. This list was limited and answers to this question would have been better suited to free-text input.
- 4) Prior exposure to learning anatomy – this included a scale of 1-5 with an accompanying description to help students quantify their learning (See Appendix A)

Moderating variables:

For experiment 2, an additional question was included. The multimedia excerpt was from a full length lecture, presented online to the cohort earlier in the year. Therefore, students were asked if they had seen the lecture. This variable was included in the split plot Anova during analysis.

Outcome variables:

The primary outcome variable used in the study is performance on a 6-item multiple choice quiz. Additionally, 4 likert type questions on self-rated confidence in understanding anatomy and the muscles of the back, and self-rated difficulty when learning anatomy and the muscles of the back were included. The scales in these questions are unidimensional i.e., for difficulty questions, a rating of 1 equals not difficult through to 5 equaling extremely difficult.

### **Theoretical orientation for the study**

The cognitive theory of multimedia learning is an information processing theory of how people learn from words and pictures. Empirical studies have yielded 10-12 principles for effective multimedia instruction which are grouped in 3 categories. The theory and its development depends on several key works and assumptions in cognitive science. Therefore, this review will begin with a brief history of its influences.

#### **Dual Coding theory**

Dual coding theory is an explanation of our cognitive representational systems developed by Allan Paivio (Paivio, 2007). It posits that there are two subsystems dedicated to processing verbal and nonverbal objects and events. These two systems are structurally and functionally distinct, however, they are partially interconnected. Thus, activity in one system may initiate function in the other.

The symbolic verbal and nonverbal systems are derived from and retain functional properties of our sensorimotor systems. For example, activation of the verbal system may result from visual, auditory, articulatory and other modes of presentation that deal with language. As a result, a defining feature of this system is that it processes information serially or sequentially. In contrast, the nonverbal system may be activated by images, environmental sounds, haptic stimuli, and other nonlinguistic objects and events. Unlike the verbal system, the nonverbal system processes information in parallel or simultaneously. The relationship between the sensorimotor stimuli and, the nonverbal and verbal symbolic systems is therefore described as orthogonal. For example, a visual presentation may include printed words or visual objects. In both cases the visual sense modality is being activated, however, the symbolic representations of the two stimuli are verbal and nonverbal, respectively. Similar orthogonal contrasts are evident in all other sense modalities with the exception of gustatory,

olfactory and affective modalities which are all nonverbal (Paivio, 2007).

The smallest representational units in the verbal and nonverbal symbolic systems are referred to as logogens and imagens. These terms are used to distinguish the underlying structural representations of consciously experienced inner speech and images. The activation of logogens and imagens may proceed directly and indirectly from the unit level to generate recognition, recall and meaning. The propagation of activity often begins with a direct sensory stimulus such as an object activating imagens and words activating logogens. This activity is defined as representational processing in the DCT scheme. When activity propagates between the nonverbal and verbal systems, i.e., from the word to its referent, it is defined as referential processing. Finally, associative processing is defined as within-system activity from logogen to logogen and imagen to imagen. All the aforementioned processes and activations are dependent on the stimulus, contextual stimuli and individual differences.

The development of DCT has roots in studies of memory and recall. The key idea from which the theory grew is the conceptual peg hypothesis (Paivio, 1991, 2007); it takes cues from a mnemonic technique where an ordered sequence of numbers and associated rhyming words (one-bun, two-shoe, three-tree, etc) form conceptual pegs for items that need to be remembered for e.g., if a pencil is the first item to remember, one may image a pencil inside a hotdog bun. Thus, when cued with the number one, the association is bun which in turn triggers the newly formed image including the pencil. In DCT terms, this example illustrates representational, referential and associative processing between and within the verbal and nonverbal systems. However, the theory does not explicitly detail the structure and function of memory. Therefore, this topic is the next area of focus. (see Appendix G for: further detail about DCT and an exploratory multimedia project applying mnemonic techniques to anatomy).

## **Memory**

In 1956 Miller published a paper that established 7 plus or minus 2 “chunks” as the average span of our immediate memory. The memory he was referring to is similarly termed the short term store, short term memory and most recently working memory. Over time, the capacity of working memory has been challenged and evidence for a shorter span of 3-4 chunks has also been proposed (Cowan, 2001). The experimental paradigms that test span rest on several assumptions about the structure of memory. Therefore a brief genealogy of the concepts are in order.

Initially, it was proposed that memory consisted of two components: a short term store and long term store. The assumption under this model is that information flows from our sensory system via the short term store into long term memory (Atkinson and Shiffrin, 1968). Information is stored and processed in the short term store before transfer to long term memory is possible, and retrieval from long term memory also depends critically on the short term store. However, neuropsychological evidence from patients with damage to their short term store did not show significant deficits in long term learning (Baddeley, 2003). The shortcomings of a two component model of memory led to the proposal of a multicomponent system of memory. The model developed by Baddeley and Hitch (1974) includes a tripartite working memory with a central executive and two slave systems, the phonological loop and the visuospatial sketchpad. The central executive functions as an attentional controller which is limited in capacity and uses storage from both short term and long term memory. The phonological loop is time-limited but not necessarily capacity limited. It stores speech based information for a few seconds and is the “mind’s ear” i.e., where articulatory rehearsal processes occur (Baddeley, 2003). The visuospatial sketchpad serves to integrate visual and spatial information from a range of modalities (Baddeley, 2001). Like the phonological loop it

can be subdivided into components capable of storage, maintenance and manipulation (Repos and Baddeley, 2006). More recently another component has been added to the model - the episodic buffer (Baddeley, 2000). It is a limited capacity component that is controlled by the central executive, and functions as an intermediary between subsystems. As an interface between the systems it provides a temporary modelling space for integrating the different modes of experience into a unitary multidimensional representation; the integrative function also extends to long term memory stores. The working memory system as conceived by Baddeley is therefore crucially involved in cognitive processes that extend beyond memory (Baddeley, 2000). Given the outline of the components of memory, the question of capacity may now be addressed.

Early tests of working memory span used serial recall of lists of words or a string of digits (Jacobs, 1887 cited in Baddeley, 1990). The number of items would increase gradually and recalling the list with 50% accuracy was deemed the limit of a participant's short term memory. Miller's magical number 7 quantified the limit in terms of meaningful chunks of information i.e., a unit of ideas that have strong associations with each other, but not to other units. However, what constitutes a chunk varies across experiments which, in part, led Cowan (2001) to propose a 4 chunk limit instead. Although Cowan presents convincing arguments and evidence for a smaller span, complex questions arise about the distinction between processing capacity, storage capacity and methods that obtain a 'true' measure of immediate memory span. Ultimately, Cowan equates the limited capacity STM with the focus of attention. In Baddeley's view, this definition accords with the central executive in the multicomponent model of working memory (Baddeley, 2001). The debate about capacity continues with several alternative models of memory and cognition proposed including: global workspace theory (Baars, 2001), ACT-R architecture (Tiitinen, 2001) and levels of

processing ( Craik and Lockhart, 1972). Simon Grondin (2001) even suggests that the question of capacity, which connotes volume, may better be replaced with a focus on temporal duration and processing. In the model proposed by Baddeley and Hitch, capacity limits are revealed indirectly as a function of time (Cowan, 2015). For example, in experiments testing the phonological loop, results yield the word length effect; memory for word sequences decline with longer words (Repovs and Baddeley, 2006). The explanation is that longer words require longer subvocal rehearsal time which results in more decay in the phonological store. It may be the case that the visuospatial sketchpad functions similarly using a visual 'cache' for storage and a spatial 'inner scribe' to rehearse and manipulate information (Logie 1995, 2001 cited in Baddeley, 2012). What remains critical is the attentional control exerted by the central executive and the binding function of the episodic buffer. To reiterate, the working memory system functions across a range of cognitive tasks. As an interface between long term memory, perception and action the system is implicated in theories of intelligence, learning and consciousness. Indeed, Sweller et al. (1998) write 'working memory can be equated with consciousness'. Therefore, understanding the constraints and structure of working memory leads to effective models of knowledge acquisition. Cognitive load theory is one such model.

### **Cognitive load theory**

What is learning and how might we define it? From a neuropsychological perspective, perhaps learning is the interplay between temporary electrical activation of the working memory system, and neural growth that characterises long term memory (Hebb, 1949 cited in Baddeley, 2003). In the history of psychology, the framework for understanding learning has shifted from observations of behaviour and a stimulus-response paradigm, to a focus on cognitive processes in individual learners. Cognitive load theory introduces several constructs to guide effective instruction based on human cognitive architecture (Moreno and Park,

2010).

In cognitive load theory, learning is defined as the acquisition of schemas in long term memory. A novice borrows information from the long term memory of an expert or teacher to build their own network of connected ideas (Sweller, 2010). When a network of ideas are understood as a unit - a schema- refinement may continue, connections to other schemas may be formed, and with that expertise may grow. The canonical example is the skill of chess grand masters. What differentiates them from novices and less practiced players is their long term memory of board configurations from real games (Sweller, 2010). Further, a well developed schema eventually becomes automated and conscious processing is not required. In order to effectively construct a schema the new information needs to be processed in the working memory system. As discussed previously, this system is limited in its capacity and processing abilities. The cognitive load that the system is subjected to is the main construct elucidated by cognitive load theory.

Sweller et al. (1998) propose 3 types of cognitive load: intrinsic, extraneous and germane, and how they may be managed during learning. Intrinsic load is defined in terms of the complexity of the content. When there are several interacting components that require simultaneous processing by a learner the intrinsic load is considered high. However, learner prior knowledge influences this variable since content deemed high in intrinsic load for a novice, will not exert this load for an expert. A contested quality of this variable is whether it is amenable to manipulation and change (de Jong, 2010). The original conception maintains that it is not.

A type of load that is subject to adjustment is extraneous load. It is defined as information that does not contribute to learning in terms of schema construction and automation. As a result, several instructional design techniques are suggested which help limit the load put on



the working memory system. For example, presenting text and images in an integrated form rather than separate reduces extraneous load by removing the need to keep isolated pieces of information in working memory; this would otherwise split attention. The modality effect follows from this principle by encouraging the use of both channels (verbal and nonverbal) of working memory to distribute the load. However, in some cases separate pieces of information may be understood in isolation such as, a paragraph of text describing a process and an image illustrating the same process. In this instance the information is redundant and increases extraneous cognitive load. However, experiments showed that for novice learners integrated and redundant information was beneficial, whereas for expert learners in a domain the integrated redundant information resulted in performance decline. There are several more cognitive load effects described by Sweller (2011) with a significant proportion aimed at reducing extraneous cognitive load (Plass et al., 2010).

The relationship between intrinsic and extraneous load is additive. If the compound amount exceeds working memory capacity, then learning is negatively impacted. When extraneous load is successfully reduced, and intrinsic load is managed free working memory capacity may be directed at tasks relevant to schema acquisition and automation. The redirection of free capacity is considered germane load and adds to the total cognitive load. Managing the ratio of the three types of load therefore leads to effective learning.

There is, however, difficulty in precisely defining and measuring cognitive load. This is a considerable conceptual problem according to de Jong (2010); the multidimensional cognitive load construct was defined by Paas (1992 in de Jong 2010) as consisting of mental load and mental effort. Mental load is associated with the characteristics of the learning material or task while mental effort is associated with the amount of mental resources allocated to the learning material or task. The concepts of mental load and mental effort are

inconsistently aligned with extraneous or germane load. Similarly, the role of the subjects' characteristics, and the task demands on mental effort also shifts; initially mental load was ascribed to task demands independent of subject characteristics, and later it was referenced as the interaction between the two variables. Adding to the construct conundrum, is the question of measurement.

Measures of cognitive load are usually obtained indirectly via self-report questionnaires. The scales used vary from a single item 9 point scale, to several items. The anchoring terms vary from: examining perceived difficulty of the task to examining mental effort. In some cases, single questions examine difficulty and understanding, and in others the terms are given separate questions. De Jong (2010) notes that single item measures of cognitive load do not distinguish between the types of load and, in cases where such attempts are made the results are inconsistent. However, Deleeuw and Mayer (2008) showed that three different measures could differentiate the three types of cognitive load. Their experimental results suggested that extraneous load could be measured via reaction time on a secondary task, intrinsic load could be measured by mental effort questions during a task, and germane load could be measured via a difficulty rating after the task. Further, physiological measures of cognitive load include heart rate, which has proven less reliable/valid, brain activity, and pupil dilation which is described as an accurate indicator with the advantage of temporal information before, during and after a task. Several of these methods are intrusive, costly and may interact with the constructs of interest (Paas et al., 2010). Therefore self-report questionnaires are useful as inexpensive measures of cognitive load, when their terms are carefully defined to reduce the ambiguity of meaning. However, the measurement of cognitive load in these instances assumes that exceeding an individual's working memory capacity is detrimental to learning without including measures of capacity in the first place

(Moreno, 2009; de Jong, 2010). We have come full circle by returning to the issue of working memory capacity and its measurement. Given that the topic is still under active debate, attention may be refocused on the notion of cognitive processing instead.

### **Cognitive Theory of Multimedia Learning**

Several theories have been introduced as a cast of supporting players for the lead protagonist of this thesis - the cognitive theory of multimedia learning. The theory was developed by Richard Mayer (2009) and colleagues, and combines the science of instruction with the science of learning (Mayer, 2008). An explanation of the assumptions, structure and functions of the model follows.

The first assumption of the theory is that the human mind processes information in two channels. The strength of this assertion rests on Allan Paivio's dual coding theory and Alan Baddeley's model of working memory. To recall, according to Paivio, stimuli may be classified as verbal or nonverbal and this dichotomy extends to internal representations. Further, the stimuli may engage the visual, auditory, haptic, gustatory, and other sensory systems in a variety of modes. For example, on screen text is classified as a verbal stimulus, which engages the visual system and activates the verbal symbolic system. Baddeley's model of working memory similarly proposes two systems but distinguishes between the speech based phonological loop and the visuospatial sketchpad (Baddeley, 1990). The focus in this model is on the sensory system that the stimulus engages. Therefore, on screen text is classified as a visual stimulus which, depending on task demands, may be converted to sound that is rehearsed in the phonological loop. Drawing on both these models, Mayer proposes two channels: the visual/pictorial channel and the auditory/verbal channel. Paivio (1991) cautions against contrasting the visual channel, which is considered a sensory modality and the verbal channel, which is considered a symbolic system. The compromise Mayer proposes

is to distinguish stimuli according to the sensory modality they initially engage, while acknowledging separate symbolic systems when learners are constructing a pictorial and/or verbal model in working memory (Mayer, 2014a). Figure 3 is a diagram of the theory (excerpt from Mayer and Moreno, 2003) which shows the visual and auditory channels, the processes engaged and the key memory structures involved in learning. Working memory is conspicuous due to the considerable cognitive processing that takes place in its bounds.

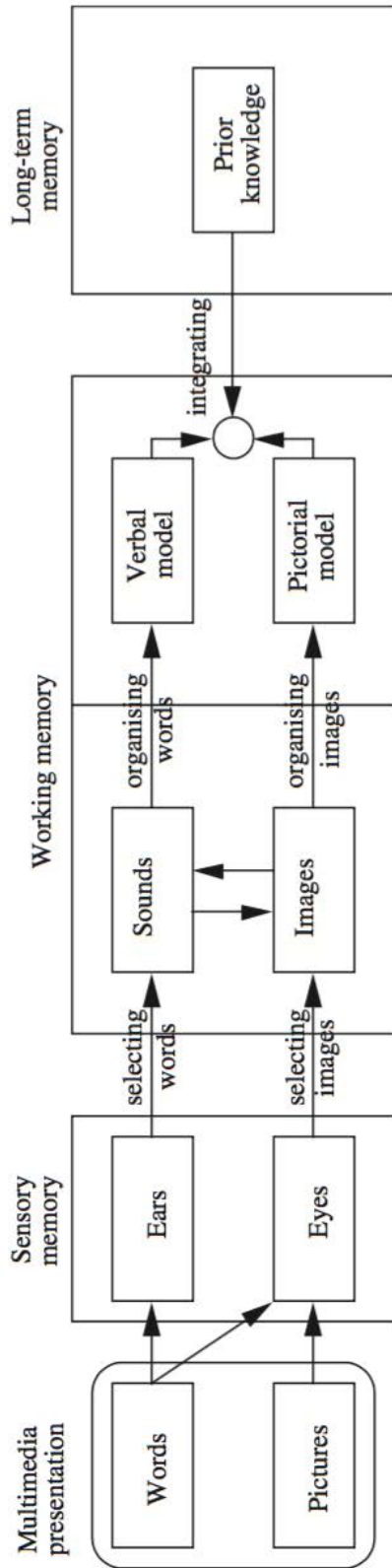


Figure 3. The cognitive theory of multimedia learning from Mayer and Moreno, 2003

The second assumption is that the working memory system is limited in the amount of information it can process in each channel. The human visual and auditory sensory systems are subject to bottlenecks. For instance, consider the flow of information through the visual/pictorial channel in Figure 3. Pictures and words as text arrive at the retina and as the information passes to the visual cortex it degrades significantly. Some estimates state that of the unlimited information in the world around us, approximately 10 billion bits per second arrive at the eye, and given that the optic nerve has about a million output connections, 6 million bits p/s leave the retina and finally about 10 000 bits p/s make it to the visual cortex (Raichle, 2010). Surprisingly, a mere 100 bits p/s are thought to be involved with conscious processing and perception. In a novel study Kaczmarzyk et al. (2013) set out to quantify stimulus items and their recollection in terms of bits of information. They presented a series of charts with four items that varied in their bit value and calculated the bit value of each item using Shannon's formula. Their results suggested that the capacity of working memory is 30 bits of information regardless of the bit size of each item. However, Cowan (2015) describes quantifying working memory capacity from an information theory perspective as an issue that ended with Miller's original article on the magical number 7 citing the fact that bits of information are not the basic units of memory, but rather meaningful chunks. Although the issue is somewhat unsettled, recent explorations using graph theory, network neuroscience and fMRI data are promising (Bola & Borchardt, 2016; Caeyenberghs et al., 2016).

As discussed in the section on memory, quantifying the capacity of working memory is not straightforward. The stance that CTML takes is that working memory is limited to 5-7 chunks of information. It follows, that given the processing limitations, the learners' cognition requires guidance and stimulation in appropriate forms. Drawing on cognitive load theory, the CTML proposes 3 working memory demands that require management for

effective learning: extraneous, essential and generative processing. The adaptation in CTML is that the demands are framed in terms of processing. This is due to the focus on learners' cognitive processes as opposed to the focus on instructional design.

The third assumption of the theory is that learning involves active processing, which was influenced by Wittrock's generative model of learning (Wittrock, 2010; Mayer, 2010b). To build an effective model, learners must engage in selecting relevant words and pictures, organising information into respective verbal and pictorial models and finally integrating the models with each other and with prior knowledge from long term memory. As is shown in Figure 3, these cognitive processes occur in working memory but do not necessarily occur in a linear way. The critical factor is the learners' management of these processes which is equivalent to the executive controller in Baddeley's working memory model (also known as metacognitive strategies). In this respect, the theory reveals the influence of modern conceptions of learning as knowledge construction as opposed to prior ideas focused on knowledge and response acquisition (Mayer, 1992); although Mayer does acknowledge that there are instances where latter approaches are appropriate (Mayer, 2014b).

Given the assumptions of a dual channel, limited capacity system which engages in active processing to make sense of multimedia presentations the CTML yields 10-12 principles for effective learning. The principles are grouped according to three aims: reducing extraneous processing, managing essential processing and fostering generative processing. Table 1 shows the principles, their descriptions, median effect size and the amount of experiments that have tested the principles.

Table 1

*Principles of the cognitive theory of multimedia learning*

Cognitive aim	Principle	Description	Median effect size
Reduce extraneous processing			
	Coherence	Remove words and pictures that are not relevant to instructional objective	0.86 (23/23 exp)
	Signalling	Add cues that highlight the organization of essential material	0.41 (24/28 exp)
	Redundancy	Do not add printed text to spoken text.	0.86 (16/16 exp)
	Spatial contiguity	Place corresponding words and pictures near each other rather than far apart	1.10(22/22 exp)
	Temporal contiguity	Present animation and narration simultaneously rather than successively	1.12(9/9 exp)
Manage essential processing			
	Segmenting	Present information in learner paced segments instead of a continuous unit	0.79(10/10 exp)
	Modality	Present words in a multimedia lesson as spoken rather than printed.	0.76(53/61 exp)
	Pre-training	Present the names and characteristics of main concepts before a narrated animation.	0.75(13/16 exp)
Foster generative processing			
	Multimedia	Learning with words (verbal content) and pictures (visual content) is better than learning with words alone	*1.39 (11/11 exp)
	Personalisation	Present words in conversational rather than formal style	0.79 (14/17 exp)
Advanced techniques			
	Voice	Use human rather than machine voice	**0.74 5/6)
	Embodiment	Give on-screen characters humanlike gestures	
	Guided discovery	Provide hints and feedback as learners solve problems	
	Self-explanation	Ask learners to explain a lesson to themselves	



drawing	Ask learners to make drawings for the lesson	
---------	--	--

\*from Mayer, 2010a, all other effect sizes are from Mayer, 2014a

\*\*from Mayer, 2014b

## Experiment 1

### Methods

**Participants.** A total of 83 students in their 2nd year of the Sydney Medical Program (SMP) agreed to participate in an evaluation of learning and teaching multimedia for anatomy. Students who did not follow instructions and missing data for pre and post test were excluded leaving a final sample of 76 participants. The students were randomly assigned to control and experiment conditions. Table 2 shows participant age and sex by group.

Table 2

*Age and sex of participants in each group*

Group	Sex			Age		
	Male	Female	Other	Mean	SD	Range
Control (n=36)	19	17	0	25.89	4.214	19
Experiment (n=40)	23	17	0	24.74	2.935	11

**Sampling procedure.** A population of 294 students in their 2<sup>nd</sup> year of the SMP were invited to participate in an evaluation of learning and teaching multimedia for anatomy. An email bulletin was sent to the cohort announcing an anatomy video trial which involved: answering questions, watching a 5 minute video and completing a follow-up test after 7 days. In preparation they were encouraged to register on a sign-up sheet on the university LMS - Blackboard. They were informed that the trial would take place on the day of their scheduled practical session in an adjacent room. There were several practical sessions to which students were assigned in the course.

On the day of the experiment one of the senior lecturers (supervising this study) and myself stood by the entry of the experiment location and invited students to participate in our video evaluation. Due to the multiple practical sessions, some students participated before their lesson, while others took part afterwards. Twenty-eight percent (N=83) of the target population volunteered. The data were collected electronically in the university LMS - Blackboard. Each group had a page designated to them and the availability of the page was restricted to the day of the experiment. The students accessed the LMS on campus in a computer lab.

The scope of the study fell within the disciplines' ethics approval for course evaluations. Students provided verbal informed consent and implicit consent when beginning the experiment. They could withdraw or elect to not answer questions if they wished.

**Sample size, power and precision.** The effect sizes of the principles for instructional design of multimedia lessons frequently yield scores above  $d=0.5$  (Clark and Mayer, 2011; Mayer and Moreno, 2003; Mayer, 2010). Mayer suggests that scores at or above 0.5 indicate practical significance and are worth applying (Clark and Mayer, 2011; Mayer et al., 2002).

Given that a combination of principles were applied to a short multimedia lesson, an effect size larger than 0.5 was expected.

A power analysis using G\*Power 3.1 (Faul and Erdfelder, 1998) indicated that a total sample of 90 people would be needed to detect an effect of  $d=0.6$  (at minimum) with 80% power using a t test between independent means with alpha at .05. Similarly, power analysis of a paired t test indicated that an individual sample of 24 people would be needed to detect an effect of  $d=0.6$  with 80% power and alpha of .05.

The sample achieved was 83 participants total. In light of the power analyses, group sizes of 25-45 were considered reasonable.

**Measures and covariates.** The primary outcome measure was performance on a 6 item multiple choice quiz. Secondary outcome measures included 4 self-rated questions about level of confidence and level of difficulty when learning anatomy and the muscles of the back. Two open ended questions were included for freeform input; they asked participants to specify what was difficult about learning anatomy and the muscles of the back. Additional population variables collected were: age, sex, prior degree and previous experience in learning anatomy. All data were collected electronically via the university LMS (Blackboard).

A number of methods were used to ensure the quality of measurements. First, the quiz questions were written by an anatomist who is a senior lecturer. The type of questions included recall and a transfer question. Second, the measures were repeated across three time periods. Finally, a self-rating question on the post test was reverse coded and worded to serve as a consistency check.

The instrument used to quantify performance was a set of 6 selected response format (SRF) questions. The questions were automatically scored and the order of possible answers (5) were randomly arranged for each individual; there was no immediate feedback indicating

correct or incorrect answers. The type of questions ranged from recall to understanding and applying information from the lesson that was presented (transfer).

The use of SRF or multiple choice questions has been shown to be a valid method for assessing anatomy. Shaibah and Vleuten (2013) found that SRF performance correlated well with that of free response formats. Other benefits of this form include: ease of administering, no bias during marking and low cost. However, undesirable effects such as cueing, and retrieval-induced forgetting require managing.

