



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Exploring Spatial Ability in Veterinary Students and the Relationship to Teaching Methods**

**Julie Dickson**

# Abstract

Anatomy is the foundation to most medical disciplines, and a sound understanding is required to underpin many aspects ranging from routine physical examination to complex surgical procedures. For qualified veterinary surgeons, anatomy knowledge is pivotal. The vast number of species dealt with, along with the fact that immediately after graduation veterinary graduates are permitted as a ‘Day One’ competency to perform surgical procedures further emphasises the necessity for strong anatomy knowledge. Anatomy by its very nature is a spatial subject; the human or animal body lives in a three-dimensional space and is, in itself, three-dimensional. It requires the mental manipulation of complex structures and an understanding of their topographical relationships. This spatially demanding aspect of anatomy is challenging to veterinary students, yet, despite the importance of the subject and the known challenges of learning anatomy, limited studies have researched the possible relationship of spatial ability to anatomy learning in veterinary medical students. The overall aim of this project was to investigate the possible relationship of spatial ability to the learning of anatomy, and the influence of different teaching methods on this learning in first-year veterinary medical students.

Three well-validated tests of spatial ability (Card Rotation Test, Mental Rotation Test, and Surface Development Test) were given to four cohorts of undergraduate first-year students. Of the four cohorts, two cohorts were first-year veterinary medical students from the same academic institution but different academic year (University of Edinburgh first-year veterinary medical students cohort 1 (UoE Vet 1) and cohort 2 (UoE Vet 2)), one cohort of first-year veterinary medical students from a different academic institution to allow for between academic institution comparisons (University of Bristol first-year veterinary medical students (UoB Vet)), and lastly, one control cohort of first-year students studying psychology (University of Edinburgh first-year psychology students (UoE Psych)) to account for the re-test

effect. All four cohorts were given the exact same spatial ability tests at the start of the academic year and 15-16 weeks later. The cohorts UoE Vet 1 and UoE Vet 2 additionally received a two-dimensional teaching method and a novel three-dimensional spatial teaching method respectively, and scores on an in-course spatial MCQ assessment and their end-of-course examinations were collected for comparison.

The first study of this project aimed to investigate the baseline spatial ability of veterinary students to assess how consistent this ability is within one academic institution (UoE Vet 1 and UoE Vet 2), across institutions (UoB Vet), and to a control cohort of students who do not learn anatomy (UoE Psych). The second study compared a two-dimensional teaching method (UoE Vet 1) to a novel teaching method purposefully designed to teach anatomy spatially (UoE Vet 2), with the aim of improving anatomy knowledge and understanding. The third study involved the design and validation of a multiple choice question (MCQ) assessment to examine anatomy knowledge spatially and non-spatially and examined whether teaching spatially impacted on performance on the MCQ (UoE Vet 1 and UoE Vet 2). The fourth study investigated whether spatial ability improved in students who learn anatomy from two academic institutions (UoE Vet 1, UoE Vet 2, and UoB Vet) to a control cohort of psychology students (UoE Psych) who do not learn anatomy to account for the re-test effect observed with spatial ability tests. The fourth study also investigated whether the novel spatial teaching method had any additional significant impact on spatial ability improvement. The fifth study of this project qualitatively analysed student views and experiences of anatomy learning, the MCQ assessment, and spatial ability to provide a more in-depth qualitative insight (UoE Vet 1 and UoE Vet 2).

The novel results of this project are as follows. An understanding that spatial ability appears to be relatively consistent across first-year veterinary medical students from the same academic institution and two different institutions (UoE Vet 1, UoE Vet 2,

and UoB Vet) ( $p > 0.05$ ). Comparison of spatial ability test scores of veterinary students to a control group of psychology students showed veterinary students scored higher on the Surface Development Test and exhibited a ceiling effect (OR = 1.85 – 1.69,  $p \leq 0.004$ ). The Mental Rotation Test exhibited gender differences with males scoring higher than females ( $p < 0.01$ ) except for the UoB Vet cohort. The UoE Psych cohort exhibited a gender difference for all three spatial ability tests ( $p < 0.05$ ). No statistical differences were observed for the demographic parameters handedness or age for each cohort.

The successful design and delivery of a novel spatial teaching method resulted in improved student experience and improved anatomy test scores for short answer questions (OR = 1.18,  $p = 0.040$ ) and an in-course oral exam (OR = 1.26,  $p = 0.005$ ) compared to a two-dimensional teaching method. While the two-dimensional teaching method showed improved scores for interpretation style questions (OR = 1.35,  $p < 0.001$ ) and in-course workbooks documenting dissection practicals (OR = 1.44,  $p < 0.001$ ). The successful design of a novel MCQ containing items testing anatomy spatially, with the MCQ significantly predicting student performances on end-of-course examinations (OR 0.86 – 1.09,  $p < 0.05$ ), and thus providing useful formative information to students on their progress.

Student spatial ability scores for cohorts UoE Vet 2 and UoB Vet improved for the Card Rotation Test (RR = 1.05,  $p = 0.049$  and RR = 1.06,  $p = 0.047$ , respectively). No improvement in spatial ability test scores was identified with the Mental Rotation Test for all four cohorts ( $p > 0.389$ ). While cohorts UoE Vet 1 and 2 exhibited improvement for the Surface Development Test (OR = 1.46,  $p = 0.014$  and OR = 1.86,  $p < 0.001$ , respectively). Overall indicating the 3D spatial teaching method improved spatial ability more than the 2D teaching method for the Card Rotation Test and Surface Development Test. However, post-hoc Tukey analysis directly comparing the post test scores of the two teaching methods identified no statistically significant differences. Further research should be carried out to investigate the 3D

spatial teaching methods effect to improve spatial ability. The last novel finding of this project is the first identification and proposal, through student views and experiences, that spatial thinking is a threshold concept for anatomy learning.

Overall, this research makes a novel contribution to veterinary anatomy education by exploring spatial ability in first-year veterinary medical students and relating it to their learning of anatomy both quantitatively and qualitatively. As one of the first detailed investigations into this aspect of cognitive ability in the context of Veterinary Medical Education, this work highlights the potential for this area of research to provide valuable insights into veterinary students learning and furthermore to inform curriculum and assessment development accordingly.



# Lay Summary

Veterinary surgeons treat and perform surgery on a variety of different species, all of which have differing anatomy. For instance, the cow has a large stomach with four different chambers (while the stomach of the dog has only one chamber), the horse has an extra pouch connected to the ear, called the guttural pouch, which houses many important structures such as vessels and nerves. The study of anatomy is fundamental for a veterinary surgeon to be knowledgeable of these differences and to diagnose and treat diseases appropriately. However, the study of the three-dimensional (3D) topic of anatomy is challenging. One hypothesised reason why this could be is related to a person's spatial ability. A person's spatial ability is their ability to think of objects and 3D space, and to be able to manipulate these structures mentally. Individual people vary in this ability and it has been shown that this ability can be improved through training. Limited studies have specifically investigated the spatial ability of veterinary students either on its own or in relation to the learning of anatomy, despite the need to know the anatomy of various species.

This project investigated this and found a 3D spatial teaching method, involving 3D printing and 3D computer models, improved students' experiences of anatomy learning, and marginally improved anatomy knowledge compared to a two-dimensional teaching method. A new multiple choice assessment designed to assess anatomy knowledge spatially and non-spatially was successfully designed which could predict students' performances on subsequent course examinations, thus providing the students with informative feedback on their progress with the subject knowledge and understanding. This project also proposes that approaching the topic of anatomy with a 3D sense is an important learning step to grasp.





## **Declaration**

“I have read and understood The University of Edinburgh guidelines on Plagiarism and declare that this written dissertation is my own work except where I have indicated otherwise by use of quotes and references.”

**Julie Dickson**



# Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Lay Summary</b> .....	<b>vii</b>
<b>Declaration</b> .....	<b>ix</b>
<b>Table of Figures</b> .....	<b>xvii</b>
<b>Table of Tables</b> .....	<b>xxiii</b>
<b>List of Abbreviations</b> .....	<b>xxvii</b>
<b>Acknowledgements</b> .....	<b>xxix</b>
<b>Chapter 1 Literature Review Part 1: Anatomy Education and Spatial Ability</b> ....	<b>1</b>
1.1 Introduction to Chapter 1 .....	1
1.2 Overview of Veterinary Education and Anatomy Education .....	2
1.2.1 A Chronology of Veterinary Anatomy Teaching .....	2
1.3 Spatial Ability and Anatomy Education .....	5
1.3.1 How are Spatial Ability and Anatomy Linked? .....	5
1.3.2 Spatial Ability and Anatomy Learning in Medical Students: Evidence for a Link .....	7
1.3.3 Spatial Ability and Anatomy Learning in Veterinary Medical Students...	14
1.3.4 3D and 2D Anatomy Teaching Methods .....	17
1.3.5 Assessment of Spatial Anatomy: Design of Multiple Choice Questions ..	27
1.4 Summary of Gaps Identified in the Literature .....	35
<b>Chapter 2 Literature Review Part 2: Spatial Ability and Human Intelligence</b> ..	<b>39</b>
2.1 Introduction to Chapter 2 .....	39
2.2 Brief Introduction to Human (General) Intelligence Research.....	40
2.2.1 Discovery of ‘g’ and the Structure of Human Intelligence.....	41
2.3 Brief Introduction to Spatial Ability Research .....	52
2.3.1 What is Spatial Ability?.....	52
2.3.2 Development of Spatial Ability Tests.....	53
2.3.3 Identification of a Spatial Ability Factor and Sub-Factors .....	55
2.3.4 Spatial Ability and Gender .....	60
2.3.5 Spatial Ability and the Re-test/Practice Effect .....	62
2.3.6 The Malleability of Spatial Ability and the Relationship of Spatial Ability to STEM .....	64

2.4 The Importance of Spatial Ability.....	66
2.5 Future Directions for Anatomy Education.....	69
2.6 Research Questions.....	70
<b>Chapter 3 Materials and Methods.....</b>	<b>71</b>
3.1 Introduction to Chapter 3.....	71
3.1.1 Overview of Research Study Design.....	71
3.2 Pilot Study.....	72
3.3 Testing of Participants' Spatial Ability.....	75
3.3.1 Spatial Ability Test Protocols.....	76
3.4 Demographic Details of Participants.....	80
3.4.1 Participants Numbers.....	81
3.5 Overview of The Animal Body 1 Course.....	83
3.5.1 Lecture Overview.....	88
3.5.2 Practical Class Overview.....	88
3.6 Assessment of The Animal Body 1 Course.....	89
3.6.1 End-of-Course Assessment.....	90
3.6.2 In-Course Assessment.....	91
3.7 2D Teaching Method.....	93
3.7.1 Lectures (2D).....	93
3.7.2 Practical Classes (2D).....	95
3.8 3D Spatial Teaching Method.....	96
3.8.1 Development of 3D Printed Models.....	98
3.8.2 Development of 3D Computer Models.....	104
3.8.3 Development of 3D Lectures.....	110
3.8.4 Development of Tutorials.....	118
3.8.5 Practical Classes (3D).....	123
3.9 Quantitative Data Analysis.....	123
3.9.1 Descriptive Statistics.....	124
3.9.2 Univariable Analysis.....	124
3.9.3 Multivariable Analysis.....	125
<b>Chapter 4 Quantitative Analysis of Baseline Spatial Ability and Demographic Parameters.....</b>	<b>129</b>

4.1 Introduction to Chapter 4 .....	129
4.2 Data Analysis .....	131
4.2.1 Descriptive Statistics .....	131
4.2.2 Within Cohort Comparison.....	132
4.2.3 Between Cohort Comparison.....	133
4.3 Results.....	134
4.3.1 Descriptive Statistics .....	134
4.3.2 Within Cohort Comparison.....	144
4.3.3 Between Cohort Comparison.....	155
4.4 Discussion.....	159
<b>Chapter 5 Comparison of a 2D and 3D Spatial Anatomy Teaching Method....</b>	<b>167</b>
5.1 Introduction to Chapter 5 .....	167
5.2 Data Analysis .....	171
5.2.1 Descriptive Statistics .....	171
5.2.2 Between Teaching Method Comparison .....	172
5.3 Results.....	174
5.3.1 Descriptive Statistics .....	174
5.3.2 Between Teaching Method Comparison .....	176
5.4 Discussion.....	184
5.4.1 Does teaching anatomy using a traditional 2D method or 3D spatial method improve students' anatomy knowledge and understanding? .....	185
5.4.2 Is spatial ability a predictor of success in anatomy examinations? .....	191
<b>Chapter 6 Assessing Non-spatial and Spatial Anatomy Knowledge and Understanding.....</b>	<b>195</b>
6.1 Introduction to Chapter 6 .....	195
6.2 Method of Spatial MCQ Design .....	197
6.3 Data Analysis .....	204
6.3.1 Descriptive Statistics .....	204
6.3.2 Between Teaching Method Comparison on Spatial MCQ .....	204
6.3.3 Spatial MCQ Score and Relationship to End-of-Course Examination Scores .....	206
6.3.4 Post-examination Item Analysis of Spatial MCQ .....	208

6.4 Results .....	209
6.4.1 Descriptive Analysis .....	209
6.4.2 Between Teaching Method Comparison on Spatial MCQ.....	215
6.4.3 Spatial MCQ Score and Relationship to End-of-Course Examination Score .....	218
6.4.4 Post-examination Analysis of Spatial MCQ .....	230
6.5 Discussion .....	233
6.5.1 Design of a spatial MCQ.....	233
6.5.2 Post-examination Analysis of the Spatial MCQ .....	235
6.5.3 Between Teaching Method Comparison on Spatial MCQ.....	240
6.5.4 Spatial MCQ Score and Relationship to End-of-Course Examination Scores .....	241
<b>Chapter 7 Comparison of Changes in Spatial Ability Following Two Different Anatomy Teaching Methods .....</b>	<b>245</b>
7.1 Introduction to Chapter 7 .....	245
7.2 Data Analyses.....	249
7.2.1 Descriptive Statistics .....	249
7.2.2 Baseline Comparisons .....	250
7.2.3 Comparison of Post Spatial Ability Test Scores .....	251
7.3 Results .....	252
7.3.1 Baseline Comparisons.....	252
7.3.2 Correlations .....	263
7.3.3 Comparison of Post Spatial Ability Test Scores .....	269
7.4 Discussion .....	276
7.4.1 Does spatial ability improve with the learning of anatomy?.....	276
7.4.2 Is there a difference between a 2D or 3D spatial teaching method on the improvement of spatial ability? .....	283
7.4.3 Study Design .....	285
<b>Chapter 8 Qualitative Analysis of Veterinary Student Experiences of a Spatial MCQ and a 2D or 3D Anatomy Teaching Method .....</b>	<b>289</b>
8.1 Introduction .....	289
8.2 Methods of Qualitative Analysis.....	289

8.2.1 Thematic Analysis Approach .....	290
8.2.2 Importance of Reflexivity .....	291
8.2.3 Researcher's Background .....	292
8.2.4 Researcher's Perspective .....	293
8.3 Thematic Analysis of UoE Vet 1 (2D) .....	295
8.3.1 Main Themes Identified.....	295
8.3.2 Theme 1: Thinking Spatially and having a 3D Sense of Anatomy is Difficult and Challenging .....	296
8.3.3 Theme 2: 2D Helps for Knowledge, Not for Figuring Out Anatomy .....	308
8.4 Thematic Analysis of UoE Vet 2 (3D spatial).....	312
8.4.1 Main Themes Identified.....	312
8.4.2 Theme 1: Thinking Spatially and having a 3D Sense of Anatomy is Difficult and Challenging .....	313
8.4.3 Theme 2: 2D Helps for Knowledge and Memorisation.....	328
8.5 Discussion.....	328
8.5.1 Comparison with Other Qualitative Research Findings .....	329
8.5.2 Familiarity of Structures .....	332
8.5.3 Discussion of an Anatomy Threshold Concept .....	333
<b>Chapter 9 General Discussion .....</b>	<b>337</b>
9.1 Overview of Chapter 9.....	337
9.2 Discussion of Findings in Relation to the Literature .....	337
9.3 Limitations of Research .....	342
9.4 Future Directions of Research .....	344
<b>References.....</b>	<b>347</b>
<b>Appendix 1 – Participant Consent Forms .....</b>	<b>371</b>
<b>Appendix 2 - Participant Demographic Form .....</b>	<b>377</b>
<b>Appendix 3 - Participant PhD Information Sheet .....</b>	<b>379</b>
<b>Appendix 4 - Anatomy Workbook Marking Sheet.....</b>	<b>383</b>
<b>Appendix 5 - 3D Teaching Method Tutorials .....</b>	<b>385</b>
<b>Appendix 6 - Focus Group Interview Questions.....</b>	<b>401</b>
<b>Appendix 7 - Spatial Multiple Choice Questions.....</b>	<b>403</b>
<b>Appendix 8 – Abstracts and Awards .....</b>	<b>423</b>





# Table of Figures

Figure 2.1 Example of group factors or cognitive domains from the WAIS-IV intelligence test. \*All of the group factors, designated by a circle, will correlate to one another (not shown on diagram). .....43

Figure 2.2 Venn diagram explaining Spearman's two-factor theory of intelligence. g = general ability, s = specific abilities. ....45

Figure 2.3 Diagram of the fluid-crystallised model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005. ....47

Figure 2.4 Diagram of Carroll's three-stratum model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005. ....48

Figure 2.5 Diagram of Vernon's verbal-perceptual model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005. ....49

Figure 2.6 Diagram of the VPR model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005. ....50

Figure 2.7 Box summarising the salient points from human intelligence research literature. ....51

Figure 3.1 Flow diagram of research protocol for each cohort, SA = spatial ability, UG = undergraduate. ....71

Figure 3.2 Histograms of each spatial ability test used in the pilot study, A = CRT, B = MRT, C = SDT, and D = PFT. ....74

Figure 3.3 Example practice question from Card Rotation Test. Copyright © 1976 Educational Testing Service. [www.ets.org](http://www.ets.org). Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred. ....77

Figure 3.4 Example of practice question of 3D Mental Rotation Test. Copyright permission granted by Peters et al. (1995). ....78

Figure 3.5 Example of Surface Development Test. Copyright © 1976 Educational Testing Service. [www.ets.org](http://www.ets.org). Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication

is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.....	79
Figure 3.6 Number of participants for UoE Vet 1. ....	81
Figure 3.7 Number of participants for UoE Vet 2. ....	82
Figure 3.8 Number of participants for UoB Vet. ....	82
Figure 3.9 Number of participants for UoE Psych.....	83
Figure 3.10 Curriculum outline for the R(D)SVS 5 year BVM&S programme.....	84
Figure 3.11 Division of The Animal Body 1 course.....	84
Figure 3.12 Lecture slide from PowerPoint presentation handwritten during 2D teaching. ....	93
Figure 3.13 Diagram from 2D teaching method of stifle joint highlighting dashed lines to represent depth and spatial relations. ....	94
Figure 3.14 Example radiographs with orthogonal views from 3D spatial teaching method.....	96
Figure 3.15 Example of diagrams showing different orientations initially considered for use for the 3D teaching method.....	97
Figure 3.16 Photograph of 3D printer printing canine shoulder, stifle and hip models. ....	99
Figure 3.17 Half scale 3D printed stifle model of option 2 flexibility. Photograph courtesy of Richard Collins.....	102
Figure 3.18 Photographs of the shoulder (top left), elbow (top right), hip (bottom left) and stifle (bottom right) 3D printed joint models. Photographs courtesy of Richard Collins. ....	103
Figure 3.19 Screenshot of Sketchfab viewer embedded in LEARN.....	104
Figure 3.20 Shoulder joint model showing extra features not on 3D printed model. ....	105
Figure 3.21 Example of colour rendering on small carpal bones. ....	107
Figure 3.22 Example of translucent structure in 3D model to aid learning. ....	108
Figure 3.23 Example of key used to identify more structures due to limitations of annotation numbers per model. ....	109
Figure 3.24 Comparison of order of teaching. ....	112

Figure 3.25 Example of 2D image drawn (A), flat 3D images (B and D), and 3D computer model (C) used to help explain the pharynx.....	115
Figure 3.26 Comparison of introduction lectures. Left is 2D and right is 3D.....	116
Figure 3.27 Comparison of head section. * = also 2D diagram. Left is 2D and right is 3D. ....	116
Figure 3.28 Comparison of introduction to joints section. * = also 2D diagram. Left is 2D and right is 3D.....	116
Figure 3.29 Comparison of forelimb section. * = also 2D diagram. Left is 2D and right is 3D. ....	117
Figure 3.30 Comparison of pharynx section. * = also 2D diagram. Left is 2D and right is 3D. ....	117
Figure 3.31 Comparison of hindlimb section. * = also 2D diagram. Left is 2D and right is 3D. ....	118
Figure 3.32 Logistics of 3D tutorials.....	119
Figure 3.33 Ground rules for tutorial 1.....	120
Figure 3.34 Photograph of students using coloured cord to trace nerves on an articulated forelimb skeleton. ....	121
Figure 3.35 Histogram showing skewed distribution of SDT Pre for each cohort. .	126
Figure 4.1 Distribution of participant scores with grey printed MRT test papers compared to rest of cohort. ....	134
Figure 4.2 Comparison of histograms for each cohort by spatial ability test.....	136
Figure 4.3 MRT Pre score distribution divided by handedness. ....	139
Figure 4.4 SDT Pre score distribution divided by gender, F = female, M = male. ..	141
Figure 4.5 SDT Pre score distribution divided by handedness.....	142
Figure 4.6 CRT Pre score by gender for each cohort, bar = median, dot = mean (points jittered). ....	145
Figure 4.7 MRT Pre score by gender for each cohort, bar = median, dot = mean (points jittered). ....	146
Figure 4.8 SDT Pre score by gender for each cohort, bar = median, dot = mean (points jittered). ....	147
Figure 4.9 Comparison of handedness for CRT Pre for each cohort, bar = median, dot = mean (points jittered). ....	148

Figure 4.10 Comparison of handedness for MRT Pre scores for each cohort, bar = median, dot = mean (points jittered). .....	149
Figure 4.11 Comparison of handedness for SDT Pre scores for each cohort, bar = median, dot = mean (points jittered). .....	150
Figure 4.12 Age by spatial ability test for cohort UoE Vet 1, regression line and 95% confidence interval included. ....	151
Figure 4.13 Age by spatial ability test for cohort UoE Vet 2, regression line and 95% confidence interval included. ....	151
Figure 4.14 Age by spatial ability test for cohort UoB Vet, regression line and 95% confidence interval included. ....	152
Figure 4.15 Age by spatial ability test for cohort UoE Psych, regression line and 95% confidence interval included. ....	152
Figure 4.16 CRT Pre score comparison between all four cohorts. ....	155
Figure 4.17 Effect of gender with CRT Pre score for each cohort. ....	156
Figure 4.18 MRT Pre score comparison between all four cohorts. ....	157
Figure 4.19 Effect of gender with MRT Pre score for each cohort. ....	158
Figure 4.20 SDT Pre score comparison between all four cohorts. ....	159
Figure 5.1 Comparison of exam parameter between teaching method/cohorts. ....	176
Figure 5.2 Comparison of AB2 score between teaching method/cohort. ....	177
Figure 5.3 Exam parameter score as a function of CRT Pre score, regression line with 95% confidence interval included.....	182
Figure 5.4 Exam parameter score as a function of MRT Pre score, regression line with 95% confidence interval included.....	183
Figure 5.5 Exam parameter score as a function of SDT Pre score, regression line with 95% confidence interval included.....	184
Figure 6.1 Example of spatial MCQ containing real photograph of canine hindlimb cross section. ....	202
Figure 6.2 Example spatial MCQ containing diagrammatic representation of canine hindlimb cross section.....	202
Figure 6.3 Diagrammatic representation of Diff score. ....	206
Figure 6.4 Histograms of nSNQ score for cohort UoE Vet 1 and UoE Vet 2. ....	209
Figure 6.5. Histograms for SNQ for cohort UoE Vet 1 and UoE Vet 2. ....	210

Figure 6.6 Histograms of spatial MCQ Diff scores for UoE Vet 1 and UoE Vet 2.	211
Figure 6.7 Correlation matrix graph, the colour of the tile indicates the size of the correlation, x = not statistically significant.	213
Figure 6.8 Correlation matrix graph, the colour of the tile indicates the size of the correlation, x= not statistically significant.	213
Figure 6.9 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) nSNQ scores.	215
Figure 6.10 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) SNQ scores.	216
Figure 6.11 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) Diff scores.	217
Figure 6.12 Gender vs examination parameter score.	220
Figure 6.13 Handedness vs examination parameter score.	221
Figure 6.14 Total spatial MCQ score as a predictor of short answer questions, regression line with 95% confidence interval included.	223
Figure 6.15 Total spatial MCQ score as a predictor of total exam score, regression line with 95% confidence interval included.	224
Figure 6.16 Total spatial MCQ score as a predictor of spot exam score, regression line with 95% confidence interval included.	225
Figure 6.17 Total spatial MCQ score as a predictor of interpretation exam score, regression line with 95% confidence interval included.	226
Figure 6.18 Total spatial MCQ score as a predictor of MCQ exam score, regression line with 95% confidence interval included.	227
Figure 6.19 Total spatial MCQ score as a predictor of oral score, regression line with 95% confidence interval included.	228
Figure 6.20 Total spatial MCQ score as a predictor of in-course score, regression line with 95% confidence interval included.	229
Figure 6.21 Question 6 from the spatial MCQ	237
Figure 6.22 Question 11 from the spatial MCQ.	237
Figure 6.23 Question 22 from the spatial MCQ.	237
Figure 6.24 Question 26 from the spatial MCQ.	238
Figure 6.25 Question 28 from the spatial MCQ.	239
Figure 7.1 Diagrammatic representation of Diff score.	250
Figure 7.2 CRT Pre distribution per cohort divided by handedness.	255
Figure 7.3 Comparison of MRT Pre distributions divided by handedness.	256