A number of studies that have tested the CTML make use of free response format questions (Mayer et al., 2002; Issa et al., 2013); the primary reason being avoidance of cueing effects. The subject matter often focuses on systems and, understanding of cause and effect. As a result, the FRF was deemed most appropriate for assessing transfer and problem solving. The expectations and methods of assessment at institutional, course, and unit of study levels are varied. The problems of this misalignment are highlighted by Lodge and Bosanquet (2014) who note the longstanding difficulty in assessing learning outcomes and the dependence on proxy measures to measure changes in cognition. It is unrealistic to seek a universal method of assessment, however, evidence collected from the science of learning (including CTML) may be leveraged to select the most appropriate method to present and test material.

The characteristics of anatomy include a high volume of material, and demands on visuo-spatial skills. In this context, retrieval practice and testing effects from frequent quizzing are beneficial for learning. Additionally, seeking feedback from students on their learning experience provides useful information on the relationship between performance and self-perception (Leutner, 2014; Park et al., 2014; Tsai et al., 2011; Burgoon, 2008).

Four likert-style self-rating questions were included in pre and post test questionnaires. They probed students' self-perceived confidence in understanding anatomy and the muscles of the back, and difficulty in learning anatomy and the muscles of the back. As a method to check consistency the post test questionnaire asked students about their ease in learning about muscles of the back. The use of single questions as measures of mental effort and task perception have been shown to have psychometric properties similar to multi-item scales (Yeo and Neal, 2004 cited in van Gog et al., 2012). The 9 point mental effort rating scaled developed by Paas (1992) has been adapted and used in numerous studies (Stuijzand et al., 2016; Kok Ng, 2014; Homer et al., 2008). This measure may be combined with performance measures to create an index of instructional efficiency. Another reason for considering this measure in the realm of multimedia instruction is in its capabilities as a tool for adaptive learning (van Gog and Paas, 2008).

**Experimental design.** A mixed methods pretest/posttest/delayed post test design with one factor between subjects was used. The independent variable was the instructional multimedia (with and without application of Mayer's CTML principles à la Issa et al, 2013). The instructional multimedia was an excerpt from an existing online lecture about the vertebral column; the section excerpted was about the muscles of the back. The primary dependent variable was performance on a quiz as per paradigm used in Mayer's experiments (Mayer, 2014a) and others testing the CTML (Issa et al., 2011; Issa et al., 2013). Additional dependent variables collected included self-rating questions about level of confidence and level of difficulty when learning anatomy, and the muscles of the back; these were scored using likert-type questions. Open ended questions were included to seek feedback about learning difficulties. Additional covariates collected were: age, sex, previous degree, and

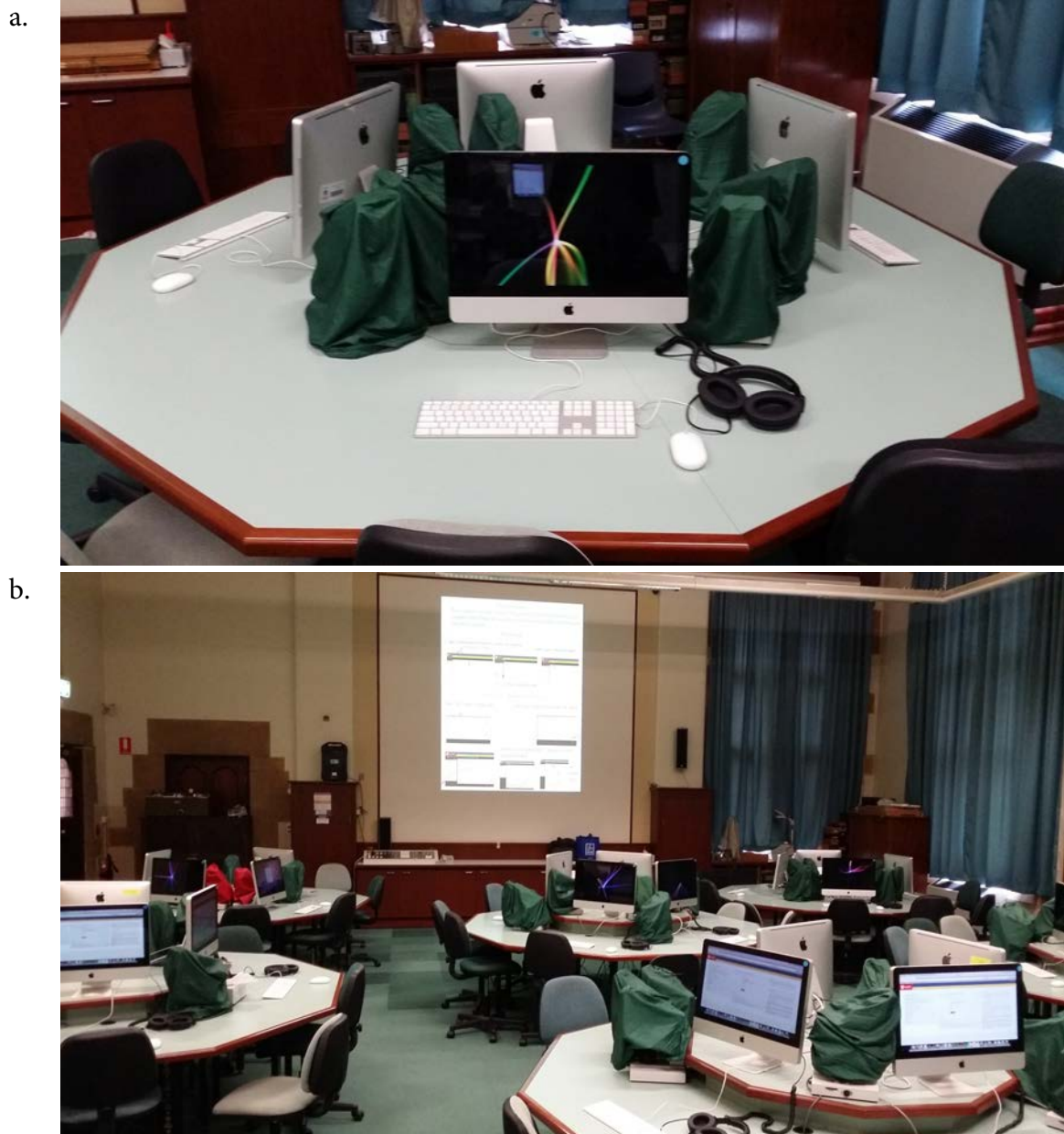
previous experience in learning anatomy (see appendix A and B for the pretest and posttest questionnaires).

Participants were recruited via a convenience sample of students in their 2<sup>nd</sup> year of the Sydney Medical Program. They were randomly allocated to control and treatment groups upon entry to the room where the study took place. A Randomization Plan From <http://www.randomization.com> was generated with 96 subjects randomized into 6 blocks (control and treatment). This list was transferred to coloured slips of paper where yellow stood for control and green was the experiment group.

**Materials.** The materials used were digital videos embedded in the University LMS. All questionnaires were also digitally administered.

**Apparatus.** The apparatus used to conduct the experiments included 20 x 21.5 inch iMac computers and 20 x Sennheiser HD380 Pro headphones. Instructions for beginning the experiment were displayed via a projector (Appendix C). Figure 4 shows the configuration of computers where the experiment was conducted.





*Figure 4.* The computer lab where the experiment was conducted. a) Experiment station set up. b) Instructions for experiment shown projected at the front of the room.

**Experimental manipulation.** The content of the experimental manipulation was a lesson about the muscles of the back; this was excerpted from an existing full length lecture. This topic area was selected based on instructors' feedback that it was an area of poor performance in exams. It was presented and narrated by a senior lecturer/anatomist (Deborah Bryce) with a runtime of 4min38s. The control group viewed the excerpted segment with no adjustments made to the audio or visual material. The video for the treatment group was edited using Adobe After Effects CC. A combination of principles from CTML were applied, while the audio was kept identical.

The control video had a total of 2 slides containing relevant images about the muscles of the back. The mouse cursor served as a laser pointer to draw attention to structures being identified. The treatment video was significantly extended to approximately 27 stills. The process of making adjustments began with importing the audio and images (Moore's anatomy) from the control video. The images were placed in sync with the narration (temporal contiguity) and short descriptive sentences were added to slides; some verbatim quotes, some slightly different (desirable difficulty). As structures were identified, coloured overlays served as highlights (signaling) and labels also appeared in synchrony near the structure (spatial contiguity). A number of additional visual devices were used to complement the principles. For example, images were faded at times to serve the signaling principle. Placement of the images was kept in the centre of the screen to assist essential processing in the visual channel; positioning also served to graphically reinforce narrated points (see still #7 in Appendix F). The treatment group video included 3 extra images, 2 of which were animated to demonstrate function and 1 that was paired with the narration to maintain the modality, multimedia and coherence principles.

The videos may be viewed at:

<http://bit.ly/controlmedia> (p/w: anatomy1)

<http://bit.ly/experimentmedia> (p/w: anatomy2)

One of the difficulties with testing the CTML is that the redesign of lecture slides from their default to modified form cannot be precisely quantified. A commentary by Rachel Ellaway (2011) notes the importance of carefully documenting the redesign of slides to generate insight into the design process and how decisions are made. Similarly, it introduces questions about performance increase and how that may be attributed to any kind of change i.e., the Hawthorne effect; this limitation was acknowledged by Issa et al (2013). A precise method to examine performance increase as a function of multimedia style could be to generate 3 conditions as opposed to the two in the current thesis. One condition could be a misaligned style of presentation, another could be presented as default and one could be an optimised version (i.e., CTML applied). In doing so the adjustments from one example to another could stand out or be more easily distinguished. However, the difficulty with this suggestion would be creating a condition that intentionally disrupts learning; this could be ethically questionable when conducted in classes, and with potentially assessable content.

The media were embedded in the university LMS and all data was acquired by downloading excel data via the grade centre. The experiment occurred on campus in a computer lab and students participated individually. Originally 3 sessions were intended: pretest, post test and delayed post test (7 days), however, due to low numbers of participation in the delayed post test the data was omitted for analysis. The delivery was estimated to take between 15-20 minutes total, with an average length of ~20mins.

**Procedure.** Students who volunteered to participate in the experiment were handed a coloured slip of paper as they entered the computer lab (~20 computer stations were available at any one time). Upon login to the Blackboard LMS, they were verbally instructed

to click the link with their assigned colour (yellow or green). Instructions for navigation and login were also displayed at the front of the room via projector (see Appendix C). Screen brightness and volume levels were adjusted to 50% at each of the computers. There was no time restriction for completion of the surveys and watching the video; the subjects were given no other restrictions. Students had the freedom to maximise the video to full screen or leave the video at its default setting. They could also pause and re-watch parts of the video if they desired. On average the students took between 15-25 minutes to complete the experiment.

After completing the lesson, students were instructed to logout and were free to leave the computer lab. This process occurred over the course of a day as revision classes occurred in an adjacent room. Between classes when there were no incoming participants each of the computer stations were checked and reset to their starting points. The experiment pages were made available from 9am to 5pm.

Date of experiment: 12/9/14

**Analysis.** The total quiz scores obtained on the pretest and post-test, for both groups were summarised using descriptive statistics including means and standard deviations (SD). The pretest quiz scores were analysed using independent t-tests to establish baseline performance. Age, sex, prior area of study and prior experience of learning anatomy were analysed to check for differences between groups and establish equivalence. A mixed design (split plot) ANOVA was used to test for differences between the groups across time. Subjective ratings were analysed using Wilcoxon Signed-Rank tests for changes within groups, and Mann-Whitney U tests for changes between groups. Effect sizes for the changes were calculated and assessed using Cohen's criteria. One pair of self-rating questions were included as a consistency check. This pair was analysed using a Mantel-Haenszel test of trend. Finally, Jacobson's Reliable Change Index (RCI) was calculated based on quiz scores to assess

improvement statistically and 'clinically'. This measure allows for a categorization of students who made large gains. The data of students who made significant improvements were summarized with descriptive statistics.

## Results

**Participants.** A sample of 83 students in their 2nd year of graduate medicine volunteered for Experiment 1. They were randomised into control and experiment groups. Seven cases were removed due to incomplete data and failure to follow instructions. The final sample size was 76 (control  $n = 36$ , experiment  $n = 40$ ). There were no significant differences in age, sex, prior degree and exposure to learning anatomy between groups. The analysis of each of these covariates follow.

**Age.** A Mann-Whitney U test was run to determine if there were differences in ages between control and experiment groups. Distributions of ages for control and experiment were similar, as assessed by visual inspection. Ages were not statistically significantly different between control ( $M = 25.89$ ,  $SD = 4.21$ ) and experiment ( $M = 24.68$ ,  $SD = 2.93$ ),  $U = 590$ ,  $z = -1.37$ ,  $p = .17$ .

**Sex.** A chi-square test for association (with Yates Continuity Correction) was conducted between sex and groups. All expected cell frequencies were greater than five. Both groups had slightly more males (control  $n=19$ , experiment  $n = 23$ ) than females (control = 17, experiment = 17), but there was not a statistically significant association between sex and group,  $\chi^2(1, n = 76) = 0.03$ ,  $p = .86$ ,  $\phi = -0.47$ .

**Prior area of study.** The SMP is a graduate entry course, therefore, students come from a variety of academic backgrounds. Figure 5 shows that among the nine areas of prior study, the most commonly reported were medical science, and science for both groups. A chi-square test for independence indicated no significant association between group and previous

degree. There was a violation of expected cell counts, 7 cells with counts less than 5,  $\chi^2(6, n = 76) = 8.06, p = .23, \phi = .33$ .

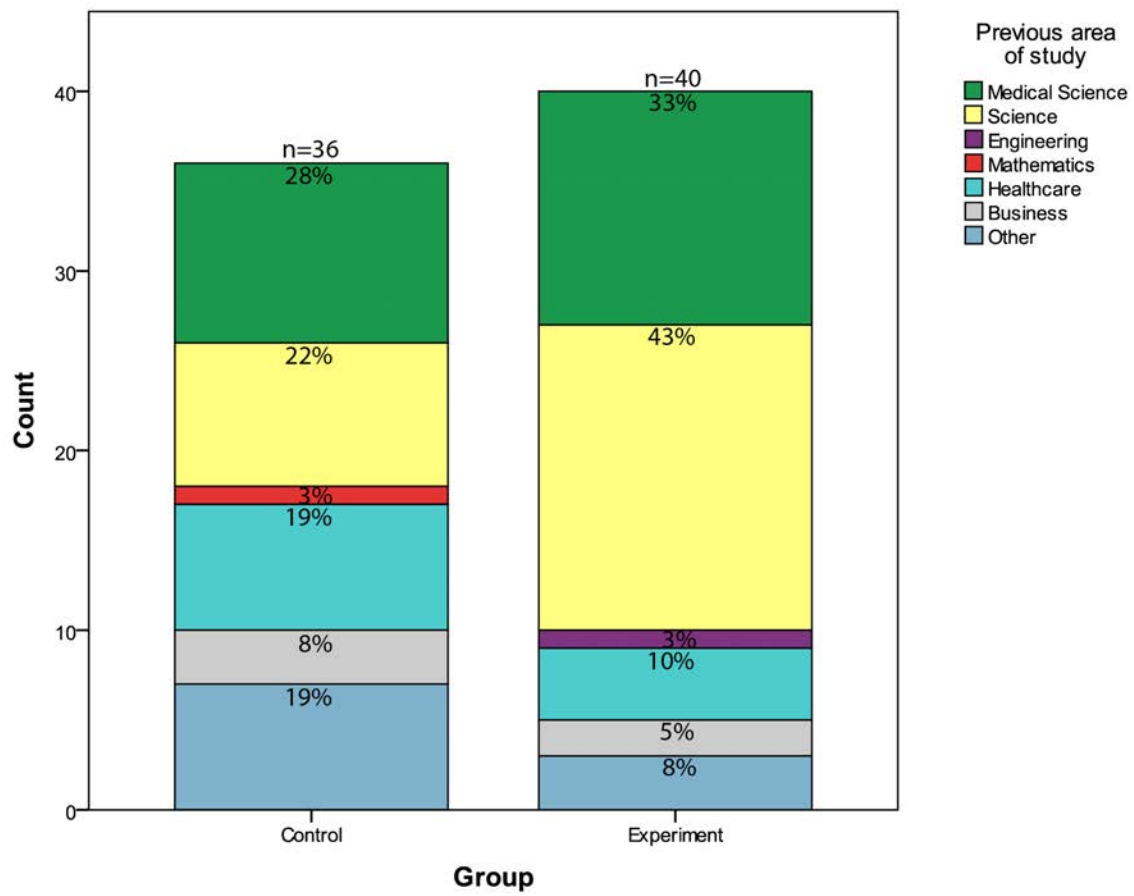
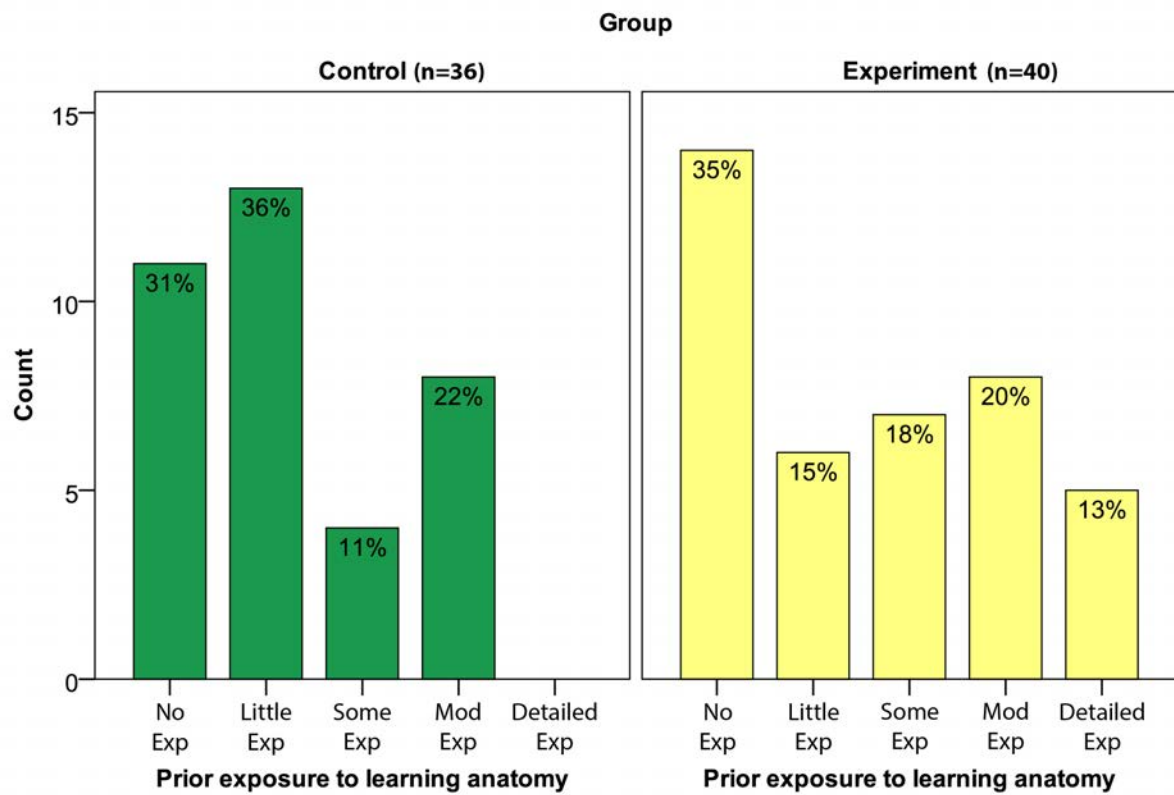


Figure 5. Area of study in which students completed their previous degree for each group.

Technology and computing are omitted due to zero counts.

***Students' prior learning of anatomy.*** A Mann-Whitney U test was run to determine if there were differences in self-rated levels of exposure to learning anatomy between control and experiment groups. The ratings were not similarly distributed for both groups, as assessed by visual inspection (see Figure 6). However, the differences in self-ratings were not statistically significant between control (mean rank = 36.28) and experiment (mean rank = 40.50),  $U = 640.0$ ,  $z = -.86$ ,  $p = .39$ .





*Figure 6.* Students' self-rated level of exposure to learning anatomy prior to entry in the SMP by group.

**Pretest comparisons.**

***Pretest self-ratings.*** Figure 7 shows the frequency of responses across the Likert scale for each of the self-rated questions. The rating distributions for each question were similar between groups and Mann-Whitney U tests indicated no significant differences (Table 3).

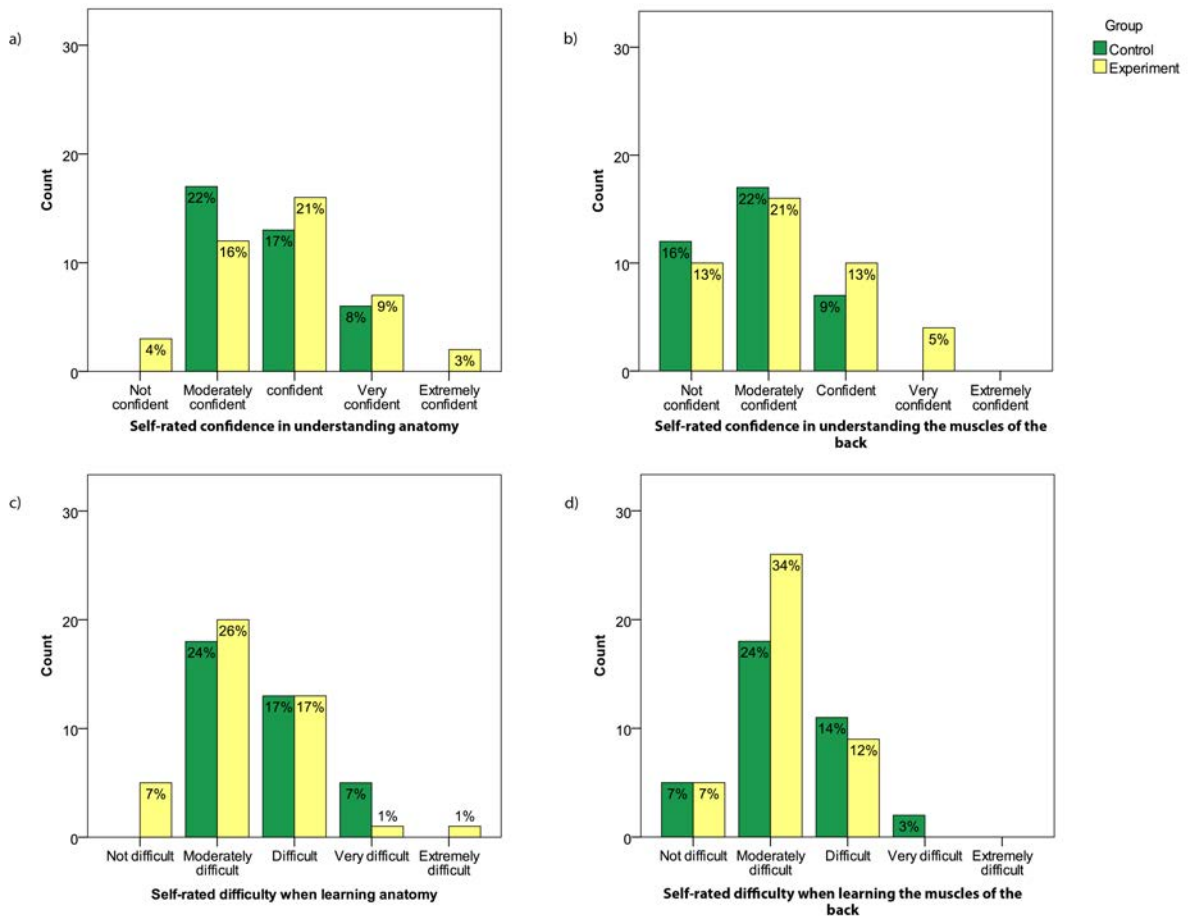


Figure 7. Pretest self-ratings of both groups for: a) Confidence in understanding anatomy b) Confidence in understanding the muscles of the back c) Difficulty when learning anatomy d) Difficulty when learning the muscles of the back.

Table 3

*A comparison of median pretest self-ratings between groups.*

Self-rated:	Median Response		U	z	p
	Control	Experiment			
Confidence					
understanding					
Anatomy	Confident	Confident	659.0	-.68	.50
The muscles	Mod. Confident	Mod. Confident	583.0	-1.52	.13
of the back					
Difficulty					
learning					
Anatomy	Mod. Difficult-	Mod. Difficult	568.0	-1.73	.08
	Difficult				
The muscles	Mod. Difficult	Mod. Difficult	633	-1.02	.31
of the back					

Both groups answered the same Likert questions after the intervention. However, the question asking 'How would you rate the difficulty of learning the muscles of the back?' was modified in the posttest; the word *difficulty* was replaced with *ease*.

Also, two open ended questions asked students to specify what they found difficult about learning anatomy and the muscles of the back. The most frequent terms occurring were memorisation and volume of content.

**Pretest quiz results.** An independent-samples t-test was run to determine if there were differences in pretest scores between groups. Data are mean  $\pm$  standard deviation, unless otherwise stated. There were 36 control and 40 experiment participants. There were two outliers in the data, as assessed by inspection of a boxplot (see Figure 8); one case in the control group scoring one on the pretest and one case in the experiment group scoring six on the pretest. The two data points were included in the analysis. Pretest scores for each group were not normally distributed, as assessed by Shapiro-Wilk's test ( $p < .05$ ), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p = .98$ ). The difference between pretest scores for the control group ( $3.4 \pm 1.0$ ) and experiment group ( $3.6 \pm 1.0$ ), were not statistically significant (95% CI, -0.67 to 0.25),  $t(74) = -0.91$ ,  $p = 0.37$ .

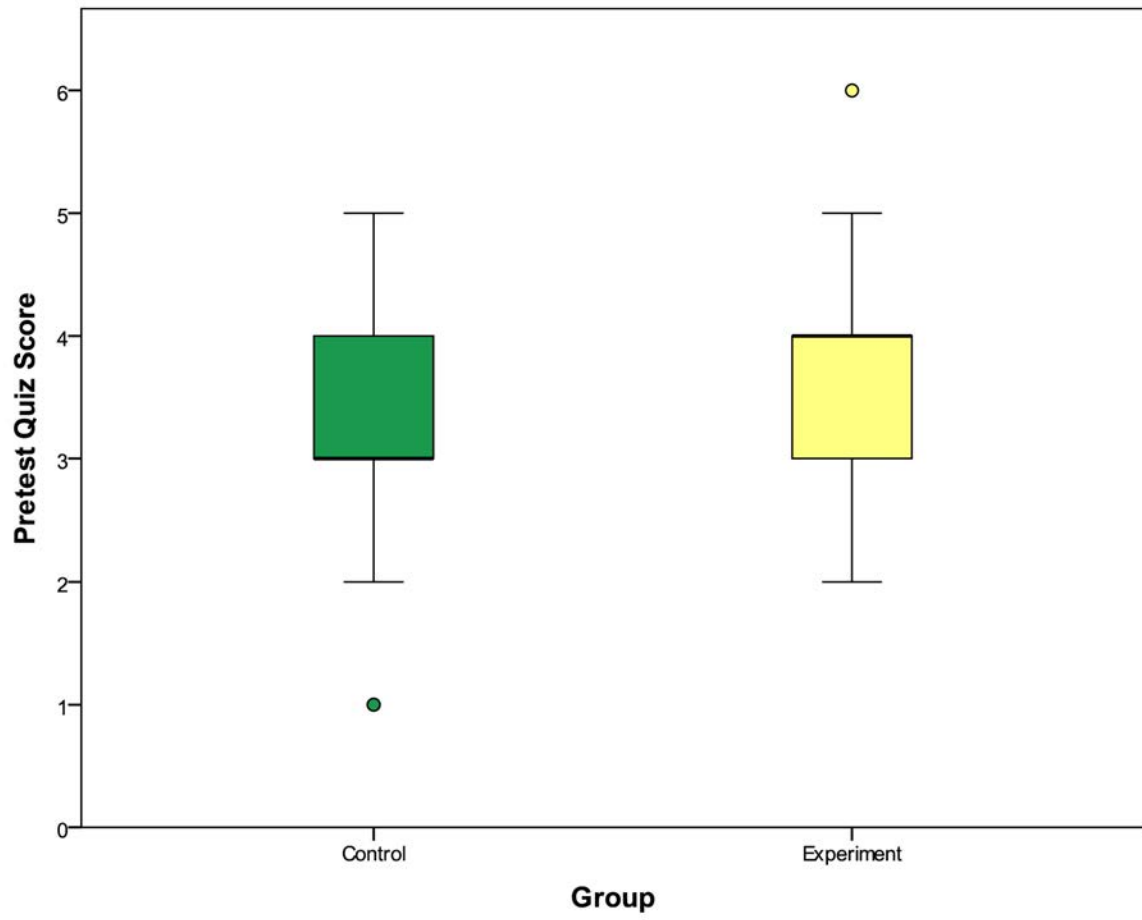


Figure 8. Pretest Quiz Scores for both groups.

**Posttest comparisons.**

**Changes in self-ratings.** Table 4 shows a summary of the shifts in self-ratings for both groups including z, p and r values. Wilcoxon Signed-Rank Tests showed significant changes in confidence ratings when learning about the muscles of the back in both groups. After exposure to the multimedia presentation the change in confidence yielded effect sizes of  $r=0.4$  for the control group and  $r=0.3$  for the experiment group.

Table 4

a) Shifts in self-rating scores for the control group (n=36).

Self-rated:	Rating change			z	p	r
	Increase	Unchanged	Decrease			
Confidence understanding						
Anatomy		31	5	-2.24	.03	-0.26
Muscles of the back	21	12	3	3.64	<.001	0.43
Difficulty learning						
Anatomy	2	25	9	-2.11	.04	0.25

b) Shifts in self-rating scores for the experiment group (n=40)

Self-rated:	Rating change			z	p	r
	Increase	Unchanged	Decrease			
Confidence understanding						
Anatomy <sup>a</sup>	3	34	2	.71	.48	0.08
Muscles of the back	19	17	4	2.91	.004	0.33
Difficulty learning						
Anatomy	4	33	3	.38	.71	0.04

<sup>a</sup>n=39



For the control group, self-rated confidence when learning about anatomy changed significantly with 5 individuals indicating decreased confidence and the remainder (n=31) remaining unchanged; an effect size of  $r=-0.3$ . Similarly, ratings of difficulty when learning about anatomy changed significantly with 9 participants indicating a decrease in difficulty, 2 indicating an increase and the remainder (n=25) unchanged; an effect size of 0.2.

The experiment group, however, did not have a significant change in confidence when learning about anatomy with an increase in confidence for 3 participants, a decrease for 2 and the remainder (n=34) unchanged; an effect size of  $r=0.08$ . Likewise, ratings of difficulty when learning about anatomy did not change significantly with 4 participants indicating an increase in difficulty, 3 indicating a decrease and the remainder (n=33) the same; an effect size of  $r=0.04$

The pretest included a self-rated question of difficulty when learning about the muscles of the back. In the post test this question was modified by replacing the term 'difficult' with 'ease' at each point in the scale. Figure 9 shows the frequency of responses for this pair of questions. A Mantel-Haenszel chi-square test was run to test if there was a linear trend between the two variables. The pretest question had a scale from 1 to 5 with lower ratings indicating low difficulty. The post test question also had a scale of 1 to 5, however low ratings indicated high difficulty i.e., 'not easy' and higher ratings indicating low difficulty 'extremely easy'. For the experiment group, the Mantel-Haenszel test of trend showed a statistically significant linear association between ratings of difficulty and ratings of ease, Experiment:  $\chi^2(1) = 16.48$ ,  $p < .0005$ ,  $r = -.65$ . The control group did not show a significant linear association Control:  $\chi^2(1) = 0.69$ ,  $p = 4.06$ ,  $r = -.14$ .

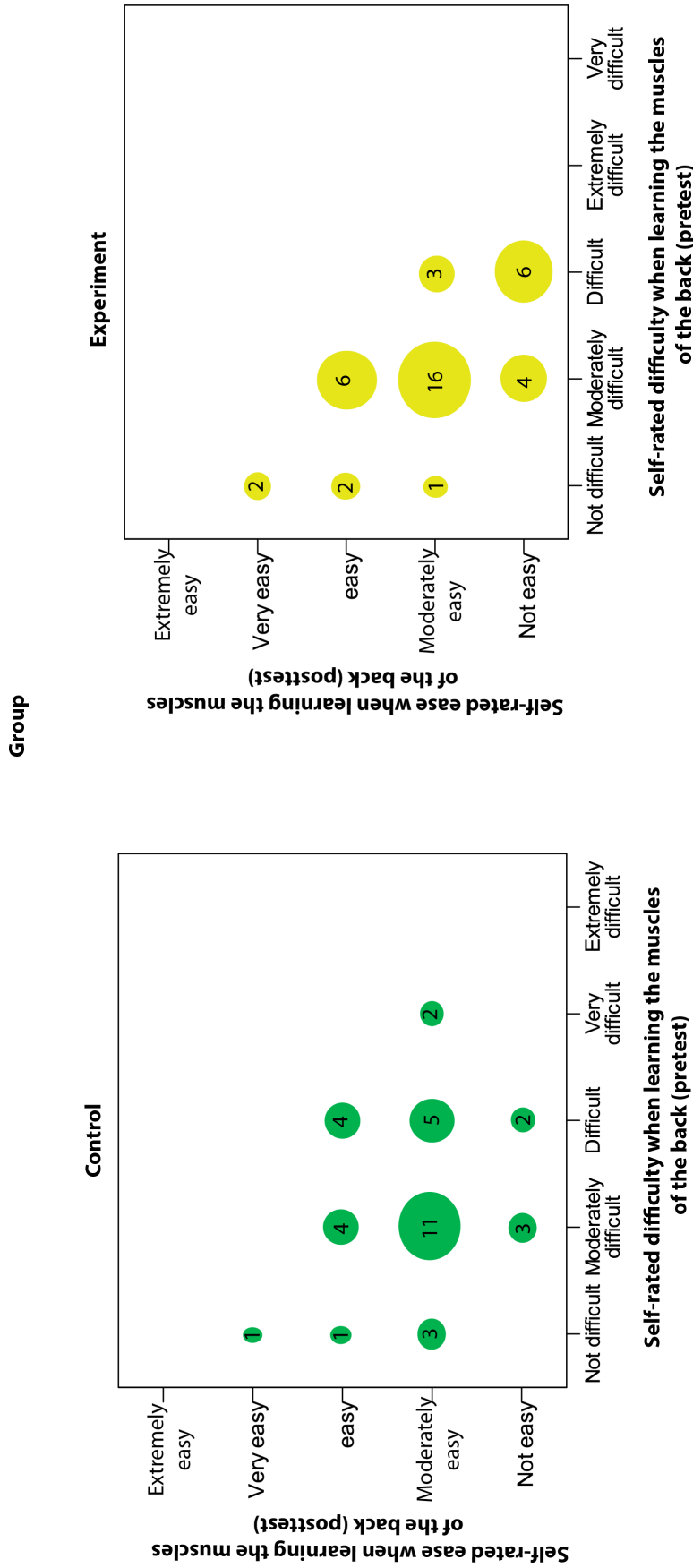


Figure 9. Frequency of responses across scales for self-rated difficulty when learning the muscles of the back (pretest) and self-rated ease when learning the muscles of the back (posttest).

**Changes in quiz results.** The mean and standard deviation of posttest scores were:  $4.5 \pm 0.9$  for the control and  $4.8 \pm 0.6$  for the experiment.

A mixed between-within subjects analysis of variance was conducted to assess the impact of the two interventions on participant quiz scores, across two time periods (pretest and posttest). There was no significant interaction between group and time, Wilks Lambda = .997,  $F(1,74) = .25$ ,  $p = .62$ , partial eta squared = 0.003.

There was a significant main effect for time, Wilks Lambda = .46,  $F(1,74) = 85.42$ ,  $p = .000$ , partial eta squared = .54, with both groups showing an increase in quiz score across the two time periods. The main effect comparing the two methods of instruction was not significant,  $F(1,74) = 2.77$ ,  $p = .100$ , partial eta squared = 0.04, suggesting no difference in the effectiveness of the two multimedia presentations.

Figure 10 shows the gain scores of the participants in both groups. The experiment group shows a higher frequency of large gain scores.

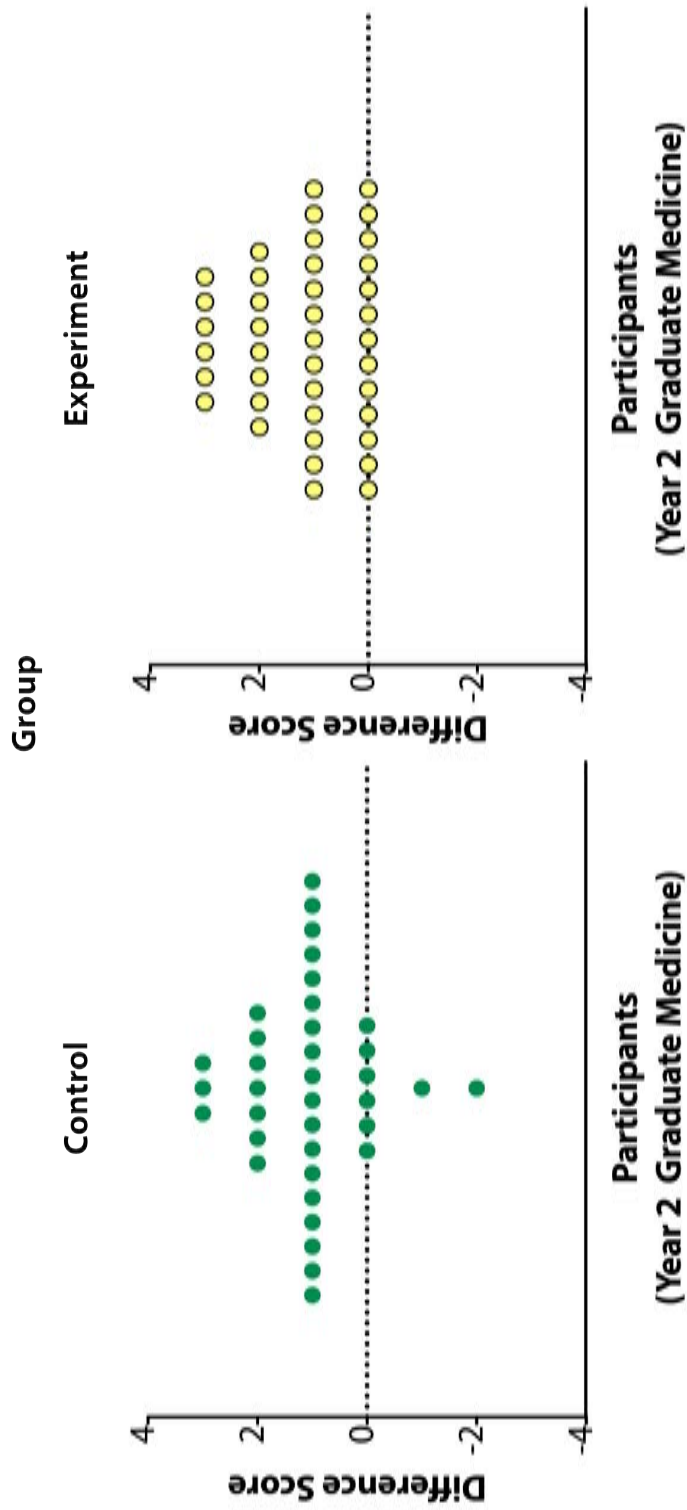


Figure 10. Gain scores for the control and experiment groups.

Jacobson's Reliable Change index was calculated to categorise the subset of students who made large gains in the quiz. The value calculated as an indicator of 'clinically' significant change was 4.7 on the posttest. There were three participants in control group with a RCI over 1.96 of those 2 scored above 4.7. There were 14 participants in the experiment group with a RCI over 1.96, and of these 13 scored above 4.7. The pretest and posttest scores and the boundary of 'clinically' significant change are shown in Figure 11. The characteristics of these participants is explored and summarised in the Improvement Profiles section.

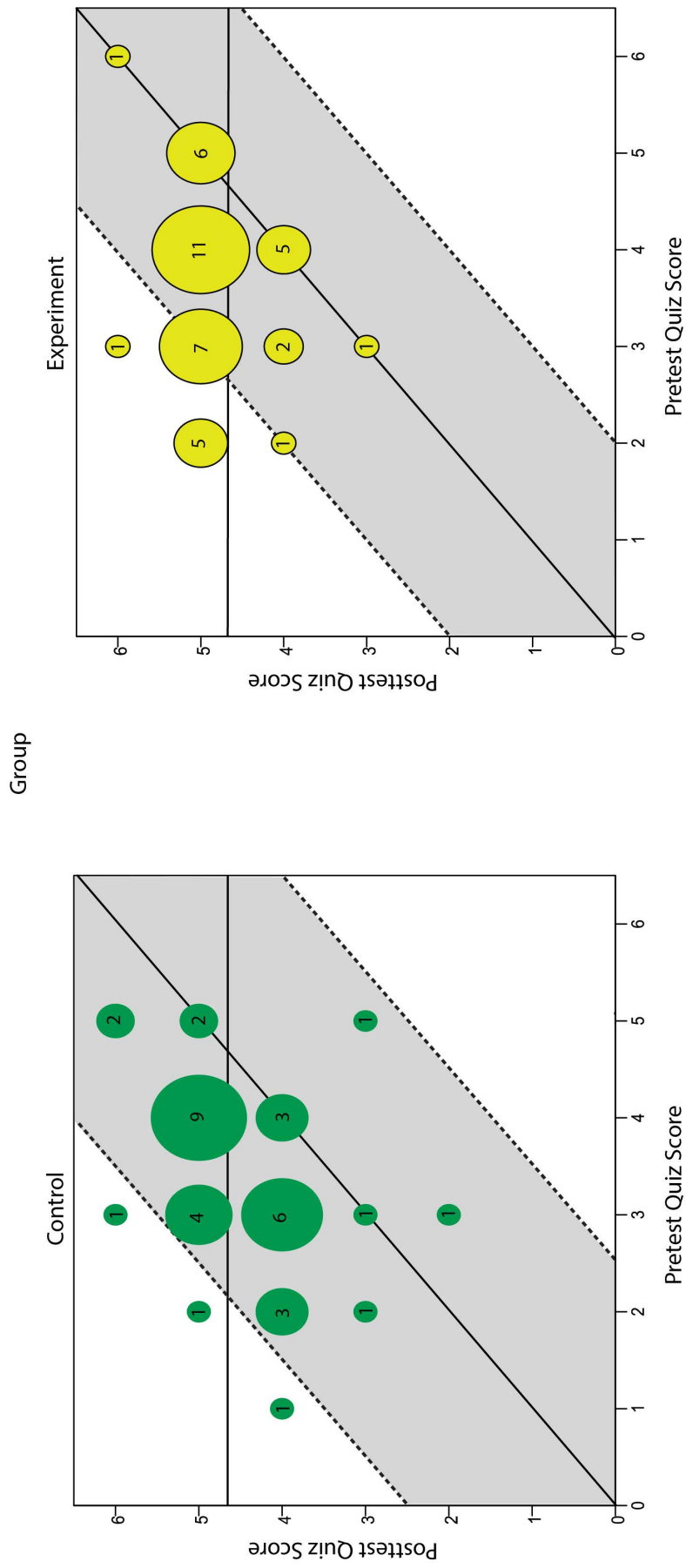


Figure 11. Pretest and posttest quiz scores for both groups. The reliable change index band is indicated by the dotted diagonal lines.

**Improvement profiles.**

**Control group.** Of the 36 participants in the control group, two (5.5%) had a reliable change index above 1.96 and a posttest score above the significant cutoff 4.7. A summary of their characteristics compared to the group is shown in Table 5.

The 2 participants' answers to the self-rated questions dominated the low and mid range of the Likert scales. Figures 12 & 13 show the answers selected for each self-rated question and how they compare to the group. For all questions, except pretest confidence in MOB and posttest ease in MOB, their answers corresponded to the mode of the group.

Table 5

*Students with a significant RCI compared to those without between groups*

Participants	n	Sex		Age M±SD	Prior Degree (mode)	Anat Exp (median)	Mean Quiz Score	
		Male	Female				Pretest	Posttest
<b>Experiment</b>								
Non Sig	27	17(63.0%)	10(37.0%)	25.0±3	Science	Little	4.11±0.8	4.67±0.62
Sig RCI	13	6(46.2%)	7(53.8%)	25.0±3.0	Science	Some	2.62±.51	5.08±.28
<b>Control</b>								
Non sig	34	19(55.9%)	15(44.1%)	26±4	Med Sci	Little	3.47±0.99	4.41±0.86
Sig RCI	2	0	2	25±4.0	Med Sci	Little	2.5±.71	5.5±0.71



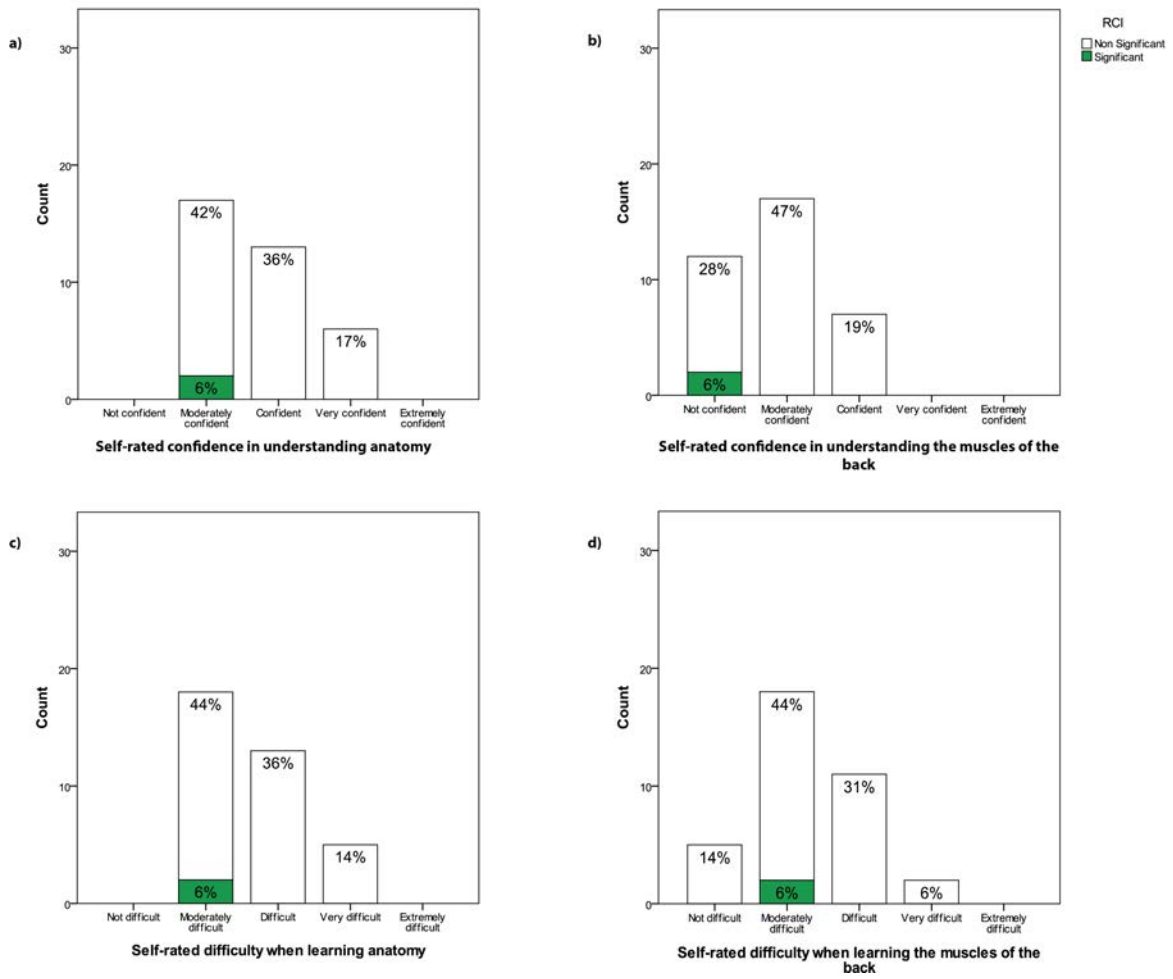


Figure 12. Distribution of pretest self-rated questions for the control group.