Figure 7.4 Comparison of SDT Pre distribution divided by gender. ....	257
Figure 7.5 Comparison of SDT Pre distributions divided by handedness. ....	258
Figure 7.6 Distribution of grey printed MRT Pre test papers for UoE Vet and UoE Vet 2 for chapter 7. ....	259
Figure 7.7 Comparison of spatial ability Pre scores. ....	260
Figure 7.8 A = Comparison of CRT Post score for each cohort. B = CRT Post score vs CRT Pre score for each cohort. C= CRT Diff score vs CRT Pre score for each cohort. The key for all three graphs is as figure A.....	271
Figure 7.9 A= Comparison of MRT Post score for each cohort. B= MRT Post score vs MRT Pre score for each cohort. C= MRT Diff score vs CRT Pre score for each cohort. The key for all three graphs is as figure C.....	273
Figure 7.10 A= Comparison of SDT Post score for each cohort. B= SDT Post score vs SDT Pre score for each cohort. C= SDT Diff score vs SdT Pre score for each cohort. The key for all three graphs is as figure A.....	275
Figure 7.11 Regression slopes for CRT Post vs CRT Pre per cohort.. ....	280
Figure 7.12 Regression slopes for MRT Post vs MRT Pre per cohort. ....	281
Figure 7.13 Regression slopes for SDT Post vs SDT Pre per cohort.....	281
Figure 8.1. Example of a mind map I created from lectures while a veterinary student. ....	294
Figure 8.2. Diagram of the main themes and sub-themes identified from thematic analysis of UoE Vet 1 focus groups.....	295
Figure 8.3. Two main themes and associated sub-themes identified for UoE Vet 2. .....	312

## Table of Tables

Table 2.1 Summary of Spatial Ability Factors, as described in Hegarty and Waller (2005) and the addition of Uttal et al (2013). .....	57
Table 2.2 Linn and Petersen spatial ability categories used for the research of this thesis. ....	59
Table 3.1 Mean ( $\bar{x}$ ), median (M), Standard Deviation (SD) and range of scores for each spatial ability test used in the pilot study. ....	73
Table 3.2 Number and duration of anatomy lectures and practicals within AB1 per body region. Topics highlighted in blue are the topics modified for this research. ...	86
Table 3.3 Breakdown and proportional weight of the AB1 examinations. ....	89
Table 3.4 Differences in the number of examination questions between cohort UoE Vet 1 and UoE Vet 2 for gross anatomy only.....	91
Table 3.5 List of all 3D computer models created. * = also 3D printed.....	106
Table 4.1 Summary of GLM analysis of baseline parameters. ....	133
Table 4.2 Matrix comparing CRT Pre distributions for all possible pairwise combinations of cohorts by Kolgomorov-Smirnov test. ....	137
Table 4.3 Number and percentage of male, female, left and right handed participants in each cohort, F = female, M = male, L = left, R = right. ....	137
Table 4.4 Matrix comparing MRT Pre distributions for all possible pairwise combinations of cohorts by Kolgomorov-Smirnov test. ....	138
Table 4.5 Matrix comparing SDT Pre distributions for all possible pairwise combinations of cohorts by Kolgomorov-Smirnov test. ....	140
Table 4.6 Matrix correlation of SA tests for all four cohorts combined. ....	143
Table 4.7 Matrix correlation of spatial ability tests for cohort UoE Vet 1.....	143
Table 4.8 Matrix correlation of spatial ability tests for cohort UoE Vet 2.....	143
Table 4.9 Matrix correlation of spatial ability tests for cohort UoB Vet.....	144
Table 4.10 Matrix correlation of spatial ability tests for cohort UoE Psych. ....	144
Table 4.11 Comparison of Gender (female vs male) CRT Pre scores for each cohort. ....	145
Table 4.12 Comparison of Gender (female vs male) MRT Pre scores for each cohort. ....	146



Table 4.13 Comparison of Gender (female vs male) SDT Pre scores for each cohort. .....	147
Table 4.14 Comparison of handedness (left vs right) scores for CRT Pre for each cohort. ....	148
Table 4.15 Comparison of handedness (left vs. right) scores for MRT Pre for each cohort. ....	149
Table 4.16 Comparison of handedness (left vs. right) scores for SDT Pre for each cohort. ....	150
Table 4.17 Mean ( $\bar{x}$ ), median (M), range, sample size (n) and standard deviation (SD) for each continuous parameter. ....	154
Table 4.18 CRT Pre GLM <sub>P</sub> output after model simplification, ref = reference level. .....	155
Table 4.19 MRT Pre GLM <sub>P</sub> output after model simplification, ref = reference level. .....	157
Table 4.20 SDT Pre GLM <sub>B</sub> output after model simplification. ....	158
Table 5.1 Summary of GLM <sub>B</sub> analysis conducted for chapter 5. ....	173
Table 5.2 Correlation matrix for exam parameters and spatial ability tests. The top half of the matrix above the grey diagonal shows 2D (UoE Vet 1) correlations and below the diagonal 3D (UoE Vet 2) correlations. Yellow box = significant at $p < 0.05$ .....	175
Table 5.3 Mean ( $\bar{x}$ ), median (M), range, and standard deviation (SD) for each exam parameter as a proportion and AB2 score. ....	177
Table 5.4 Output of GLM <sub>B</sub> analyses after model simplification by highest p-value. Ref = reference level for categorical explanatory variable. ....	179
Table 6.1 Example spatial MCQ questions based on each spatial ability test. ....	201
Table 6.2 Summary of GLM <sub>P</sub> analyses for between cohort comparisons on spatial MCQ in chapter 6.....	205
Table 6.3 Summary of GLM <sub>B</sub> analyses with spatial MCQ score as a predictor of end-of-course examination parameter.....	207
Table 6.4 Correlation matrix of nSNQ, SNQ, CRT Pre, MRT Pre, and SDT Pre for UoE Vet 1 above the diagonal line, and for UoE Vet 2 below the diagonal line. Yellow = significant at $p < 0.05$ .....	212

Table 6.5 Dependent correlation comparisons for 2D (UoE Vet 1).	214
Table 6.6 Dependent correlation comparisons for 3D (UoE Vet 2).	214
Table 6.7 Median (M), range, SD = standard deviation, max score = maximum score, and n = number of students for nSNQ, SNQ and Diff scores on the spatial MCQ.	215
Table 6.8 Summary of GLM <sub>P</sub> output for nSNQ and SNQ, ref = reference level.	216
Table 6.9 Summary of GLM <sub>B</sub> output for spatial MCQ score predicting exam parameter score. a = extremely small values for OR and CI, ref = reference level.	219
Table 6.10 Updated summary of GLM <sub>B</sub> output for spatial MCQ score predicting oral score.	222
Table 6.11 KR-20, mean, standard deviation (SD), standard error of measurement (SEM) and total number of students (n) of the spatial MCQ for each cohort.	230
Table 6.12 Item-difficulty index, item-discrimination index and 2 proportion test result for each spatial MCQ question for each cohort. Questions 1 – 15 are non-spatial and questions 15 – 30 are spatial.	232
Table 7.1 Summary of GLM <sub>P/B</sub> analysis comparing baseline Pre-test score for each spatial ability test.	251
Table 7.2 Summary of GLM <sub>P/B</sub> analysis comparing Post test scores for each spatial ability test.	252
Table 7.3 Number and percentage of female, male, left and right-handed participants in each cohort, with Pre and Post test scores.	253
Table 7.4 Number (n), mean ( $\bar{x}$ ), median (M), standard deviation (SD) and range of Pre, Post and Diff test scores for each spatial ability test for each cohort.	254
Table 7.5 GLM <sub>P</sub> analysis for CRT Pre cohort comparison.	261
Table 7.6 GLM <sub>P</sub> analysis for MRT Pre cohort comparison.	261
Table 7.7 GLM <sub>B</sub> analysis for SDT Pre cohort comparison.	262
Table 7.8 Correlation matrix of each spatial ability test Pre, Post, and Diff score for cohort UoE Vet 1. Yellow = statistically significant.	264
Table 7.9 Correlation matrix of each spatial ability test Pre, Post, and Diff for cohort UoE Vet 2. Yellow = statistically significant.	265
Table 7.10 Correlation matrix of each spatial ability test Pre, Post, and Diff score for cohort UoB Vet. Yellow = statistically significant.	266

Table 7.11 Correlation matrix for each spatial ability test Pre, Post, and Diff score for cohort UoE Psych. Yellow = statistically significant. ....	267
Table 7.12 Summary of GLM <sub>P/B</sub> output for each spatial ability test Post score as the response variable.....	270
Table 7.13 Count and proportion of students with a negative, zero or positive diff score for each SA test.....	277

## List of Abbreviations

SA	Spatial ability
CRT	Card Rotation Test
MRT	Mental Rotation Test
SDT	Surface Development Test
UoE Vet 1	Cohort University of Edinburgh Veterinary Students academic year 2014/15
UoE Vet 2	Cohort University of Edinburgh Veterinary Students academic year 2015/16
UoB Vet	Cohort University of Bristol Veterinary Students Academic year 2016/17
UoE Psych	Cohort University of Edinburgh Psychology Students academic year 2017/18
SNQ	Spatial anatomy multiple choice questions
nSNQ	Non-spatial anatomy multiple choice questions



## Acknowledgements

I would like to thank the following for the support given to this research:

Prof. Ian Deary	Spatial ability tests and study design
Chloe Fawns-Ritchie	Spatial ability tests
Dr. Tobias Schwartz	CT scans
Brian Mather	3D modelling
Eoghan Clarkson	IT support
Richard Collins	3D printing and 3D modelling
Dr. Kirsty Hughes and Dr. Jill McKay	Data collection and focus groups
Dr. Darren Shaw	Statistical support
Prof. Sarah Baillie and Dr. Imogen Richens	UoB collaboration

I would also like to give a special thank you to my supervisors Prof. Susan Rhind and Dr. Andrew Gardiner for all their much appreciated support throughout the years, and also Dr. Stuart Ritchie for supporting the psychology aspect of this research. I would also like to especially thank Dr. Darren Shaw for patiently providing me with statistical support and mentoring as well as supporting me throughout the whole process.

Lastly thank you to my family for the never ending support they have given me.



# **Chapter 1 Literature Review Part 1: Anatomy Education and Spatial Ability**

## **1.1 Introduction to Chapter 1**

This chapter presents a review of the literature on spatial ability and anatomy education to provide the rationale for the research conducted in this thesis. The chapter has two sections, the first focuses on veterinary education and anatomy education, and the second section focuses on spatial ability and anatomy education.

The second section of this literature review specifically discusses how spatial ability and anatomy learning are linked, then discusses the relevant literature in the context of medical and veterinary students. The literature on two-dimensional and three-dimensional anatomy teaching methods and the literature on assessments of spatial anatomy knowledge (specifically using multiple-choice questions) are then discussed. The chapter ends with a summary of the gaps identified in the literature, which are the focus of the research presented in this thesis.

Throughout this thesis, the term ‘teaching spatially’ will be used to describe teaching that is designed to encourage learners to think spatially (use their spatial aspect of intelligence) to learn and problem-solve anatomy.



## **1.2 Overview of Veterinary Education and Anatomy Education**

In recent years, an increasing amount of research has been carried out focusing on the pedagogy related specifically to veterinary medical education (VME) thus establishing VME as a research discipline in its own right.

The Journal of Veterinary Medical Education began publication in 1974. Until 2009, educational research and development were presented either at medical education conferences, such as AMEE (the Association for Medical Education in Europe), at more general veterinary research events such as AVTRW (Association of Veterinary Teachers and Research Workers) annual conference or through Higher Education Academy sponsored events. In 2009, the first annual UK ‘VetEd’ symposium took place for veterinary educators with the aim of sharing ideas, innovations and best practice in veterinary education (Gardiner & Rhind, 2013). The symposium has been an annual event since 2009 and was hosted for the first time in continental Europe in 2018. In contrast, medical education has been established as a research area for considerably longer and as a result of this, and the relative size of the profession along with lack of funding for veterinary medical education research more has been published in medical education. Consequently, many of the articles cited in this thesis stem from medical education.

### **1.2.1 A Chronology of Veterinary Anatomy Teaching**

The history of veterinary anatomy education is vast, and a comprehensive analysis is outwith the scope of this thesis. However, an outline summary of changes to veterinary anatomy education over the years with the most salient points relevant to this thesis will be presented to frame the subsequent sections.

Anatomy has been a cornerstone of veterinary medical education since the beginning of the profession (as is the case with the sister profession of medicine). In 1929, lecturers taught 765 hours of anatomy in the Glasgow veterinary degree programme amounting to the largest number of hours for any course on the five-year programme and often resulted in students struggling to progress in the degree (Gardiner, 2010).

Traditionally the topic of anatomy in veterinary medicine and medicine has been taught using dissection and prosected specimens. Prosected specimens are anatomical specimens which are pre-dissected by a professional anatomist to demonstrate particular structures, whereas dissection involves the actual cutting of specimens by the students. Historically in veterinary medical education, the horse was the animal that students had the privilege of dissecting in small groups, due to the prominent use of the horse in everyday life at that time and thus it was the commonest patient of the veterinary surgeon. During this time in the 1920's other veterinary species were dissected, with the dog and ox close second to the horse. In the UK, the dog is now the commonest species dissected by students in most veterinary degree programmes and the base species upon which students learn the basics of anatomy before moving onto comparative anatomy. In some veterinary schools, comparative anatomy is taught from the very beginning of studies as a body systems based approach (Provo-Klimek, 2002).

In addition to the use of dissection and prosection in veterinary anatomy education, images and drawings have also been an important educational tool. Historically students would be required to draw the dissections they had performed or use drawings to help them in their studies. Over time, technology has advanced, and with the incorporation of digital cameras into mobile phones and other readily portable devices, photographs are a common way to document anatomy. However, in the researcher's experience, many students still currently attempt to draw anatomical structures to be used as study and revision aids. Furthermore, anatomy lectures are typically filled with drawings and diagrams to help explain anatomical concepts.

Drawings may partly be used because they can be easily altered or simplified to help show complex areas of anatomy and can be drawn 'on demand' in a classroom setting. In contrast, a photograph is less malleable, and it requires time to alter the original image using a computer.

Many factors have influenced both the approach to teaching and the overall status and position of anatomy in the curriculum over time (Drake *et al.*, 2009; McMenemy *et al.*, 2014; Yamine & Violato, 2015). Such factors include: the ethical sourcing of specimens; the health and safety issues of the embalming chemical formaldehyde (EU, 2012); the introduction of clinical and professional skills into degree programmes (typically associated with a reduction, or perceived reduction in time given to traditional 'pre-clinical' subjects); and the rapid advancement in learning technologies.

Additionally, with the advent of problem-based curricula in medical education, concern has arisen regarding the extent of anatomy knowledge achieved by students in such curricula (Langlois *et al.*, 2009). Traditional problem based learning curricula uses clinical problems for students to learn the relevant anatomy or physiology specific to the problem, with advantages such as improving problem-solving skills, learning in context, and developing self-directed learning skills. However, there are concerns that this style of curriculum does not give students a wide and deep enough transferable anatomy and physiology knowledge (Norman, 2009).

The introduction of equally important clinical and professional skills through integrated curricula (Drake *et al.*, 2009; Böckers, Mayer & Böckers, 2014; Bergman *et al.*, 2008; Drake, 2014) has resulted in a reduction in anatomy teaching hours, and the implication of students spending less time on anatomy is beginning to be recognised. Dissection is diminishing or disappearing in some courses and students are relying heavily on two-dimensional (2D) anatomical diagrams from textbooks

and the internet. This has spawned a new area of anatomy education utilising three-dimensional (3D) technologies to develop the students' understanding of anatomy without the possible need for dissection or prosected specimens (Lim *et al.*, 2016; McMenamin *et al.*, 2014; Preece *et al.*, 2013; Smith *et al.*, 2018; Garg, Norman & Sperotable, 2001; Küçük, Kapakin & Göktaş, 2016). During dissection students can manually manipulate and visually appraise the anatomy in front of their eyes - less mental generation or imagination is required. However, 2D textbook diagrams and 3D computer models require the students to generate the anatomical images mentally - an ability some are more naturally inclined towards than others. This difference along with the concern voiced by some that students lack sufficient anatomical knowledge (Bergman *et al.*, 2008) has resulted in several studies investigating spatial ability and anatomy learning in different medical disciplines.

## **1.3 Spatial Ability and Anatomy Education**

### **1.3.1 How are Spatial Ability and Anatomy Linked?**

Anatomy is an inherently spatial subject. The human or animal body occupies a three-dimensional space and is, in itself, three-dimensional. Anatomy is the study of the human and animal body concerned with the structure of systems, organs, and tissues both grossly and histologically. Learning anatomy involves developing an understanding of the positions and relationships of these structures and comparing these across species. The study of anatomy matches psychologist John Carroll's (1993:p.304) definition of spatial ability:

“[A]bilities in searching the visual field, apprehending the forms, shapes and positions of objects as visually perceived, forming mental representations of those forms, shapes and positions, and manipulating such representations mentally.”

Additionally, anatomy is a dynamic topic: organs can move within body cavities; muscles can contract and relax resulting in changes in movement, direction, and biomechanics; certain organs beat or move constantly, for example, intestinal peristalsis or rhythmic cardiac contractions. Organs can also change conformation within this three-dimensional environment into obscure, often unpredictable, pathological entities; ligaments can become torn to change the movements of joints, and cancerous masses can invade and disorientate normal anatomy.

Diagnostic imaging modalities such as radiography, ultrasound, magnetic resonance imaging (MRI), and computed tomography (CT) are used to investigate anatomical normalities and abnormalities. Most of these imaging modalities present 2D representations of internal 3D anatomical structures. These representations can be in the form of cross-sectional images taken at different angles and levels (ultrasound, CT and MRI), shadowgraphs showing structures lying on top of one another (radiography) and real-time cross-sectional images (echocardiogram).

Veterinarians and doctors perform physical examinations on almost every patient attended. Routinely physical examinations include observation, palpation, auscultation, and percussion - techniques that rely on the use of the clinician's spatial ability, sometimes augmented by additional equipment (for example an otoscope or ophthalmoscope to examine the internal ear and eye respectively). These initial physical examinations most commonly lead onto further investigation of patient anatomy, often involving some of the imaging modalities described above, as well as many other tests and procedures; biopsy, fine needle aspirate (FNA), electrocardiogram (ECG), blood sample, skin scrape, and dynamic tests to name a few.

All of these procedures require knowledge and appreciation of anatomy, and a few of them require extremely detailed anatomical awareness. For example, collecting a

simple jugular blood sample from a dog relies on an appreciation of where the jugular vein lies, where to apply occlusion of the vessel, and where to correctly position the needle (regarding both depth and orientation). A more complex skill, such as ultrasound-guided fine needle aspiration of the liver would require knowledge of the location of the liver in relation to other organs and body cavities, namely the negative pressure of the thorax. Additionally, an appreciation of the dynamic use of the ultrasonography imaging modality in relation to actions performed with biopsy tools, patient positioning, and an ability to cope with individual patient variations (e.g. differences in patient size or different species in veterinary patients) are required. Along with ensuring students know relevant anatomical structures to avoid.

### **1.3.2 Spatial Ability and Anatomy Learning in Medical Students: Evidence for a Link**

The majority of research in this field has been conducted with medical students (Garg, Norman & Sperotable, 2001; Lufler *et al.*, 2012; Sweeney, Hayes & Chiavaroli, 2014; Langlois *et al.*, 2009; Vorstenbosch *et al.*, 2013a), with only sporadic research reported with veterinary students (Provo, Lamar & Newby, 2002; Chatterjee, 2011; Gutierrez *et al.*, 2017; Preece *et al.*, 2013). Admission into dentistry schools in the United States requires students to take an entrance examination including a Perceptual Aptitude Test, assessing spatial ability (Hegarty *et al.*, 2007). Studies have also been conducted to investigate the spatial ability of graduates within postgraduate medical specialties such as surgery, which demand high spatial skills (Graham & Deary, 1999; Keehner *et al.*, 2004).

It has been reported that a student's spatial ability is an important predictor of success in learning anatomy (Garg, Norman & Sperotable, 2001; Lufler *et al.*, 2012; Rochford, 1985). Rochford (1985) investigated the spatial ability of second-year medical students and discovered that students with a low spatial ability (as measured

by a range of spatial ability tests) performed poorly on practical anatomy examinations, whereas students with a high spatial ability performed better.

If the underlying spatial ability of learners is linked to success in anatomy, a related question could be: does the learning of anatomy itself have the potential to improve a learner's spatial ability? Rochford found that about one-third of medical students commencing the anatomy course had a low spatial ability. However, 8-10 months later, only 7-10% of students were spatially struggling (although he does not show this data in his study) (Rochford, 1985). More recently, Vorstenbosch et al. (2013a) carried out a study investigating the effects of studying anatomy on performance on the Mental Rotation Test (Vandenberg & Kuse, 1978). In this study, they used the Mental Rotation Test (MRT) to assess the spatial ability of 242 naïve first-year medical students learning anatomy before and after teaching with no dissection. The medical students scores were then compared to the MRT scores of 258 first year educational science students learning research methods in the social sciences. The authors found that the medical students performed significantly better on the MRT test both pre and post-teaching, that both groups showed improvements between the two testings, but the improvement by the medical students was significantly higher than the educational sciences students. This study is consistent with the findings of Rochford (1985), and also demonstrates the 're-test effect' observed with spatial ability assessments (section 2.3.5): the practice of the spatial ability test improves subsequent scores (Goldberg *et al.*, 2015; Hausknecht *et al.*, 2007).

Before the study by Vorstenbosch et al. (2013a), Lufner et al. (2012) performed a similar study and tested the spatial ability of two first-year medical student cohorts (2008 and 2009) before and after teaching. The study did not include a naïve comparison control group nor had any exclusion criteria. Again, the Vandenberg and Kuse MRT was used, and the results of this were compared to the students' practical and written examination performances given at the end of each course section which involved dissection of a cadaver. The study showed that students in the highest

quartile of the MRT were 2.2 times more likely to score above 90% on practical examinations. The practical examination required students to identify structures on their cadavers, and removed from their cadavers, on prosected specimens and skeletons - all tasks that are spatially demanding. Additionally, Lufler et al. (2012) compared the results of the MRT to scores on the written examination and found similar results. These questions, although written, could still be spatially demanding, requiring students to mentally visualise and manipulate structures in order to construct an answer. Lufler et al.'s (2012) findings show that spatial ability is not only used to answer practical examination questions involving 3D specimens but also used to answer questions presented in a 2D format.

Rochford (1985) found that spatially aware students, as expected, tend to do well on spatially-demanding questions, but that these same students do not perform any better than others on non-spatially demanding questions. What exactly defines an anatomy question as spatial or non-spatial is, however, often open to debate (see section 1.3.5). Lufler et al. (2012) found that the students' MRT scores improved between the pre and post teaching tests, although there appears to be no control group to adjust for a re-test effect. It can, therefore, be concluded from both Lufler's and Vorstenbosch's studies that there could be a reciprocal relationship between spatial ability and the learning of anatomy (i.e., spatial ability helps to learn anatomy and the learning of anatomy helps to enhance spatial ability). Although, as Lufler et al. (2012) states, the phenomenon of anatomy learning itself improving spatial ability is unexplored and requires further investigation.

In addition to these studies with medical students, spatial ability has been explored in the context of dental education. Hegarty et al. (2009) investigated the reciprocal relationship between spatial ability and anatomy learning in dentistry students with the two research questions: does spatial ability enhance learning in dentistry?; and does studying dentistry enhance spatial ability? The authors conducted two studies comparing the spatial ability and cross-sectional interpretation ability of first-year



dentistry students, second-year dentistry students, fourth-year dentistry students, and psychology students (control group). Two spatial ability tests were used, and two cross-section tests were used; one of a novel random object, and one on teeth. The first study compared a group of second-year dentistry students to a group of fourth-year dentistry students using the two spatial ability tests and the novel object cross-section test. The authors found no relationship between the students' scores on the spatial ability tests and the scores on anatomy classes, leading the authors to conclude that spatial ability was not related to anatomy learning. There was also no evidence to support that learning dentistry enhanced spatial ability or the ability to imagine cross-sections of novel objects.

The second study compared first-year dentistry students, fourth-year dentistry students, and psychology students on the two spatial ability tests and both cross-section tests (of a novel object and of teeth). Again no relationship between spatial ability scores and scores on anatomy classes was found; indicating spatial ability does not enhance anatomy learning in dentistry (as was found in study 1). Furthermore, comparisons of the three groups scores on the tooth cross-section test revealed the fourth-year dentistry students performed better than the first-year dentistry students and the psychology students, and the psychology students had lower scores on the novel object cross-section test compared to the two dentistry groups. From these results, the authors concluded that spatial ability was not related to the learning of anatomy in dentistry, but that the learning of dentistry specifically improved the ability to imagine cross-sections of teeth and thus improved domain-specific knowledge (Hegarty *et al.*, 2009).

Interestingly, in the postgraduate context, studies have been conducted examining spatial ability in relation to specialties such as surgery. Surgeons are thought to naturally have a higher spatial ability due to the nature of the specialism. Keehner *et al.* (2004) compared the spatial ability, operative skill, and videoscopic experience of two groups of surgeons: highly experienced surgeons and beginners to laparoscopic

surgery. Laparoscopy surgery (also known as keyhole surgery or minimally invasive surgery) uses small skin incisions to allow surgical instruments and a laparoscopic camera to be inserted into the abdomen. The laparoscopic camera permits visualisation of the procedure on a TV monitor (videoscopic technique), rather than the direct visualisation of structures through a larger incision and is thus deemed to be more spatially challenging.

Keehner et al. (2004) found that in the novice group, a high spatial ability was a significant predictor of success in videoscopic technique compared to the more advanced surgeons, with the advanced surgeons having no significant difference between spatial ability and videoscopic technique. The experienced surgeons performed below average on the spatial ability test compared to a population of college students. This highlighted the fact that even individual surgeons with a low initial spatial ability can achieve the required skills to become proficient in surgery, suggesting caution about the use of spatial ability testing for admission to specialisms and undergraduate medical, veterinary, and dental professions. This difference also highlights the generation differences known as the Flynn effect - where scores on intelligence tests or components of intelligence tests such as spatial ability are improving over generations (Flynn, 2014).

Keehner et al. speculated that initially when learning a new surgical task, spatial ability can help, but experience takes over once the skill is acquired and learned, although this could vary depending on the difficulty or anatomical location of the surgical procedure. The authors suggest that teaching could, therefore, be tailored to accelerate the process of skill acquisition. Teaching spatially may especially help those with limited initial spatial ability, up to the point at which experience takes over.

It is significant that veterinarians graduate as surgeons and will be performing technically quite advanced surgery (e.g., laparotomy) from an early, even 'day 1' stage, when entering clinical practice. In human medicine, surgical training is specialty-driven and incorporates a prolonged period of supervision and skill acquisition regarding numbers of procedures carried out. What then are the implications of spatial ability for trainee veterinarians who do not have such structured postgraduate training programmes, and the time to develop? Bringing spatial ability training early into veterinary teaching may have value concerning acquiring basic competencies.