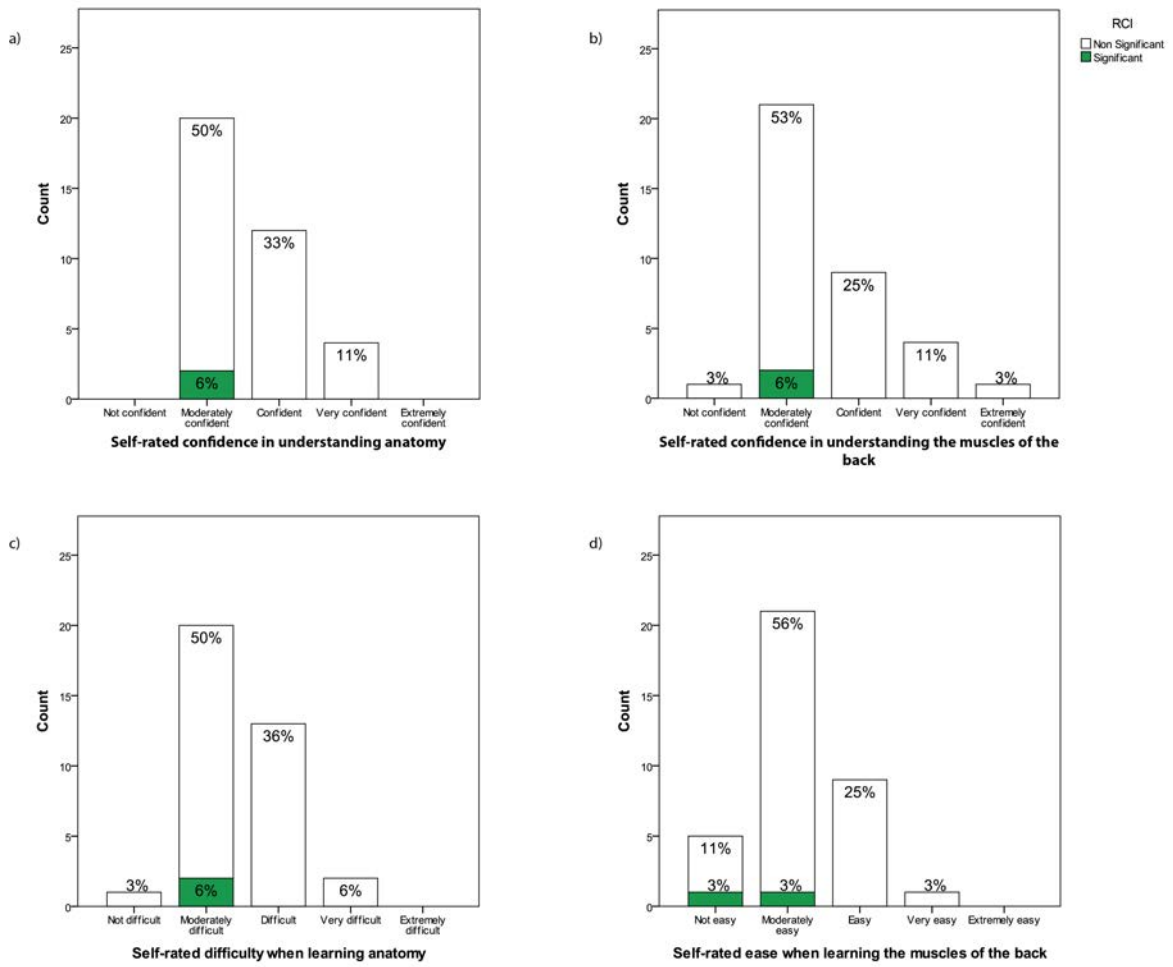


Figure 13. Distribution of posttest self-rated questions for the control group.

**Experiment group.** Of the 40 participants in the experiment group, 13 (32.5%) had a reliable change index above 1.96 and a post test score above the significant cutoff 4.7. Figure 14 shows the profiles of participants in a parallel plot. Their characteristics are heterogeneous across the variables of age, sex, prior degree and level of exposure to learning anatomy. The summary statistics in table 5 show that the subgroup with significant RCI's closely represented the group. In addition, their answers to the self-rated questions are distributed across the scales as seen in Figures 14 and 15.

\

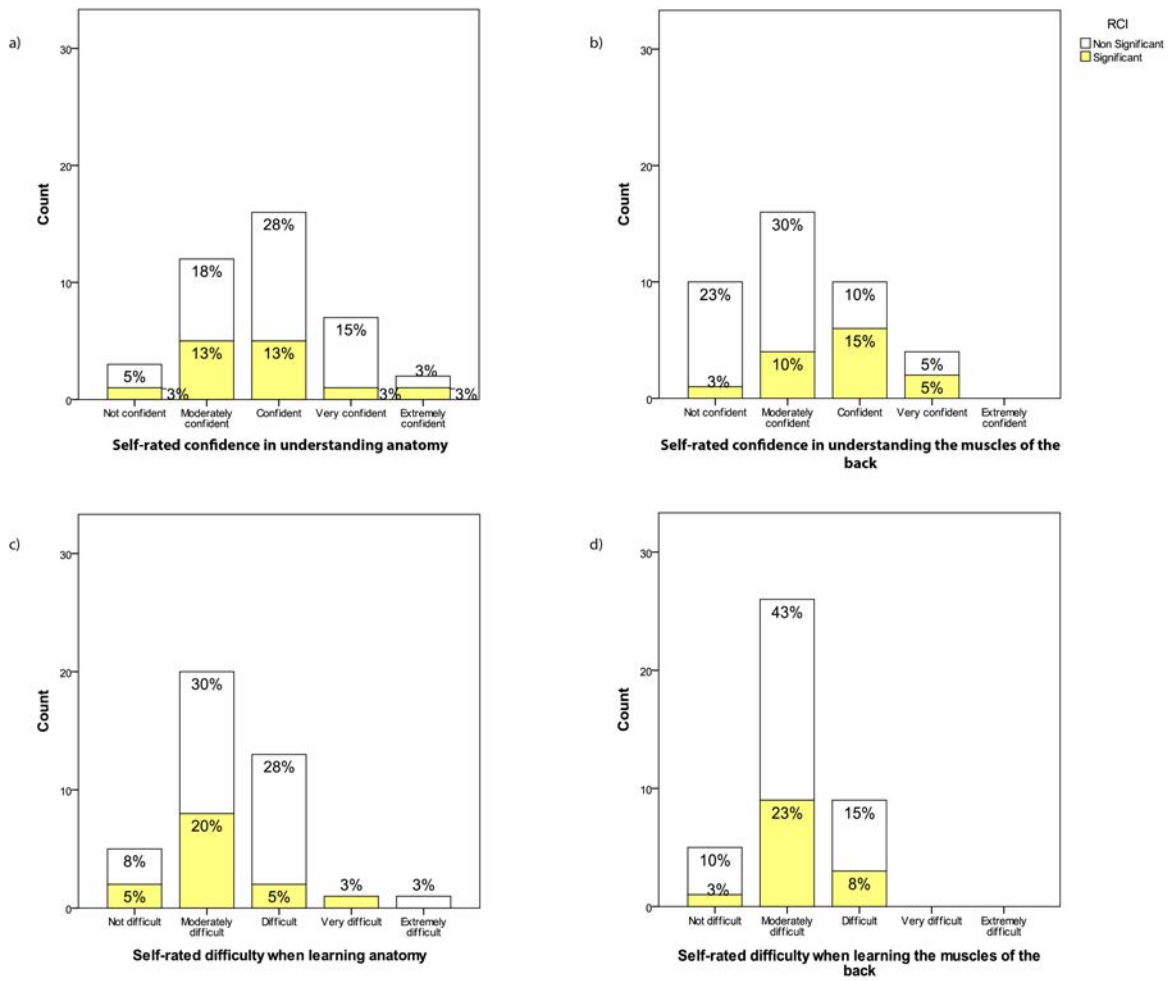


Figure 14. Distribution of pretest self-ratings for the experiment group.

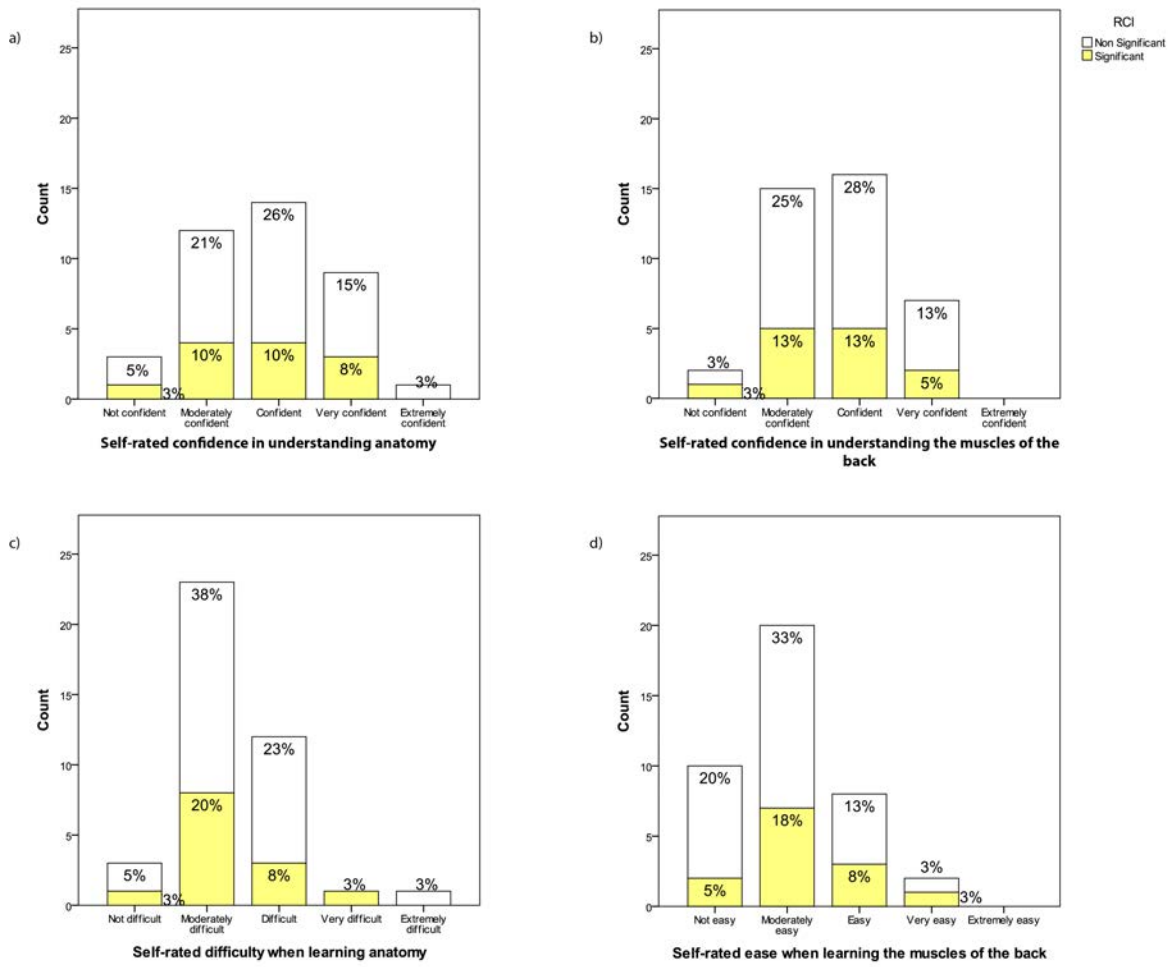


Figure 15. Distribution of posttest ratings for the experiment group.

## Experiment 2

### Methods

**Participants.** A total of 123 students in their 1<sup>st</sup> year of the Sydney Medical Program (SMP) agreed to participate in an evaluation of learning and teaching multimedia for anatomy. Students who did not follow instructions and those missing data for pre and/or post test were excluded leaving a final sample of 116 students. The participants were randomly assigned to control and experiment conditions using rules created in the LMS. Table 6 shows participant age and sex by group.

Table 6

*Age and sex of participants in each group*

Group	Sex			Age		
	Male	Female	Other	Mean	SD	Range
Control (n=60)	33	26	1	24.3*	3.1	13
Experiment (n=56)	36	20	0	24.8	3.5	15

\*n=57

**Sampling procedure.** A population of 297 students in their 1st year of the Sydney Medical Program (SMP) were invited to participate in an evaluation of learning and teaching multimedia for anatomy. An email bulletin was sent to the cohort announcing an anatomy video trial which involved: answering questions, watching a 5 minute video and completing a follow-up test after 7 days. They were informed that the trial would take place on the day of their scheduled practical session in an adjacent room. There were several practical sessions to which students were assigned in the course.

On the day of the experiment one of the senior lecturers (supervising this study) and myself stood by the entry of the experiment location and invited students to participate in our video evaluation. Due to the multiple practical sessions, some students participated before their lesson, while others took part afterwards. Forty-one percent (N=123) of the target population volunteered. The data were collected electronically in the university LMS - Blackboard. Each group had a page designated to them and the availability of the page was restricted to the day of the experiment. The students accessed the LMS on campus in a computer lab.

The scope of the study fell within the disciplines' ethics approval for course evaluations. Students provided verbal informed consent and implicit consent when beginning the experiment. They could withdraw or elect to not answer questions if they wished.

**Sample size, power and precision.** Calculations were done as per Experiment 1. The total sample size of 116 participants gives sufficient power to detect effect sizes above  $d=0.6$ .

**Measures and covariates.** Identical to experiment 1, with one addition. The experimental media was taken from an online lecture from earlier in the year (Introduction to the vertebral column). As a result, there was a question asking students if they had seen the



lecture. This question was included in analysis as a covariate since many students would be watching the lecture for a 2<sup>nd</sup> or even 3<sup>rd</sup> time.

**Experimental design.** The design of experiment 2 was identical to experiment 1. The only change was that students were randomly assigned to groups based on rules created in the LMS and the extra question mentioned previously.

**Materials.** The materials used were digital videos embedded in the University LMS. All questionnaires were also digitally administered.

**Apparatus.** The apparatus used to conduct the experiments included 40 x 21.5 inch iMac computers, 20 x Sennheiser HD380 Pro and 20 x Sennheiser HD280 Pro headphones. Instructions for beginning the experiment were displayed via a projector. Figure 4 shows the configuration of computers where the experiment was conducted.

**Experimental manipulation.** Identical to experiment 1.

**Procedure.** Students who volunteered to participate in the experiment were randomly assigned to groups via a randomising rule in blackboard. Instructions for navigation and login were also displayed at the front of the room via projector (see Appendix D). Screen brightness and volume levels were adjusted to 50% at each of the computers. There were three sections the students needed to complete: the pretest, watching the video, and the immediate posttest. After each part, they confirmed completion by selecting a button which read 'mark reviewed'. Only after selecting this option would the next part appear. There was no time limit for completion of the surveys and watching the video; the participants were given no other restrictions. Students had the freedom to maximise the video to full screen or leave the video at its default setting. They could also pause and re-watch parts of the video if they desired. On average the students took between 15-25 minutes to complete the experiment.

After completing the experiment, students were instructed to logout and were free to leave the computer lab. This process occurred over the course of a day as revision classes occurred in an adjacent room. Between classes when there were no incoming participants each of the computer stations were checked and reset to their starting points. The experiment pages were made available from 9am to 5pm.

Date of experiment: 06/11/14

**Analysis.** The total quiz scores obtained on the pretest and post-test, for both groups were summarised using descriptive statistics including means and standard deviations (SD). The pretest quiz scores were analysed using independent t-tests to establish baseline performance. Age, sex, prior area of study and prior experience of learning anatomy were analysed to check for differences between groups and establish equivalence. A mixed design ANOVA, with 'Seen previous lecture' as a covariate, was used to test for differences between the groups across time. Subjective ratings were analysed using Wilcoxon Signed-Rank tests for changes within groups, and Mann-Whitney U tests for changes between groups. Effect sizes for the changes were calculated and assessed using Cohen's criteria. One pair of self-rating questions were included as a consistency check. This pair was analysed using a Mantel-Haenszel test of trend. Finally, Jacobson's Reliable Change Index (RCI) was calculated based on quiz scores to assess improvement statistically and 'clinically'. This measure allows for a categorization of students who made large gains. The data of students who made significant improvements were summarized with descriptive statistics.

## Results

**Participants.** A sample of 123 students in their first year of graduate medicine volunteered for experiment 2. Cases with incomplete quiz score data were removed leaving a total of 116 participants. They were randomised into control (n=60) and experiment (n=56)

groups. There were no significant differences in age, sex, prior degree and exposure to learning anatomy between groups. The analysis of each of these covariates follows.

**Age.** A Mann-Whitney U test was run to determine if there were differences in ages between control and experiment groups. Distributions of ages for control and experiment were similar, as assessed by visual inspection. Median ages were not statistically significantly different between control (24, n=57) and experiment (24, n=56)  $U = 1703.5$ ,  $z = .62$ ,  $p = .53$ .

**Sex.** A chi-square test for association was conducted between sex and groups. Two cells had expected frequencies below five and the minimum expected count was .48. There was no statistically significant association between sex and group,  $\chi^2(2, n = 116) = 1.78$ ,  $p = .41$ , Cramer's  $V = .12$ .

**Prior area of study.** The students in this experiment represented a variety of prior academic backgrounds as was also found in experiment 1. Figure 16 shows that among the nine areas of prior study, the most commonly reported was science for both groups. A chi-square test for independence indicated no significant association between group and previous degree. There was a violation of expected cell counts, 6 cells with counts less than 5,  $\chi^2(6, n = 116) = 7.07$ ,  $p = .31$ ,  $\phi = .33$  Cramer's  $V = .25$

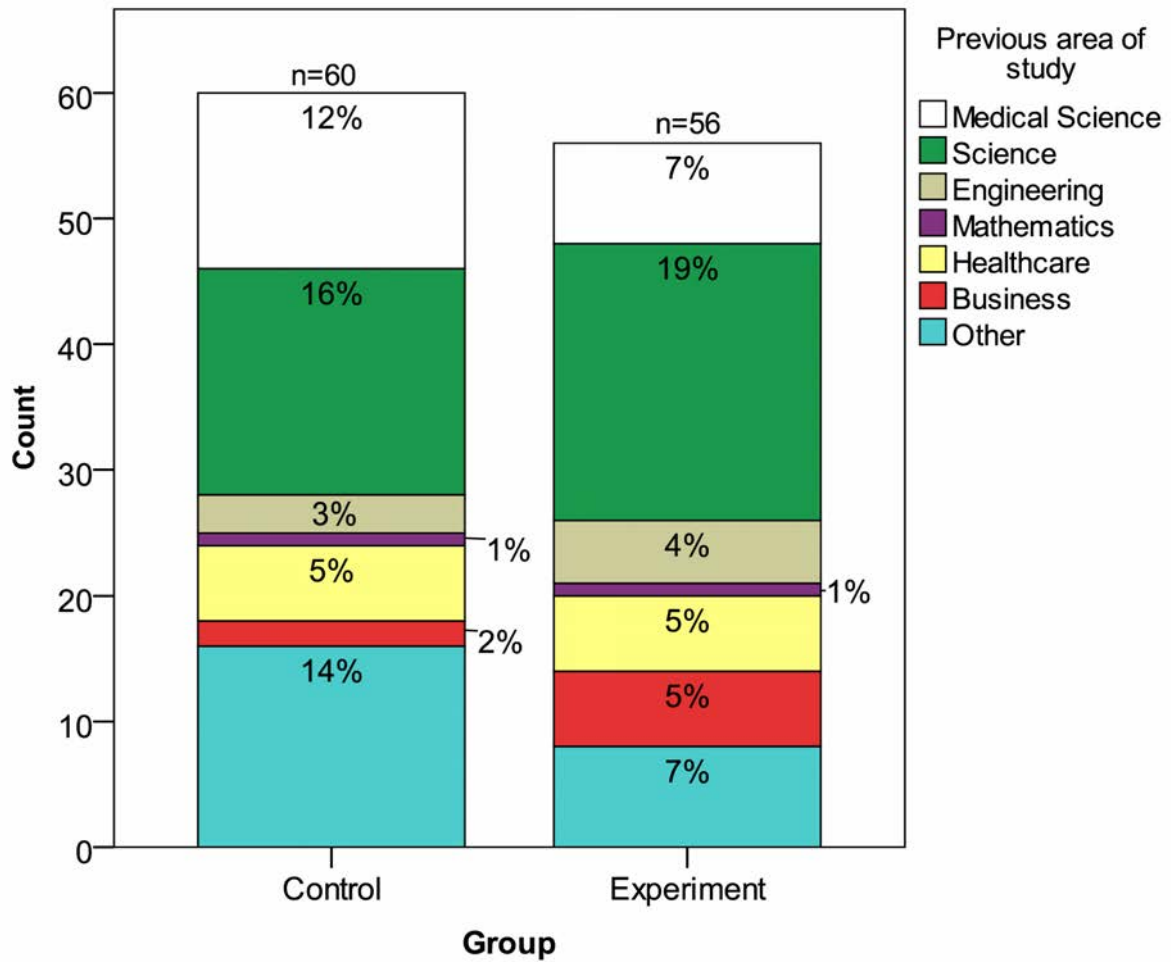


Figure 16. Area of study in which students completed their previous degree for each group. Technology and computing are omitted due to zero counts.

***Students' prior learning of anatomy.*** A Mann-Whitney U test was run to determine if there were differences in self-rated levels of exposure to learning anatomy between control and experiment groups. The ratings were similarly distributed for both groups, as assessed by visual inspection (see Figure 17). The differences in self-ratings were not statistically significant between control (mean rank = 59.25) and experiment (mean rank = 56.69),  $U = 1578.0$ ,  $z = -.43$ ,  $p = .67$ .

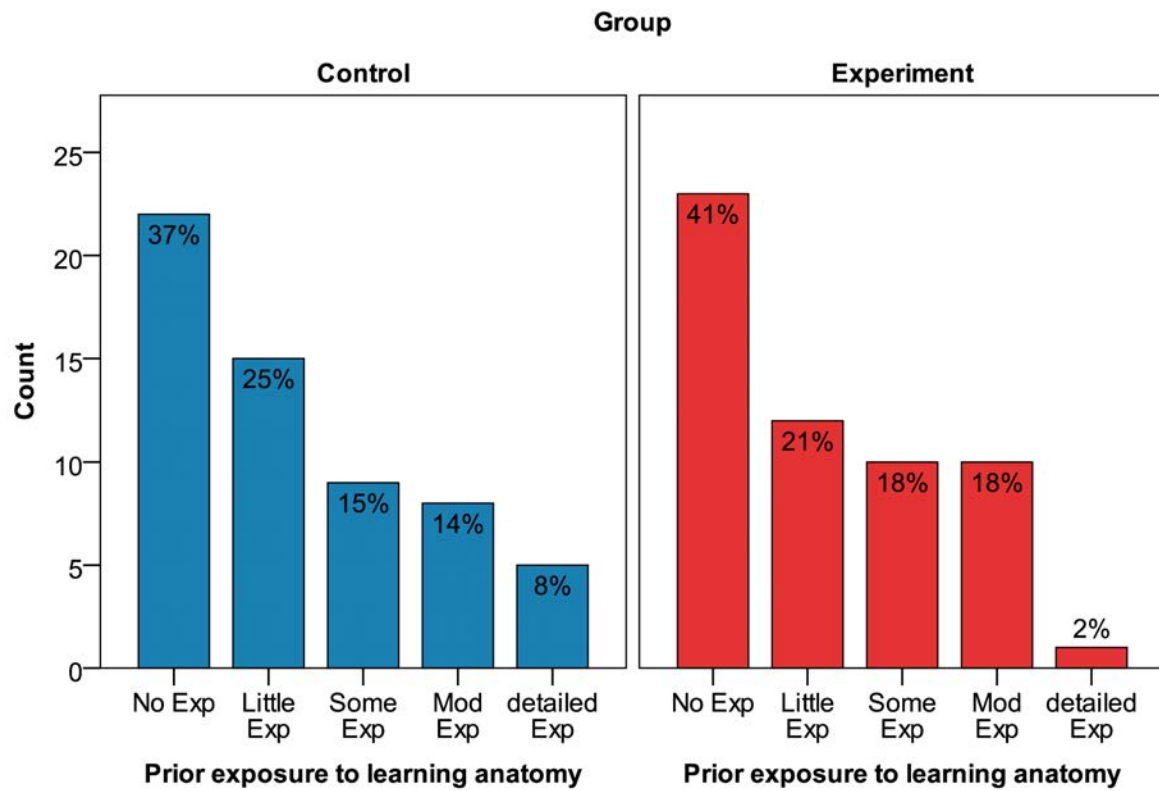


Figure 17. Students' self-rated level of exposure to learning anatomy prior to entry in the SMP by group.

**Lecture watched.** The multimedia lesson in this experiment was a 5 minute segment taken from a full length lecture. The lecture had been presented earlier in the year for this cohort of students. Therefore, an additional question was included which asked students if they had seen the online lecture 'Introduction to the vertebral column'. A chi-square test for association was conducted between group and if participants had seen the lecture. All expected cell frequencies were greater than five. There was not a statistically significant association between group and having seen the lecture,  $\chi^2(1) = 0.43$ ,  $p = .51$ ,  $\phi = 0.06$ ,  $p = .51$ . (Control group: seen lecture yes=73.3%, no= 26.7%. experiment: yes=78.6%, no=21.4%)

### **Pretest comparisons.**

**Pretest self-ratings.** Participants were asked 4 Likert scale questions about their level of confidence and perceived difficulty when learning anatomy and the muscles of the back. Mann-Whitney U tests were run to determine if there were differences between the groups on pretest ratings of: Confidence in learning anatomy, confidence in learning muscles of the back, difficulty learning anatomy and difficulty learning the muscles of the back. For both groups, all questions had similar distributions, as shown in Figure 18. Median responses and U, z, and p values are shown in Table 7.

They were also asked 2 open ended questions which probed for answers about what they found difficult about learning anatomy and what they found difficult when learning about the muscles of the back. The most frequent terms occurring were memorization and volume of content.

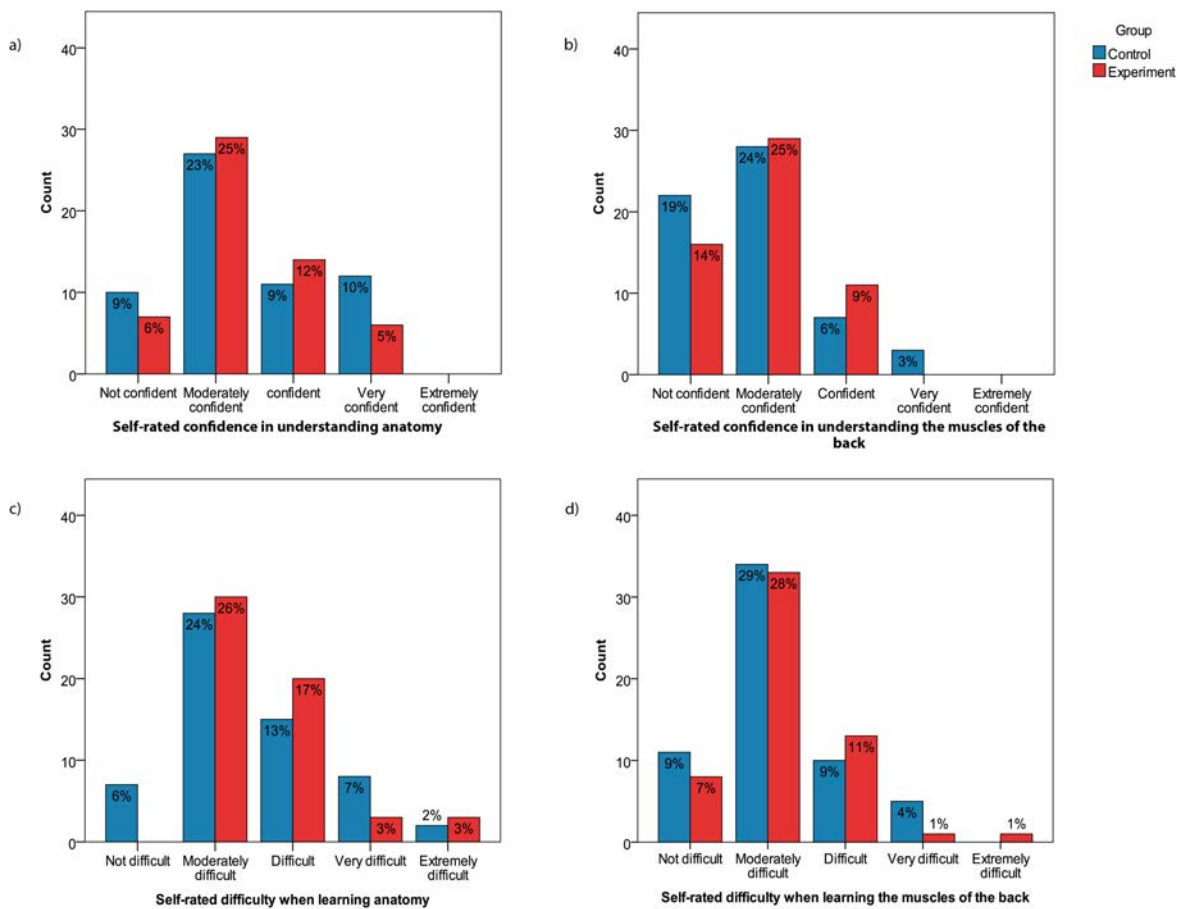


Figure 18. Pretest self-ratings of both groups for: a) Confidence in understanding anatomy b) Confidence in understanding the muscles of the back c) Difficulty when learning anatomy d) Difficulty when learning the muscles of the back.



Table 7

*A comparison of median pretest self-ratings between groups*

Variable	Median Response		U	z	p
	Control	Experiment			
Confidence					
understanding					
Anatomy	Confident	Confident	1635	-.26	.79
The muscles	Confident	Confident	1808.5	.77	.44
of the back					
Difficulty					
learning					
Anatomy	Moderately	Moderately	1819.0	.84	.40
	Difficult	Difficult			
The muscles	Moderately	Moderately	1735.0	.34	.73
of the back	Difficult	Difficult			

**Pretest quiz results.** A quiz with 6 questions (including one 'transfer' style question) was administered to establish existing knowledge of anatomy and muscles of the back (max score=6). An independent-samples t-test was run to determine if there were differences in pretest scores between groups before participants viewed the multimedia lesson. Data are mean  $\pm$  standard deviation, unless otherwise stated. There were 60 control and 56 experiment participants. There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box (See Figure 19). Pretest scores for each group were not normally distributed, as assessed by Shapiro-Wilk's test ( $p < .05$ ), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p = .31$ ). The difference between pretest scores for the control group ( $3.8 \pm 1.3$ ) and treatment group ( $3.6 \pm 1.4$ ), were not statistically significant  $t(114) = 840.0$ ,  $p = .40$ .

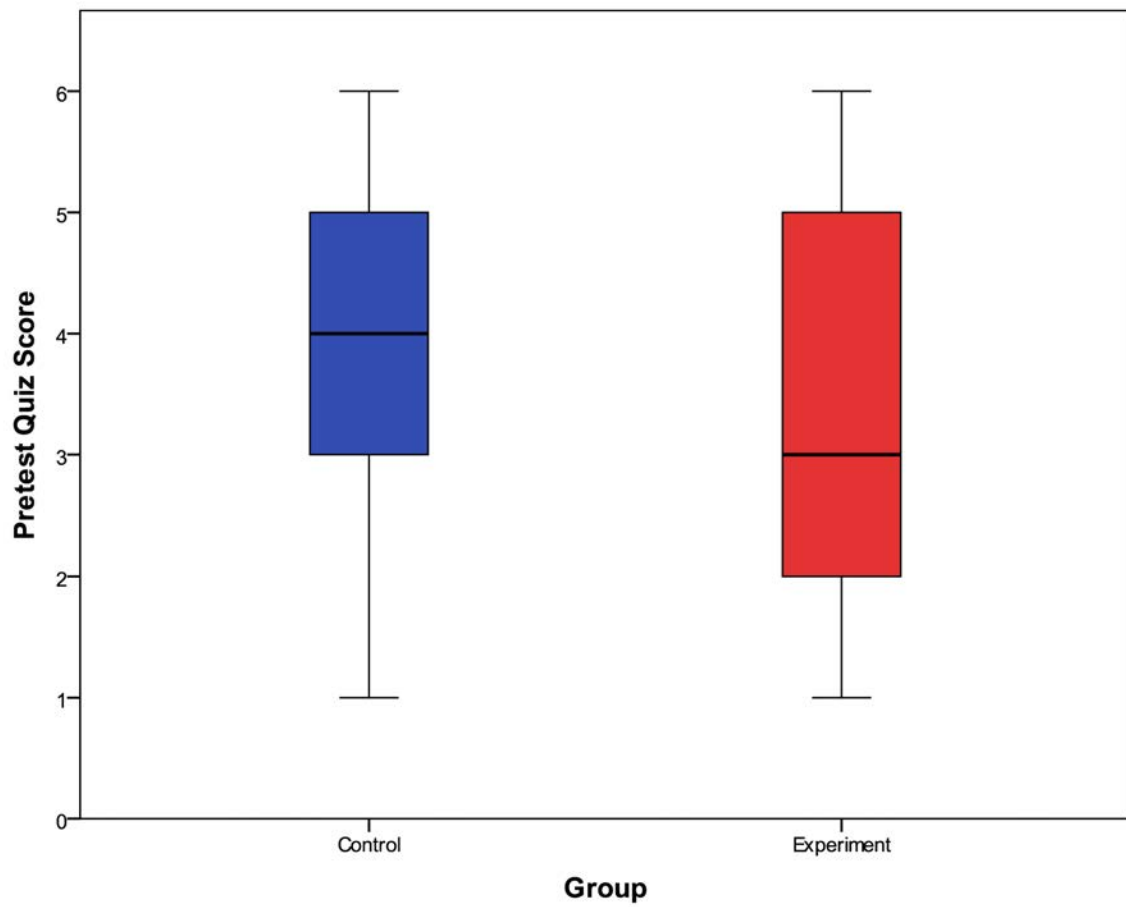


Figure 19. Pretest quiz scores for both groups

**Posttest comparisons.**

***Changes in self-ratings.*** A summary of shifts in self-ratings is shown in Table 8.

Wilcoxon Signed-Rank Tests showed significant changes in confidence ratings when learning about the muscles of the back in both groups. After exposure to the multimedia presentation the change in confidence yielded effect sizes of  $r=0.38$  for the control group and  $r=0.53$  for the experiment group. For both groups, there were no significant changes in self-rated confidence when learning about anatomy. The changes in self-rated difficulty when learning about anatomy were not significant for the control group, who had a small effect size ( $r = -0.10$ ). However, the changes in difficulty rating were significant for the experiment group who had a medium effect size of  $r = -0.24$ .

Table 8

a) Shifts in self-rating scores for the control group (n=60).

Self-rated:	Rating change			z	p	r
	Increase	Unchanged	Decrease			
Confidence understanding						
Anatomy	6	48	6	0	1.0	0
The muscles of the back	27	31	2	-4.16	<0.05	-0.38
Difficulty learning						
Anatomy	4	47	9	-1.13	.26	-0.10

b) Shifts in self-rating scores for the experiment group (n=56)

Self-rated:	Rating change			z	p	r
	Increase	Unchanged	Decrease			
Confidence understanding						
Anatomy	5	48	3	-.71	.48	-0.07
The muscles of the back	35	21	0	-5.65	<0.05	0.53
Difficulty learning						
Anatomy	3	40	13	-2.56	.01	-0.24

A Mantel-Haenszel chi-square test was run to test if there was a linear trend between self-rated difficulty when learning the muscles of the back (pretest) and self-rated ease when learning the muscles of the back (posttest). For both groups, the Mantel-Haenszel test of trend showed a statistically significant linear association between ratings of difficulty and ratings of ease, Experiment:  $\chi^2(1) = 11.33$ ,  $p < .0005$ ,  $r = -.44$ , Control:  $\chi^2(1) = 4.32$ ,  $p < 0.05$ ,  $r = -.28$ . Ratings of high difficulty in the pretest were associated with ratings of low ease in the post test (see Figure 20).

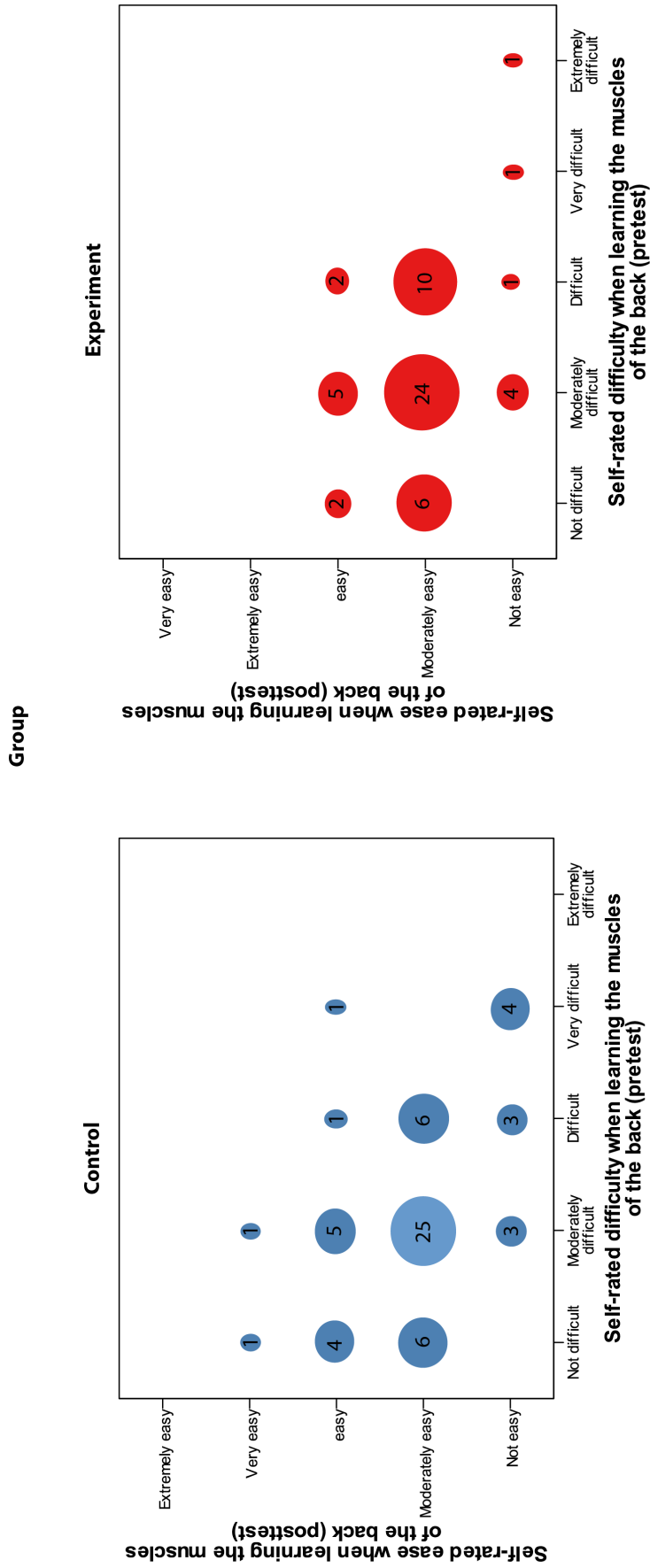


Figure 20. Frequency of responses across scales for self-rated difficulty when learning the muscles of the back (pretest) and self-rated ease when learning the muscles of the back (posttest).

**Changes in quiz results.** The mean and standard deviation of posttest scores were:  $4.9 \pm 1.2$  for the control and  $5.0 \pm 0.1$  for the experiment (See Figure 21 for distribution of gain scores between groups).

A mixed between-within subjects analysis of variance was conducted to assess the impact of the two interventions on participant quiz scores, across two time periods (pretest and posttest). There was no significant interaction between group and time, Wilks Lambda = .99,  $F(1,114) = 1.56$ ,  $p = .22$ , partial eta squared = 0.01.

There was a significant main effect for time, Wilks Lambda = .57,  $F(1,114) = 85.89$ ,  $p = .000$ , partial eta squared = .43, with both groups showing an increase in quiz score across the two time periods. The main effect comparing the two methods of instruction was not significant,  $F(1,114) = .04$ ,  $p = .84$ , partial eta squared = 0.0, suggesting no difference in the effectiveness of the two multimedia presentations.



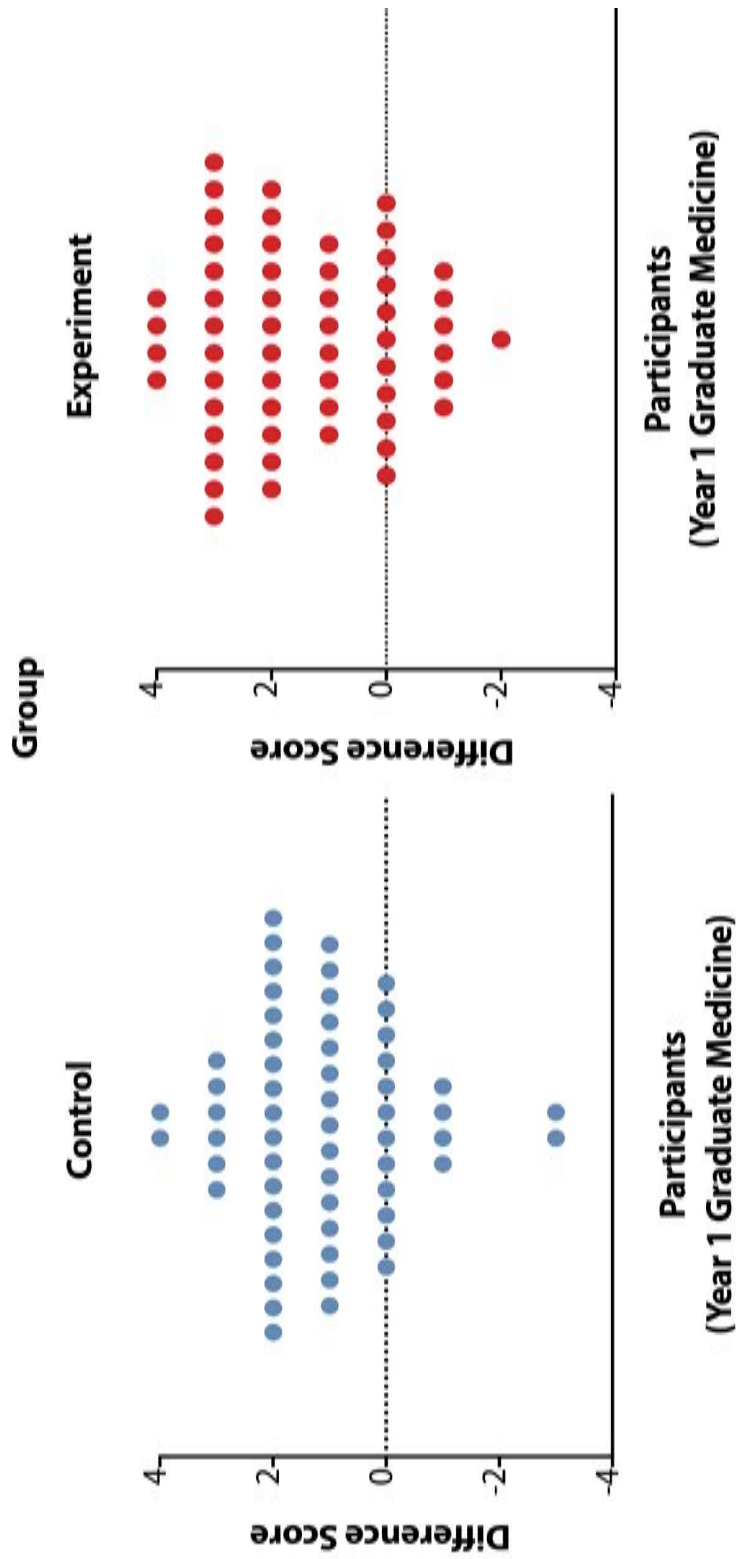


Figure 21. Gain scores for the control and experiment groups.

Jacobson's Reliable Change index was calculated to categorise the subset of students who made large gains in the quiz. The value calculated as an indicator of 'clinically' significant change was 5.0 on the posttest. There were two participants in control group with a RCI over 1.96 of those 1 scored above 5.0. There were 38 participants in the experiment group with a RCI over 1.96, and of these 19 scored above 5.0. The pretest and posttest scores and the boundary of 'clinically' significant change are shown in Figure 22. The characteristics of these participants is explored and summarised in the Improvement Profiles section.

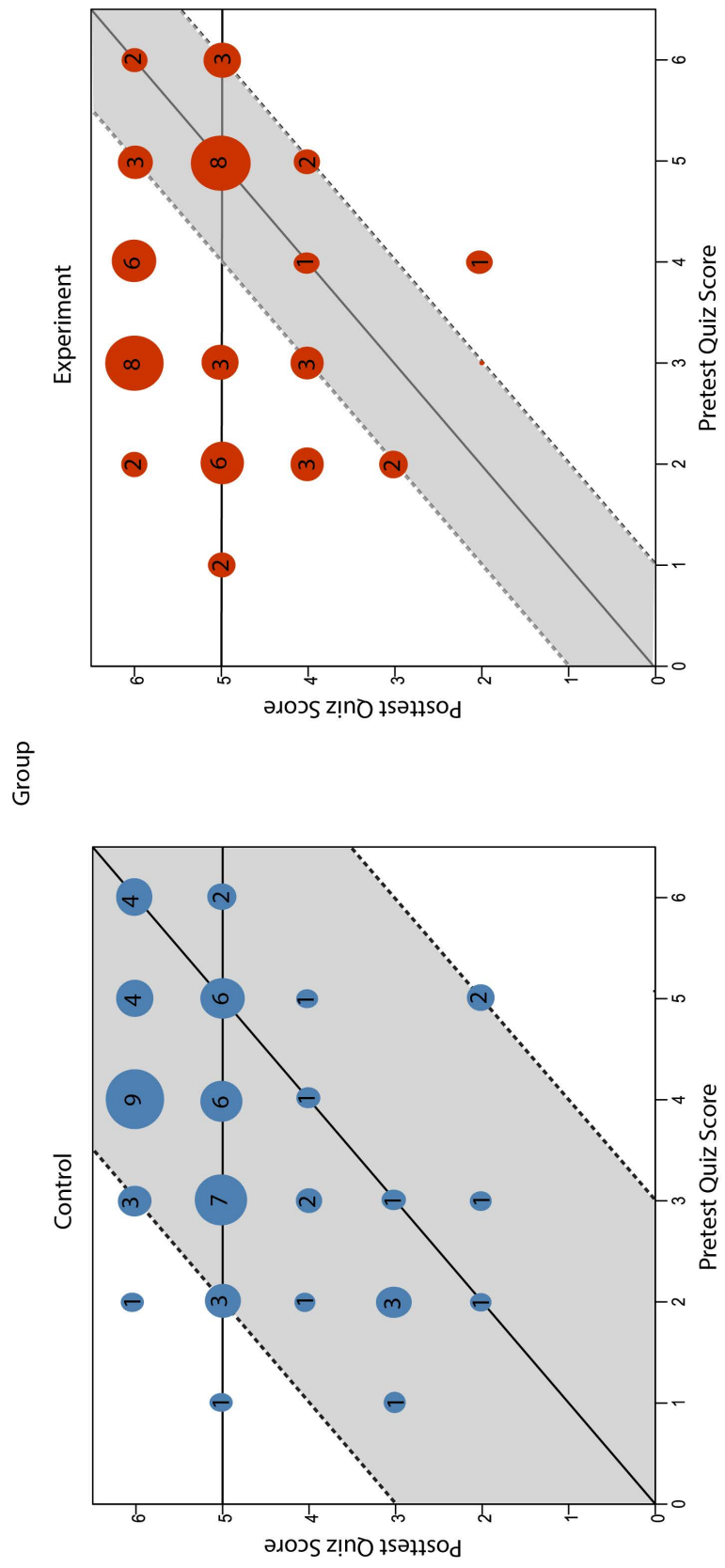


Figure 22. Frequency of pretest and posttest scores between groups. Dotted diagonal lines indicate the reliable change index band.

**Improvement profiles.**

**Control group.** Of the 60 participants in the control group, one (1.6%) had a reliable change index above 1.96 and a posttest score above the significant cutoff 5.0. A summary of their characteristics compared to the group is shown in Table 9.

Table 9

*Students with a significant RCI compared to those without between groups*

Participants	n	Sex			Age M+SD	Prior Degree (mode)	Anat Exp (median)	Mean Quiz Score	
		Male	Female	Other				Pretest	Posttest
Experiment									
Non Sig	37	24(64.9%)	13(35.1)	0	24±3	Science	Little	3.59±1.61	4.54±0.90
Sig RCI	19	12 (63.2%)	7 (36.8%)	0	26±3	Science	Little	3.53±.90	6±0
Control									
Non Sig	59	32 (54.2%)	26 (44.1%)	1(1.7%)	24±3.0	Science	Little	3.81±1.31	4.88±1.18
Sig RCI	1	100%			32	Science	No Exp	2	6

The profile of the participant with a significant RCI indicates no prior experience in learning anatomy. Figures 23 and 24 show that their pretest and posttest self-ratings tend toward low confidence and higher perceived difficulty when learning anatomy and the muscles of the back. However, the student made significant gains in their quiz score; it is possible that in this case a student with no prior experience in learning anatomy would invest greater effort during learning. It is difficult to extrapolate further without speculation.