In the specific context of anatomy, studies have also been carried out examining the spatial ability of advanced and novice anatomists to help investigate whether spatial ability improves with the learning of anatomy. Fernandez, Dror, and Smith (2011) studied the spatial ability of novice, intermediate, and advanced anatomists. Novice anatomists were defined as first-year medical students, intermediates as fourth and fifth-year medical students, and advanced anatomists as lecturers with more than five years of teaching experience. The authors tested four components of spatial ability, and the intermediate and experienced anatomists were only marginally significantly better at one of these components. It could be argued that, initially, spatial ability improves with anatomy learning (Vorstenbosch *et al.*, 2013a; Lufler *et al.*, 2012), then as experience and cognitive practice are developed through the learning process, spatial ability becomes less important as mentally generated models of structures are created.

Additionally, teaching a spatially-demanding subject such as anatomy may further train the spatial ability of the teacher as different strategies are required to explain the same 3D concept to different students. Similarly, students starting in anatomy learning may approach a spatial problem differently to the lecturer and each other; some may mentally change their perspective of the object or move the object itself to gain a different perspective. The anatomy lecturer needs to transform the anatomical

spatial relations created in their head to convey concepts to the students in tangible ways.

Nguyen et al. (2013) looked at the techniques high and low spatial ability students use to approach answering spatial task problems. Students were either from a science or social science program, and the spatial ability of the students was measured using the Vandenberg and Kuse MRT (1978) together with a novel test developed by the authors called the Spatial Anatomy Task (Nguyen, Nelson & Wilson, 2012). The students answered a questionnaire detailing how they answered the spatial anatomy task questions, and it was found the answers were so varied that no one strategy could be said to be used (Nguyen *et al.*, 2013). Suggesting individuals vary and will take different approaches or strategies when tackling spatial anatomy.

Despite studies such as those described earlier highlighting links between spatial ability and anatomy learning, others have proved less conclusive. Lischka and Gittler (1997) studied the spatial ability of first-year medical students and compared this to performance on a Multiple Choice Question (MCQ) examination. They used the 3DC5 spatial ability test, which is infrequently cited in the literature, and therefore has questionable validity (see section 2.3.2). Their paper did not indicate the extent of students' anatomy knowledge before testing. They found a low positive correlation ( $r = 0.11$ ,  $p < 0.05$ ) between the 3DC5 spatial ability test and the MCQ examination score, but this only held for female students.

More recently, Sweeney, Hayes, and Chiavaroli (2014) looked at the relationship between spatial ability and anatomy examination performance across a curriculum. Their subjects were biomedical science students studying an anatomy course. Prior anatomy knowledge was not taken into account nor excluded. They measured the students' spatial ability with the 3DAT spatial ability test developed in 2011 (The University of Newcastle Australia, n.d.) a new test compared to other well-validated

spatial ability tests (see section 2.3.2). No significant correlation was found between the students' 3DAT test scores and anatomy examination results. In the teaching curriculum, this study did not use cadaver dissection but used models and digital images. These may already be sufficiently 3D to improve students' spatial ability during the early stages of anatomy learning, as reported by Berney et al. (2015). Hoyek et al. (2009) also found a non-significant correlation between scores on the Vandenberg and Kuse MRT and anatomy test scores. Furthermore, Keedy et al. (2011) found a similar result with the Vandenberg and Kuse MRT when given before a 30-minute teaching session, compared to a nine-item MCQ given straight after the teaching session.

On balance, whilst the literature is not definitive on the relationship between spatial ability and anatomy assessment performance, the methodological flaws described above (no control group, non-naïve participants), and the novel nature of the tests used, along with the possibility that the use of spatial ability diminishes as domain-specific knowledge develops, highlight the importance of understanding what is already known about spatial ability from the psychology literature (chapter 2), and designing adequate studies to reflect this.

### **1.3.3 Spatial Ability and Anatomy Learning in Veterinary Medical Students**

Despite the variety of species dealt with in veterinary medicine and the range of possible clinical procedures undertaken by general practising veterinarians, very little research has been conducted on the spatial ability of veterinary students concerning gross anatomy learning.

Provo, Lamar, and Newby (2002) looked at the use of cross-sections in the teaching of anatomy to first-year veterinary students in order to “enhance three-dimensional anatomical understanding.” Cross-sections were used in this study because of the increasing clinical use of cross-sectional imaging modalities. During the second study reported in this paper, the authors tested the spatial ability of students and gave the experimental group of first-year veterinary students a different resource for studying the canine head during dissection. The different teaching resource involved the use of a coloured photograph showing a cross-section of a real canine head, with an inset indicating the level at which the cross-section was taken. The students had to identify a list of thirteen structures on the cross-section by making a tracing on an acetate sheet; this was then marked and feedback received.

The control groups were instead given lateral and ventrodorsal radiographs of the thorax or abdomen, and again were required to identify structures and trace these onto an acetate sheet. A “head test” was given for the students to voluntarily take regardless of whether they had received the intervention or not. The test involved three parts: asking students to identify structures that would be visible on a cross-section of the head at a particular level and to sketch these on a cross-sectional diagram, and lastly to draw on two orthogonal views where a list of structures would be located. The students’ spatial ability was tested with the Purdue Visualisation of Rotations Test in August and again in April. This test appears to come under the category of mental rotation as defined by Linn and Peterson (1985) but is not identical to the MRT (Vandenberg & Kuse, 1978) (see section 2.3.2).

The study found no significant differences between the groups, although a significant positive correlation was found between spatial ability and examination performance. The authors hypothesise that this was because the intervention was not powerful enough because radiographs taken at orthogonal views require a spatial appreciation of structures being radiographed (particularly of the unsymmetrical thorax and abdomen). Also, the curriculum implemented a ‘big sibling’ system where students

shared test files, meaning students outwith the intervention group could have inadvertently seen the intervention used. Another possible reason could be the teaching intervention was not long enough. The questionnaires answered by the students highlighted the students felt they had gained new insights into spatial relationships. In this study by Provo, Lamar, and Newby female students had the largest increase in their spatial ability score. However no control group was incorporated into the study design.

More recently Gutierrez and colleagues (2017) tested the spatial ability of eighty-one first-year veterinary students with two spatial ability tests and one abstract reasoning test; 1. Guay's visualisation of views test (spatial ability test), 2. Mental rotation test (spatial ability test), and 3. Raven's advanced progressive matrices test (abstract reasoning test). The students were tested twice thirty-two weeks apart, and descriptive statistical analysis showed that student scores increased between the first and second sitting across each of the three tests, but this increase was only significant for the mental rotation and Raven's test. Importantly there was no control group to take account of the 're-test effect' (see section 2.3.5).

The study presented in this thesis involves three spatial ability tests assessing spatial ability across two sub-categories to assess the domain of spatial ability further, and to gain an understanding of the baseline spatial ability of veterinary students. A novel teaching method is also implemented to investigate whether different teaching practices can improve spatial ability, and a novel anatomy assessment is designed to examine anatomy knowledge and understanding spatially. Overall, this approach aims to develop the research and understanding of spatial ability and anatomy education in veterinary students.

### 1.3.4 3D and 2D Anatomy Teaching Methods

The teaching of anatomy has traditionally utilised textbooks with written descriptions of the relationships between organs and tissues, illustrated by flat, two-dimensional diagrams (often in black and white) and cadaver dissection. Recognition that anatomy teaching time is declining, and students vary in their spatial ability (Rochford, 1985; Garg, Norman & Sperotable, 2001; Lufler *et al.*, 2012; Bergman *et al.*, 2008; Drake *et al.*, 2009) has led to the development of three-dimensional teaching methods, with the aim of supporting the understanding of spatial relationships along with providing learning resources outwith the traditional dissection room (Berney *et al.*, 2015; Nguyen, Nelson & Wilson, 2012; Garg *et al.*, 1999a; Keedy *et al.*, 2011).

Anatomy education research has increasingly focused on the pedagogical value of 3D technology (Garg *et al.*, 1999b; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Küçük, Kapakin & Göktaş, 2016; Plumley *et al.*, 2013; Smith *et al.*, 2018). Several of these studies (Nguyen, Nelson & Wilson, 2012; Garg *et al.*, 1999b; Berney *et al.*, 2015) have also investigated the links of anatomy learning to a learner's spatial ability. Advantages of 3D technology include better visualisation of areas too difficult to be seen, unlimited viewing perspectives of spatial depth, non-permanent destruction of material, 24/7 access to material, and the creation of bespoke teaching aids (Tan *et al.*, 2012; Berney *et al.*, 2015).

A recent meta-analysis of thirty-six studies looked at the effectiveness of 3D technology used in anatomy teaching and concluded that 3D technology resulted in improved anatomy knowledge both factually and spatially, but particularly so for spatial knowledge (Yammine & Violato, 2015). Furthermore, regarding the qualitative experiences of the technology, several studies have shown that students have enjoyed learning from 3D technology and rated the technology highly (Preece *et al.*, 2013; Smith *et al.*, 2018; Tan *et al.*, 2012; Yammine & Violato, 2015).



Certain studies investigating the educational value of 3D technology have also designed spatial anatomy assessments as part of the study design (section 1.3.5). The introduction of 3D technology might be expected to improve students' anatomy knowledge and has been introduced amidst concerns of a decline in students' anatomy knowledge due to a reduction in anatomy teaching hours as discussed earlier (section 1.2.1). Three-dimensional teaching methods have utilized different approaches such as stereoscopic displays, 3D printing, 3D computer models, and augmented reality which are discussed further below. Studies on the educational value of 3D technology have taken different approaches with mixed results as evidenced below (Berney *et al.*, 2015; Garg *et al.*, 1999b; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a; Nguyen, Nelson & Wilson, 2012; Keedy *et al.*, 2011; Roach, Mistry & Wilson, 2014; Al-Khalili & Coppoc, 2014).

### **3D Computer Models**

The incorporation of 3D computer models has become an increasingly popular approach to provide a 3D teaching method with the benefit of access to relevant models outside the dissection laboratory. Studies have identified that learner-controlled models are significant for the learning of anatomy (Garg *et al.*, 1999a). Three of the most referenced studies on 3D computer models are by Garg *et al.* (1999b, 1999a; 2001). The researchers conducted a series of studies investigating the use of 3D computer models on the learning of the human carpal bones. The 3D models used in the studies exhibited varying degrees of rotation ( $10^0$ ,  $90^0$ , or  $180^0$ ), and they were either computer-controlled or learner-controlled.

The researcher's first study compared the  $90^0$  to the  $10^0$  learner-controlled models and found no significant difference (no p-value or test statistics provided). Next, the  $10^0$  to the  $180^0$  computer controlled models were compared with the two groups having no significant difference in spatial carpal knowledge as measured by a 50 question MCQ ( $F = 3.09$ ,  $p = 0.08$ ) (Garg *et al.*, 1999a). In the third study, the

researchers compared the 10<sup>0</sup> to 180<sup>0</sup> learner-controlled models and found the students who received the 10<sup>0</sup> rotating model had significantly higher carpal knowledge as measured by the same 50 question MCQ on spatial carpal knowledge (with a mean difference of 6%,  $t = 2.49$ ,  $p = 0.01$ ) (Garg, Norman & Sperotable, 2001). Spatial ability scores as measured by the MRT were shown to predict student scores on the MCQ significantly ( $p < 0.001$ ) and the model viewed was a significant predictor ( $p = 0.01$ ) although, the researchers do not specify which model.

From these studies, the authors concluded whether the model was computer- or learner-controlled was an important aspect to be considered, as the previous study (Garg *et al.*, 1999a) using computer-controlled models had no significant effect on student MCQ scores. The authors also concluded that students' spatial ability was an important predictor of success to the learning of anatomy. Similarly, Nicholson *et al.* (2006) found that using a learner-controlled model of the ear improved scores on a 3D anatomy quiz compared to a control group (83% vs. 65%,  $p < 0.001$ ).

One study investigated the combination of the control (or interactivity) of a model with the dynamism (dynamic or static). Nguyen, Nelson, and Wilson (2012) looked at the relationship between the spatial ability of learners (as measured by the MRT), control (or interactivity) of the teaching intervention, and dynamism (dynamic video or static image). The 3D anatomical model in this study was of an aorta, trachea, and oesophagus and the non-anatomical control model was of a cube. The study design involved dividing participants into high or low spatial ability, then further dividing into three groups within each (cube model control group, static 3D anatomical image group, or dynamic 3D anatomical model group). The authors then further subdivided the three groups into interactive (learner-controlled) or non-interactive (computer-controlled) to create a total of six groups each for high and low spatial ability. The creation of these groups resulted in lower sample sizes which, the researchers do discuss as a limitation. The dynamic group watched a video of the anatomical model rotating around the x, y, and z-axes, the static group watched the anatomical model switch between the six possible views of the model, and the

control group watched the cube model switch between the six possible views. Participants were also given a spatially designed anatomy test (named the SAT as discussed above, max score = 30) both before and after their respective teaching interventions.

The SAT involved three different question types: mental rotation of the anatomical model; identification of a view observed when the model was cut at a certain cross-section, and identification of the level the model was cross-sectioned for a given view. A significant interaction was found between spatial ability and dynamism on post-test scores ( $F(2, 48) = 3.38, p < 0.05$ ) with, specifically, the static cube control group having improved anatomical knowledge for low spatial ability participants (low  $\bar{x} = 20.63$  vs. high  $\bar{x} = 16.91$ ), and the dynamic 3D anatomical group having improved anatomical knowledge for high spatial ability participants (low  $\bar{x} = 17.48$  vs. high  $\bar{x} = 18.55$ ). The researchers concluded that different computer visualisations could hinder or help learners, because the simpler static cube model helped low spatial ability learners and the complex dynamic anatomical model helped high spatial ability learners. However, it should be noted the dichotomizing of spatial ability into high and low may exacerbate any differences because the exact granularity in spatial ability test scores is lost by reducing the data to a binary form.

Another study investigating bones, specifically the scapula and the flexion movement of the shoulder joint, found no difference between a static 2D representation and an auto-controlled 3D computer representation on questions on the scapula and shoulder joint movement ( $t < 1.695, p > 0.09$ ) (Berney *et al.*, 2015). Two cognitive tests were selected in this study to measure spatial ability, the widely used MRT to test mental rotation and the Group Embedded Figures Test (GEFT) to test spatial visualisation. However, the GEFT was designed to measure cognitive style (Witkin, Raskin & Oltman, 1971) by measuring field dependence or field independence (Colman, 2015). Although the GEFT has been shown to measure the spatial ability sub-category Flexibility of Closure (Carroll, 1993:p.339) as supposed to spatial visualisation (see section 2.3.3).

The two cognitive tests were given at the end of the teaching session meaning any improvement in spatial ability was not accounted for as no baseline Pre score before the intervention was measured, although the incorporation of a control group would allow a 'between-subjects' comparison (Uttal *et al.*, 2013). In this study neither the MRT nor GEFT were related to either the static/2D or dynamic/3D teaching methods. The anatomy knowledge assessment involved five different question types based on a six-step cognitive task analysis. The task analysis hypothesised learners' cognitive operations for tackling 3D anatomical concepts (see section 1.3.5 below for a further discussion of spatial anatomy assessments).

Another study that looked at the efficacy of a 3D computer model was by Tan *et al.* (2012). Participants of this study were randomised into two groups, one group received a computer-based tutorial with 2D images of the larynx, and the second group received the same computer-based tutorial with 3D interactive models. The 3D interactive models were static 2D images of 3D CT and MRI reconstructions of a human neck and made interactive by adding colour, audio, videos and clinical vignettes. Therefore the 3D interactive models have questionable true 3D interaction because they were non-rotatable either by the computer or the learner. Each group received the tutorial for 45 minutes and afterward was given a 20 question MCQ divided into factual and spatial questions.

The MRT was given to participants both before and after their respective tutorial. Again, there was no significant difference between MCQ scores for either the 2D or 3D group (2D  $\bar{x} = 15.5$  vs 3D  $\bar{x} = 15.7$ ,  $p = 0.722$ ), and no significant correlation between the MRT and the MCQ ( $r = 0.48$ ,  $p = 0.085$ ). However, it is not clear whether the author divided the questions into factual or spatial for the statistical analysis. They concluded that several factors could make 3D resources more effective: learner-controlled and interactive or rotatable.

Conversely, a self-controlled 3D model of the hepatobiliary system compared to static 2D images of the model improved anatomy knowledge for a post-intervention MCQ with factual and spatial questions (73.7% vs. 60.2 %,  $p = 0.03$ ). This was also found for the spatial questions only (67.4% vs. 53.5%,  $p = 0.04$ ). Nevertheless, when more accurately adjusted for pre-intervention test scores, there was no difference when compared to static 2D images of the model for the whole MCQ test ( $p = 0.330$ ) and the spatial questions ( $p = 0.320$ ) (Keedy *et al.*, 2011).

Although there are relatively few veterinary studies exploring anatomy and spatial ability in detail, some veterinary studies have explored the use of computer visualisations to enhance the spatial and factual understanding of functional anatomy. For example, Clements *et al.* (2013:p.30) developed a computer animation to depict the “spatial and factual understanding of functional anatomy of the [canine] stifle joint.”

This animation was used to teach undergraduate final year veterinary students about cranial cruciate ligament rupture and repair techniques. Students were either randomly assigned to watch an animated or non-animated version of the program (consisting of screenshots), and an assessment was given at the end of the session. The authors found that the animation was better at imparting certain aspects of cranial cruciate ligament rupture. For instance, students viewing the animated version had a better understanding of why and how the medial meniscus becomes injured following cranial cruciate ligament rupture. Furthermore, this study concluded that the bulk of students preferred 3D images regardless of whether they were animated or not. This last finding is particularly interesting as it might be expected that most final year students would be experienced in appreciating the 3D structure of a common joint such as the stifle due to familiarity, similarly to that explained in the context of dentistry students (Hegarty *et al.*, 2009). The use of 3D images to teach anatomy (rather than conventional 2D textbook images) therefore seems paramount.

## **Stereoscopic Displays, 3D-Printed Models and Augmented Reality**

Other 3D technology teaching methods to improve anatomy education have been used and researched. Stereoscopic displays, most commonly used for a 3D cinema experience, have been used to represent the vasculature of the head and neck (Cui *et al.*, 2016) and neuroanatomy (de Faria *et al.*, 2016; Plumley *et al.*, 2013), with positive and significant results to improve students' anatomy knowledge.

Stereoscopic displays have also been used in veterinary anatomy education as pre-dissection tools for the canine thorax, abdomen, and pelvis (Al-Khalili & Coppoc, 2014). A crossover research design was used for veterinary students to receive one of three pre-dissection teaching interventions (including; review laboratory manual, 2D stereoscopic video, or 3D stereoscopic video) for either the thorax, abdomen or pelvis dissection. A post-intervention quiz was given, and the researchers found the 2D groups performed better than the groups reviewing the laboratory manual ( $p = 0.028$ ), and there was no significant difference between the 2D and 3D groups ( $p > 0.05$ ). In comparison, one study found no differences in whether a monoscopic or stereoscopic display was used on scores on laparoscopic skills for low and high spatial ability learners (Roach, Mistry & Wilson, 2014).

The now widespread commercial use of 3D printers offers exciting possibilities for veterinary and medical anatomy education. 3D printing is a type of rapid prototyping and is described as an additive manufacturing process. Additive manufacturing means sequential layers of material are laid down on top of one another to build up a complete model. In contrast to subtractive manufacturing, where the model is created, and parts removed afterward to create the final product.

CT and MRI scans can be used to create a digital input to 3D printers in the form of STL files (stereolithography). The STL files created from CT or MRI scans can be modified by computer programs before printing, allowing endless possibilities

regarding modifications of the models themselves. 3D printers offer a range of invaluable resources to anatomy teaching in the form of working models, infinite obsolescence, education resources outside the dissection room, models that can be built to exact even unique specifications, and extraordinary pathological entities that can be replicated. However, how life-like are 3D printed models and do they make exact comparable replicas of good enough quality to be used for teaching?

McMenamin et al. (2014) looked at the use of 3D printed models for producing replicas of original specimens. They studied four varying anatomical structures: 1. CT scan of the upper human limb, 2. CT scan of warthog head sinuses (for negative airspace), 3. Contrast CT of coronary angiogram, 4. X-ray tomography of cochlea and vestibular apparatus. The authors concluded that highly realistic replicas could be made including small nerves and vessels and that negative air spaces such as sinuses and vessels were anatomically accurate. They also concluded that specimens needed to be an appropriate size for scanning, the scans themselves need to be of high quality, and 3D printed models should be used as an adjunct to dissection.

Smith et al. (2018) looked at the value of 3D printed models incorporated into a medical curriculum. They investigated whether 3D printed models of normal lungs (75% scale and printed in cream colour) improved test scores compared to a 2D approach (using 2D anatomical images) during a heart, lungs, and blood course module. The experiences of students and faculty members were also evaluated as well as the value students placed on being able to take 3D prints home (3D prints of the humerus, scapula, fused foot bones, and distal tibia and fibula were taken home by students), during the musculoskeletal and immunology module.

The authors found that students using the 3D printed lung models improved their post-test scores on ten short answer questions compared to the 2D group (3D  $\bar{x}$  = 7.46 vs 2D  $\bar{x}$  = 6.52,  $p = 0.001$ ) although the internal consistency of the test was a limitation (Cronbach's  $\alpha = 0.44$ ). Qualitatively it was found, as would be expected, the accuracy of the 3D printed models was important for showing structures correctly and the students suggested that augmentation with a key would

be useful. Additionally the students liked to take the models home and faculty members gave mixed reports on the 3D printed models, such as the heart model lacked the capacity to look inside the 'chambers', the accuracy of the 3D prints was thought to be better than commercial models, and the 3D prints were unable to be manipulated.

3D prints of external cardiac anatomy have been shown to be educationally beneficial. In a study reported by Lim et al. (2016), student performance on a pre-test (thirteen question MCQ of labeled prosected hearts) and a post-test (involving thirty short answer questions using labeled images to test structure identification, function, and relations), were compared across three self-directed teaching groups. The three self-directed teaching groups were: cardiac 3D prints only, cadaveric material only, and a combination of 3D prints and cadaveric material. Multivariable analysis identified that all groups performed significantly better on the post-test when adjusted for pre-test scores ( $p = 0.012$ ). The researchers do not mention whether the pre- or post-test were designed to test anatomy knowledge spatially. When analysed as a univariable analysis, only the 3D print group performed significantly better between pre- and post-tests ( $t = 3.50$ ,  $p = 0.003$ ).

The recent 3D technology of augmented reality (AR), where an image in the real world is augmented by a superimposed 3D computer image, shows initial promising results for the teaching of neuroanatomy to medical students. Küçük, Kapakin, and Göktaş (2016) found that learners' neuroanatomy knowledge, as measured by a thirty question non-spatial MCQ post-intervention, was higher when taught with supplemental AR compared to a control group (2D pictures, graphs, and text) that did not receive teaching using AR ( $F = 5.87$ ,  $p < 0.05$ ).

Interestingly, the use of physical models to teach anatomy has been shown to improve students learning and confidence in veterinary education significantly. Preece et al. (2013) investigated the use of a physical model to teach MRI anatomy



of the equine foot compared to a 3D computer model and traditional textbooks. Sixty-two third-year veterinarian students were randomly assigned to one of three teaching aids: textbooks (four most commonly borrowed anatomy textbooks from the library); 3D computer model (Glass Horse<sup>1</sup>); and physical equine foot model.

The physical equine foot model was made from an MRI scan using rapid prototyping that uses laser-cutting technology, and the model could be dismantled and rebuilt. Students spent time on their respective teaching aids, and their anatomical knowledge was tested by identifying structures on an MRI of the equine foot, with students shading various structures a specified colour on the MRI, with five MRI images used in total. The students' confidence was assessed at the beginning and end of the teaching session using a questionnaire.

The authors of the study found the students using the physical equine foot model performed significantly better on the MRI identification assessment compared to students using textbooks or the 3D computer model. Although this study did not measure the students' spatial ability, the student questionnaire on confidence included statements pertaining to spatial ability, such as: "I am confident in my ability to mentally visualise the anatomy of the equine foot in 3D" (Preece *et al.*, 2013:p.219). Surprisingly, the students level of confidence significantly improved in all three groups pre and post-teaching, but the confidence of students in the equine foot model group was significantly higher than the textbook or 3D computer model groups.

The questionnaire that was given to the students post-teaching also included questions regarding the students' enjoyment of the three different teaching aids, and these questions were all spatially orientated and included:

How helpful the students considered their respective teaching aid in facilitating:

---

<sup>1</sup> University of Georgia's College of Veterinary Medicine. 501 D. W. Brooks Drive, Athens, GA 30602, USA.

- Mental visualization of anatomical structures of the foot in 3D
- Orientation with regards to the MRI planes and slices
- Mental visualization of the MRI images in 3D

All three statements were significantly more positive for the MRI model group. However, for the computer model group, the second and third statements were slightly less positive, suggesting it did not help with MRI interpretation. Students agreed they found the equine foot model and computer model strongly enjoyable, easy to use and stimulating. Feedback from the student group using the textbooks was less positive.

These results suggest that physical models could allow better development of spatial skills since students using the model scored significantly higher on the MRI quiz when identifying structures such as ligaments and tendons (these are deemed more spatially-demanding compared to bones). An important comment made by the authors of this study is that despite the 3D computer model group performing poorly on the MRI assessment, their confidence levels improved, potentially giving a false positive impression to the students, highlighting the relatively subjective nature of such self-evaluation scales and the importance of mixed methods study designs.

### **1.3.5 Assessment of Spatial Anatomy: Design of Multiple Choice Questions**

Within the literature on spatial ability and anatomy learning there are discussions on the distinction of what anatomy knowledge is specifically spatial and what is purely factual with no associated spatial component. Defining the exact nature of non-spatial and spatial anatomy knowledge is challenging because of the inherent 3D nature of anatomy. Factual, non-spatial anatomy knowledge has been previously described as identification and simple organisation of structures (Yammine &

Violato, 2015), with other descriptions including anatomical terminology such as the names and functions of structures (Nguyen, Nelson & Wilson, 2012).

The term non-spatial knowledge implies no requirement to construct mental images or to spatially manipulate structures to acquire the knowledge, i.e., it is known and stored in memory. In contrast, spatial anatomy knowledge has previously been described as understanding complex organ relationships (Yammine & Violato, 2015), and the knowledge of visuospatial information such as the size, 3D shape, orientation and spatial location of structures (Nguyen, Nelson & Wilson, 2012). However, the categorisation of anatomy knowledge into non-spatial or spatial knowledge is not black and white.

Arguably, understanding the function of some anatomical structures (which above is categorised as non-spatial knowledge) could be considered, at least in part, to include a spatial element i.e. an appreciation of a structure's 3D shape and orientation could help a student to mentally manipulate and 'figure out' what the functions are if not already known. For example, simplistically the function of the triceps brachii muscle is to extend the elbow joint. If this knowledge was unknown to a learner they could mentally visualise the triceps brachii muscle's topographical location, the origins, and insertions of the muscles, and then the motion created upon muscular contraction, and use this information to deduce that its function is to extend the elbow joint. This newly acquired knowledge could then be stored in memory and retrieved when necessary, potentially shifting the knowledge to the non-spatial category. Alternatively, it is possible this knowledge was read and memorised from a book or a lecture involving no mental spatial manipulation of the muscle. Another possibility is that anatomy knowledge could be transient moving between the two categories during the learning process (Ackerman, 1988).