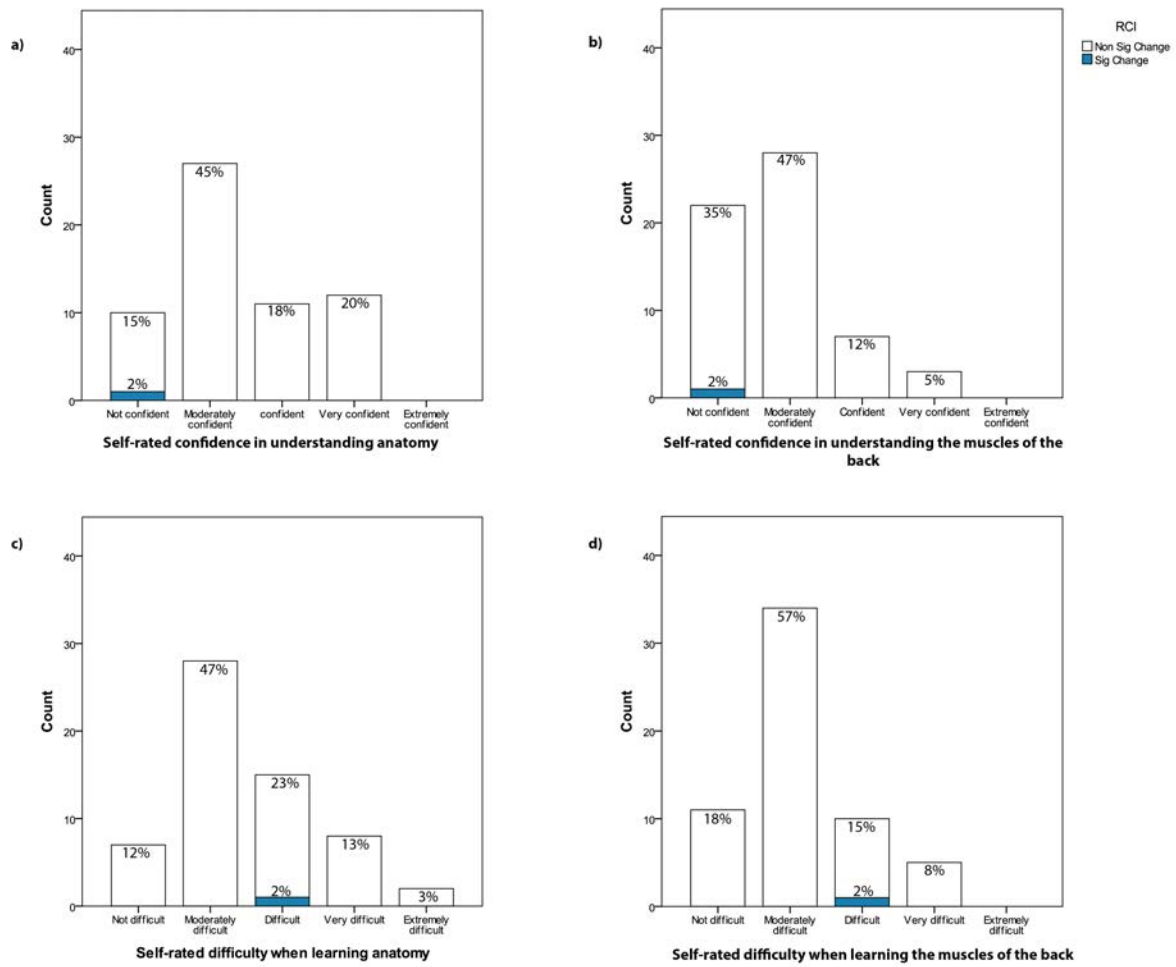


Figure 23. Distribution of pretest self-ratings for the control group.

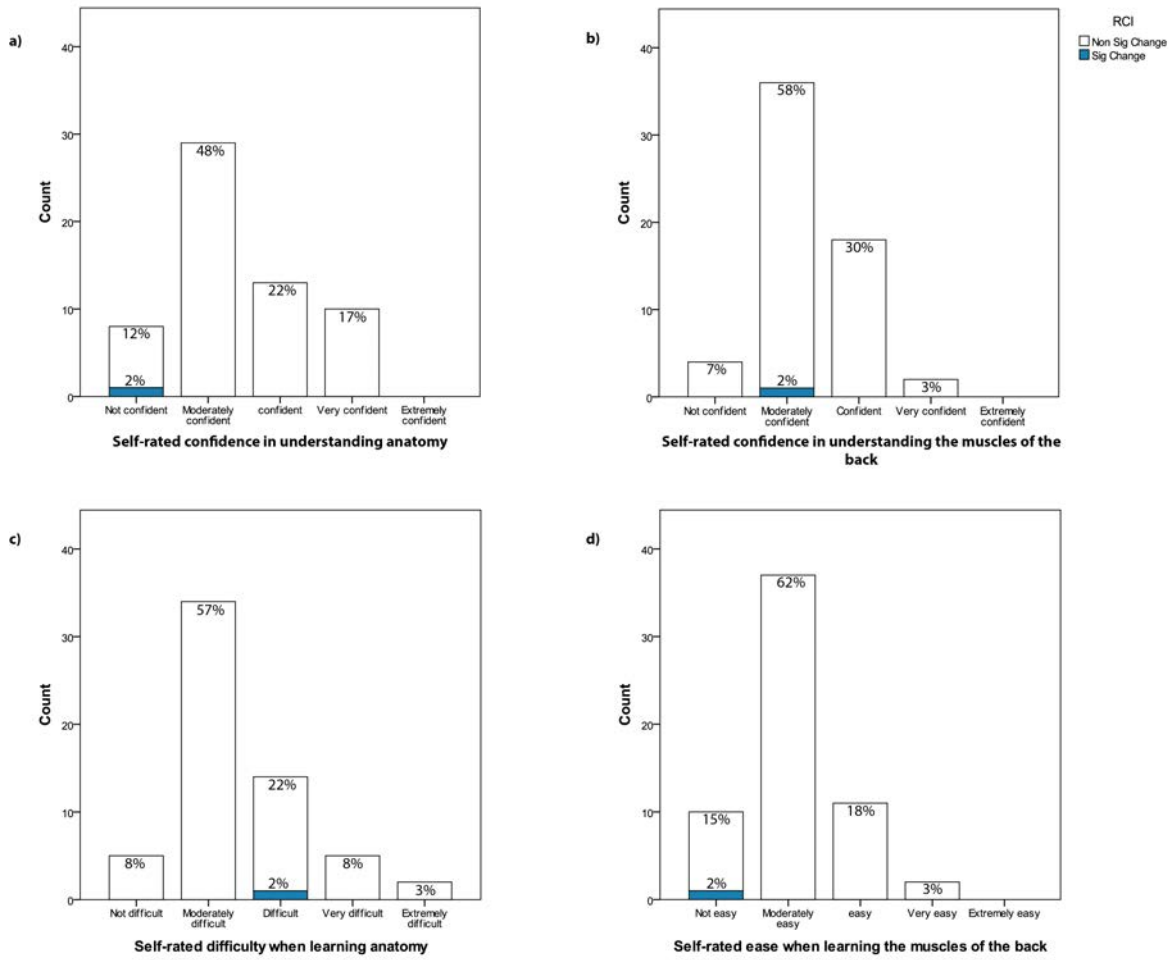


Figure 24. Distribution of posttest self-ratings for the control group.



**Experiment group.** Of the 56 participants in the experiment group, 19 (28.57%) had a reliable change index above 1.96 and a post test score above the significant cutoff 5.0. Their characteristics are heterogeneous across the variables of age, sex, prior degree and level of exposure to learning anatomy. The summary statistics in Table 9 show that the participants with significant RCI's closely represented the group. This is similar to the findings in experiment 1. The distribution of self-ratings in the pretest and posttest are shown in Figures 25 and 26 respectively.

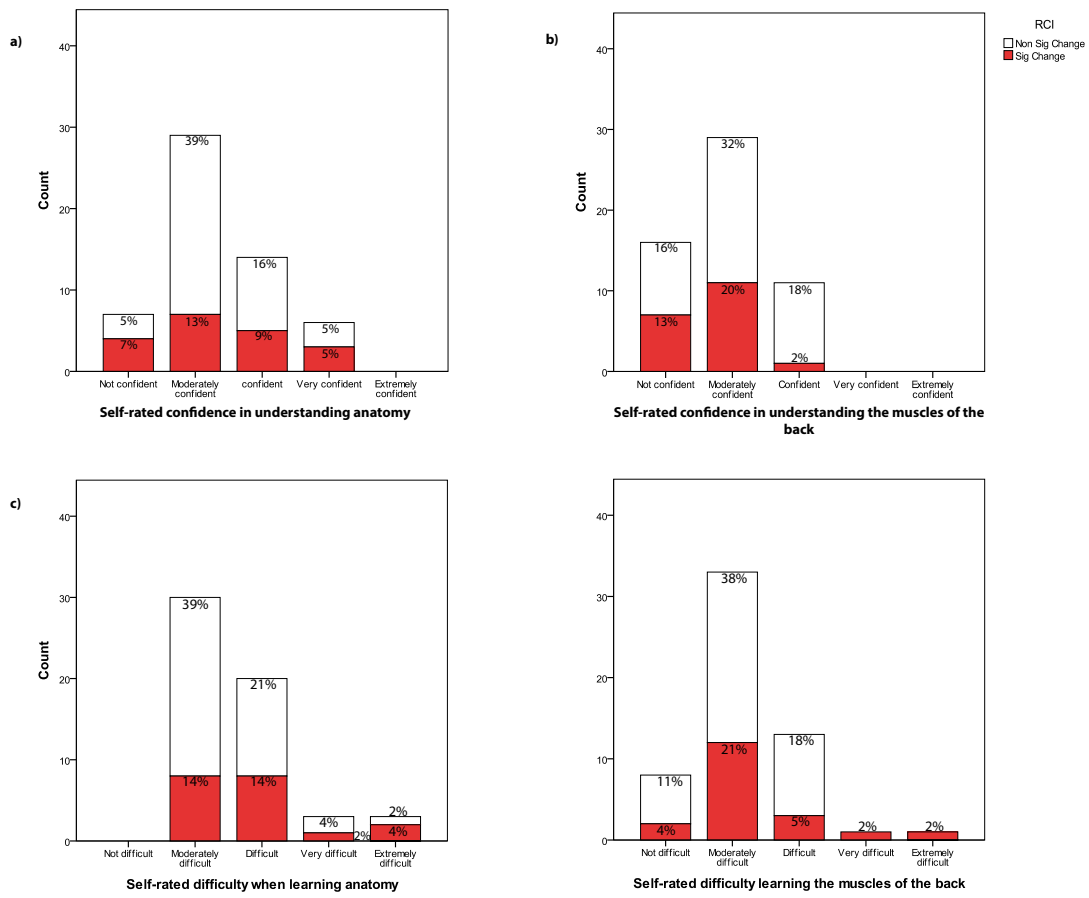


Figure 25. Distribution of pretest self-ratings for the experiment group.

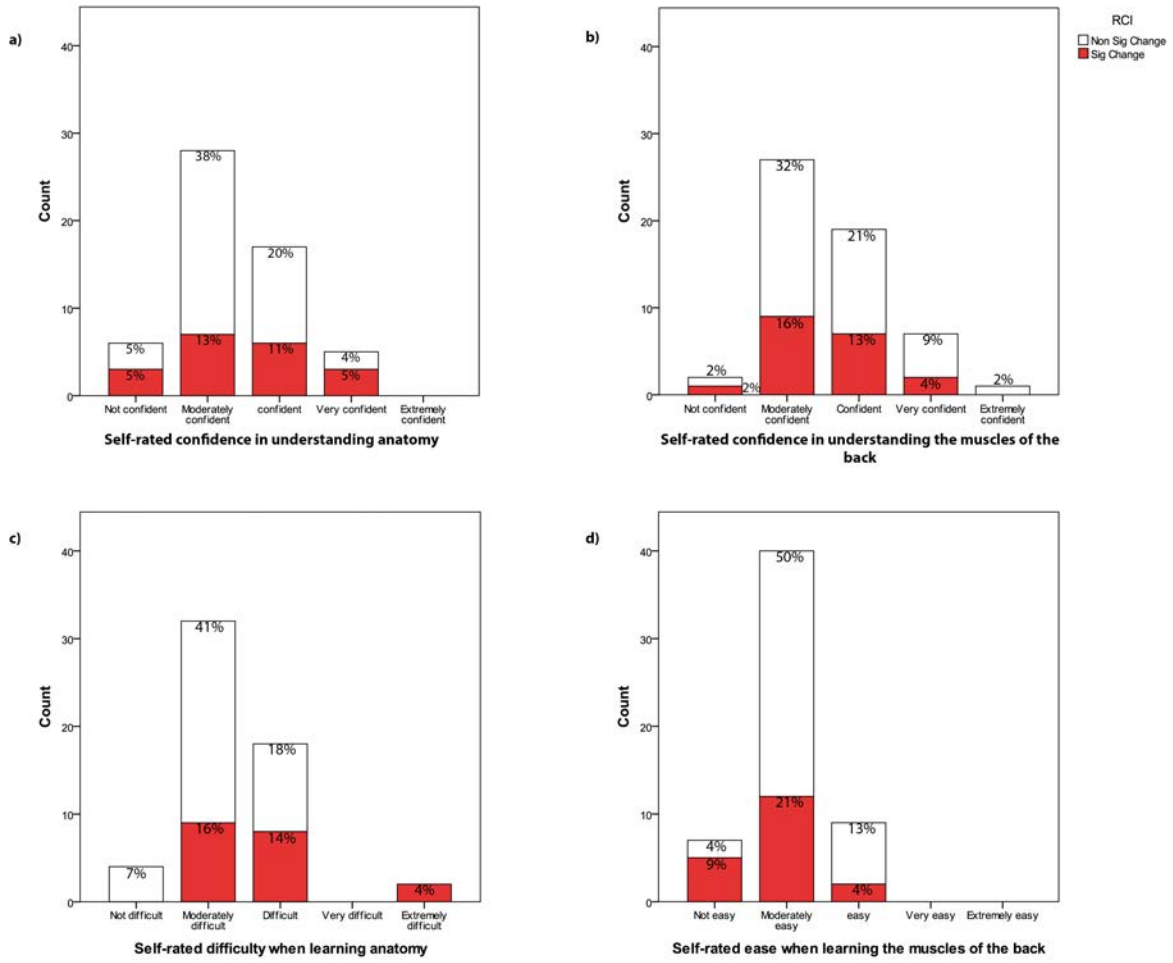


Figure 26. Distribution of posttest self-ratings for the experiment group.

## Discussion

### Conclusions

In the present study an excerpt from an online lecture on the vertebral column was delivered as a learning intervention for students in their first and second year of the Sydney Medical Program. For the experiment groups, the excerpt was visually modified to align with principles from the CTML while the control groups watched the unchanged excerpt. Learning outcomes were measured via a 6 item multiple choice quiz, and cognitive-affective outcomes were measured using likert-style self-rating questions.

The primary research question was, 'Will an online anatomy lesson consistent with the CTML versus an unmodified 'default' online lesson yield larger learning gains for graduate medical students?' The results from both experiments did not indicate a significant difference in learning gains between the two groups, however, for both experiments there was a main effect of time indicating significant improvement in performance from pretest to posttest. Although the results from both experiments suggested no significant difference between the groups, follow up analysis using Jacobson's RCI showed that more students in the experiment groups had larger performance gains in comparison to the control groups.

The secondary research question was 'What are the metacognitive ratings of medical students when learning anatomy and the muscles of the back and do they change depending on the lesson format?' Four likert style questions asked students to rate their confidence in understanding anatomy, and the muscles of the back, and their difficulty learning anatomy, and the muscles of the back. The first three questions were repeated in the posttest, while the final question was matched with its inverse to serve as a consistency check. Thus, the analysis of rating shifts was only applicable to the first three questions.

The self-rating about confidence when learning about the muscles of the back changed significantly for all groups in both experiments. The shifts indicated increased confidence with effect sizes of 0.4 and 0.3 for the control and experiment groups in experiment 1, and 0.4 and 0.5 for the control and experiment groups in experiment 2. Significant change in self-rated confidence when learning anatomy only occurred for the control group in experiment 1, with shifts indicating a decrease in confidence with an effect size of 0.3. The changes for self-rated difficulty when learning anatomy were significant for the control group in experiment 1, with results showing a decrease in difficulty and an effect size of 0.3. Conversely, in experiment 2, it was the experiment group who showed significant change in their self-rated difficulty when learning anatomy with shifts indicating a decrease in difficulty and an effect size of 0.2.

### **Theoretical Implications**

Although results from the statistical analysis did not show significant differences between the groups, Jacobson's RCI indicated that in the experiment groups a subgroup of medical students had learning gains of a larger magnitude. These results occurred for both experiments and have several implications.

First, the results lend support to predictions by CTML and CLT. By applying the techniques outlined by the CTML, cognitive processes were more effectively guided. Specifically, the technique used most extensively was signaling via coloured highlights. In the CTML framework this technique functions to reduce extraneous processing. Likewise, from a CLT perspective, working memory capacity is relieved of extraneous load and consequently free capacity may be redirected to relevant processes for schema construction such as selecting, organizing and integrating information. Further, although there may have been ceiling effects impacting higher performing students, approximately 60% of students in both

experiments reported little to no prior experience in learning anatomy. In this respect the students may be characterized as novices, therefore, the principles from CTML and CLT are well suited. However, as a student progresses in the medical program the same principles may induce the expertise reversal effect. In this instance including techniques to stimulate germane or generative processing may help in mitigating negative learning outcomes. Indeed, there are several instances where instructional approaches may impede learning due to learner characteristics or differences.

The use of animation for instruction has yielded varied results and discussion in terms of its use cases (Schnotz and Lowe, 2008). For example, Paik and Schraw, 2013 found that adding animation to multimedia presentations affects metacognitive monitoring in complex ways. They tested the illusion of understanding hypothesis i.e., that the addition of animation makes learners perceive a presentation as easier to understand and therefore develop optimistic metacomprehension; a phenomenon consistent with the expertise reversal effect. To test their hypothesis they used the sum of several questions to establish students' judgement of difficulty during learning and their judgement of comprehension. They found that for learners with low proficiency, the illusion of understanding was induced; their judgement of difficulty was low, but their judgement of comprehension was high. In the present experiments, the likert questions about confidence in understanding the muscles of the back, and anatomy, and difficulty learning anatomy and the muscles of the back may be loosely aligned with the judgement of difficulty and judgement of comprehension metrics in the study by Paik and Schraw (2013). In the present studies, the control group in experiment 1 had a significant increase in their confidence in understanding the muscles of the back and a significant decrease in their difficulty rating when learning about anatomy after the intervention. However, they also had a significant decrease in their confidence in

understanding anatomy. Similarly, the experiment group in experiment 2 showed significant increase in their confidence in understanding the muscles of the back and a significant decrease in their difficulty rating when learning anatomy. These patterns may be indicative of an illusion of understanding. Interestingly, the control group in experiment 1 also indicated a decrease in their confidence in understanding anatomy. A possible reason is that since all structures were visible across only two slides cognitive load may have been adversely affected. Although attention was directed via the use of a laser pointer to relevant regions of the slide during narration, the presence of all the information may have interfered with the process of selecting information and exerted extraneous cognitive load. The experiment group in experiment 2, had shifts in their judgement of difficulty when learning anatomy and confidence in understanding the muscles of the back but no significant shifts in their confidence in understanding anatomy. Like the control group in experiment 1 this may suggest that an illusion of understanding was induced, however, additional variables may be influencing their metacognitive judgement and associated performance. As Paik et al., 2013 note, the causal mechanism behind changes in these metacognitive judgments is not clearly understood.

In the current experiments, the modified lesson consisted primarily of directive animation techniques i.e., signaling/cueing/highlighting. An additional technique defined by Paik and Schraw (2013) is representational animation where, for example, dynamic systems are shown. Their predictions were that representational as opposed to directive techniques would induce the illusion of understanding, however they found evidence for the effect from both techniques. Similarly, Schnotz and Rasch (2008), assign two functions to animations and their effect on cognitive load. Their framework assigns animations as either enabling or facilitating; where the enabling function makes processes that would be impossible, possible

and the facilitating process makes processes normally requiring high effort, become possible with less effort. As with the illusion of understanding and several principles in CTML and CLT, these techniques are ideally aligned with a students' ability. For example, students with high learning prerequisites benefit most from the enabling function while those lower in prerequisites benefit from the facilitating function. Misapplication of the techniques are likely to induce detrimental learning effects such as the expertise reversal effect, or extraneous cognitive load.

It is interesting that the suitability of instructional techniques often varies inversely with the expertise of the learner. What is more intriguing are the instances that highlight the boundary conditions. For example, Park et al (2011) found that cognitive load scenarios have a moderating effect on seductive details. Normally, seductive details – that is details not directly related to a learning outcome such as music, or eye catching graphics tangentially related to content – are considered extraneous details that induce extraneous processing and as a result impede effective learning. In their study they found that for low cognitive load conditions seductive details can assist learning possibly due to their motivational or arousing role. These results highlight the utility of incorporating affect and motivation into frameworks for multimedia instruction.

Omitting the relationship between constructs like cognitive load, affect and motivation is a limitation in CLT pointed out by Moreno (2009), who argues that, for example, the difficulty of material can mediate students' motivation to learn which in turn affects the amount of effort invested during learning. The strength of Moreno's cognitive affective theory of learning with media is that it predicts that the amount of effort invested in learning will depend on students' beliefs about themselves and the learning task.



There is evidence to suggest that techniques that induce desirable difficulties are useful strategies for effective learning. Yue et al., (2013) examined how a discrepancy between onscreen text and narration in a multimedia lesson impacts recall and transfer. According to the CTML and CLT including visual text in concert with identical narration creates redundancy which hampers learning; the reason being that working memory resources end up unnecessarily taxed in order to co-ordinate the sources of information (Kalyuga and Sweller, 2014). However, in the studies by Yue et al., (2013) they found that a slight but not significant discrepancy between narration and on screen text is beneficial for learning, possibly due to the generative processing that it encourages. As such they describe their findings as an extension to the redundancy principle.

In light of the findings by Yue et al., (2013) and Moreno (2009), the results from the current experiments may also support the benefits of discrepant onscreen text during narration. In experiment 2, perhaps the patterns indicating the illusion of understanding i.e., increased confidence and decreased difficulty, for students in the experiment group, reflect affective and motivational changes based on the effective guidance of cognitive processes and the effects of text discrepancy. The discrepancy in text may serve to activate germane processes and maintain engagement. This appears to be a useful technique in a subject such as anatomy where there is a high volume of factual knowledge.

The results of this thesis lend support to the benefits of principles from CTML and CLT. Given that both frameworks are based on our understanding of the human information processing system and have strong empirical bases, their strategies are especially relevant given the explosion of online learning. Additionally, the results suggest the moderating effects of constructs such as affect, motivation and metacognition on learning. Omitting these components, although simpler, results in an incomplete picture of learning.

### **Practical implications**

One of the most valuable practical implications from conducting studies, as in the present case, is gaining experience in the application of principles from the CTML and testing their effectiveness. Similar studies include: a thesis by Richard Alan Lamb (2013) which looks at the process of transforming existing videos to comply with CTML, Hong Kok Ng's thesis (2014) which explores the use of tracing in animations and its implications for CLT, Chris Lawrence's thesis (2014) which explores the design of lecture presentations using CLT and principles from visual communication, and Derek Muller's thesis (2008) exploring the design of multimedia for physics education. All these cases assist in setting evidence based standards for multimedia instruction.

In terms of instruction in anatomy, the present student may help determine how to design multimedia given a heterogeneous group of students. For example, Kalyuga et al., (2000) found interesting learner dependent effects when including text with a diagram, audio with a diagram and a diagram alone. They found that for less experienced learners, well integrated text and a diagram were superior to the other conditions, however, as expertise increased they found that diagrams alone were most effective. In the present studies, it would be straightforward to convert the modified video into a visual only resource, perhaps with increased playback speed as a resource for students at more advanced levels. The benefit is that the visual principles align with CTML, and that the process is easily achievable from a technical standpoint. Further, the same visual only video may also be used as a form of pre-training, i.e., introducing the names and characteristics of elements prior to a more complex lesson. Pre-training is one of the principles in CTML with fewer applications in experiments (Mayer and Pilegard, 2014) and in freely available videos online (Yue et al., 2013). In a subject matter with a high volume of content, such as anatomy, using this principle is one way to

manage essential processing and, its application would assist in determining boundary conditions.

Another area with interesting implications is the integration of desirable difficulties during learning. The previous discussion, included this subject in terms of a slight discrepancy between on screen text and narration during a multimedia lesson. As an extension to the redundancy effect, this strategy could prove useful in delivering a lesson to a group of students of varying experience. The reason for this is that its key strategy is in encouraging generative or active learning. A method to extend this effect is also to include tests throughout a lesson. This strategy was applied in Dobson and Linderholm's (2015) study with kinesiology students studying anatomy, and they found superior performance in immediate and delayed recall.

### **Future directions**

Future explorations of this topic in anatomy would benefit from several strategies, some of which were limitations in the current studies. For example, long term assessment of recall and transfer learning, and incorporating measures of spatial ability may help shed light on individual differences between students and thus how to adjust instruction to meet their needs. A more direct extension would be closer investigation of the technique of signaling or directive animation across various dimensions, for example illustrations, photos, live video and computer generated imagery. In cases of computer generated imagery there is the added utility of being able to construct, distort or exaggerate structure which can serve to reinforce key messages in a lesson. As Tversky et al, (2008) highlight, there is much discussion about the contrasts between static images and animations, which often point to static images being superior for learning. In the current studies the images presented remained static, with directive techniques such as highlighting, fading, and movement as some of the key

adjustments. Beginning with this form of presentation and augmenting analysis via new methodology, such as eye tracking, may assist in establishing how directive techniques differ to representational techniques, or even how best to combine of the two. Similar analysis was done in terms of the spatial contiguity effect by Johnson and Mayer (2012). The strength of the tool is the ability to accumulate more direct data on the processing occurring during a learning event which may otherwise be overlooked. Similarly, analysis of video playback is a similar approach to discern how students navigate a resource (Lowe, 2008) and to indicate problem areas, where additional explanation is required. Conversely, the style and format of presentation may incorporate techniques from narrative film and animation to purposely induce cognitive and affective processes suitable for learning.

Moreno's CATLM is a framework including the core tenets of CTML and CLT, while also incorporating cognitive, affective and motivational factors. It is a valuable perspective for future research because it forms a more complete prediction of the learning experience by including several variables. Thus, future research would benefit from including assessment of affective and metacognitive variables and the continued development of validated instruments.

Generative activities and peer assessment are methods which may assist in creating a bridge between in person activities and the advantages (and idiosyncracies) of online learning or communication. For instance, Hubscher-Younger and Narayanan (2008) had students generate multimedia which required them to explain algorithms. The videos were rated and judged by their peers across several domains. In their first study students began to converge towards a uniform style that may have been established via an example from the teacher or based on ratings that were displayed. After anonymizing the review process, obscuring ratings and including a rating of originality rather than contiguity, there was a greater variety of

approaches to presenting the subjects. Further there is research showing that teaching is an effective strategy for learning. Some creative applications include a generative form of learning with medical students who were required to use clinical skills and reasoning to construct a patient case, based on a diagnosis rather than learning via a single case example (Philip et al., 2008). The value of this approach highlights perspectives such as Rand Spiro's Cognitive Flexibility theory i.e, multiple representations are a strength which emphasise the variation in complex topics.

Finally, a more speculative suggestion is the combination of students from different schools and disciplines while learning common content in anatomy. For instance, students in art often learn anatomy as part of their training. Although an artist's focus is on issues such as: proportions of skeleton and the muscles that give the body its form (i.e., they would not necessarily learn details such as deep muscles of the back), a significant amount of musculoskeletal anatomy overlaps with lessons in the sciences. Given theories like connectivism, and new forms of networking, the question would be: what would happen if students in art learned anatomy alongside science students or students in early year med? There are several points of interactions, for example would it be advantageous to students of medicine to explain anatomy to art students who are learning for a very different purpose? Likewise, for student of art explaining their anatomical understanding to medical or science students could likely be an effective strategy for learning. This scenario echoes cognitive flexibility theory where several perspectives on a topic bring greater depth and nuance in understanding.

### References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes (Vol. 2, pp. 89–195). Elsevier. [http://doi.org/10.1016/S0079-7421\(08\)60422-3](http://doi.org/10.1016/S0079-7421(08)60422-3)
- Baars, B. J. (2001). A biocognitive approach to the conscious core of immediate memory. *Behavioral and Brain Sciences*, 24(1), 115–116. <http://doi.org/10.1017/S0140525X01233925>
- Bacro, T. R. H., Gebregziabher, M., & Ariail, J. (2013). Lecture recording system in anatomy: Possible benefit to auditory learners. *Anatomical Sciences Education*, n–a–n–a. <http://doi.org/10.1002/ase.1351>
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Baddeley, A. (2001). The magic number and the episodic buffer. *Behavioral and Brain Sciences*, 24(1), 117–118. <http://doi.org/10.1017/S0140525X01253928>
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*.
- Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1–29. <http://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. D. (1990). Human Memory. Allyn and Bacon.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, 47–89. [http://doi.org/10.1016/S0079-7421\(08\)60452-1](http://doi.org/10.1016/S0079-7421(08)60452-1)
- Barry, D. S., Marzouk, F., & Oglu, K. C. (2016). Anatomy education for the YouTube generation. *Anatomical Sciences Education*. <http://doi.org/10.1002/ase.1550>
- Bednar, A. K., Cunningham, D., Duffy, T. M., & Perry, J. D. (1992). Theory into Practice:

How do we link? In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the Technology of Instruction*.

Bergman, E. M., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2011). Why don't they know enough about anatomy? A narrative review. *Medical Teacher*, 33(5), 403–409.

<http://doi.org/10.3109/0142159X.2010.536276>

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition*, (108).

Bola, M., & Borchardt, V. (2016). Cognitive Processing Involves Dynamic Reorganization of the Whole-Brain Network's Functional Community Structure. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, 36(13), 3633–3635.

<http://doi.org/10.1523/JNEUROSCI.0106-16.2016>

Brünken, R., Plass, J. L., & Moreno, R. (2010). Current Issues and Open Questions in Cognitive Load Research. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 1–20). Cambridge University Press.

<http://doi.org/10.1017/CBO9780511844744.014>

Burgoon, J. M. (2008, July 25). *An Investigation of the Self-Efficacy of Medical Students for the Anatomy Curriculum: Role of Gender and Prior Experience, and Self-Efficacy's Influence on Academic Achievement* (Doctoral Thesis, University of North Carolina, Chapel Hill, USA). Available from ProQuest Dissertations and Theses Full Text database. (UMI No. 3315735)

Caeyenberghs, K., Metzler-Baddeley, C., Foley, S., & Jones, D. K. (2016). Dynamics of the Human Structural Connectome Underlying Working Memory Training. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, 36(14), 4056–4066.

<http://doi.org/10.1523/JNEUROSCI.1973-15.2016>

- Choudhury, B., & Gouldsbrough, I. (2012). The use of electronic media to develop transferable skills in science students studying anatomy. *Anatomical Sciences Education*, 5(3), 125–131. <http://doi.org/10.1002/ase.1259>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <http://doi.org/10.1007/BF01320076>
- Clark, R. C., & Mayer, R. E. (2011). *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* [E LEARNING & THE SCIENCE OF IN] [Hardcover]. Pfeiffer & Company.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity (vol 23, pg 87, 2001). *Behavioral and Brain Sciences*, 24(3).
- Cowan, N. (2015). George Miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review*, 122(3), 536–541. <http://doi.org/10.1037/a0039035>
- Craik, F., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [http://doi.org/10.1016/s0022-5371\(72\)80001-x](http://doi.org/10.1016/s0022-5371(72)80001-x)
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38(2), 105–134. <http://doi.org/10.2307/23372530?ref=search-gateway:5296538c0494289dd2a7ec947de22d33>
- DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223–234. <http://doi.org/10.1037/0022-0663.100.1.223>
- Dobson, J. L., & Linderholm, T. (2015). The effect of selected “desirable difficulties” on the



- ability to recall anatomy information. *Anatomical Sciences Education*, 8(5), 395–403.  
<http://doi.org/10.1002/ase.1489>
- Drake, R. L., McBride, J. M., Lachman, N., & Pawlina, W. (2009). Medical education in the anatomical sciences: The winds of change continue to blow. *Anatomical Sciences Education*, 2(6), 253–259. <http://doi.org/10.1002/ase.117>
- Ellaway, Rachel (2011). Reflecting on multimedia design principles in medical education. *Medical Education*, 45(8), 766-767. doi:10.1111/j.1365-2923.2011.04064.x
- Estai, M., & Bunt, S. (2016). Best teaching practices in anatomy education: A critical review. *Annals of Anatomy - Anatomischer Anzeiger*. doi:10.1016/j.aanat.2016.02.010
- Faul, F., & Erdfelder, E. (1998) G\*Power (3.1). Retrieved from  
<http://www.gpower.hhu.de/en.html>
- Fei, J., Mather, C., Elmer, S., Allan, C., Chin, C. & Chandler, L. (2013). Use of Echo360 generated materials and its impact on class attendance. In Carter, H., Gosper, M., and Hedberg, J. (Eds.), *Electric Dreams. Proceedings ascilite 2013 Sydney*. (pp.288-292)
- Fyfe G., Fyfe, S., Dye, D., Crabb, H. (2013). Use of Anatomage tables in a large first year core unit. In Carter, H., Gosper, M., and Hedberg, J. (Eds.), *Electric Dreams. Proceedings ascilite 2013 Sydney* (pp.298-302)
- Gupta, A., & Saks, N. S. (2013). Exploring medical student decisions regarding attending live lectures and using recorded lectures. *Medical Teacher*, 35(9), 767–771.  
<http://doi.org/10.3109/0142159X.2013.801940>
- Grondin, S. (2001). A temporal account of the limited processing capacity. *Behavioral and Brain Sciences*, 24(1), 122–123. <http://doi.org/10.1017/S0140525X01303928>
- Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social

presence in multimedia-learning. *Computers in Human Behavior*, 24(3), 786–797.

<http://doi.org/10.1016/j.chb.2007.02.009>

Hubscher-Younger, T., & Narayanan, N. H. (2008). Turning the tables: Investigating characteristics of efficacy of student-authored animations and multimedia representations. In R. Lowe & W. Schnotz (Eds.), *Learning with Animation*. Cambridge University Press.

Issa, N., Mayer, R. E., Schuller, M., Wang, E., Shapiro, M. B., & DaRosa, D. A. (2013).

Teaching for understanding in medical classrooms using multimedia design principles.

*Medical Education*, 47, 388–396.

Issa, N., Schuller, M., Santacaterina, S., Shapiro, M., Wang, E., Mayer, R. E., & DaRosa, D. A.

(2011). Applying multimedia design principles enhances learning in medical education.

*Medical Education*, 45(8), 818–826.

Jaffar, A. A. (2012). YouTube: An emerging tool in anatomy education. *Anatomical Sciences*

*Education*, 5(3), 158–164. <http://doi.org/10.1002/ase.1268>

Johnson, E. O., Charchanti, A. V., & Troupis, T. G. (2012). Modernization of an anatomy

class: From conceptualization to implementation. A case for integrated multimodal–

multidisciplinary teaching. *Anatomical Sciences Education*, 5(6), 354–366.

<http://doi.org/10.1002/ase.1296>

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity

effect in multimedia learning. *Journal of Experimental Psychology. Applied*, 18(2), 178–

191.

Kaczmarzyk, M., Francikowski, J., Łozowski, B., Rozpędek, M., Sawczyn, T., & Sułowicz, S.

(2013). The bit value of working memory. *Psychology & Neuroscience*, 6(3), 345–349.

<http://doi.org/10.3922/j.psns.2013.3.11>

- Kalyuga, S. (2014). The Expertise Reversal Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 576–597). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781139547369.028>
- Kalyuga, S., & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. Mayer, *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 247–262). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139547369.013
- Kok Ng, H. (2014, June 20). *The use of tracing to reduce transience in instructional animations: a cognitive load theory perspective*. (J. Sweller & S. Kalyuga, Eds.). UNSW. Phd
- Lamb, R. A. (2013, April 1). *A makeover for the captured lecture: Applying Multimedia Learning Principles to Lecture Video*. Nova Southeastern university. Phd
- Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, 29(C), 174–175. <http://doi.org/10.1016/j.learninstruc.2013.05.004>
- Lowe, R. (2008). Learning from animation. In R. Lowe & W. Schnotz (Eds.), *Learning with Animation*. Cambridge University Press.
- Lawrence, Chris (2013). *Designing Lecture Presentations Using Cognitive Load Theory and Visual Communication Principles*. Uni newcastle. MEd
- Lazarus, M., Rascoe, A., Iskand, K., & Black, K. (2015, December). Vertical integration of anatomy into clinical year medicine improves student knowledge and retention of essential clinical anatomy. *12<sup>th</sup> Annual conference of the Australian and New Zealand Association of Clinical Anatomists*. Sydney.
- Lodge, J. M., & Bonsanquet, A. (2014). Evaluating quality learning in higher education: re-examining the evidence. *Quality in Higher Education*, 20(1), 3–23. <http://doi.org/10.1080/13538322.2013.849787>

- MacLean, J. E., Scott, K. M., Marshall, T., & Asperen, P. V. (2011). Evaluation of an e-learning teaching resource: What is the medical student perspective? *Focus on Health Professional Education: a Multi-Disciplinary Journal*, 13(2), 53.
- Mayer, R. (Ed.). (2014a). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781139547369>
- Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84(4), 405–412.  
<http://doi.org/10.1037/0022-0663.84.4.405>
- Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *The American Psychologist*, 63(8), 760–769.
- Mayer, R. E. (2009). *Multimedia Learning*. Cambridge University Press.
- Mayer, R. E. (2010a). Applying the science of learning to medical education. *Medical Education*, 44(6), 543–549.
- Mayer, R. E. (2010b). Merlin C. Wittrock's Enduring Contributions to the Science of Learning. *Educational Psychologist*, 45(1), 46–50.  
<http://doi.org/10.1080/00461520903433547>
- Mayer, R. E. (2014b, July 8). Research-based principles for multimedia learning. *Harvard Initiative for Learning and Teaching*. Harvard University. Retrieved from <https://www.youtube.com/watch?v=AJ3wSf-ccXo>
- Mayer, R. E., & Moreno, R. (2002). Animation as an Aid to Multimedia Learning. *Educational Psychology Review*, 14(1), 87–99. <http://doi.org/10.1023/A:1013184611077>
- Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52.  
[http://doi.org/10.1207/S15326985EP3801\\_6](http://doi.org/10.1207/S15326985EP3801_6)

- Mayer, R. E., & Sims, V. K. (1994). For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning. *Journal of Educational Psychology, 86*(3), 389–401.
- Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology. Applied, 8*(3), 147–154.
- Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638–643. <http://doi.org/10.1037/0022-0663.91.4.638>
- Mayer, R. E., & Pilegard, C. (2014). Principles for Managing Essential Processing in Multimedia Learning: Segmenting, Pre-training, and Modality Principles. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 316–344). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781139547369.016>
- McMenamin, P. G., Quayle, M. R., McHenry, C. R., & Adams, J. W. (2014). The production of anatomical teaching resources using three-dimensional (3D) printing technology. *Anatomical Sciences Education, 7*(6), 479–486. <http://doi.org/10.1002/ase.1475>
- McNulty, J. A., Sonntag, B., & Sinacore, J. M. (2009). Evaluation of computer-aided instruction in a gross anatomy course: A six-year study. *Anatomical Sciences Education, 2*(1), 2–8. <http://doi.org/10.1002/ase.66>
- Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.
- Moreno, R. (2005). Instructional Technology: Promises and Pitfalls. In L. M. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-Based Education Bringing Researchers and Practitioners Together*. Information Age Publishing.

- Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology*, 21(6), 765–781. <http://doi.org/10.1002/acp.1348>
- Moreno, R. (2009). Cognitive load theory: more food for thought. *Instructional Science*, 38(2), 135–141. <http://doi.org/10.1007/s11251-009-9122-9>
- Moreno, R., & Park, B. (2010). 1. Cognitive Load Theory: Historical Development and Relation to Other Theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 1–20). Cambridge University Press. <http://doi.org/10.1017/CBO9780511844744.003>
- Narayan, R., Rodriguez, C., Araujo, J., Shaqlaih, A., & Moss, G. (2013). Constructivism-- Constructivist Learning Theory. In B. J. Irby, G. Brown, R. Lara-Alecio, & S. Jackson (Eds.), *The Handbook of Educational Theories*.
- Nieder, G. L., & Borges, N. J. (2012). An eight-year study of online lecture use in a medical gross anatomy and embryology course. *Anatomical Sciences Education*, 5(6), 311–320. <http://doi.org/10.1002/ase.1289>
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <http://doi.org/10.1037/0022-0663.84.4.429>
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2010). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), 63–71. [http://doi.org/10.1207/S15326985EP3801\\_8](http://doi.org/10.1207/S15326985EP3801_8)
- Paik, E. S., & Schraw, G. (2013). Learning With Animation and Illusions of Understanding. *Journal of Educational Psychology*, 105(2), 278–289.
- Paivio, A. (2007). *Mind and Its Evolution*. Lawrence Erlbaum Associates.

- Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior*, 27(1), 5–10. <http://doi.org/10.1016/j.chb.2010.05.006>
- Park, B., Plass, J. L., & Brünken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction*, 29(C), 125–127. <http://doi.org/10.1016/j.learninstruc.2013.05.005>
- Persaud, T. V. N. (1984). Early history of human anatomy. Charles C. Thomas Publisher.
- Philip, C. T., Unruh, K. P., Lachman, N., & Pawlina, W. (2008). An explorative learning approach to teaching clinical anatomy using student generated content. *Anatomical Sciences Education*, 1(3), 106–110. <http://doi.org/10.1002/ase.26>
- Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). Cognitive Load Theory. Cambridge University Press.
- Raichle, M. E. (2010). The Brain's Dark Energy. *Scientific American*, 302(3), 44–49. <http://doi.org/10.1038/scientificamerican0310-44>
- Raikos, A., & Waidyasekara, P. (2014). How Useful is YouTube in Learning Heart Anatomy? *Anatomical Sciences Education*, 7(1), 12–18. <http://doi.org/10.1002/ase.1361>
- Regehr, G. (2004). Trends in medical education research. *Academic Medicine*, 79(10), 939–947. <http://doi.org/10.1097/00001888-200410000-00008>
- Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–21. <http://doi.org/10.1016/j.neuroscience.2005.12.061>
- Rohs, M., & Ganz, M. (2015). MOOCs and the claim of education for all: A disillusion by empirical data. *International Review of Research in Open and Distance Learning*, 16(6), 1–19.

- Ruiz, J. G., Cook, D. A., & Levinson, A. J. (2009). Computer animations in medical education: a critical literature review. *Medical Education*, 43(9), 838–846.  
<http://doi.org/10.1111/j.1365-2923.2009.03429.x>
- Schnotz, W., & Lowe, R. (2008). A Unified View of Learning from Animated and Static Graphics. In R. Lowe & W. Schnotz (Eds.), *Learning with Animation*. New York: Cambridge University Press.
- Schnotz, W., & Rasch, T. (2008). Functions of Animations in Comprehension and Learning. In R. Lowe & W. Schnotz (Eds.), *Learning with Animation*. New York.
- Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In S. Tobias & T. Duffy (Eds.), *Constructivist Instruction: Success or failure*. Routledge.
- Shaibah, H. S., & van der Vleuten, C. P. M. (2013). The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. *Anatomical Sciences Education*, 6(3), 149–156. <http://doi.org/10.1002/ase.1325>
- Siemens, G. (2013). Massive Open Online Courses: Innovation in Education? In R. McGreal & S. Marshall (Eds.), *Open Educational Resources: Innovation, research and practice*. Vancouver: Commonwealth of Learning and Athabasca University.
- Siemens, G. (2004, December 12). Connectivism: A Learning Theory for the Digital Age. Retrieved November 18, 2016, from <http://www.elearnspace.org/Articles/connectivism.htm>
- Stuijzand, B. G., Van Der Schaaf, M. F., Kirschner, F. C., Ravesloot, C. J., Van Der Gijp, A., & Vincken, K. L. (2016). Medical students' cognitive load in volumetric image interpretation: Insights from human-computer interaction and eye movements. *Computers in Human Behavior*, 62, 394–403. <http://doi.org/10.1016/j.chb.2016.04.015>



- Sugand, K., Abrahams, P., & Khurana, A. (2010). The anatomy of anatomy: A review for its modernization. *Anatomical Sciences Education*, NA–NA. <http://doi.org/10.1002/ase.139>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). Explorations in the Learning Sciences, Instructional Systems and Performance Technologies - Cognitive Load Theory. (J. M. Spector & S. P. Lajoie, Eds.) (pp. 1–280). New York: Springer.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296.
- Swinnerton, B. J., Morris, N. P., Hotchkiss, S., & Pickering, J. D. (2016). The integration of an anatomy massive open online course (MOOC) into a medical anatomy curriculum. *Anatomical Sciences Education*. <http://doi.org/10.1002/ase.1625>
- Tiitinen, H. (2001). How to interface cognitive psychology with cognitive neuroscience? *Behavioral and Brain Sciences*, 24(1).
- Topping, D. B. (2013). Gross anatomy videos: Student satisfaction, usage, and effect on student performance in a condensed curriculum. *Anatomical Sciences Education*, 7(4), n–a–n–a. <http://doi.org/10.1002/ase.1405>
- Trelease, R. B. (2002). Anatomical informatics: Millennial perspectives on a newer frontier. *The Anatomical Record*, 269(5), 224–235. <http://doi.org/10.1002/ar.10177>
- Trelease, R. B. (2016). From chalkboard, slides, and paper to e-learning: How computing technologies have transformed anatomical sciences education. *Anatomical Sciences Education*. <http://doi.org/10.1002/ase.1620>
- Tsai, C.-C., Jessie Ho, H. N., Liang, J.-C., & Lin, H.-M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*, 21(6), 757–769. <http://doi.org/10.1016/j.learninstruc.2011.05.002>

- van Gog, T., & Paas, F. (2008). Instructional Efficiency: Revisiting the Original Construct in Educational Research. *Educational Psychologist*, 43(1), 16–26.  
<http://doi.org/10.1080/00461520701756248>
- van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied Cognitive Psychology*, 26(6), 833–839. <http://doi.org/10.1002/acp.2883>
- Vorstenbosch, M., Bolhuis, S., van Kuppeveld, S., Kooloos, J., & Laan, R. (2011). Properties of publications on anatomy in medical education literature. *Anatomical Sciences Education*, 4(2), 105–114. <http://doi.org/10.1002/ase.210>
- Wilhelmsson, N., Dahlgren, L. O., Hult, H., Scheja, M., Lonka, K., & Josephson, A. (2009). The anatomy of learning anatomy. *Advances in Health Sciences Education*, 15(2), 153–165. <http://doi.org/10.1007/s10459-009-9171-5>
- Wittrock, M. C. (2010). Learning as a Generative Process. *Educational Psychologist*, 45(1), 40–45. <http://doi.org/10.1080/00461520903433554>
- Yuan, L., & Powell, S. (2013). *MOOCs and Open Education: Implications for Higher Education A white paper. JISC Cetis* (pp. 1–21). University of Bolton.
- Yue, C., Kim, J., Ogawa, R., Stark, E., & Kim, S. (2013). Applying the cognitive theory of multimedia learning: an analysis of medical animations. *Medical Education*, 47(4), 375–387. <http://doi.org/10.1111/medu.12090>
- Zemsky, R. (2014). With a MOOC MOOC Here and a MOOC MOOC There, Here a MOOC, There a MOOC, Everywhere a MOOC MOOC. *The Journal of General Education*, 63(4), 237–243. <http://doi.org/10.1353/jge.2014.0029>

**Appendix A. Pretest Questionnaire and Quiz**

1. What is your current stage in the Sydney Medical Program?
  - a) Stage 1
  - b) Stage 2
  - c) Stage 3
2. Sex
  - a) Male
  - b) Female
  - c) Other
3. What is your age? [ Free text input]
4. Which of the following best describes the field in which you received your highest degree?
  - a) Medical Science
  - b) Science
  - c) Technology
  - d) Engineering
  - e) Computing
  - f) Mathematics
  - g) Healthcare
  - h) Business
  - i) Other
5. How would you describe your level of exposure to anatomy prior to the Sydney Medical Program?
  - a) No Exposure (i.e., I had never studied anatomy before)
  - b) Little Exposure (i.e., Anatomy was a component of a unit of study I took)

- c) Some Exposure (i.e., I took a single unit of study in anatomy)
  - d) Moderate Exposure (i.e., I took several units of study in anatomy)
  - e) Detailed Exposure (i.e., I majored in anatomy)
6. How would you rate your confidence in understanding anatomy?
- a) Not confident
  - b) Moderately confident
  - c) Confident
  - d) Very confident
  - e) Extremely confident
7. How would you rate your confidence in understanding the muscles of the back?
- a) Not confident
  - b) Moderately confident
  - c) Confident
  - d) Very confident
  - e) Extremely confident
8. How would you rate the difficulty of learning anatomy?
- a) Not difficult
  - b) Moderately difficult
  - c) Difficult
  - d) Very difficult
  - e) Extremely difficult
9. What in particular do you find difficult about learning anatomy? [Free text answer]
10. How would you rate the difficulty of learning the muscles of the back?
- a) Not difficult

- b) Moderately difficult
- c) Difficult
- d) Very difficult
- e) Extremely difficult

11. What in particular do you find difficult about learning the muscles of the back? [Free text answer]

12. [EXPERIMENT 2 ONLY] Have you watched the online lecture 'Introduction to the vertebral column'?

- a) Yes
- b) No

### 13. CLASSIFICATION

Intrinsic back muscles include:

- A. Trapezius
- B. Latissimus dorsi
- C. Longissimus (answer)
- D. Levator scapulae
- E. Rhomboids

### 14. INNERVATION

Intrinsic back muscles are innervated by:

- A. Spinal nerves
- B. Anterior roots of spinal nerves
- C. Posterior roots of spinal nerves
- D. Anterior rami of spinal nerves

- E. Posterior rami of spinal nerves (answer)

#### 15. ORGANISATION OF

Which of the following is the deepest of the transversospinalis muscles group?

- A. Suboccipital muscles
- B. Splenius capitis muscle
- C. Semispinalis muscles
- D. Rotatores muscles (answer)
- E. Multifidus muscles

#### 16. RELATIVE POSITION

The most superficial of the true back muscles in the cervical region is:

- A. Suboccipital muscles
- B. Splenius capitis muscle (answer)
- C. Semispinalis muscles
- D. Trapezius
- E. Levator scapulae

#### 17. FUNCTION

The true back muscle that rotates the head to the same side is

- A. Trapezius
- B. Levator scapulae
- C. Splenius capitis (answer)

D. Semispinalis capitis

E. Spinalis

18. TRANSFER

Bending your trunk to the right hand side to pick up your keys requires which muscles?