Perhaps the definitions are not purely based on anatomical topics but on how an anatomist (novice or expert, see section 1.3.2 above) thinks about the anatomy or

could think about the anatomy. Do they, for example, problem-solve and use their intelligence to figure out whether that structure is deep, or rotated, or related to other structures? This ‘figuring out’ will be different among individuals although an individual’s intelligence will naturally be used as part of the problem-solving process. Therefore, a better definition of spatial anatomy knowledge could incorporate definitions involving problem-solving processes such as spatial ability and intelligence to reflect the cognitive tasks required.

Anatomy education studies investigating spatial ability and anatomy learning have attempted to incorporate some form of spatial anatomy assessment into study designs. The method of spatial anatomy assessments has been varied including essays (Rochford, 1985), MCQs (Rochford, 1985; Guillot *et al.*, 2006; Hoyek *et al.*, 2009; Keedy *et al.*, 2011; Garg, Norman & Sperotable, 2001), practical examinations (Rochford, 1985; Provo, Lamar & Newby, 2002; Lufler *et al.*, 2012), 3D synthesis from 2D views (Provo, Lamar & Newby, 2002), drawing of views (Provo, Lamar & Newby, 2002), cross-sections (Hegarty *et al.*, 2009; Provo, Lamar & Newby, 2002), or a combination of these methods (Provo, Lamar & Newby, 2002; Tan *et al.*, 2012; Berney *et al.*, 2015).

A recent systematic review of twenty-one studies investigating spatial ability tests and anatomy knowledge assessments found no evidence of a significant relationship between spatial ability tests and students’ scores on essays or MCQs (Langlois *et al.*, 2017). However, significant relationships (with correlations ranging between 0.31 – 0.67) were noted between spatial ability tests and practical examinations, 3D synthesis from 2D views, drawing of views, and cross-sections (Langlois *et al.*, 2017). In this review, a total of twenty different spatial ability tests were used in studies with 85% using the Mental Rotation Test (MRT). The authors concluded that anatomy knowledge could be assessed both spatially and non-spatially, but the relationship between spatial ability tests and MCQs was unclear particularly for non-spatial MCQs, because the spatial ability tests used in studies were not always shown

to have a significant relationship to the MCQs (mainly shown by correlations between spatial ability test scores and MCQ scores).

Several studies have attempted to design MCQs to assess spatial anatomy knowledge (Rochford, 1985; Guillot *et al.*, 2006; Hoyek *et al.*, 2009; Keedy *et al.*, 2011; Schubert, Schnabel & Winkelmann, 2009; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Tan *et al.*, 2012). Below a brief overview is provided of each study to explain and contrast the approach taken to the design of a spatial MCQ presented in this thesis.

Nguyen, Nelson, and Wilson's (2012) MCQ design, called the Spatial Anatomy Test (SAT), involved thirty questions based on a 3D anatomical computer model of the aorta, trachea, and oesophagus (as previously mentioned). The SAT was divided into three sections of ten questions, and no non-spatial MCQs were designed. The first section involved mental rotation of the 3D model, the second section identification of the model in 2D cross-section, and the last identification of planes/levels corresponding to a selected cross-section. Spatial ability was measured by the MRT, and they found it correlated moderately and significantly with the Pre SAT scores (i.e., before students received a teaching intervention) ( $r = 0.64$ ,  $p < 0.05$ ). The authors concluded from this finding that spatial ability (as measured by the MRT) is related to spatial anatomy knowledge in general, and so the significant correlation to the MRT supports the spatial design of the SAT and highlights that as MRT score increases SAT score increases. Furthermore, the relationship of MRT score to the time spent on the SAT was negative ( $r = -0.67$ ,  $p < 0.05$ ) indicating students with a higher spatial ability spent less time on the SAT. Participants were divided into high and low spatial ability with high spatial ability participants found to score significantly higher on the SAT (high  $\bar{x} = 18.03$  vs. low  $\bar{x} = 12.04$ ,  $t = 4.54$ ,  $p < 0.05$ ), and to be significantly quicker at answering (high  $\bar{x} = 467$  vs. low  $\bar{x} = 521$ ,  $t = -4.50$ ,  $p < 0.05$ ).

Keedy et al. 's (2011) study exploring the educational value of a 2D or 3D teaching intervention on the hepatobiliary system included a post-test spatial MCQ. The authors designed a nine-item post-intervention MCQ test to assess anatomy knowledge factually (or non-spatially) and spatially, and the questions were of a higher difficulty than a pre-test (the pre-test was a ten-item MCQ assessing baseline abdominal and hepatobiliary knowledge and the authors do not give a reason for the difference). The distinction between the non-spatial and spatial MCQs was based on the subjective opinions of two expert radiologists on a five-point Likert scale (mean score of 1, 2, or 3 = factual knowledge and mean score of 4 or 5 = spatial knowledge). From this scale five questions were considered to assess spatial knowledge while four questions were considered to assess factual knowledge, all questions were text-only with no images. The internal consistency of the post-test ten items MCQ was 0.69.

The post-test questions were not significantly correlated to spatial ability (as measured by the MRT,  $r = 0.18$ ,  $p = 0.23$ ), even when divided into factual ( $r = 0.13$ ,  $p = 0.41$ ) or spatial ( $r = 0.17$ ,  $p = 0.27$ ). The 3D group performed significantly better on the post-test questions compared to the 2D group, however, a more accurate multivariable analysis incorporating pre-test scores, on a ten question MCQ to assess baseline basic knowledge, identified no significant differences for either the whole post-test questions or the spatial questions only ( $p = 0.33$  and  $p = 0.32$ , respectively as discussed above in section 1.3.4). The non-spatial questions were not analysed separately. The participants of this study were first-year and fourth-year medical students and the mixed knowledge of the sample may affect the results; the authors do not appear to have analysed the two years separately.

Hegarty et al. (2009) developed a test called the "Tooth Cross-section Test" designed to assess dentistry students' abilities to slice a tooth mentally. The Tooth Cross-section Test is not explicitly stated to be an MCQ, although the questions require participants to select the best response from a list of five options. The test involved twenty items, and non-spatial questions were not designed. To obtain the students'

spatial ability the authors administered the MRT, Guay's Visualisation of Views test (assesses the ability to visualize a 3D object from different perspectives), and obtained the students' entrance scores on the Perceptual Ability Test (entry to dentistry in the U.S requires taking the Perceptual Ability Test). The main aim of this study was to investigate whether spatial ability enhances the learning of dentistry and vice versa (as discussed above section in 1.3.2). However, the Tooth Cross-section Test was shown to correlate positively and significantly with the MRT ( $r = 0.37, p < 0.01$ ), Guay's Visualisation of Views ( $r = 0.29, p < 0.01$ ), and the Perceptual Ability Test ( $r = 0.45, p < 0.01$ ). The general reasoning ability of the dentistry students was also measured by the Abstract Reasoning Test, and the correlations remained significant when controlling for reasoning ability. These findings confirmed the Tooth Cross-section Test assessed dentistry anatomy spatially.

Two of Garg et al.'s studies (2001; 1999a) investigating 3D computer models and carpal knowledge involved a fifty item MCQ (internal consistency = 0.88), with spatial ability measured by the MRT. The fifty items MCQ assessed spatial carpal knowledge by questions requiring the identification of the carpal bone intersected by a pin, at different angles, going through the skin of the carpal region. The authors found a significant relationship between student scores on the MRT and performance on the carpal MCQ ( $p < 0.001$ ) for both studies. In another study by Garg et al. (1999b) a 36 item MCQ was designed with three different question types 'connect,' 'real,' and 'rod.' The connect questions involved a rod connecting two carpal bones, the real questions an image of a real hand with a probe pointing to a carpal bone, and rod questions depicting a rod intersecting one carpal bone. The questions also showed the carpus at different angles on the vertical axis. No differences in the MCQ scores were found between students given either a 10<sup>0</sup> or 90<sup>0</sup> learner-controlled carpal bone computer model. No statistical analysis indicating whether the MCQ examined anatomy knowledge spatially or not were presented.

Spatial MCQs have also been designed based on cognitive task analysis. Cognitive task analysis is the study of how people think and reason about complex problems

that require a great deal of cognitive activity (Crandall & Hoffman, 2013). Therefore it can be used to provide details of the cognitive processes used to solve problems such as learning functional anatomy. Berney et al. (2015) developed an MCQ with spatial MCQs, true/false, and sequence order questions based on steps two to six of a six-step cognitive task analysis, which the authors explain was developed based on literature within medical education.

The assessment involved the scapula bone and shoulder joint flexion with picture-only questions and was designed to assess students' mental representations of functional anatomy. Specifically, questions were one of five types; 1. 2D image in one anatomical plane involving feature identification on the scapula; 2. Complete 3D structure involving rotation of the scapula and comparison to a target figure; 3. Relationship to topography involving rotation of the scapula to a human body; 4. Dynamic function in one anatomical plane involving comparing two video recordings of shoulder joint flexion; and 5. 3D dynamic function involving re-ordering static images depicting different phases of shoulder joint flexion. Significant relationships were identified between the MRT and the 1<sup>st</sup> question type on feature identification ( $p = 0.003$ ), 4<sup>th</sup> question type on dynamic function ( $p = 0.003$ ), and the 5<sup>th</sup> question type on movement order ( $p = 0.008$ ).

The Group Embedded Figures Task test (GEFT, assesses cognitive style and analytic ability) was also used in this study as a measure of spatial ability, particularly spatial visualisation. However, the GEFT has been identified by factor analysis to load onto the Closure Flexibility sub-category of spatial ability (Carroll, 1993:p.339), therefore is questionable for measuring the spatial visualisation sub-category (see section 2.3.3). The GEFT was significantly related to the 1<sup>st</sup> question type on feature identification ( $p = 0.027$ ) and 2<sup>nd</sup> question type on rotation ( $p = 0.041$ ). This study compared a 2D (static image of the 3D model) and 3D (dynamic video of the 3D model) teaching method with neither method significantly affecting scores.



Schubert, Schnabel, and Winkelman (2009) designed a spatial MCQ with a combination of text-only MCQs and MCQs involving real 3D objects such as bones or specimens called the 3D-MC. The purpose of creating the 3D-MC was to combine the advantages of text-only MCQs, such as reliability and consistency, with the advantages of traditional anatomy 'spot' type questions that often use specimens. Although the 3D-MC questions were deemed to assess spatial anatomy knowledge no spatial ability tests were taken by students to assess whether spatial ability was related to the 3D-MC. Essentially, Schubert, Schnabel, and Winkelman (2009) converted a traditional spot exam with a free text answer into a multiple-choice answer format, with potentially no spatial ability component to the questions.

Tan et al. (2012) incorporated an MCQ anatomy test into their investigation comparing a 2D versus 3D anatomy teaching method of the larynx. The anatomy test was composed of thirteen factual questions and seven spatial questions designed to assess 3D spatial relationships. The MRT was given and found to have a moderately insignificant relationship ( $r = 0.48$ ,  $p = 0.085$ ) with the MCQ. It is unclear whether this relationship was to the whole MCQ or only the seven spatial questions as this level of analysis was not presented. If the questions were divided into non-spatial and spatial for the analysis, the differences could be analysed.

Interestingly, not all spatial MCQ anatomy assessments such as those described in the various studies above have involved an image in the question. A study comparing the use of an image versus a text response for extended match questions of the thorax, abdomen, and pelvis found the use of images affected question difficulty either making the question easier or more difficult (Vorstenbosch *et al*, 2013b). The author of this thesis proposes this could be due to exactly what the question is asking and the quality of the image. For instance, in Vorstenbosch et al.'s (2013b) study questions on foetal circulation, a complex area of anatomy, were all shown to be influenced by the use of an image or text, whereas all questions on the intestines were not. Vorstenbosch et al. (2013b) also found the students with a high spatial ability (measured by the MRT with scores trichotomised into low, medium, and high

spatial ability) scored higher on extended match questions for three areas of anatomy on the thorax, abdomen, and pelvis, compared to low spatial ability students. Furthermore, the authors found a significant interaction between spatial ability and question format (text or images); for three questions, using an image was advantageous for low spatial ability students. However, the trichotomising of data may exacerbate the differences because of the loss of students' exact scores and therefore the exact numerical (rather than categorical) differences in their scores, as discussed earlier.

## 1.4 Summary of Gaps Identified in the Literature

The spatial ability test primarily used in studies is the Vandenberg and Kuse (1978) Mental Rotation Test or the re-drawn version by Peters et al (1995), and has shown some positive and encouraging associations (Berney *et al.*, 2015; Cui *et al.*, 2016; Garg *et al.*, 1999b; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a; Guillot *et al.*, 2006; Gutierrez *et al.*, 2017; Hegarty *et al.*, 2009; Hoyek *et al.*, 2009; Keedy *et al.*, 2011; Langlois *et al.*, 2009; Nguyen *et al.*, 2013; Nguyen, Nelson & Wilson, 2012; Tan *et al.*, 2012; Vorstenbosch *et al.*, 2013a). The re-drawn MRT by Peters et al. (1995) will be used in the research presented in this thesis to align with the literature and measure the 3D mental rotation sub-category of spatial ability (see section 2.3.3). Studies have used other spatial ability tests such as the Purdue Visualisation of Views Test or the modified version Guay's Visualisation of Views Test (Chatterjee, 2011; Gutierrez *et al.*, 2017; Hegarty *et al.*, 2009; Provo, Lamar & Newby, 1998) which could be classified to test spatial perception and/or mental rotation (Linn & Petersen, 1985). Studies have also used lesser known spatial ability tests such as the 3DC5 test (Lischka & Gittler, 1997) and the 3DAT (Sweeney, Hayes & Chiavaroli, 2014), and a smaller number of studies have used more than one test (Berney *et al.*, 2015; Guillot *et al.*, 2006; Gutierrez *et al.*, 2017; Hegarty *et al.*, 2009).

To further address whether different sub-categories of spatial ability are used for anatomy learning this study uses the Mental Rotation Test (Peters *et al.*, 1995) and two further well validated spatial ability tests; the Card Rotation Test and the Surface Development Test (Ekstrom *et al.*, 1976). With the aim of investigating the sub-categories 2D mental rotation (Card Rotation Test) and spatial visualisation (Surface Development Test). Another point to note is that the marking criteria for the tests may vary between studies, with some researchers employing negative marking to discourage guessing amongst participants. This study does not employ negative marking to avoid an individual's propensity for guessing to influence the data.

It should also be borne in mind that the studies above are conducted within different academic institutions with varying course designs, with some still incorporating dissection into the anatomy course. With dissection being considered by many to be the gold standard (Theoret, Carmel & Bernier, 2007) and spatially-demanding, this may provide an advantage to students experiencing this form of teaching. However, in some respects dissection could be considered to be 'destructive' with regards to developing a spatial understanding as successive layers of material are taken away and, potentially, limbs and organs removed.

The recent acknowledgement that learning anatomy improves spatial ability and students of higher spatial ability perform better on anatomy (Rochford, 1985; Lufler *et al.*, 2012; Garg, Norman & Sperotable, 2001; Hegarty *et al.*, 2009; Vorstenbosch *et al.*, 2013a) there is a design weakness in studies which have not recruited for anatomically naïve students nor excluded students with possible previous anatomy knowledge when investigating spatial ability and anatomy learning. This study addresses this by recruiting anatomically naïve first-year veterinary students across two academic cohorts and two veterinary academic institutions and excludes for prior anatomy knowledge as identified by previous degree(s) studied.

Furthermore, as yet, no studies have compared the spatial ability of veterinary (or medical) students entering schools to ascertain how stable a trait spatial ability is across veterinary/medical students, or of students in another academic discipline. This study addresses this by recruiting first-year veterinary students from two academic institutions and across two academic years for one institution. Additionally this study also recruits a cohort of psychology students to compare to students of another academic discipline. While some of the papers in the current literature have looked at spatial ability for specific professions/specialties (Hegarty *et al.*, 2009; Keehner *et al.*, 2004) and have looked at spatial ability across a whole anatomy curriculum (Lufler *et al.*, 2012), this study specifically focuses on spatial ability in the context of veterinary students in the early years of a UK veterinary programme.

So far 3D teaching methods have involved the use of 3D technology (computer models, stereoscopic displays, 3D printing, and augmented reality) as a way of representing the anatomy in 3D, and as an alternative to dissection and the use of specimens. However, given the differences in spatial ability between students and, given that students of a higher spatial ability potentially perform better, a 3D anatomy teaching method designed to improve spatial ability could enhance students learning of anatomy. A study by Hoyek *et al.* (2009) looked at training spatial ability through generic spatial training tasks. The study presented in this thesis further investigates the link between spatial ability and anatomy learning by designing a 3D anatomy teaching method designed to encourage the spatial thinking of anatomy by teaching anatomy spatially, with the aim of improving students' spatial thinking.

Three-dimensional teaching interventions have primarily been given out with the core curriculum as extra optional tutorials (Garg *et al.*, 1999b; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a; Nguyen, Nelson & Wilson, 2012; Tan *et al.*, 2012). This has primarily been done for ethical reasons, as a new intervention may hinder or be disadvantageous to learning, and this has meant varied contact time with the intervention (ranging roughly between 3 minutes to 5 hours). Some studies have

incorporated 3D teaching methods and spatial ability into curriculums (Peterson & Mlynarczyk, 2016; Lufler *et al.*, 2012). To further contribute to this growing research specifically within veterinary education this study investigates a new spatial teaching method across a veterinary anatomy curriculum.

In the literature on spatial ability and anatomy learning studies have investigated the link by using students' anatomy end-of-course examination results (Lufler *et al.*, 2012; Sweeney, Hayes & Chiavaroli, 2014; Rochford, 1985; Provo, Lamar & Newby, 2002) or with students' results on spatial anatomy assessments (Garg *et al.*, 1999b; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a; Berney *et al.*, 2015; Nguyen, Nelson & Wilson, 2012; Provo, Lamar & Newby, 2002; Hegarty *et al.*, 2009). The results of these studies have shown evidence for a link between spatial ability and anatomy, and also provided evidence against. To further investigate this in the context of veterinary students the study of this thesis explores the relationship of spatial ability to both end-of-course examination results and a spatial anatomy assessment. Furthermore, this study will also investigate this in the context of a 2D and a 3D anatomy teaching method.

This chapter has presented the literature specifically on spatial ability and anatomy education with the majority of studies conducted in medical education and little within veterinary medical education. The next chapter expands on this literature by presenting the findings on spatial ability within the context of psychology with the aim of providing further insight into spatial ability to better inform the research of this thesis.

# Chapter 2 Literature Review Part 2: Spatial Ability and Human Intelligence

## 2.1 Introduction to Chapter 2

This chapter follows on from the literature review of anatomy education and spatial ability presented in chapter 1, by presenting the psychology literature on spatial ability and human intelligence.

“...you must either turn him or your eye so as to examine him from different aspects, from below from above and from the sides, turning the subject around and investigating the origin of each member, and in this way, satisfying yourself as to your knowledge of the actual anatomy.”

-Leonardo Da Vinci, c.1489 (Schuman, 1952:p.32)

The ability to look at the space around oneself, take in different objects regarding form, shape, and position, develop mentally-generated models of these, and manipulate them in ‘the mind’s eye,’ has been considered fundamental to the learning of anatomy for generations. Anatomy educators have researched the links between spatial ability and anatomy learning for approximately thirty years. These research efforts have investigated whether anatomy learning is linked to spatial ability (Rochford, 1985; Fernandez, Dror & Smith, 2011; Lufner *et al.*, 2012; Hegarty *et al.*, 2009; Guillot *et al.*, 2006; Berney *et al.*, 2015; Peterson & Mlynarczyk, 2016; Nguyen, Nelson & Wilson, 2012; Nguyen *et al.*, 2013; Chatterjee, 2011; Pedersen, 2012; Vorstenbosch *et al.*, 2013a). The use of three-dimensional teaching resources on the learning of anatomy and the influence of spatial ability (Al-Khalili & Coppoc, 2014; Yammine & Violato, 2015; Cui *et al.*, 2016; Lim *et al.*, 2016), and the

development of spatial anatomy assessments (Langlois *et al.*, 2017; Schubert, Schnabel & Winkelmann, 2009; Vorstenbosch *et al.*, 2014).

The concept of spatial ability has chiefly developed within human intelligence research, specifically within differential psychology. To fully investigate the link between spatial ability and anatomy learning, an understanding of the background of spatial ability and its relation to human intelligence are fundamental. However, the implications from studies in this field are often not considered in relation to the development of anatomy education research. This chapter aims to address this by presenting a new and integrated perspective of spatial ability to anatomy education and is divided into three sections. Firstly, a summary explaining what is known about human intelligence is presented. Secondly an explanation of what is known about spatial ability (while providing a perspective of how this fits with the anatomy education literature), and lastly, a discussion on the novel and informative future directions for the field of anatomy education in relation to spatial ability.

## **2.2 Brief Introduction to Human (General) Intelligence Research**

Investigating the intricacies of human intelligence is a cornerstone of differential psychology research, and serves as a foundation for our discussion of spatial ability. One of the best and most well-known definitions of human intelligence, agreed by 52 leading intelligence researchers, is the following (Gottfredson, 1997: p.13):

“Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it

reflects a broader deeper capability for comprehending our surroundings- ‘catching on’, ‘making sense’ of things, or ‘figuring out’ what to do.”

Research on human cognitive abilities (or intelligence) began in the late 1800s when Sir Francis Galton began collecting data from the general public who attended his ‘anthropometric laboratory’ (1885). Cognitive abilities are defined as “[a]n ability to perform any of the functions involved in cognition” (Colman, 2015) with cognition defined as “[t]he mental activities involved in acquiring and processing information” (Colman, 2015). In other words, cognitive abilities are the abilities the brain uses to process and acquire information.

Since Galton’s anthropometric laboratory experiments, the field of cognitive abilities/intelligence has expanded into the development of intelligence tests, the development of theories on the structure of intelligence, developmental trajectories of intelligence, exploration of how people of varying abilities solve problems, and the potential malleability of cognitive abilities. These advances have more recently taken place alongside allied research in neuroscience (Deary, Penke & Johnson, 2010; Deary, 2000), genetics (Lee *et al.*, 2010; Payton, 2009) and epidemiology (Deary, Weiss & Batty, 2010).

### **2.2.1 Discovery of ‘g’ and the Structure of Human Intelligence**

From 1885, Galton researched human cognitive abilities using a range of tests. Some, such as the recording of body dimensions, would not be regarded in the present day as tests of cognitive ability as no cognition was required, while others such as the reaction time test could be regarded as proto-cognitive. It was not until the beginning



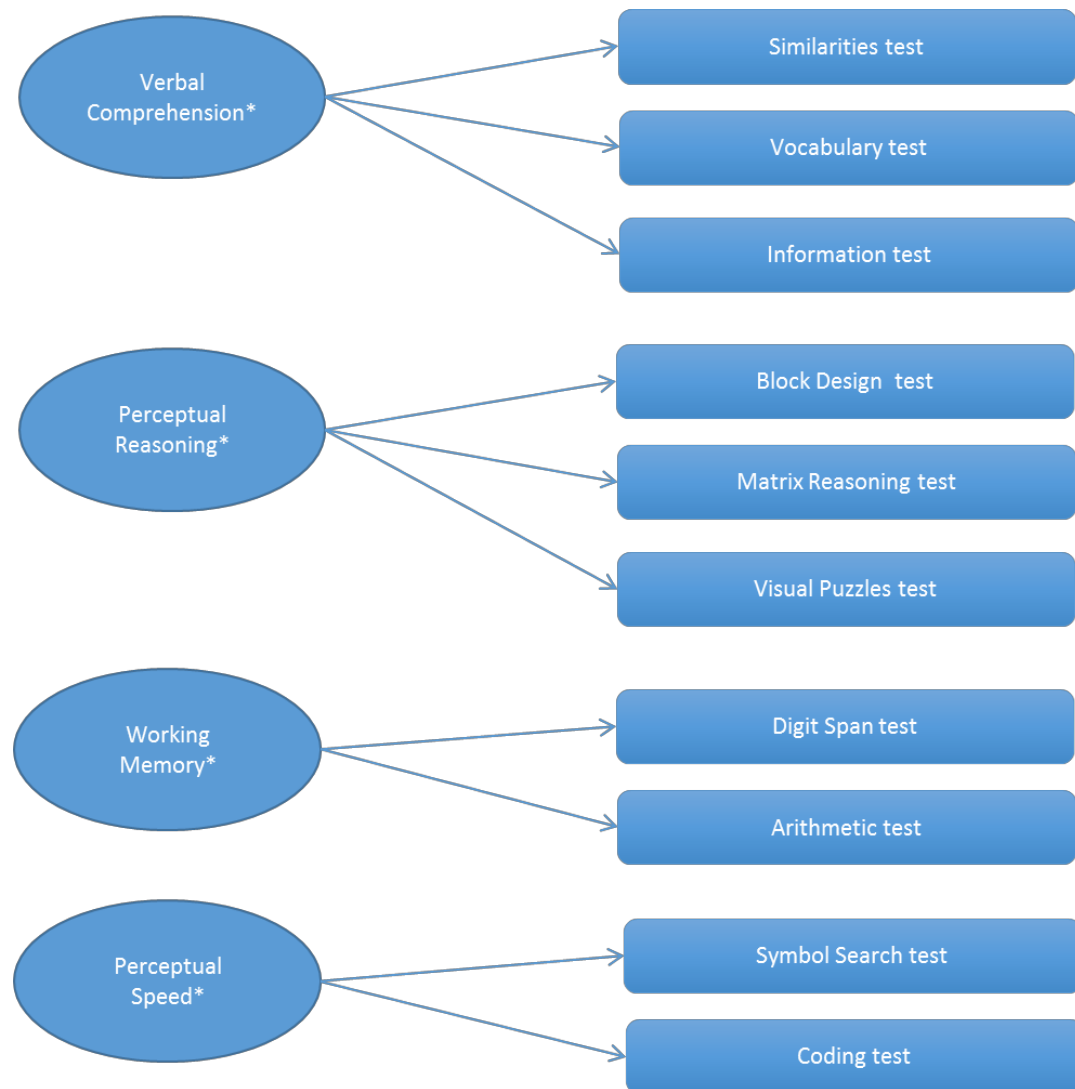
of the 20<sup>th</sup> Century that the first ‘true’ intelligence tests were developed (as discussed in Ritchie 2015). In 1903, Alfred Binet studied underperforming French school students and hypothesised intelligence tests needed to involve more complex tasks, and designed tests similar to the cognitive tasks required of school students, such as tasks involving language and reasoning ability (as discussed in Carroll, 1993:chap.2)

A typical test of human intelligence, for example, the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) (Wechsler, 1958), uses a battery of validated and standardized cognitive ability tests to assess a person’s overall cognitive capability or intelligence. Specifically, the WAIS-IV scale includes tests such as *Similarities* (participants have to describe how two words are similar), *Vocabulary* (participants give a definition to a word presented in a picture), *Block Design* (participants re-arrange coloured and patterned blocks to match a specific pattern), and *Digit Span* (participants are to recall a sequence of numbers in the correct order).