A. Spinalis

B. Trapezius

C. Iliocostalis (answer)

D. Rotatores

E. Splenius capitis

**Appendix B. Post test Questionnaire and Quiz**

1. How would you rate your confidence in understanding anatomy?
  - a) Not confident
  - b) Moderately confident
  - c) Confident
  - d) Very confident
  - e) Extremely confident
  
2. How would you rate your confidence in understanding the muscles of the back?
  - a) Not confident
  - b) Moderately confident
  - c) Confident
  - d) Very confident
  - e) Extremely confident
  
3. How would you rate the difficulty of learning anatomy?
  - a) Not difficult
  - b) Moderately difficult
  - c) Difficult
  - d) Very difficult
  - e) Extremely difficult
  
4. How would you rate the ease of learning the muscles of the back?
  - a) Not ease
  - b) Moderately easy
  - c) easy



- d) Very easy
- e) Extremely easy

## 5. CLASSIFICATION

Intrinsic back muscles include:

- F. Trapezius
- G. Latissimus dorsi
- H. Longissimus (answer)
- I. Levator scapulae
- J. Rhomboids

## 6. INNERVATION

Intrinsic back muscles are innervated by:

- F. Spinal nerves
- G. Anterior roots of spinal nerves
- H. Posterior roots of spinal nerves
- I. Anterior rami of spinal nerves
- J. Posterior rami of spinal nerves (answer)

## 7. ORGANISATION OF

Which of the following is the deepest of the transversospinalis muscles group?

- F. Suboccipital muscles
- G. Splenius capitis muscle
- H. Semispinalis muscles
- I. Rotatores muscles (answer)

- J. Multifidus muscles

#### 8. RELATIVE POSITION

The most superficial of the true back muscles in the cervical region is:

- F. Suboccipital muscles
- G. Splenius capitis muscle (answer)
- H. Semispinalis muscles
- I. Trapezius
- J. Levator scapulae

#### 9. FUNCTION

The true back muscle that rotates the head to the same side is

- F. Trapezius
- G. Levator scapulae
- H. Splenius capitis (answer)
- I. Semispinalis capitis
- J. Spinalis

#### 10. TRANSFER

Bending your trunk to the right hand side to pick up your keys requires which muscles?

- F. Spinalis
- G. Trapezius
- H. Iliocostalis (answer)

I. Rotatores

J. Splenius capitis

### Appendix C. Instructions for Experiment 1.

#### Video Evaluation

This evaluation includes 3 parts. The parts are revealed sequentially as you complete them. Please be sure that you have clicked on the colour that you have been assigned.

#### Walkthrough

1. Login to Blackboard > Anatomy Online Learning Hub

bb shortcut in toolbar

2. Go to Video Evaluation page

3. Select your colour and begin

#### Nitty Gritty - Blackboard 'Gotchyas'

Part 1 & 3 open in new tabs

After you submit, you click 'OK' twice

When you complete Part 2, be sure to click 'Mark Reviewed'

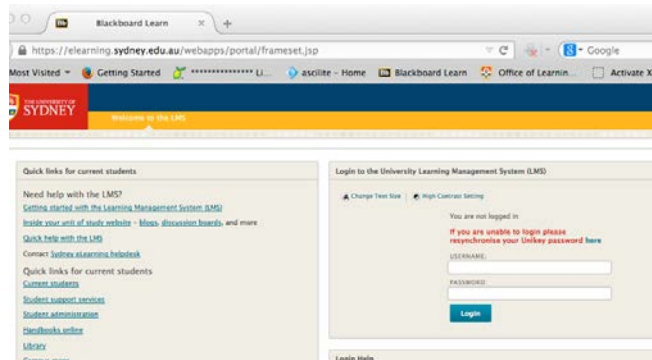
All three Parts

**Appendix D. Instructions for Experiment 2.**

**VIDEO EVALUATION**

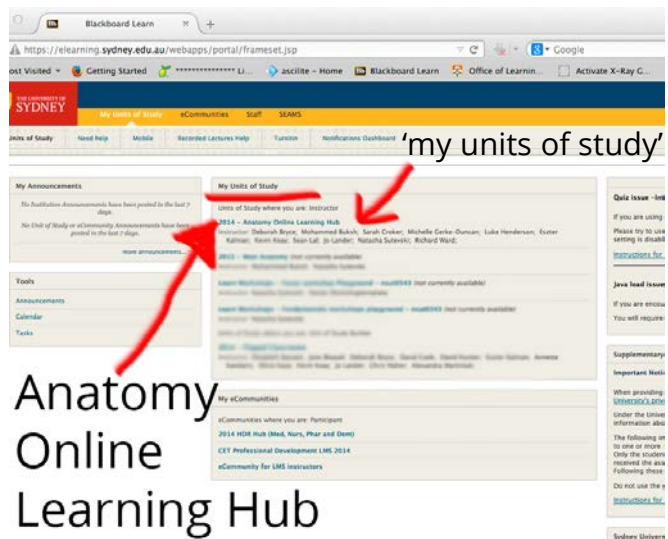
**STEP 1**

Login to  
BLACKBOARD



**STEP 2**

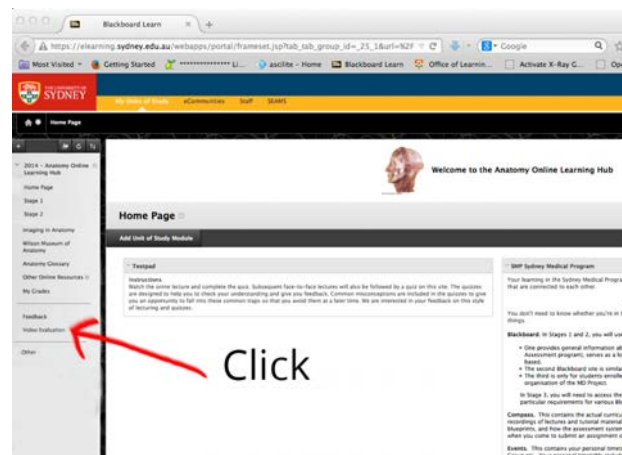
ENTER  
'Anatomy  
Online  
Learning Hub'



Anatomy  
Online  
Learning Hub

**STEP 3**

Click on '**Video  
Evaluation**'  
link in LEFT hand  
menu  
Then begin...



Click

Appendix E. Control Video Stills

## Intrinsic Back Muscles

**Superficial Group**

**Splenius Group**  
Splenius capitis  
Splenius cervicis

Extension, rotation to same side

**Intermediate Group**

**Erector spinae Group:**  
Iliocostalis, Longissimus, Spinalis

Extension, lateral flexion (not spinalis)

**Deep Group**

**Transversospinalis Group:**  
Semispinalis mm. (more superficial, C° & T°)  
Multifidus (all regions)  
Rotatores (deepest, only in T°)

Extension, rotation to opposite side  
Multifidus stabilises lumbar spine (like guy ropes)

**Others...**

**Suboccipital mm**  
Control movements at AO and AA Jts (extension, rotation, LF)

**Others include...**  
Interspinalis mm  
Intertransverse mm

**Cross-section of back muscles**

Rotatores, Multifidus, Semispinalis → **Transversospinalis**  
Spinalis, Longissimus, Iliocostalis → **Erector spinae**

Moore, K.L. & Dalley, A.F. 2006, *Clinically Oriented Anatomy*, 5th Edn, Lippincott Williams & Wilkins, Philadelphia, except intermuscular group image: Netter, F. (1997), *Interactive Atlas of Anatomy on CD*, Celba-Geigy Corporation, New Jersey. Adapted by: Ross, L.M. & Langmaid, E.D. (eds), 2006 *Thieme Atlas of Anatomy: General Anatomy and Microanatomical System*, Thieme, Stuttgart.

## Back Muscles

**Extrinsic back mm.**

**Extrinsic (move UL)**

e.g., lat. dorsi, rhomboids, trapezius, serratus posterior superior & inferior mm.  
More superficial position on the back.  
Move the *upper limb & rib cage*.

Innervated by **anterior rami** of spinal nerves (except trapezius – cranial nerve XI)

**Intrinsic back mm.**

Splenius capitis, Semispinalis capitis, Erector spinae, Suboccipital mm (headaches)

e.g., erector spinae, semispinalis, splenius mm.  
Move the *back & head*.  
Longer muscles important for movement

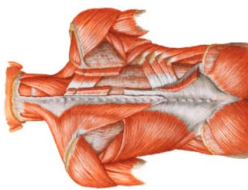
Innervated by **posterior rami** of spinal nerves

Shorter, deeper muscles more important for proprioception.

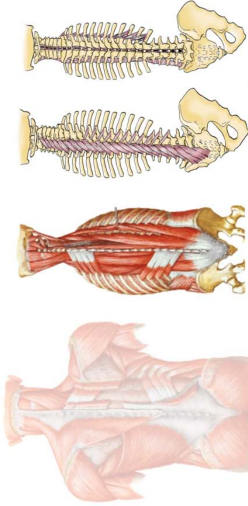
Netter, F. (1995), *Interactive Atlas of Anatomy on CD*, Celba-Geigy Corporation, New Jersey. Moore, K.L. & Dalley, A.F. 2006, *Clinically Oriented Anatomy*, 5th Edn, Lippincott Williams & Wilkins, Philadelphia.

**Appendix F. Experiment Video Stills**

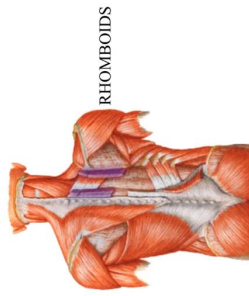
1. The muscles of the back can be divided into two groups: Extrinsic & Intrinsic



2. The muscles of the back can be divided into two groups: Extrinsic & Intrinsic



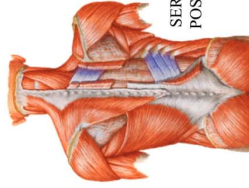
3. Extrinsic muscles move the upper limb and rib cage



1. MOVE THE UPPER LIMB

RHOMBOIDS

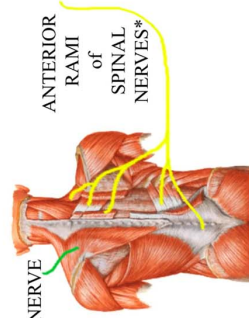
4. Extrinsic muscles move the upper limb and rib cage



2. MOVE THE RIBCAGE

SERRATUS POSTERIOR

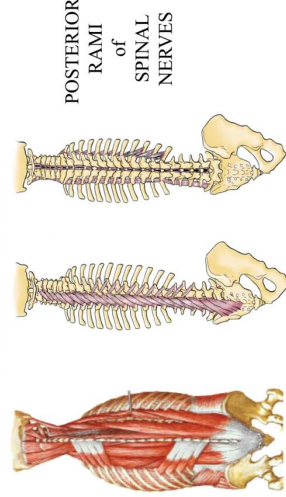
5. EXTRINSIC Innervations



CRANIAL NERVE XI

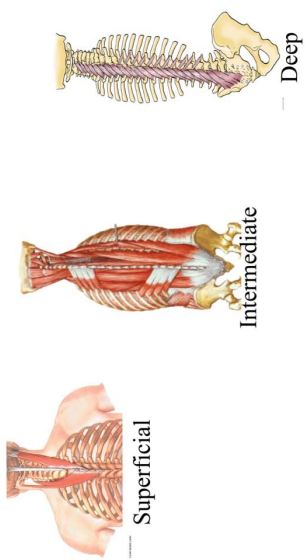
ANTERIOR RAMI of SPINAL NERVES\*

6. INTRINSIC Innervations

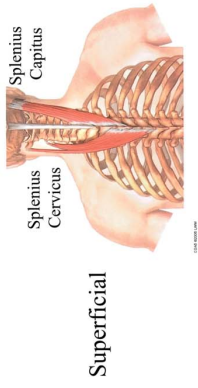


POSTERIOR RAMI of SPINAL NERVES

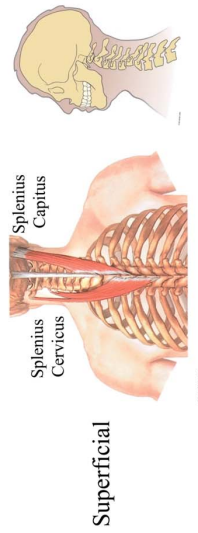
7. Intrinsic muscles move the vertebral column & head



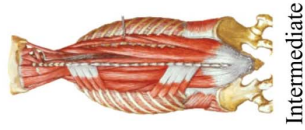
8. Intrinsic muscles move the vertebral column & head



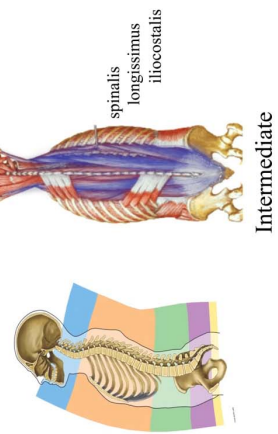
9. Intrinsic muscles move the vertebral column & head



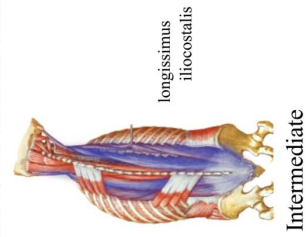
10. Intrinsic muscles move the vertebral column & head



11. Intrinsic muscles move the vertebral column & head



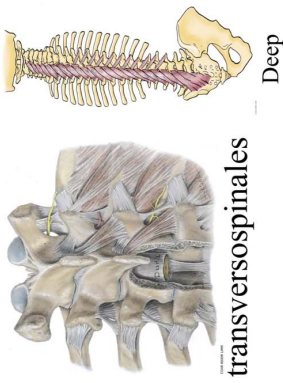
12. Intrinsic muscles move the vertebral column & head





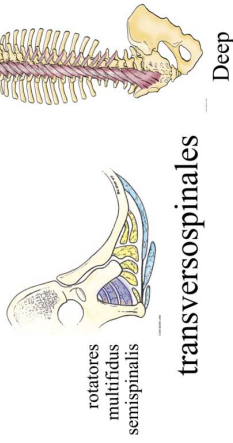
13

Intrinsic muscles move the vertebral column & head



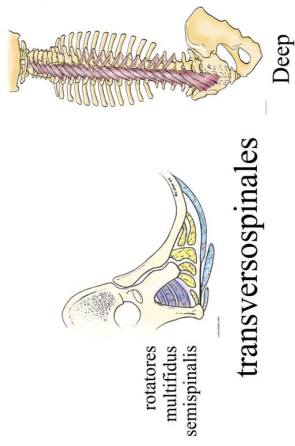
14

Intrinsic muscles move the vertebral column & head



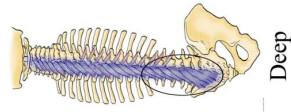
15

Intrinsic muscles move the vertebral column & head



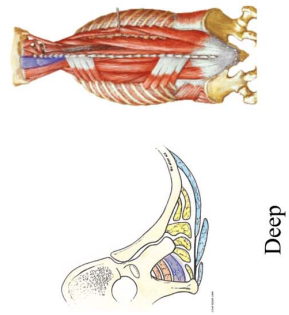
16

Intrinsic muscles move the vertebral column & head



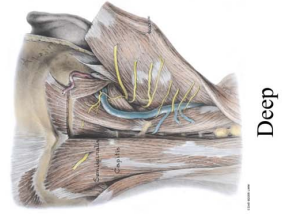
17

Intrinsic muscles move the vertebral column & head



18

Intrinsic muscles move the vertebral column & head



### **Appendix G. Mnemonic exploration**

A keystone idea in the evolution of DCT is the conceptual-peg hypothesis. The hypothesis tests memory and recall based on the concreteness of words. Historically, the use of imagery in memorisation may be traced back to Simondes (500BC) and a technique named the method of loci (Paivio, 2007). The idea is to image a path through a familiar place and visualise ideas/concepts at points along this path. By forming concrete visualisations, ideally that interact in unusual ways, one is capable of recalling the content with ease. Overtime, as outlined by Paivio (2007), the technique extended to creating numerically ordered objects which serve as pegs on which to 'hang' items for recall. The most popular technique is one that uses concrete words which rhyme with numbers (one-bun, two-shoe, etc). The content to be remembered is then imaged interacting with the concrete word or peg e.g., to pair scissors with shoe one might imagine scissors cutting a shoe in half and straining against the leather. Thus, the number two, will trigger the association shoe, which would trigger the association of scissors. This mnemonic technique was experimentally tested by Paivio with special emphasis on concreteness and imagery as key variables. The first experiment conducted in 1965 tested 16 pairs of words. The pairs consisted of a stimulus word and a response word in various combinations; concrete-concrete, concrete-abstract, abstract-concrete, abstract-abstract. It was found that recall of the response word was strongest with a

concrete stimulus word and weakest with an abstract stimulus word. Further experiments retested this effect while accounting for variables such as meaningfulness, imagery value and context and, with methods that included cued-recall and free recall. When experiments extended to using picture-word and picture-picture pairs as stimulus and response items, a picture superiority effect was consistently found, with an exception in that sometimes recall was higher for picture-word than picture-picture pairs for children.

Understanding these effects using DCT also lead to the integration hypothesis. Where a pair of items need to be memorised, it is more effective in both cued and free recall if the two items are interacting or integrated when visualised than if they are not. Begg (1973, cited in Paivio 2007) defined this effect in terms of redintegration, that is, the degree that a stimulus triggers recall versus free recall. He found that recall increased from free to cued recall when participants were instructed to visualise integration between word pairs as opposed to visualising them as separate entities. Similarly, using word pairs that could be imaged concretely as a whole e.g., white horse versus abstract pairs e.g., basic truth yielded advantages in cued recall. The reasoning being that the concrete word pairs are imaged as an integrated whole so that cueing redintegrates more effectively than abstract pairs. The probability of retrieving an item from memory is therefore increased with dual encoding or referential processing.

Given the ancient technique, an idea of creating a memory palace by using locations in the Anderson Stuart building was explored. Although such a technique makes for vivid imagery the cost in time of producing the resource was high and the externally presented associations may not assist memory as much as a personally generated version would. A storyboard featuring characters travelling through the building follows.

1.

### Muscles of the Back A story of pairs and trios

2.

Adapt instructions

*"Method of Loci"*  
The Method of Loci has been proven to significantly increase the effectiveness of memory. In this method, memory is established from pictures and images. If we wish to remember an object, we find that object in our image, and then place it in a location. If we wish to remember a list of objects, then we must make a path and the many locations. This method may be used to imagine a familiar environment and place the imagined objects inside. Thus, you can give an object an image and place the object like a code."  
From *LOOPS*, ©2012, art. 6. (2012)

3.

The following events & characters  
are based on real parts of the  
human body

4.



5.



6.



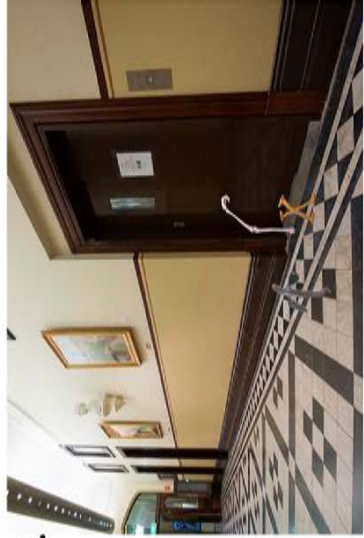
7.

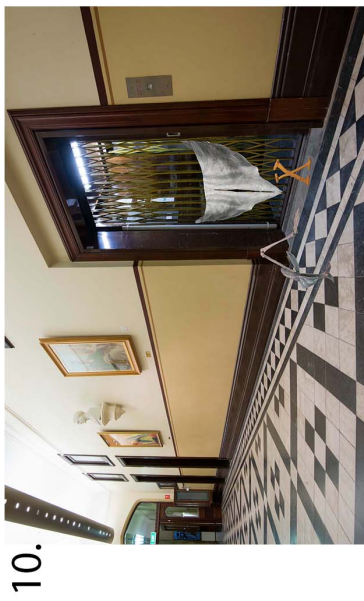


8.



9.





10.



11



12



13.



14



15



16.



17



18



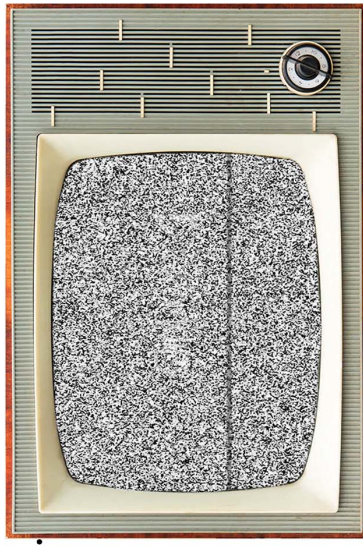
21



20



19.



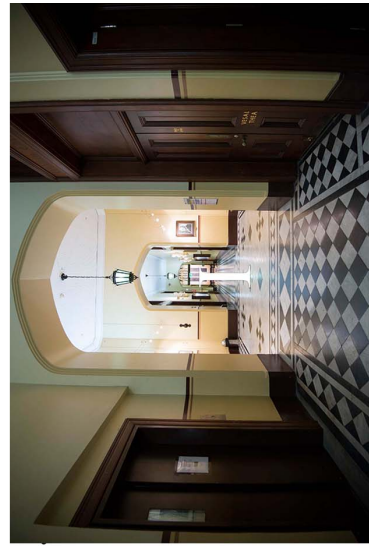
24.



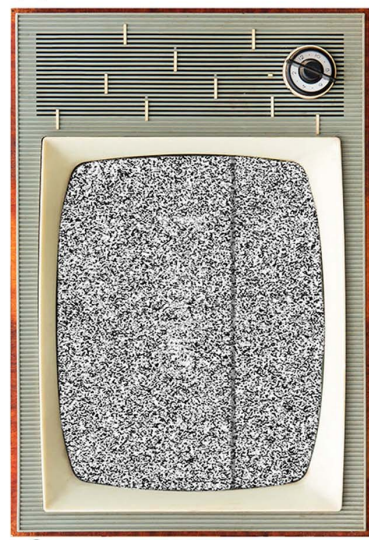
23



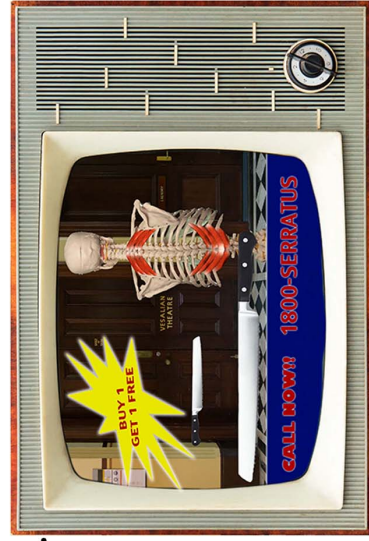
22.



27



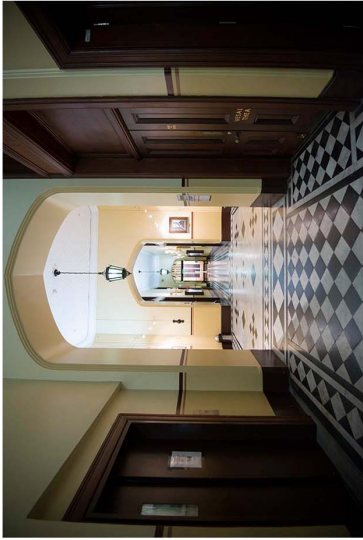
26



25.



30.



29.



28.



33.



32.



31.



36.

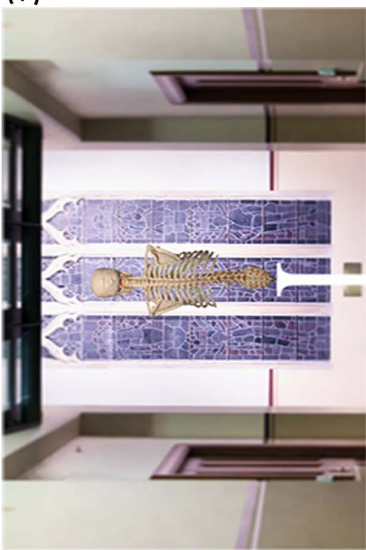


35.

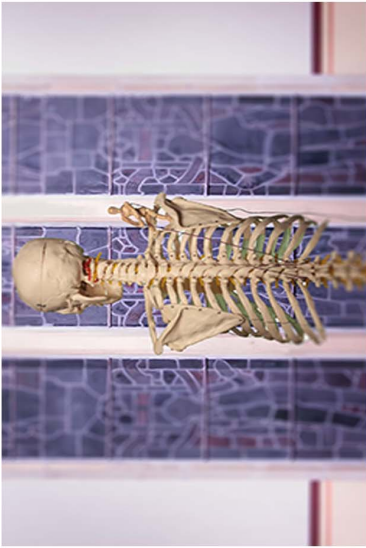


34.





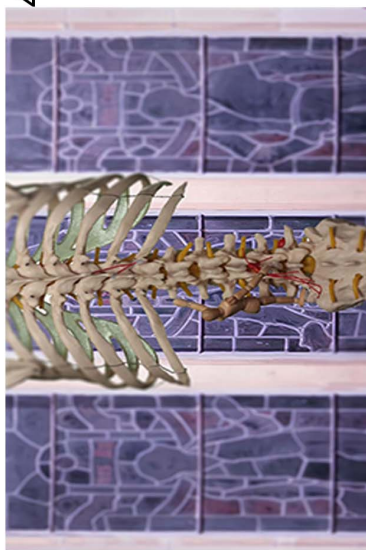
37.



38



39



40.

41

The three wooden figures climb into the skull, skull snaps shut.  
The End.