Through the statistical technique of factor analysis on the scores of cognitive ability tests from groups of participants, psychologists have proposed theoretical models on the structure of intelligence (with ‘structure’ meaning the psychometric and statistical relationships between cognitive ability tests). The objective of factor analysis is to reduce the number of possible directions in a correlation matrix by representing measured variables as linearly related to latent unobservable factors, i.e. to group correlations directly measuring the same latent trait. Overall, this aims to reduce the measured variables into a smaller number of factors, known as group factors or group domains, such as spatial ability, memory, or verbal ability, making interpretation of correlation patterns easier.

For example with the WAIS-IV the *Similarities*, *Vocabulary*, and *Information* tests can be grouped into one factor called ‘Verbal Comprehension’, *Block Design*, *Matrix*

*Reasoning, and Visual Puzzles into ‘Perceptual Reasoning’, Digit Span and Arithmetic into ‘Working Memory’, and Symbol Search and Coding into ‘Processing Speed’ (Figure 2.1).*



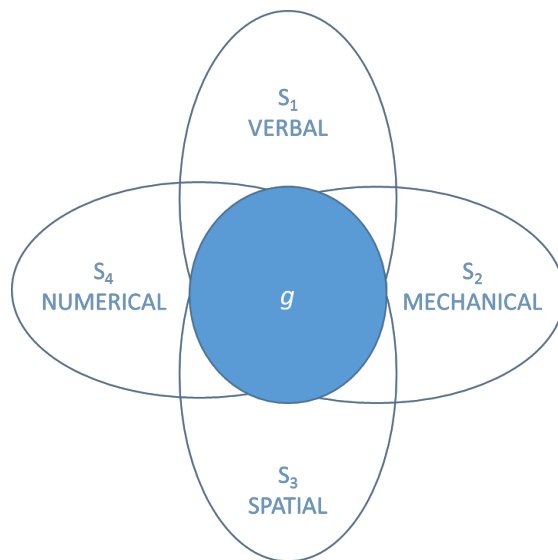
**Figure 2.1 Example of group factors or cognitive domains from the WAIS-IV intelligence test. \*All of the group factors, designated by a circle, will correlate to one another (not shown on diagram).**

The latent unobservable factors extracted from factor analysis are hypothetical and subjectively postulated by the researcher based on what the cognitive ability tests are asking the testee to do (such as ‘Verbal Comprehension’ or ‘Working Memory’) and

thus the latent factors extracted are also guided by the type of cognitive ability tests present in an intelligence test battery. This has resulted in much controversy among researchers (Deary, 2012) because the name, number, and content of the cognitive domains can vary depending on the battery of cognitive ability tests used in an intelligence test.

By the use of factor analysis, it was also discovered that all the individual cognitive ability tests in an intelligence test correlated positively and significantly with each other, and this became known as the positive manifold effect (Spearman, 1904). The positive manifold effect is one of the best replicated and convincing results from the field of intelligence research (such as Carroll's (1993) large factor analysis study on over 400 data sets of intelligence tests), and means that individuals who perform well on one cognitive ability test, such as a test on mental rotation, also perform well on tests of a seemingly unrelated category, such as vocabulary. It also means it is not strictly true to state that for example a test of spatial ability such as the Card Rotation Test, purely measures that one single ability.

The positive manifold effect was first hypothesized by Spearman (1904), and he proposed that for all these tests to positively correlate with one another they likely have some underlying latent common, or general, factor to them all which he called  $g$  (Figure 2.2). It has since been found that  $g$  accounts for approximately half (40 – 50%) of the variability in cognitive ability in a sample of the human population (as explained in Deary, 2001).



**Figure 2.2 Venn diagram explaining Spearman's two-factor theory of intelligence. *g* = general ability, *s* = specific abilities.**

Not all psychologists agreed with Spearman's *g* hypothesis (Thurstone, 1957; Gardner, 1983). An alternative theory, which is different from *g*, is that different cognitive abilities are distinct and independent. Howard Gardner (1983) described multiple intelligences, such as 'logical-mathematical' intelligence, 'musical' intelligence, and 'linguistic' intelligence. However, there are no peer-reviewed published studies or other empirical evidence to support the multiple intelligences theory (Waterhouse, 2006). Another opposing theory to '*g*' is Thurstone's (1957) "Primary Mental Abilities" with intelligence proposed as being multi-factorial with no common general factor. Thurstone believed the *g* factor extracted was not general to every possible test but an average of the specific test batteries used.

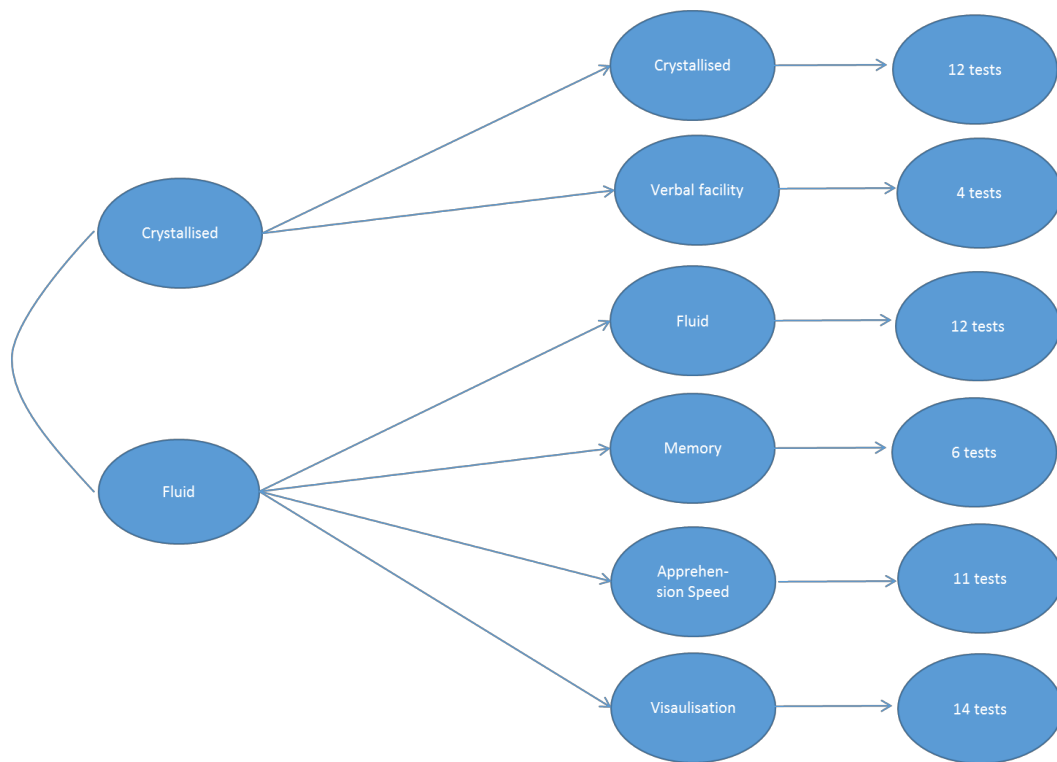
However, factor analysis of Thurstone's data has shown a strong *g* factor (Johnson & Bouchard, 2005), and statistical analysis of three different intelligence test batteries taken by 400 participants resulted in each battery having a strong *g* factor.

Furthermore, each *g* factor correlated near perfectly and significantly to one another (correlations of 0.99, 0.99, and 1.00 for each test battery respectively), providing

evidence that *g* is general to batteries of tests (Johnson *et al.*, 2004). This finding was further confirmed using datasets from five test batteries with similar results (Johnson, Nijenhuis & Bouchard, 2008), overall these high correlations mean that with a large number of tests in a test battery the same *g* factor is ‘tapped’ into when factor analysed.

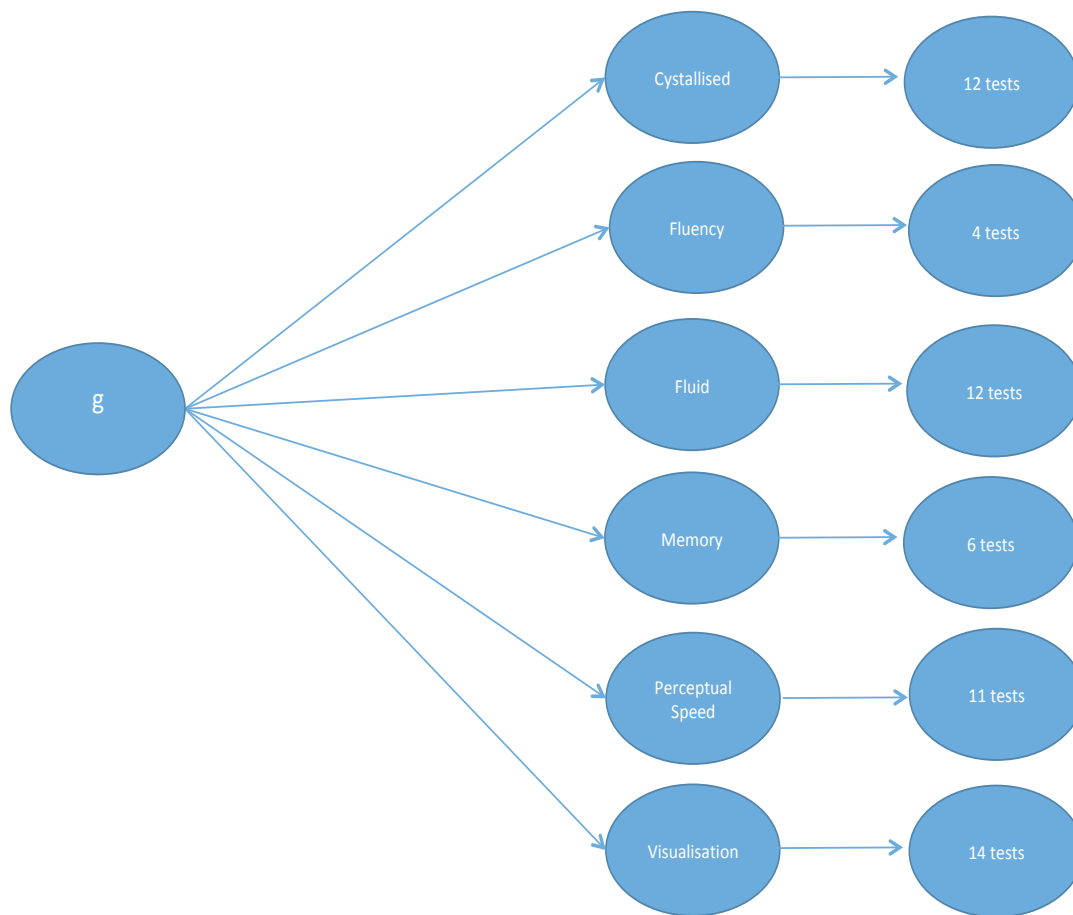
From the use of factor analysis psychologists have proposed various theories on the structure of human intelligence, with some incorporating a *g* factor while others do not. Currently, the main empirically supported theories on the structure of human intelligence are the Cattell-Horn “fluid-crystallised” theory of intelligence (Horn, 1989); Carroll’s three-stratum theory (Carroll, 1993); a combination of the fluid-crystallised theory and the three-stratum theory (McGrew, 2009); the verbal perceptual model proposed by Vernon (1950); and the more recently proposed visual, perceptual, and image rotation (VPR) model proposed by Johnson and Bouchard (2005).

The Cattell-Horn ‘fluid-crystallised’ theory of intelligence (Horn, 1989), divides intelligence into two broad categories (Figure 2.3): *fluid* (*gf*, the use of logical reasoning to solve newly-encountered problems where prior experience, knowledge, and skills do not help, and can be described as intelligence as a process) and *crystallised* (*gc*, using learned knowledge gained from education, cultural information, and experience to solve problems, and can be described as intelligence as product). This theory does not include an overall *g* factor despite *gf* and *gc* being highly correlated.



**Figure 2.3** *Diagram of the fluid-crystallised model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005.*

John Carroll (1993) collected over 400 data sets of intelligence test results and performed factor analysis on each to confirm the theory of *g*. Carroll aimed to construct a psychology equivalent to the Periodic Table of Elements to help psychologists come to a consensus. Carroll found similar results to Spearman's 'g' theory, and he also found many other latent group factors (or cognitive domains) because the same battery of cognitive tests was not used in each data set. From these results, Carroll proposed another model of human intelligence called the three-stratum theory (Figure 2.4).

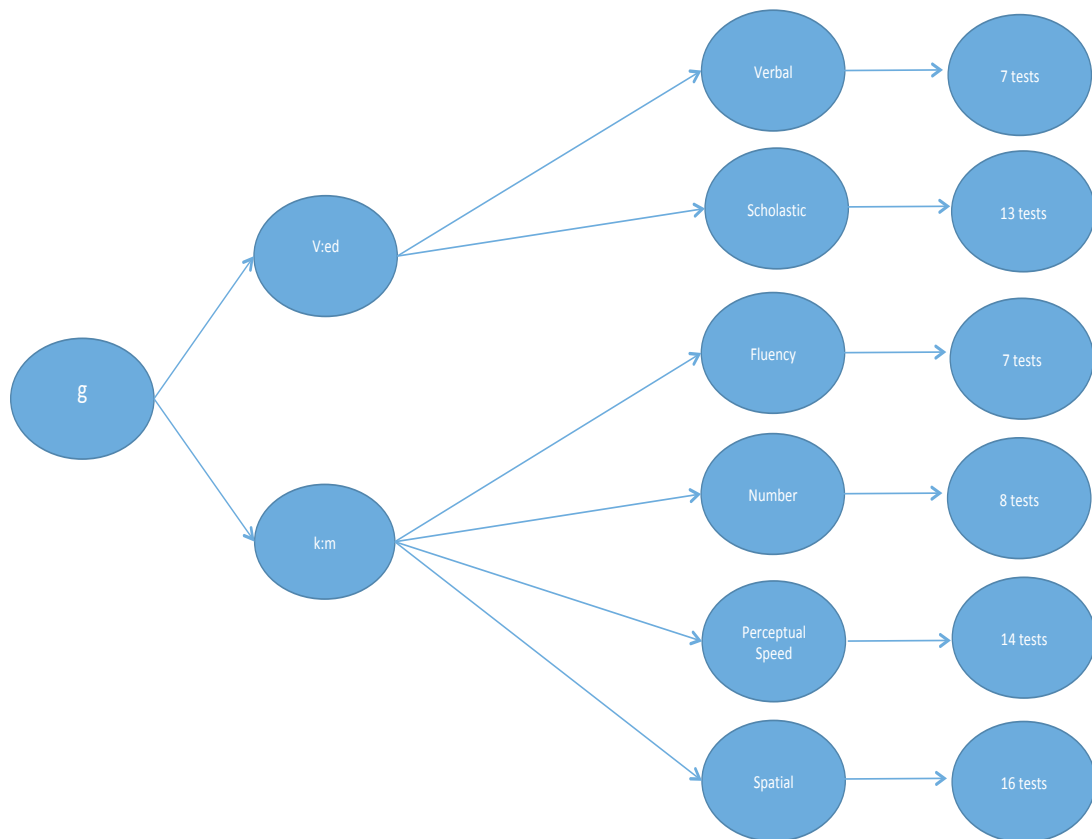


**Figure 2.4** Diagram of Carroll's three-stratum model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005.

The first stratum, or level, in Carroll's theory comprises the cognitive ability tests administered; the second comprises latent group factors or cognitive domains (such as spatial ability, or memory, or verbal ability), and the third comprises the general intelligence factor (*g*). Individuals vary at each of these three levels. The three-stratum theory and the fluid-crystallised theory are often combined and termed the Cattell-Horn-Carroll theory of cognitive abilities or CHC theory due to their similarities (McGrew, 2009).

Another model including a *g* factor is Vernon's verbal-perceptual model (1950). Vernon proposed that the latent group factors (equivalent to Carroll's second

stratum) can be divided into one of two broad categories, namely *v:ed* for verbal and educational abilities, and *k:m* for spatial, practical, and mechanical abilities (Figure 2.5).

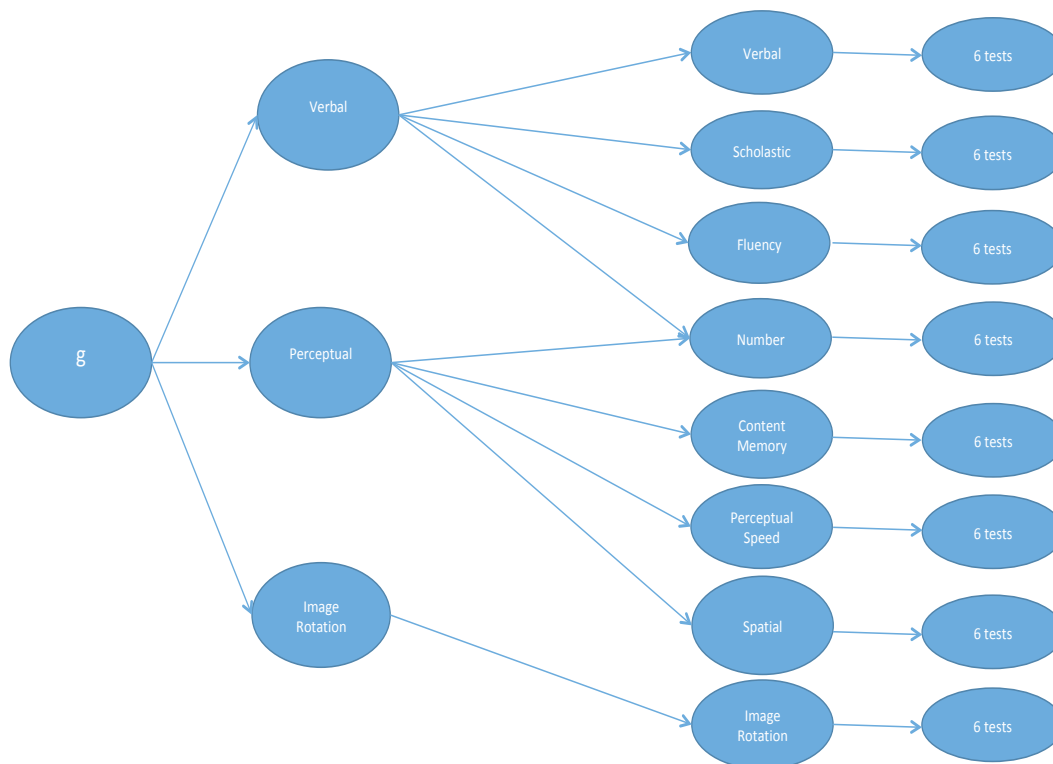


**Figure 2.5** Diagram of Vernon's verbal-perceptual model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005.

Despite decades of research on the structure of human intelligence, no one model is agreed. As explained above several models have been proposed from the results of factor analysis and other theories, and this has generated much debate within the field that continues to the present day (Kovacs & Conway, 2016; Deary, Cox & Ritchie, 2016). However, the most dominant model based on psychometrics is the fluid-crystallised theory (Johnson & Bouchard, 2005).



Interestingly, a recent model that highlights the significance of the spatial ability cognitive domain has been proposed. Johnson and Bouchard compared, using factor analysis, the fluid-crystallised model, Vernon’s verbal-perceptual model, and Carroll’s three-stratum model to one another using three batteries of cognitive ability tests (Johnson & Bouchard, 2005). They found Vernon’s verbal-perceptual model fitted significantly better than the other two models, and decided to improve this model by adding three factors at the third strata instead of Vernon’s two factors (v:ed and k:m) (Figure 2.6).

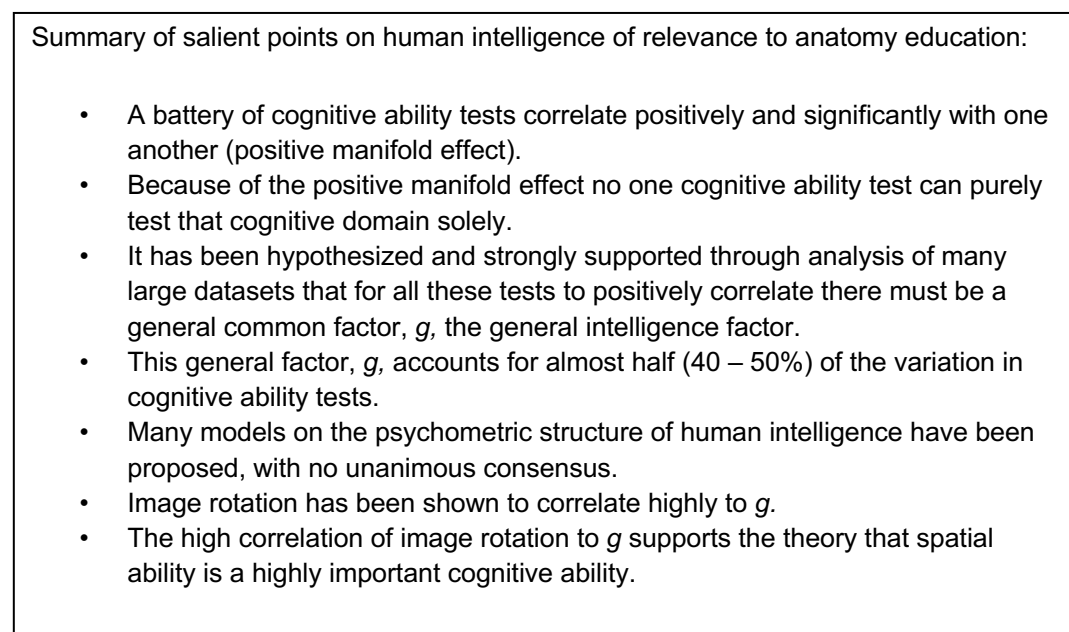


**Figure 2.6** Diagram of the VPR model of intelligence. Cognitive test examples from 3 test batteries as described in Johnson and Bouchard 2005.

The three factors being verbal, perceptual, and image rotation (a sub-category of spatial ability, see section 2.3.3 below). They found that each of the three factors correlated near perfectly, positively, and significantly with *g* (0.96, 0.99 and 0.97 respectively). They concluded that this model, known as the VPR model, demonstrates the significance of image rotation in human general intelligence. From

this, it has been supported that spatial ability is a prominent cognitive ability when it comes to human general intelligence.

The first section of this chapter aimed to explain the main findings of human intelligence research to provide a basis for discussions in the subsequent sections on spatial ability. Figure 2.7 summarises the main points of human intelligence research discussed in this first section that are relevant to anatomy education studies involving cognitive abilities.



***Figure 2.7 Box summarising the salient points from human intelligence research literature.***

## 2.3 Brief Introduction to Spatial Ability Research

Summaries on the history of spatial ability research are available within psychology and allied disciplines (Miller, 1996; Mohler, 2008; Harle & Towns, 2011; Carroll, 1993; Eliot & Smith, 1983; Hegarty & Waller, 2005; McGee, 1979; Smith, 1964). One of the earliest investigations (which was mistakenly taken to measure spatial ability) was Sir Francis Galton's breakfast table experiment (1880). This involved participants imagining the contents of their breakfast table but instead measures imagery ability (Hegarty & Waller, 2005). Since the advances in intelligence research, such as the development of factor analysis, the definition of spatial ability and its sub-categories has been refined, and tests to measure it developed. The question of whether spatial ability is a malleable ability has also been researched.

### 2.3.1 What is Spatial Ability?

Carroll (1993:p.304) defines spatial ability as:

“abilities in searching the visual field, apprehending the forms, shapes and positions of objects as visually perceived, forming mental representations of those forms, shapes and positions, and manipulating such representations ‘mentally’.”

Wright, Frier, and Deary (2009:p.1503) define spatial ability as:

“[t]he ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment.”

Another earlier definition by McFarlane (1925:p.56) is:

“Like literary or mathematical ability, practical ability involves analysis and synthesis, judgment and conception: its uniqueness lies in the fact that those persons possessing it in a high degree analyse and judge better about concrete spatial situations than do other individuals who perhaps excel in dealing with more highly abstract symbols.”

In common parlance spatial ability is the ability to think of objects and three-dimensional spaces, and to be able to manipulate these structures (such as cutting, folding, slicing, merging, changing perspective, and rotating, etc.) in your ‘mind’s eye.’ Spatial ability can be divided into subcategories, such as *spatial visualisation*, *spatial perception*, *spatial relations*, *closure speed*, and *visual memory*, to name a few. However, an exact definition of spatial ability and the identification (and definition) of sub-categories remains contentious due to the subjective nature of factor analysis (Uttal *et al.*, 2013; Hegarty & Waller, 2005:chap.4), and is an area worthy of further research.

### **2.3.2 Development of Spatial Ability Tests**

Following on from Binet’s work on the development of intelligence tests, as mentioned in the above section, research on intelligence focussed on predicting academic and vocational success using standardised tests. As explained by Hegarty and Waller (2005:chap.4) in “The Cambridge Handbook of Visuospatial Thinking”, this work led on to investigating individual differences, and for spatial abilities, testing arose from attempts to measure a practical ability as a means of predicting success in technical occupations. Therefore, early spatial ability tests often involved assembly or manipulation of actual objects.

In 1925 one of the earliest investigations of ‘practical’ ability was performed. In Margaret McFarlane’s doctoral thesis, “*A Study of Practical Ability*”, a variety of different ‘practical’ or ‘motor’ ability tests were administered to children in order to define practical ability (McFarlane, 1925:p.56). She concluded:

“Like literary or mathematical ability, practical ability involves analysis and synthesis, judgment and conception: its uniqueness lies in the fact that those persons possessing it in a high degree analyse and judge better about concrete spatial situations than do other individuals who perhaps excel in dealing with more highly abstract symbols.”

McFarlane’s conclusion suggests that she was referring more to a spatial ability than a practical/motor one. This finding, along with the growing popularity of Binet’s intelligence tests to measure individual differences, resulted in spatial ability tests changing design from manual manipulation of objects to paper-and-pencil exercises. Paper-and-pencil tests allowed large groups to be assessed at once and thus provided large data sets for analysis, contributing to the investigations on the structure of human general intelligence (as discussed above) and spatial ability.

Binet’s work aimed to identify struggling individuals and thus his tests were based on the mental activities required of students at school and consequently consisted primarily of verbal material. Smith (1964:chap.2) discusses that the popularity of Binet’s intelligence tests led to the assumption that intelligence was best measured by verbal questions and this led to a “bias to associate all aspects of intellectual ability with a single general (typically verbal) intelligence” (Hegarty & Waller, 2005:chap.4, page 124). During human intelligence research, hundreds of cognitive abilities tests were developed, meaning there are now numerous spatial ability tests available, of varying validity. For the study design of anatomical educational research, it is important to select tests that are well validated and statistically shown to consistently measure what psychologists believe to be the spatial ability group

factor (examples of less well known spatial ability tests used for anatomy education research include: Lischka & Gittler, 1997; Sweeney, Hayes & Chiavaroli, 2014).

### **2.3.3 Identification of a Spatial Ability Factor and Sub-Factors**

In the late-1920s and the mid-1930s, the validity of verbal-based intelligence tests was beginning to be questioned. It was no longer being assumed that intelligence tests of the verbal type had the greater authority (Smith, 1964). Eliot and Smith in “An International Directory of Spatial Tests” (1983), explains there was slow acceptance of non-verbal intelligence tests because the main reason for the initial development of such tests was to measure academic success, which was associated with reading and writing. Therefore tests with verbal material were designed to measure these abilities primarily. He further explains it was not until the United States Army in 1918 performed large scale testing to screen recruits, that the requirement to develop non-verbal tests to test the uneducated and those who had difficulty with language was apparent.

The development of non-verbal tests brought about much controversy on the connection language had with non-verbal tests and the relationship of non-verbal tests to intelligence. It was argued that non-verbal tests would be better than verbal tests as the influence of education would be removed, and thus represent a truer test of a person’s intelligence. When verbal ability was shown to be a distinct cognitive ability from general intelligence ability (i.e., *g*, as proposed by Spearman) attention was made to the development of non-verbal tests (Stephenson, 1931), and in particular spatial tests (Eliot & Smith, 1983).

Hegarty and Waller (2005) and Smith (1964) explain that studies by Kelley (1928), El Koussy (1935) and Thurstone (1957) showed, using statistical analysis (including factor analysis and Spearman's tetrad differences technique), that spatial ability was a distinct factor. It was also shown that spatial ability accounted for a significant amount of variability on intelligence test scores.

Once spatial ability was identified and deemed through factor analysis to be a separate group factor, research on spatial ability in the mid-20<sup>th</sup> century focussed on establishing the structure of spatial ability. Focussing whether spatial ability could be further sub-divided into separate sub-factors or sub-categories.

Psychologists Guilford and Lacey (1947), Zimmerman (1954), Thurstone (1950) and French (1951) looked at large batteries of spatial ability tests and found that several factors could be extracted from them. The cognitive domain of spatial ability was thought to have a mixture of correlated sub-factors by the 1960s, and with the identification of many sub-factors through factor analysis, all named by different research groups, much subjective terminology arose. How best to characterise these sub-factors continues to be undecided today and until further systematic research is done the identification of sub-factors will continue to be a problem.

Hegarty and Waller in "The Cambridge Handbook of Visuospatial Thinking" (2005:chap.4) provide a summary of the spatial ability sub-categories identified by factor analysis in earlier studies. Table 2.1 shows this summary with the addition of results from a further study (Uttal *et al.*, 2013) to provide a general overview of currently proposed spatial ability sub-categories.

Study	Subcategories Identified
Michael, Guilford, Fruchter & Zimmerman (1957)	Spatial Visualisation Spatial relations and orientation Kinesthetic imagery
McGee (1979)	Spatial Visualisation Spatial Orientation
Lohman (1988)	Spatial Visualisation Spatial Relations Spatial Orientation
Carroll (1993)	Spatial Visualisation Spatial Relations Closure speed Flexibility of Closure Perceptual Speed Visual memory
Linn and Petersen (1985)	Spatial Visualisation Spatial Perception Mental Rotation
Uttal et al. (2013)	Intrinsic Extrinsic Static Dynamic

*Table 2.1 Summary of Spatial Ability Factors, as described in Hegarty and Waller (2005) and the addition of Uttal et al (2013).*

It can be seen from Table 2.1 that spatial visualisation is the most common sub-category. McGee (1979) states that much of the confusion has arisen due to inconsistencies in the use of terminology due to the subjective element of factor analysis. The psychometric approach to determining spatial ability sub-factors typically follows exploratory factor analysis, where there is no prior theoretical model to the understanding of the relationship between variables in the analysis. However, Uttal et al. (2013) explain that spatial ability tests were not synthesised from a known description but came to fruition as a result of other tests (on practical ability), therefore states that it is not unexpected that factor analysis has not provided support for common sub-categories.



Furthermore, factor analysis assumes that all participants use the same strategy per test and that the same strategy is used per question, and this may not be the case. Additionally, there are different techniques for exploratory factor analysis including different ways to rotate solutions, resulting in many different exploratory factor analysis models (Hegarty & Waller, 2005). Therefore, although factor analysis has provided interesting insights by analysing patterns of correlations, the subjective aspect has provided mixed views and mixed levels of agreement.

However, the newer technique of confirmatory factor analysis where researchers can fit the data to a pre-determined factor model is providing insight. For example, more recently, researchers exploring the distinction between mental rotation and perspective taking ability used confirmatory factor analysis to explore the relationship between the two sub-categories. The researchers proposed that a one-factor model showing mental rotation and perspective taking loading onto one factor (i.e., spatial ability) should not fit better than a two-factor model (i.e., mental rotation and perspective taking ability separately) if the two sub-categories are distinct. The authors found a two-factor model fitted best and thus a distinction between mental rotation and perspective taking was found (Hegarty & Waller, 2004).

The three sub-categories (spatial perception, mental rotation, and spatial visualisation) proposed by Linn and Petersen's (1985) meta-analysis on gender differences in spatial ability are commonly cited and used in the literature. Uttal et al. (2013) explain that a classification system based on linguistic, cognitive and neuroscientific investigations can give the sub-categories intrinsic versus extrinsic and dynamic versus static, and a 2x2 table can be constructed with these four categories. The authors suggest that Linn and Petersen's (1985) three categories can be mapped onto this new structure.

There is therefore still ongoing debate about spatial ability sub-categories however Linn and Peterson's three simple sub-categories will be used for the research of this thesis (Table 2.2).

Spatial Ability sub-category	Cognitive processes definition
<b>Spatial perception</b>	<i>'subjects are required to determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information'</i> (Linn & Petersen, 1985:p.1482)
<b>Mental rotation</b>	<i>'the ability to rotate two- or three-dimensional figures in one's mind'</i> (Wright, Frier & Deary, 2009:p.1503)
<b>Spatial visualisation</b>	<i>'spatial ability tasks that involve complicated, multistep manipulations of spatially presented information. These tasks may involve the processes required for spatial perception and mental rotations but are distinguished by the possibility of multiple solution strategies'</i> (Linn & Petersen, 1985:p.1482).

**Table 2.2 Linn and Petersen spatial ability categories used for the research of this thesis.**

These categories were proposed using a theoretical cognitive processes approach based on the similarities of how participants answer questions and were checked for homogeneity of effect sizes as a means to try and ensure the studies were near

replicates of each other (by having similar effect sizes), thus measuring a similar construct, rather than sub-categorising spatial ability with a data-driven correlational psychometric approach.

The confusion over spatial ability sub-categories has also influenced anatomy education research with some studies referring to the Vandenberg and Kuse mental rotation test as a test of visualisation or mental rotation (such as Nguyen, Nelson & Wilson, 2012). However, in some instances, mental rotation is classified under spatial visualisation depending on the classification approach used. Since the exact sub-categories are unknown, it is important for anatomy education researchers to be aware of this when using and interpreting spatial ability tests.

### **2.3.4 Spatial Ability and Gender**

Gender differences are a well-researched area in psychology, yet an area that still generates much debate. There are many studies on the gender differences identified with spatial ability, and a brief overview of this is provided in this thesis. Whether males or females are strongest (although in general favours males) and the effect size of the differences can vary as explained below. Studies have routinely found favour for males to perform better with higher scores than females on tasks of mental rotation with effect sizes as large as 0.9-1.0 standard deviations (Masters & Sanders, 1993; Reilly & Neumann, 2013) and particularly for tasks of 3D mental rotation. Whereas spatial tests such as the Hidden Figures Test (participants are to identify a shape hidden in a complex figure) and the Paper Folding Test (participants are to identify out of five options showing unfolded pieces of paper which one correctly shows the hole punch pattern of a piece of folded paper punched with a hole) show very small effect sizes ( $d = 0.19$ ) between males and females (Voyer, Voyer & Bryden, 1995).

Gender differences are a well-researched area within psychology in general not just for spatial ability. Recently Zell, Krizan, and Teeter (2015) performed a metasynthesis (i.e., a meta-analysis on meta-analyses) on 106 meta-analyses and 386 individual meta-analytic effects to investigate the general *gender differences* hypothesis in psychology. Zell and colleagues identified mental rotation as the second largest gender difference ( $d = 0.57$ ; as identified in the meta-analysis Maeda and Yoon (2013)). Interestingly, Maeda and Yoon's (2013) meta-analysis involved 40 studies on the Purdue Spatial Visualisation Tests: Visualisation of Rotations (participants are shown the rotation of an object and are to identify the same rotation for a different object correctly), found that although males performed better than females, as mentioned above, the time limit given to answer each question influenced the effect size. When participants had more than 30 seconds per item the effect size between males and females was smaller ( $\bar{g} = 0.31$ ), whereas fewer than 30 seconds per item had a larger effect size ( $\bar{g} = 0.67$ ). This is relevant because it demonstrates the effect size is influenced by the criteria of the testing and that males are not necessarily considerably better than females.

Silverman and Eals (1992) hypothesise a hunter-gatherer theory for sex differences in spatial abilities. This theory hypothesises the division of labour between males and females in the Pleistocene era, explains males' higher development of three-dimensional mental rotation (associated with hunting activities) and females' higher development of visual-spatial memory (associated with gathering activities). Silverman and Eals (1992) found support for the female advantage in visual-spatial memory. To investigate this theory across different countries, and therefore increase the validity, Silverman and colleagues (2007) used data from a BBC internet study involving two tests: the Vandenberg and Kuse mental rotation test and an object location memory test designed by Silverman and Eals (1992). They concluded that across 40 countries, males commonly had an advantage with mental rotation, but this was not the same for females on object location memory.

The above is a very brief section on gender differences within spatial ability, and from an anatomy education viewpoint it is important that veterinary educators are aware and acknowledge that there are possible gender differences about spatial ability, as this could be one of many reasons for differences in performance involving a spatial ability element. However, spatial ability has been shown to improve with practice and training (Uttal *et al.*, 2013) (see section 2.3.6 below). Gender differences are analysed in this thesis, to see if they match the psychology literature, but not investigated in detail as the main research questions of this thesis involve the investigation of the relationship between spatial ability and anatomy teaching methods.

### **2.3.5 Spatial Ability and the Re-test/Practice Effect**

It is well known within psychology that improvement between repeated administrations of cognitive ability tests can represent a practice effect rather than true improvement in that cognitive ability. The practice effect is not always accounted for in research design but is paramount to the correct interpretation of cognitive ability test results when the same test (or an alternate version of the same test) are administered consecutively. McCaffrey and Westervelt (1995) provide an overview of the various issues that may arise with repeated assessments, such as the practice or re-test effect (previous exposure to a test improves subsequent testing), regression to the mean (if scores on a test are extreme on the first testing they will be closer to the mean on the second testing, or vice versa, i.e. scoring higher on the second testing, if low on the first testing, could be due to regression to the mean), and the test-retest correlation coefficient (measures how consistent the results of a test are over time by correlating the test scores on the same test by the same participants, a correlation of 1 = excellent reliability, 0 = no reliability).

McCaffrey and Westervelt's article also provides examples of faulty research design, such as the Kilburn, Warsaw, and Shields (1989) study on firefighters exposed to polychlorinated biphenyls. A group of fourteen firefighters exposed to polychlorinated biphenyls were given eight cognitive ability tests before (pre) and two months after (post) a detoxification program aimed to improve symptoms from exposure (such as headaches, fatigue, impaired balance, weight loss). A control group of fourteen firefighters were also tested with the same eight tests at the start of the study (pre), but not post. Comparisons were then made between the pre and post scores of the chlorinated biphenyl exposed firefighters with statistically significant improvements, and thus the researchers concluded the detoxification program worked. However, because the control group did not take the tests two months later, and so no pre-post comparison was made for the control group, the influence of the practice effect was not included in the research design. This falsely led to conclusions the detoxification intervention improved test performance.

Instances of faulty research design and resultant potential impact on research conclusions are also, unfortunately, evident in anatomy education research on spatial ability such as Gutierrez et al. (2017) and Lufler et al. (2012). A pre/post study design was adopted by both of these studies with Lufler et al. administering the MRT, and Gutierrez et al. the MRT, Raven's matrices, and Guay's Visualisation of Views tests to medical and veterinary medical students respectively. However, no control group was incorporated into the study design to adjust for the practice effect leading the researchers to conclude the improved performance in the cognitive ability tests was due to the learning of anatomy.

Lezak's "*Neuropsychological Assessment*" (2012) provides a discussion on the practice effect (Chapter 5 "The neuropsychological examination: Procedures"). It is explained that the practice effect is generally most prominent with tests involving a timed response, i.e., a speed component; with tests which use an unfamiliar or infrequently practiced form of response; or tests which have a single solution (and

this solution is easily conceptualised). Unfortunately, most spatial ability tests tend to fall into one or more of these categories. It is therefore paramount to design a study that takes into account the test practice effect. Ideally, this would involve a control group who do not receive the intervention but take the tests the same number of times and at the same interval, as the study population.

### **2.3.6 The Malleability of Spatial Ability and the Relationship of Spatial Ability to STEM**

Other research has looked at the malleability of spatial ability, including the effect of prior experience. Uttal et al. (2013) conducted a meta-analysis involving 188 studies investigating the training of spatial ability. They concluded that spatial ability was a moderately malleable skill (hedges's  $g = 0.47$ ) and that training improved performance (hedges's  $g = 0.62$ ). Additionally, the same authors analysed whether the training effect was sustained or fleeting. They did this by comparing post-test results across studies with the post-tests given immediately, less than one week, or less than one month later, and no significant difference ( $p > 0.19$ ) was found. Additionally, on comparing post-tests given immediately to all post-tests that were delayed, no significant difference was found ( $p > 0.67$ ). They concluded that because the different time intervals were not statistically different, that improvements (post-tests given immediately  $g = 0.48$ , and post-tests delayed  $g = 0.44$ ) to spatial ability were durable, although it could be argued that their longest interval used in the study (one month) was a relatively short follow-up time.

Within STEM subjects (science, technology, engineering, and mathematics) spatial ability has been shown to be highly important but often overlooked. Wai, Lubinski, and Benbow (2009) looked at the importance of spatial ability in education and the workplace by using the Project TALENT database (Wise, McLaughlin, & Steel, 1979). The Project TALENT database includes longitudinal data on mathematical

ability, verbal ability, and spatial ability at one year, five years, and eleven years after graduation from high school in the United States. The authors looked at the eleven year follow-up data of participants that included their highest degree (bachelor, masters, or PhD) and occupation. The authors scaled participants' spatial ability score at high school on a nine-point scale (i.e., stanine, where a score of one means the participant is in the bottom 4% and nine means a participant is in the top 4% of spatial ability scores). When the stanine scaled spatial ability scores were graphed against the proportion of participants' highest degree, 45% of participants with a PhD within STEM were in stanine nine, with 30% and 25% in stanine nine with a bachelors or masters respectively, indicating those students with a high spatial ability go on to achieve high academic credentials.

The researchers also used the Project TALENT database to see whether students with high spatial ability were overlooked in talent searches because these searches require students to obtain scores within the top 1% or 0.5% in mathematical or verbal ability but not spatial ability. They found that 70% of students within the top 1% of spatial ability were not in the top 1% for mathematical or verbal ability, and when comparing these 70% of students for highest degree within STEM compared to the base rate in Project TALENT, they were more than twice as likely compared to the base rate to have a bachelors, masters, or PhD within STEM, thus identifying a large pool of unrecognised talent. The authors concluded with three clear statements:

1. Spatial ability is a “salient psychological” trait for those who obtain educational and occupational achievements in STEM.
2. Spatial ability is important for educational and occupational outcomes.
3. Talent searches involving mathematical and verbal ability overlook individuals who are spatially inclined.



## 2.4 The Importance of Spatial Ability

The first section on this chapter presented the literature of human intelligence to explain the relationship of spatial ability to intelligence, and the second section moved on to present the literature specifically on spatial ability. Throughout both of those sections, evidence has been presented on the importance of spatial abilities in relation to intelligence. This current section aims to pull all of this evidence together, along with other evidence, to help explain and expand on the importance of spatial ability to intelligence. The link between spatial ability and intelligence is fundamental to the spatial teaching method designed for the research of this thesis.

To evidence the importance of spatial ability this chapter has shown that models of intelligence have involved a spatial factor at the second stratum. For instance, Vernon's verbal-perceptual model proposes the broad category of spatial, practical, and mechanical abilities ( $k:m$ ) is placed immediately below  $g$  (or general intelligence ability). With  $g$  being the highest order of intelligence. This high hierarchical position of spatial ability is further emphasised by the recent VPR (visual, perceptual, and image rotation) intelligence model proposed and again places spatial ability directly below  $g$ . Additionally, earlier studies by El Koussy (1935) found the spatial ability factor to be over and above  $g$  and accounting for more variance than other factors.

Furthermore, a large longitudinal study, looking at the TALENT research database of precious youth, identified spatial ability as an overlooked cognitive ability. Nevertheless, the study also identified spatial ability as a 'salient psychological' trait among those who go on to achieve high educational achievements and showed spatial ability to be important for education and occupational outcomes.

To help explain the link, non-psychometrically, it seems central to recall the definition of human intelligence to aid in explaining the importance of spatial ability to intelligence (Gottfredson, 1997:p.13):

“Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader deeper capability for comprehending our surroundings- ‘catching on’, ‘making sense’ of things, or ‘figuring out’ what to do.”

Thus, intelligence can be defined as abstract creative thinking and with spatial ability defined as (Wright, Frier & Deary, 2009:p.1503):

“[t]he ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment.”

It can be difficult to see how the two are related and why spatial ability is an important cognitive ability to intelligence (abstract/creative thinking) compared to verbal or mathematical ability.

David F Lohman (1996) eloquently explains this in Chapter 6, page 98, in the section the “Importance of Spatial Abilities”:

“There is a paradox in the literature on human spatial abilities. Indeed, many of those who have studied spatial abilities have noted it with reactions that range from amusement to annoyance [...] [i]t is this: On the one hand, tests of spatial abilities – especially performance tests that use blocks or form boards or pieces of paper that must be folded and unfolded – such tests are among the best measures of g (or gf). Furthermore, spatial abilities are routinely implicated

in accounts of creative and higher order thinking in sciences and mathematics [...] [o]n the other hand, spatial abilities are often equated with concrete, lower level thinking. Thus, they are used to predict success in various practical and technical occupations, such as carpentry, auto mechanics, and the like.”

Lohman (1996) further explains the creative thinking and higher order thinking spatial abilities are implicated with on page 99:

“High levels of spatial ability have frequently been linked to creativity, not only in the arts, but in science and mathematics as well ... [f]or example, on several occasions Albert Einstein reported that verbal processes seemed not to play a role in his creative thought. Rather, he claimed to achieve insights by means of thought experiments on visualised systems of waves and physical bodies in states of relative motion. Other physicists (such as James Clerk Maxwell, Michael Faraday, and Herman Von Helmholtz), inventors (such as Nikola Tesla and James Watt), and generalists (such as Benjamin Franklin, John Herschel, Francis Galton, and James Watson) also displayed high levels of spatial abilities and reported that they played an important role in their most creative accomplishments.”

From the evidence above it can be seen that spatial abilities are not purely about 3D constructs but related to the abstract, ‘figuring out’, and the higher order level of thinking of intelligence. This finding has been used to develop the spatial teaching method intervention presented in this thesis.

## 2.5 Future Directions for Anatomy Education

“Spatial ability is a set of complex, cognitive abilities about which there are still many questions.”

- (Mohler, 2008:p.20)

The concept of spatial ability originally arose from research efforts to understand human intelligence. However, this foundation literature, together with current investigations within psychology, have had limited impact on research on spatial ability in the context of anatomy education. A true understanding of anatomy involves the appreciation of an organ's complete structure, both isolated and together as part of the human or animal body. Anatomy structures are often function specific, unsymmetrical, and irregular shapes, thus mental manipulation and memory of these 3D shapes and spatial relations are necessary to be able to truly grasp an understanding of organs. So far anatomy education studies have primarily used the Mental Rotation Test (Peters *et al.*, 1995; Vandenberg & Kuse, 1978) to measure spatial ability, have not always incorporated a control group for the re-test effect, have had variable success on designing 3D teaching methods, and so far have had limited success in designing spatial MCQs correlated to spatial ability tests.

Furthermore the relationship between spatial ability and specifically veterinary anatomy knowledge is sparsely researched, and so this thesis addresses this by using three well validated tests of spatial ability (measuring two different categories of spatial ability; mental rotation and spatial visualisation), incorporates a control group of non-anatomy learning students, designs a novel teaching method specifically designed to improve students spatial ability and thus understanding of anatomy, and designs a novel MCQ with the aim of investigating whether anatomy knowledge can be assessed spatially and non-spatially.

## 2.6 Research Questions

There is little published research in the area of spatial ability and veterinary anatomy education, and little in the UK context. This research aims to contribute to the research evidence in this area by addressing the following research questions:

- Is spatial ability a predictor of success in anatomy examinations?
- Does teaching anatomy using a traditional method improve spatial ability and anatomy understanding?
- Does anatomy teaching incorporating diagnostic imaging/ 3D images and 3D printing improve students' spatial ability and anatomy understanding?
- Is there any quantifiable difference in improvement in spatial ability between the two approaches (3D and non-3D)?

# Chapter 3 Materials and Methods

## 3.1 Introduction to Chapter 3

The purpose of this research is to explore the relationship of spatial ability to anatomy learning with veterinary students. Chapter 3 provides a detailed overview of all teaching methods presented in this research to answer the research questions as stated at the end of chapter 2. A mixed methods approach was used including both quantitative and qualitative analyses to provide a richer analysis.

### 3.1.1 Overview of Research Study Design

Figure 3.1 depicts the study design for the research in this thesis.

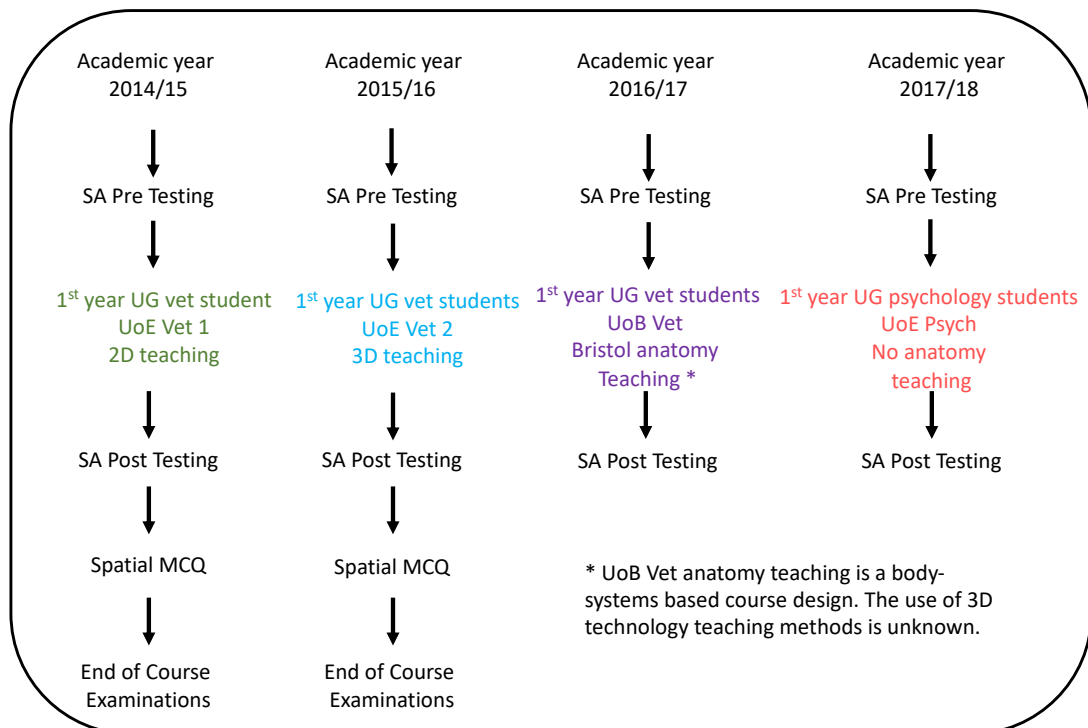


Figure 3.1 Flow diagram of research protocol for each cohort, SA = spatial ability, UG = undergraduate.

Each cohort was assessed during a different academic year to mitigate the possible ethical implications of dividing a year into two different teaching methods, which may, or may not, be advantageous to one cohort over another. Cohort University of Edinburgh Veterinary Students 1 (UoE Vet 1) were assessed during academic year 2014/15, University of Edinburgh Veterinary Students 2 (UoE Vet 2) during 2015/16, University of Bristol Veterinary Students (UoB Vet) during 2016/17, and University of Edinburgh Psychology students (UoE Psych) during 2017/18. Spatial ability testing for all cohorts was carried out before teaching or as close as logistically possible (Pre). Each cohort then went on to engage with their respective teaching method and then were re-tested with the same three spatial ability tests 15 – 16 weeks from the spatial ability Pre-testing (Post). Cohorts UoE Vet 1 and 2 were both assessed with the Spatial MCQ and end-of-course examinations at the author's institution.

## 3.2 Pilot Study

A pilot study was conducted on 10-21 graduate entry programme students (4-year BVM&S) to ensure the time allowed to complete the spatial ability tests was appropriate, and to assess how a cohort of veterinary students would perform. For the pilot study, four spatial ability tests were used:

Card rotation Test (Ekstrom *et al.*, 1976)

Paper Folding Test (Ekstrom *et al.*, 1976)

Mental Rotation Test (Peters *et al.*, 1995)

Surface Development Test (Ekstrom *et al.*, 1976)

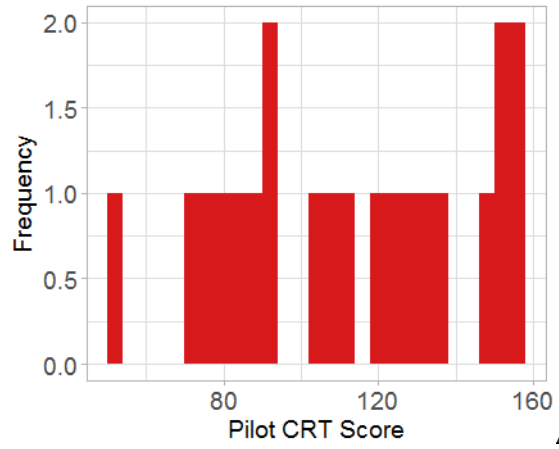
The Card Rotation Test, Mental Rotation Test, and Surface Development Test were selected from the pilot study to be used in the main study (Table 3.1 and Figure 3.2), the Paper Folding Test was not selected because this test was deemed to not involve

mental manipulation of anatomical objects as well as the three other remaining spatial ability tests, although in hindsight the inclusion of this SA test would help to explore the sub-category of spatial visualisation. The time taken to complete the main tests were not adjusted following the pilot, but the practice time for the mental rotation test was reduced to 3 minutes from 5 minutes because the students were finished before the allotted time and did not need the full time.

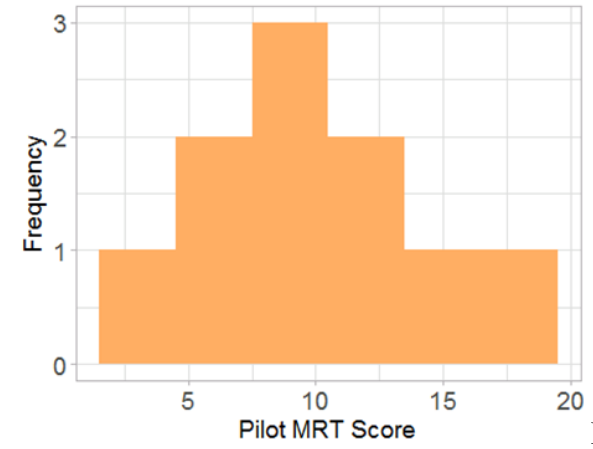
Spatial Ability Test	Mean/Median ( $\bar{x}/M$ )	SD	Range	N
CRT	$\bar{x} = 113.1$	31.39	-	21
MRT	$\bar{x} = 10.5$	4.25	-	10
SDT	M= 48.5	-	21 – 60	20
PFT	$\bar{x} = 15.0$	2.16	-	10

**Table 3.1 Mean ( $\bar{x}$ ), median (M), Standard Deviation (SD) and range of scores for each spatial ability test used in the pilot study.**

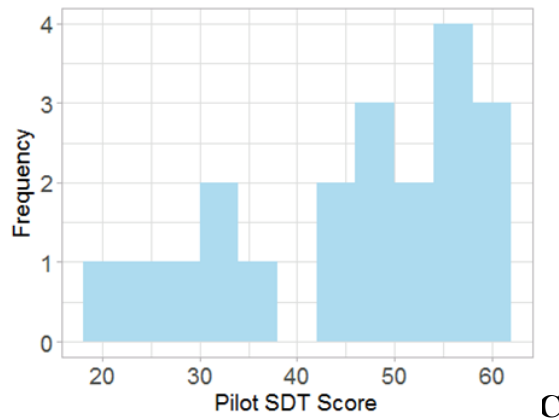




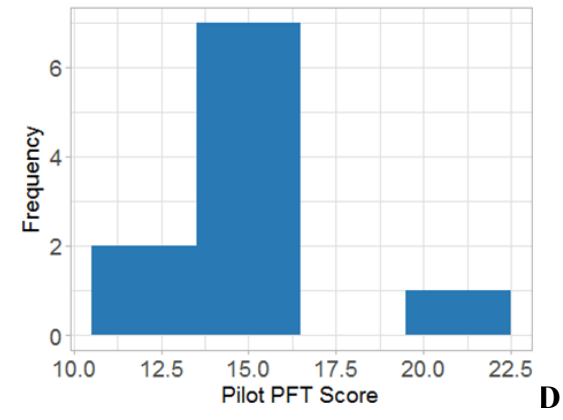
**A**



**B**



**C**



**D**

*Figure 3.2 Histograms of each spatial ability test used in the pilot study, A = CRT, B = MRT, C = SDT, and D = PFT.*

### 3.3 Testing of Participants' Spatial Ability

Ethics approval was granted from the relevant ethics committee for the population under study (The University of Edinburgh; College of Medicine and Veterinary Medicine Ethics Committee (no. SS0G/14/04), The University of Edinburgh; School of Psychology Ethics Committee (no. 95-1617/3) and the University of Bristol; Faculty of Health Sciences Research Ethics Committee (no. 2669)). Participation was voluntary, all results and data were anonymised, and participants could opt out of the study at any point.

The main purpose of this study was to assess spatial ability in relation to anatomy learning by measuring spatial ability and comparing the results of two different teaching methods. Therefore the spatial ability testing was given before (August) and after (January) the teaching received by UoE Vet 1 and UoE Vet 2. The first (Pre-testing) was measured before, or as close as logistically possible with free timetable slots and room bookings, to the start of teaching. The second (Post testing) took place at the end of the course during the start of the hindlimb section of teaching to avoid the end-of-course examination period. The period between the Pre and Post testing was 15 - 16 weeks. The exact same spatial ability tests were used for both testings.

To take into account the re-test effect (see section 2.3.5) UoE Psych students were also tested with the same three spatial ability tests 15-16 weeks apart. As these students do not learn anatomy, and thus are a control group, this testing was conducted in October and January when the first year psychology students were made familiar with the online recruitment for receiving course credit in order to increase student participation in these tests. To compare veterinary students' spatial ability across institutions the spatial ability of first-year Bristol veterinary students was measured with the same three spatial ability tests at the start of anatomy teaching (September) and then 15 – 16 weeks later (January).

A few hours or days before test administration, participants were given a brief introduction on what the test session would involve, without specifically stating that spatial ability tests were being used (however, the session was scheduled in the year timetable as 'Spatial Ability Tests'). The students were informed that their participation was contributing to research within veterinary education and education in general, the session would be roughly one hour long, it would be paper-based, it was voluntary and anonymous, and they could opt out at any point during the study.

To encourage students to participate, those attending both test sessions were entered into a prize draw to win a £50 Amazon voucher. Chocolate was also given for the Post-testing, and the Pre-testing with UoB Vet and UoE Psych cohorts, as an extra incentive to participate.

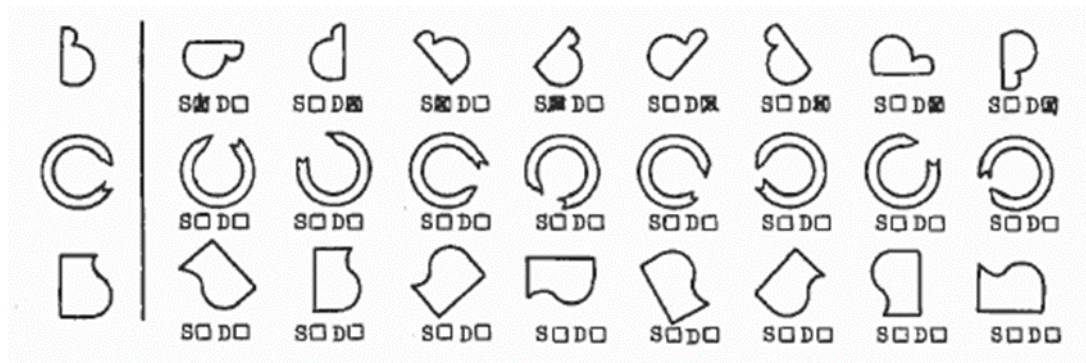
### **3.3.1 Spatial Ability Test Protocols**

Three well-validated paper and pencil tests of spatial ability were used for testing the domains of 2D mental rotation (Card Rotation Test (Ekstrom et al., 1976)), 3D Mental Rotation (Mental Rotation Test (Peters et al., 1995)) and Spatial Visualisation (Surface Development Test (Ekstrom et al., 1976)) according to the Linn and Petersen (1985) sub-categories of spatial ability.

#### **Card Rotation Test (Ekstrom et al., 1976)**

The card rotation test from the Ekstrom et al. battery of human cognitive tests (1976) come under the category of spatial orientation tests. This test involves participants viewing a primary card that has been cut into an irregular shape. There are then eight different drawings of the same card on the right of the primary card. The eight cards can be either a rotated version of the primary card or a flipped version of the primary card. Participants indicate which of these eight cards are the same or different from the primary card (Figure 3.3). The participants receive one mark for every correct

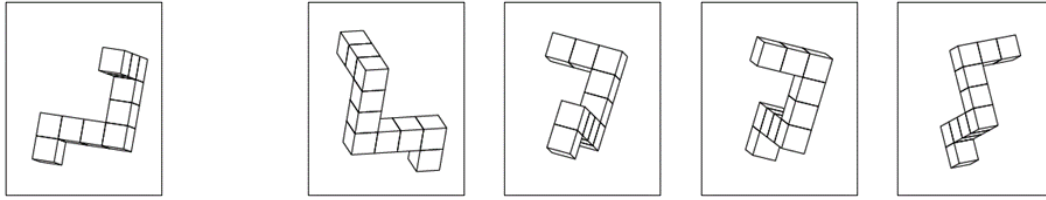
response and zero marks for every incorrect response with a maximum score of 160. There was no negative marking.



*Figure 3.3 Example practice question from Card Rotation Test. Copyright © 1976 Educational Testing Service. www.ets.org. Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.*

### **Mental Rotation Test (Peters et al., 1995)**

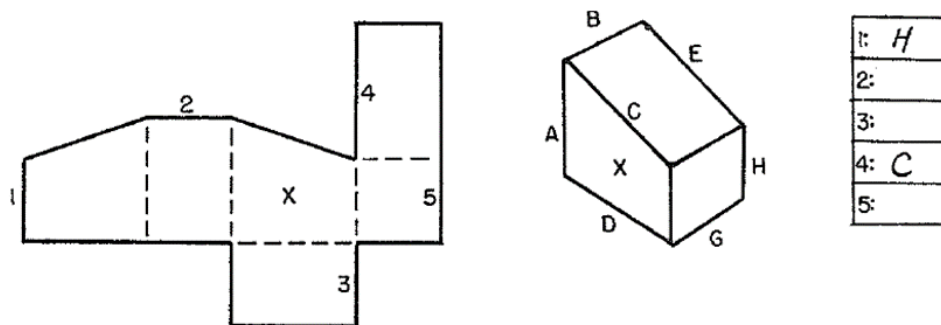
The Peters et al. (1995) redrawn version of the 1978 Vandenberg and Kuse Version A test was used (due to deteriorating physical copies). This test comprises a target figure on the left and four comparison figures on the right. The figures are composed of 3D cubes arranged into a 3D structure. The participant has to correctly identify which two comparison figures are the same as the target figure by rotating in both 2D and 3D planes on the vertical axis. The remaining two distractor comparison figures are either a mirror image of the target figure or a rotated version of a different target figure (Figure 3.4). Participants receive one mark for correctly identifying the two correct comparison figures with a maximum score of 24. Both correct comparison figures must be identified to gain the mark. There was no negative marking.



*Figure 3.4 Example of practice question of 3D Mental Rotation Test. Copyright permission granted by Peters et al. (1995).*

### **Surface Development Test (Ekstrom et al., 1976)**

The surface development test from the Ekstrom et al. battery of human cognitive tests (1976) comes under the category of tests of visualisation ability (Figure 3.5). There is a drawing of a 3D structure and a diagram of a flat ‘floor-plan’ version of the same 3D structure, showing how it may be folded or cut to make the 3D structure. One surface of the 3D structure is marked with an ‘X’ and the corresponding surface on the flat ‘floor-plan’ version is also marked with an ‘X’. Each of the sides or dotted lines of the flat ‘floor-plan’ version are given a number. The participants aim to correctly match the numbers on the flat ‘floor-plan’ drawing with the letters on the 3D structure. Participants receive a mark for each correctly paired number and letter, and zero marks for each incorrectly identified pairing with a maximum score of 60. There was no negative marking. It was noted that the instructions of the Surface Development Test were not read out consistently for the first year of testing involving Cohort UoE Vet 1. It was also noted that the answers to the practice question and the information regarding the position of ‘X’ were not read out, although these instructions were written on the test paper. Therefore to keep the reading of the instructions consistent the remaining three cohorts had the note about the ‘X’ read during the practice time and were asked to refer to the answer paragraph below the practice question.



*Figure 3.5 Example of Surface Development Test. Copyright © 1976 Educational Testing Service. www.ets.org. Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.*

For Cohorts UoE Vet 1 and 2 it was noted that after spatial ability testing some of the MRT test papers had printed in grey rather than black and white. Therefore this could affect the performance of these participants on the MRT, and this was accounted for during statistical analysis.

For this research, the tests were not negatively marked. However negative marking can be used with these three tests to factor in participant guessing.

The tests were given in this order:

1. Card Rotation Test
2. 3D Mental Rotation Test
3. Surface Development Test

Each participant received a test pack on the day of testing including; a consent form (Appendix 1), a demographic information sheet (Appendix 2) and each spatial ability test enclosed within an envelope to ensure continuity of testing. The instructions for

each test were read out to the students before the start of that test, and this was done as consistently as possible.

Each test had a practice question(s), and the students were given 2-3 minutes to work on this - 2 minutes for the Card Rotation Test and Surface Development Test, and 3 minutes for the 3D Mental Rotation Test. All participants started and ended the main test at the same time. Each test had two parts, and the students were given a one minute rest period between these parts.

### **3.4 Demographic Details of Participants**

Each participant completed a consent form and demographic form before taking the spatial ability tests (Appendix 1 and 2). Demographic details of the participants were collected, including:

- Age
- Gender
- Left or right handed
- Title of any previous degrees

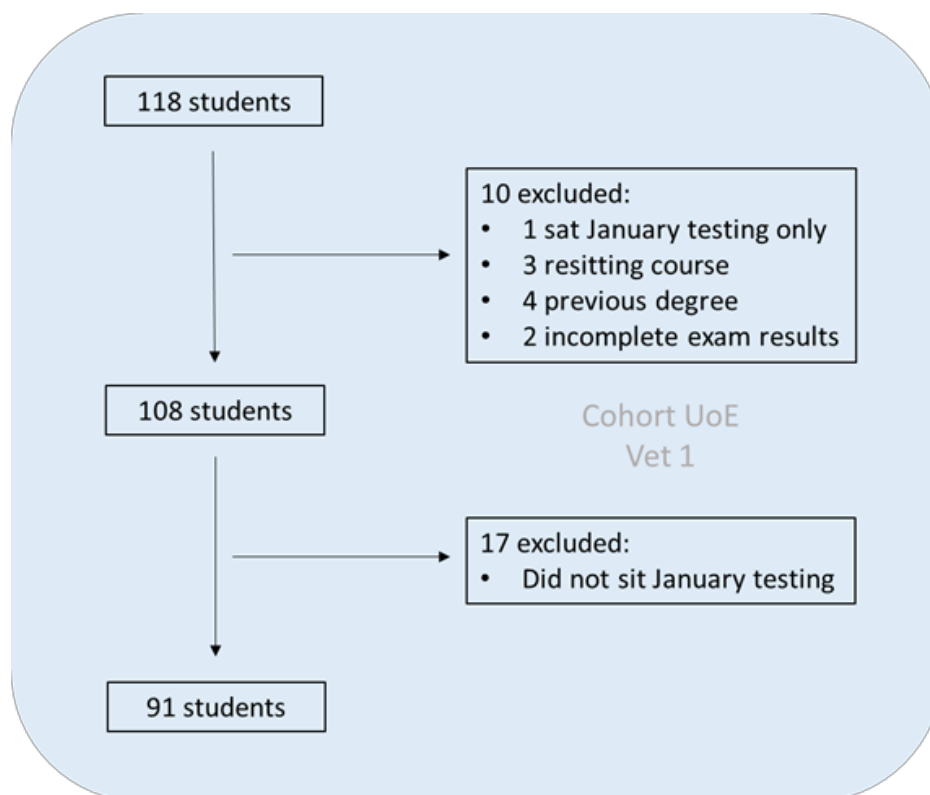
For Cohorts UoE Vet 2, UoB Vet, and UoE Psych an inclusion on video game usage (how often the participant played video games) was added to the demographic sheet (Appendix 2) for further research involving such usage, spatial ability, and anatomy learning.

Participants who were resitting a veterinary degree course (identified by course enrolment), and/or who had a previous degree relating to anatomy, and/or had incomplete exam results were excluded from the study (see section 3.4.1 'Participant

Numbers' below). The demographic detail form was included in both the Pre-teaching and Post-teaching spatial ability testing sessions to ensure correct participant identification between testings. All the demographic details were taken to be what was stated on the first spatial ability testing (as some participants may have had a birthday between the two spatial ability testings).

### 3.4.1 Participants Numbers

Figures 3.6, 3.7, 3.8, and 3.9 show the number of participants included in the study and those excluded including a reason for exclusion.



*Figure 3.6 Number of participants for UoE Vet 1.*



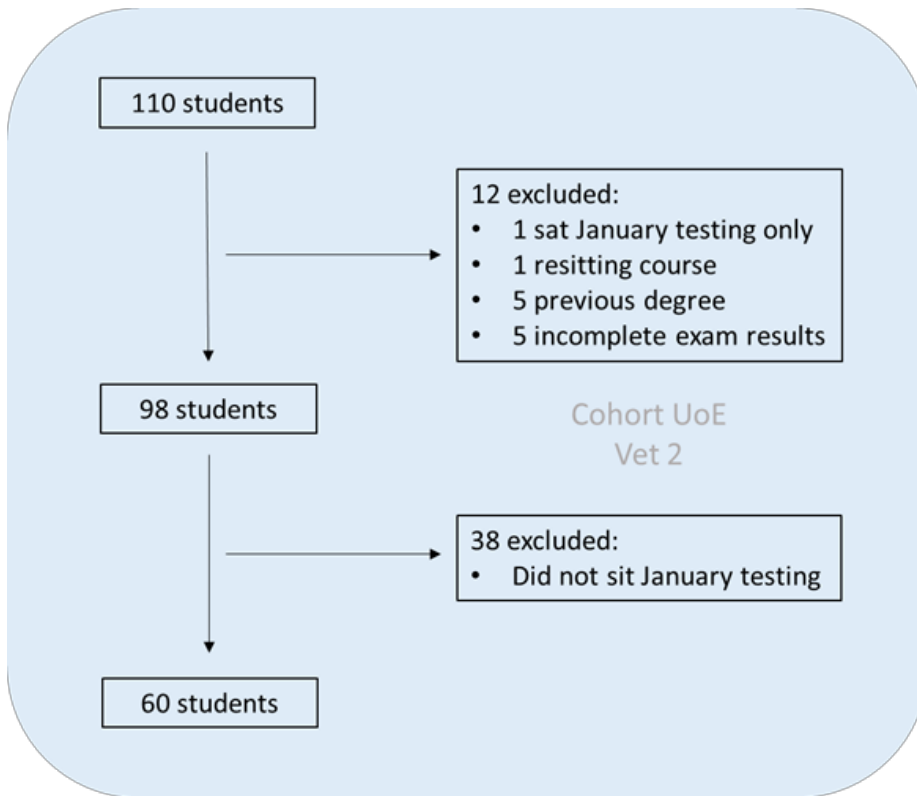


Figure 3.7 Number of participants for UoE Vet 2.

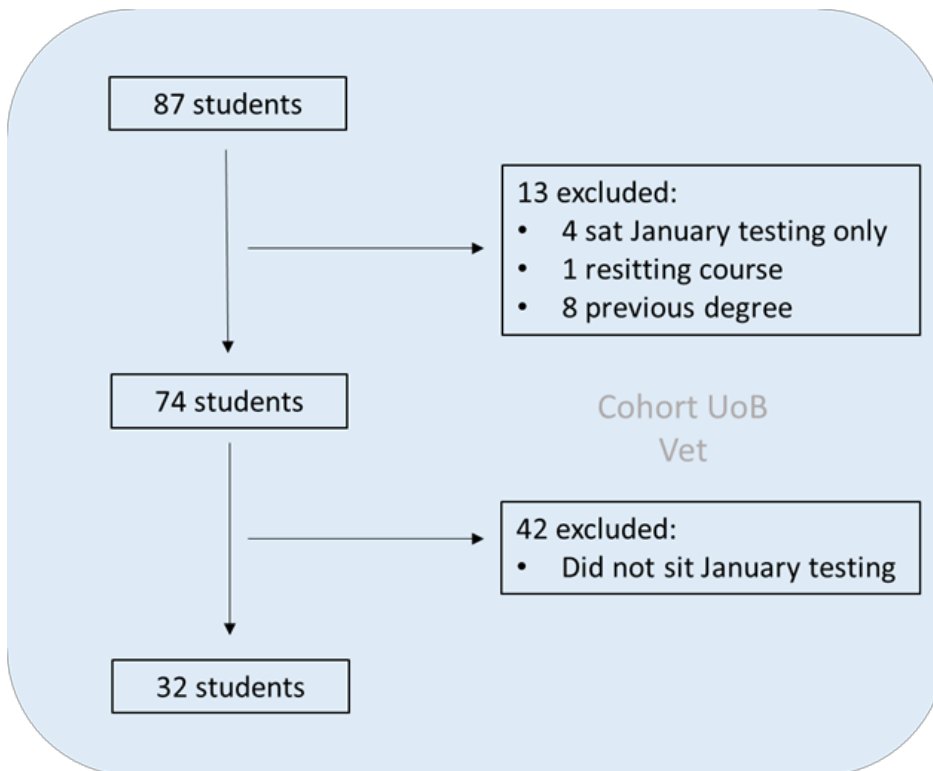
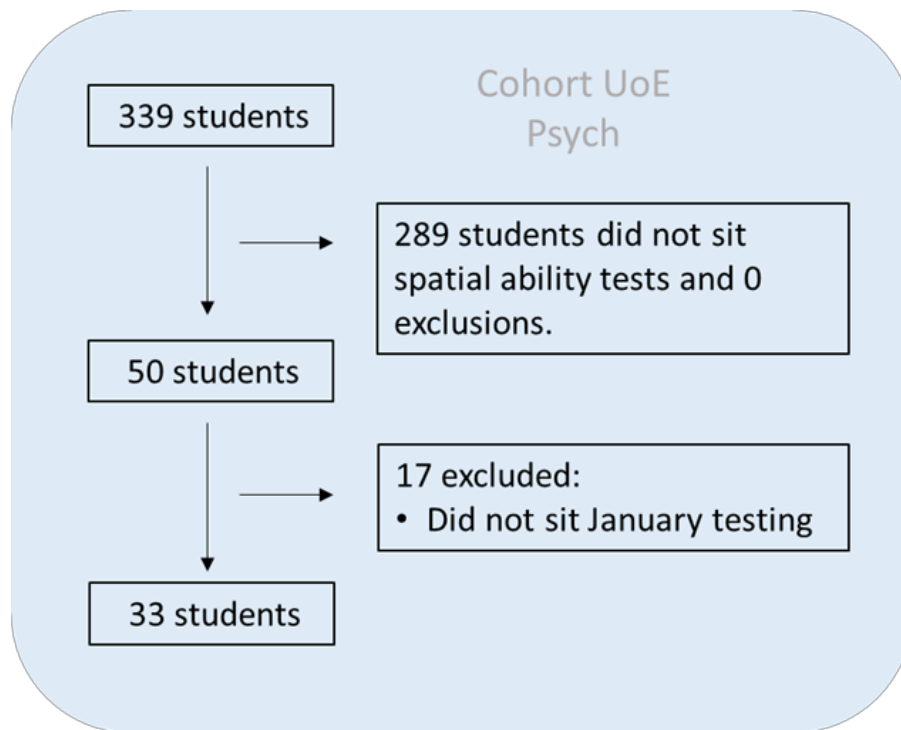


Figure 3.8 Number of participants for UoB Vet.

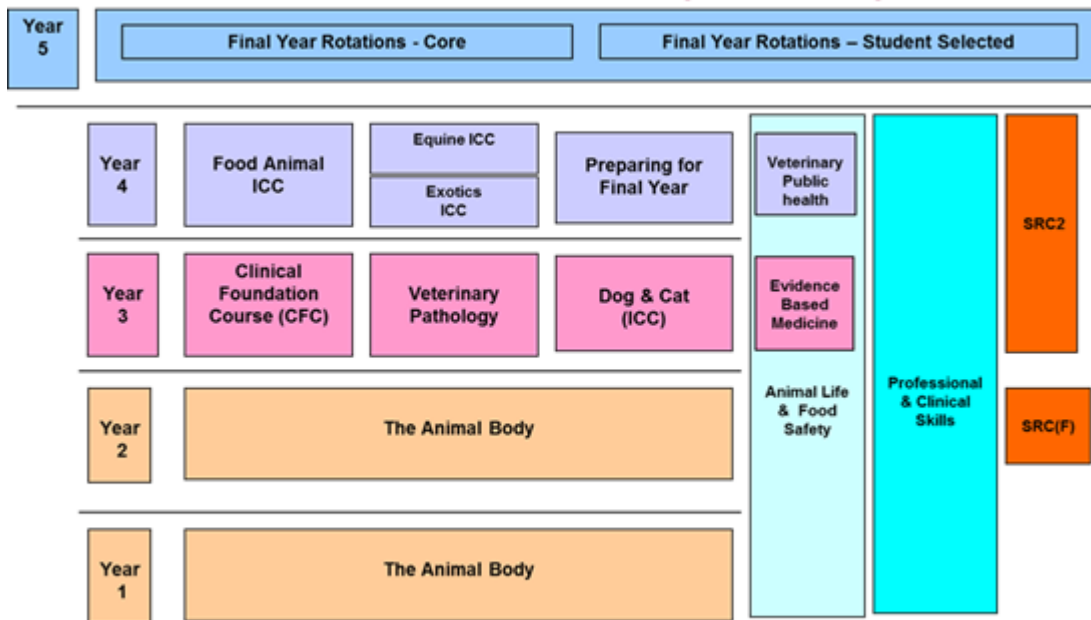


*Figure 3.9 Number of participants for UoE Psych.*

### 3.5 Overview of The Animal Body 1 Course

The Bachelor of Veterinary Medicine and Surgery (BVM&S) course at The Royal (Dick) School of Veterinary Studies, University of Edinburgh, is an integrated program, which can be studied over four or five years. In the five year programme, first-year students learn basic anatomy based on the canine during The Animal Body 1 course (AB1) which runs alongside the Animal Life and Food Safety 1, and Professional and Clinical Skills 1 courses (Figure 3.10).

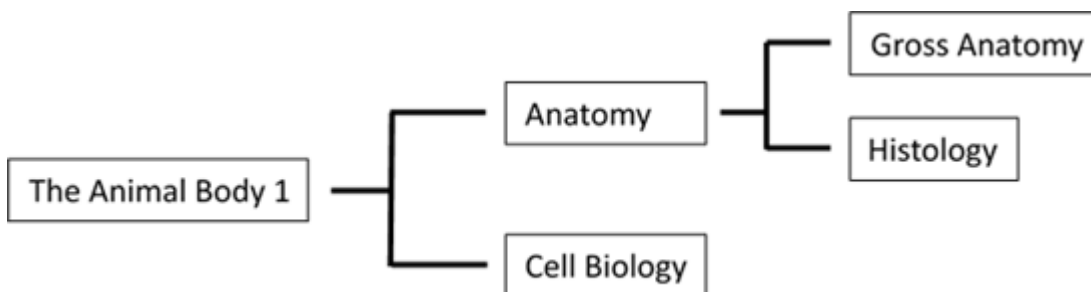
## BVM&S Curriculum (2015-16)



ICC = Integrated Clinical Course, SRC= Student Research Component

*Figure 3.10 Curriculum outline for the R(D)SVS 5 year BVM&S programme.*

The AB1 course is divided into Anatomy and Cell Biology sections and runs from September to February. The Anatomy course begins with ‘Introduction to AB1 Lectures’ which consists of three introductory lectures (1. Language of Anatomy, 2. Anatomy of the Body, 3. Physiology of the Animal Body). The rest of the anatomy course comprises histology and gross anatomy (Figure 3.11).



*Figure 3.11 Division of The Animal Body 1 course.*

The research of this thesis focuses on the gross anatomy component of the AB1 course. The gross anatomy course is based on canine anatomy delivered by body regions with lectures (of 50 minutes) and corresponding dissection practical(s) (1hr 50mins – 2hr 50mins) (Table 3.2).

Anatomical Region	Lectures		Practicals	
	n	Total Time	n	Total Time
The Language of Anatomy	1	50 mins	N/A	N/A
The Anatomy of the Adult Body	1	50 mins	N/A	N/A
The Physiology of the Adult Body	1	50 mins	N/A	N/A
Introduction to dissection and dissection technique	N/A	N/A	1	2 hrs 50 mins
Skull (Dry lab)	N/A	N/A	2	1 hr 50 mins
Head	2	50 mins	1	2 hrs 50 mins
			2	1 hr 50 mins
			3	2 hrs 50 mins
Introduction to Joints	1	50 mins	N/A	N/A
Introduction to Nervous System 1	1	50 mins	N/A	N/A
Forelimb	4	50 mins	1	2 hrs 50 mins
			2	2 hrs 50 mins
			3	2 hrs 50 mins
			4	2 hrs 50 mins
			5	1 hr 50 mins
Vertebrae (Dry lab practical)	2	50 mins	1	1 hr 50 mins
Introduction to Nervous System 2	1	50 mins	N/A	N/A
Biomechanics and the Vertebral Column	1	50 mins	N/A	N/A
Applied Clinical Anatomy 1	N/A	N/A	1	2 hrs 50 mins
			2	2 hrs 50 mins
			3	2 hrs 50 mins
Neck	1	50 mins	1	1 hr 50 mins
			2	1 hr 50 mins
Pharynx	1	50 mins	1	2 hrs 50 mins
Trunk and Body Wall	1	50 mins	1	2 hrs 50 mins
Thorax	2	50 mins	1	2 hrs 50 mins
			2	1 hr 50 mins
Abdomen	2	50 mins	1	1 hr 50 mins
			2	1 hr 50 mins
Hindlimb	4	50 mins	1	1 hr 50 mins
			2	1 hr 50 mins
			3	1 hr 50 mins
			4	1 hr 50 mins
			5	1 hr 50 mins
Pelvis	1	50 mins	1	1 hr 50 mins
Introduction to Ultrasound	1	50 mins	1	1 hr 50 mins
Introduction to Lymphatics	1	50 mins	N/A	N/A

**Table 3.2 Number and duration of anatomy lectures and practicals within AB1 per body region. Topics highlighted in blue are the topics modified for this research.**

The anatomical regions involved in the teaching intervention of this research study are highlighted in blue in Table 3.2 above, namely:

- Head
- Introduction to Joints
- Forelimb
- Pharynx
- Hindlimb
- Introduction to Ultrasound

These are the regions for which teaching methods were varied in the study design (2D Teaching Method and 3D Teaching Method). The rest of the teaching, delivered by other lecturers, remained the same for the two consecutive years of the research study.

The order of topic delivery was as shown in Table 3.2 (although this was slightly changed for the 3D teaching method see Figure 3.24), with practical sessions given roughly parallel to lectures. The Cell Biology and Histology components of the AB1 course ran alongside the Gross Anatomy section.

The Gross Anatomy section of the AB1 course was assessed with formative and summative assessments (see section 3.6 of this chapter). Formative assessment involved an in-course dissection workbook and MCQ, and summative assessment was carried out by end-of-course examinations.

### **3.5.1 Lecture Overview**

Lectures were delivered to the students on almost a daily basis in 50-minute teaching slots. Lectures were delivered in a traditional format using Microsoft PowerPoint with the content structured around learning objectives, which remained the same for the two consecutive years of this study. PowerPoint slides were available to students before their attendance at lectures via the school's VLE (virtual learning environment) platform LEARN, along with a maximum 6-page student handout most commonly consisting of salient lecture slides. A Smartboard was used to draw anatomical diagrams for the 2D teaching method. The additional use of 3D computer models, 3D printed models, and 3D diagrams were used in the 3D teaching method. Other lecture enhancing technologies such as interactive clicker questions were not used to reduce confounding variables.

### **3.5.2 Practical Class Overview**

The practical classes involved dissection groups of three students (occasionally two or four depending on year size) tasked with dissecting a formalin-fixed dog<sup>2</sup>. The dissection groups remained the same for the whole course with groups rotating to a different dog for each new practical. Rotating allowed students to experience a variety of breeds and specimens of differing quality (for example, obese specimens could be more difficult to dissect or very small specimens may not show some salient features clearly). The students were given dissection instructions at the start of each practical, which were also accessible to students via the VLE platform LEARN before the class. Textbook pages on the forelimb were also available to help students become acquainted with the recommended dissection guide (Evans and de Lahunta (2010) *Guide to the Dissection of the Dog*. Seventh edition. Elsevier, Missouri). A 10-minute live video demonstration was given at the start of each practical outlining the dissection protocol to be followed as well as identifying any salient structures on

---

<sup>2</sup> All dogs were ethically sourced.

a prosected specimen. This introduction was delivered in the same way for cohorts UoE Vet 1 and 2. The prosected specimens used for the 10-minute introduction were pinned with numbered push pins, and an accompanying key was provided, the students could refer to these specimens during the class to aid their dissection. The prosected specimens remained out during subsequent classes to facilitate revision.

### 3.6 Assessment of The Animal Body 1 Course

The assessment of the AB1 course involved both in-course and end-of-course assessments; Table 3.3 shows the breakdown and weighting of each element.

Assessment	Question Type	Percentage (%)
End-of-Course (80% of final exam mark)	MCQ	10%
	Interpretation	30%
	Short Answer	30%
	Spot	10%
In-course (20% of final exam mark)	Oral assessment	5%
	Cell biology MCQ	2.5%
	Cell biology presentation	2.5%
	Anatomy workbook 1	3%
	Anatomy workbook 2	2%
	Histology workbook	3%
	Spatial MCQ	2%

*Table 3.3 Breakdown and proportional weight of the AB1 examinations.*

The AB1 end-of-course exam covers the entire course including questions on cell biology, histology, and gross anatomy. Only the results for questions on gross anatomy were eligible for inclusion in this research. The removal of histology and cell biology from the examinations marks was conducted, with some spot questions covering a small section on histology still included. Due to pre-existing examination



structure and logistics, the oral exam score was a composite mark for gross anatomy, histology, and cell biology.

### **3.6.1 End-of-Course Assessment**

The end-of-course professional examinations include a variety of question types:

- Multiple choice questions, MCQ (~1.5 minutes/question)  
Questions where students have four to five options to choose from and give the best response.
- Interpretation (~15 minutes/question)  
Questions where students need to interpret information (data, graphs, photographs) to problem solve and correctly answer questions.
- Short Answer (~10 minutes/question)  
Questions require the student to give a short written answer displaying core knowledge.
- Spot (~10 minutes/question)  
Questions are based on pinned prosected specimens involving accurate identification of structures and application of knowledge.

The above question types are used in three examination papers for the AB1 end-of-course examination:

1. MCQ and interpretation paper
2. Short answer paper
3. Spot paper

The students are required to answer all questions in each paper.

The number of questions in the examination changed between UoE Vet 1 and UoE Vet 2 (Table 3.4). This was outwith the author's control and arose due to AB1 exam board guidance, which stated that the examination exceeded the recommended length as per University of Edinburgh assessment guidelines. The number of MCQs was increased, and the number of short answer questions was reduced for gross anatomy questions for UoE Vet 2. These differences were taken into account for statistical analysis.

Question Type	Number of gross anatomy questions	
	UoE Vet 1	UoE Vet 2
Interpretation	2	2
MCQ	6	10
Short answer	6	4
Spot	13	13
Total Exam	27	29

*Table 3.4 Differences in the number of examination questions between cohort UoE Vet 1 and UoE Vet 2 for gross anatomy only.*

### **3.6.2 In-Course Assessment**

The AB1 in-course assessment includes:

1. Oral assessment (20 minutes combined cell biology, histology, and anatomy oral examination divided between two examiners).
2. Practical workbooks (anatomy and histology).

3. Spatial Anatomy MCQ (designed for this research to replace workbook 3)
4. Cell biology MCQ and presentation.

The anatomy workbooks (point 2. above) require the students to document their dissections in the form of accurately titled, labeled, and orientated photographs. The workbooks are critiqued primarily on the quality of dissection, the accuracy of labeling, and correct use of anatomical terminology, along with general presentation (see Appendix 4 for Workbook Mark sheet). Ordinarily, the anatomy workbooks are handed in at three points during the course. Workbook 1 includes; the skull (1 & 2), Head (1-3), Forelimb (1-5), Vertebrae, Ribs & Sternum, and Applied Clinical Anatomy 1. Workbook 2 includes; Neck, Pharynx, Body Wall, Thorax (1 & 2), Abdomen (1 & 2), and Applied Clinical Anatomy 2. Workbook 3 includes; Hindlimb (1-5), Pelvis, Applied Clinical Anatomy 3, and Ultrasound.

However, for this research, the third anatomy workbook was replaced by the Spatial MCQ designed to test anatomy knowledge both spatially and non-spatially as a means to quantify the differences, if any, between the two UoE Vet cohorts receiving different teaching methods (see section 6.2). Therefore the Spatial MCQ involved questions on the topics of the Hindlimb, Pelvis, and Ultrasound (but did not include Applied Clinical Anatomy 3 because these classes were not assessed in the workbooks).

For UoE Vet 2, a workbook feedback session was introduced and given during the 6th week of teaching with the aim of providing the students with some feedback, and relieve anxieties around the anatomy workbook in-course assessment for the Gross Anatomy component of the course. This feedback session involved dissection groups submitting a labeled dissection photograph taken during their Head 2 dissection. The

photographs were then reviewed to the whole cohort by the Author and a senior anatomy lecturer demonstrating examples of strong and weak workbook presentation.

## 3.7 2D Teaching Method

### 3.7.1 Lectures (2D)

For the last four decades, the didactic component of anatomy teaching in the first year BVM&S degree has comprised of lectures in the form of handwritten notes and accompanying hand-drawn diagrams. This content was delivered in the lecture theatre with students copying in real time from an overhead projector. This method was used for the 2D teaching method with both the lecture notes written and diagrams drawn on a SmartBoard (Figure 3.12 and 3.13).

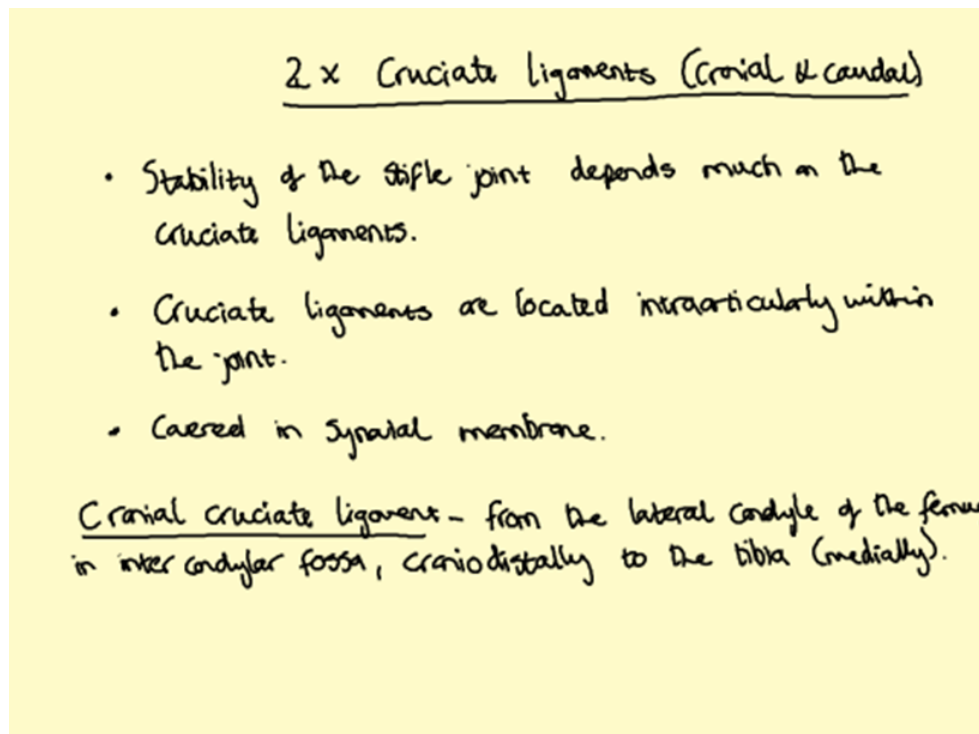
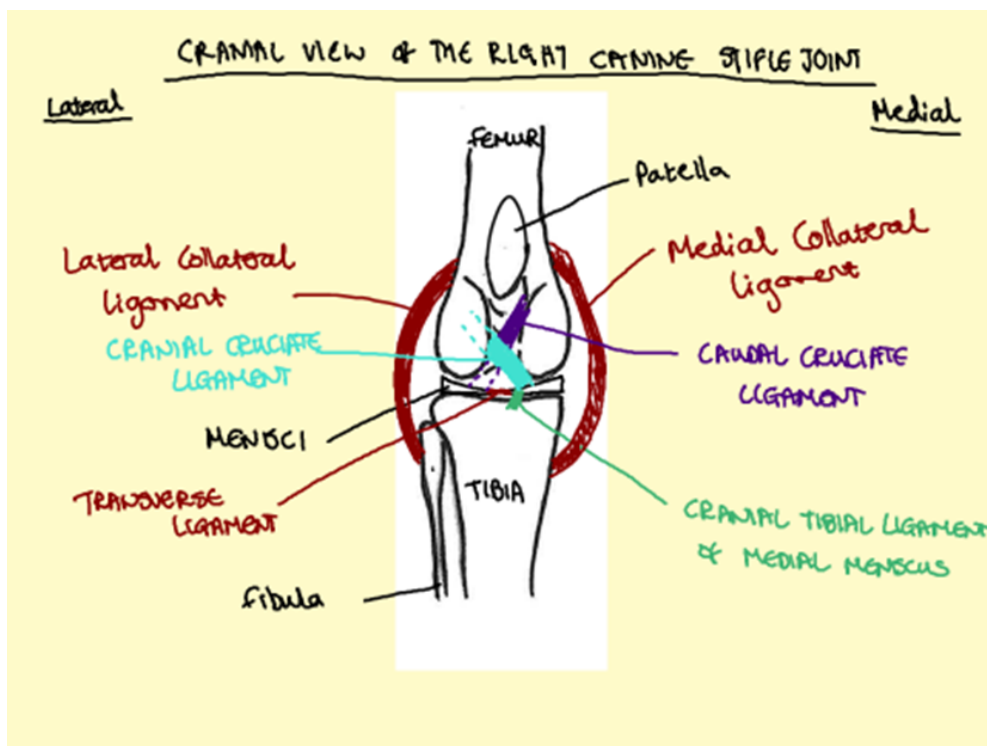


Figure 3.12 Lecture slide from PowerPoint presentation handwritten during 2D teaching.



*Figure 3.13 Diagram from 2D teaching method of stifle joint highlighting dashed lines to represent depth and spatial relations.*

The 2D teaching method has proved to be a robust form of teaching demonstrated by students' success in their anatomy exams and positive external examiners comments. This teaching method is not designed to teach or present the course content in a 3D spatially orientated format. All the in-lecture diagrams are 2D (for example, see Figure 3.13 showing the canine stifle joint), sometimes at different conventional views, but with very little to no appreciation of structural arrangement in 3D space.

In the 2D teaching method for this research, the lecture notes and diagrams were written on a smartboard and projected onto two screens in the lecture theatre for students to directly copy, and the lecture content was read out and repeated while writing for students to follow and copy down at a manageable pace. The lecturer's written notes, PowerPoint slides, and associated diagrams were uploaded onto the school's VLE (LEARN) for students to access and supplement their versions after the lecture. Some of the diagrams from the lectures were selected for the students to

complete in their own time as ‘homework’ due to the time-consuming nature of handwriting lecture notes and drawing diagrams in real time.

The 2D diagrams were drawn/annotated using the SmartBoard in the lecture theatre (Figure 3.13). Due to time constraints, and to ensure the students coherently received all lecture material, the drawn/annotated diagrams were supported by annotated PowerPoint slides with annotations listed in the same order as was drawn in the lecture for the students to copy from at home.

The 2D diagrams made it very difficult to explain how all the structures related to each other spatially, and this was often explained in more detail through the written aspect of the lecture. For example, an explanation of the origin and insertion of structures (e.g., cruciate ligaments) was better illustrated by a written explanation, compared to a drawing showing stifle ligaments (containing colours and lines running in various directions).

### **3.7.2 Practical Classes (2D)**

AB1 practical classes involve instructor-led canine cadaver dissection with labeled prosected specimens provided for student reference and an in-course assessment. For the in-course assessment, students were required to produce two workbooks cataloging their dissections containing images and appropriate annotations (see section 3.6.2 above). A third workbook that had been used for previous cohorts was replaced with the Spatial MCQ (see section 6.2) for the cohorts in this study.

Students could also borrow ‘bone boxes’ from the library to aid their learning. Bone boxes contain a complete, disarticulated canine skeleton. The students could borrow these from the library for one week at a time throughout the year.

### 3.8 3D Spatial Teaching Method

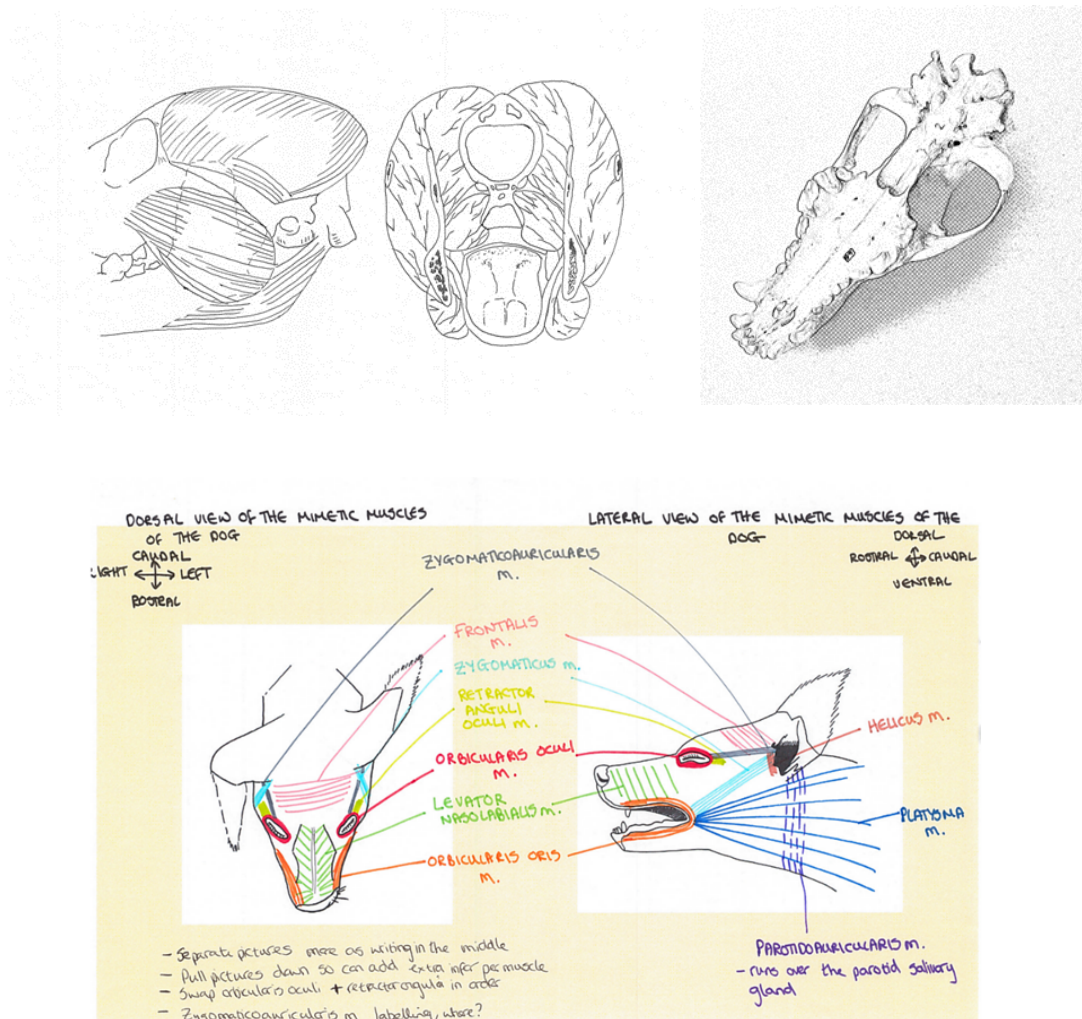
The 3D teaching method is a completely new method designed for this research with a focus on teaching anatomy spatially. This teaching method used a combination of diagnostic images (e.g., radiographs, computed tomography (CT) scans, magnetic resonance imaging (MRI) scans, and ultrasound images - see Figure 3.14), 3D printed models, 3D computer models and, for some of the lectures, textbook images. Textbook images were kept to a minimum and primarily used for teaching the anatomy of the pharynx. The images were chosen with the aim of helping students develop a 3D spatial understanding.



*Figure 3.14 Example radiographs with orthogonal views from 3D spatial teaching method.*

Initially, the 3D teaching method involved utilising 3D printed models (created using CT scans of a canine bone box and the game modeling software Blender

(Roosendaal, 2002)) and the drawing/labeling of 2D diagrams. An important distinction in the use of 2D diagrams was that in the 3D teaching method the diagrams were selected to illustrate different spatially-orientated views (Figure 3.15).



**Figure 3.15** Example of diagrams showing different orientations initially considered for use for the 3D teaching method.

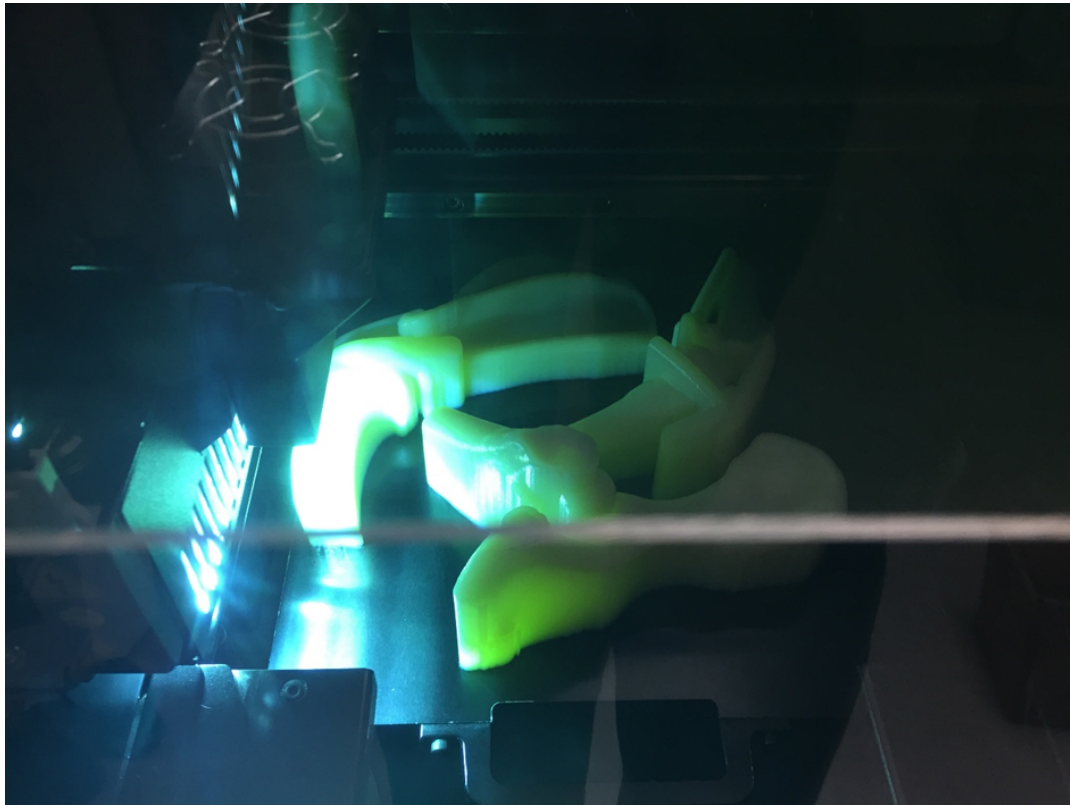
In addition, and consistent with the literature on learner-controlled 3D models (Garg, Norman & Sperotable, 2001; Barrett & Hegarty, 2016; Nguyen, Nelson & Wilson, 2012), further enhancements to emphasise the spatial aspects were developed. These developments involved uploading the 3D printed models onto a 3D viewer website called Sketchfab (Denoyel, Pinson & Passet, n.d., <https://sketchfab.com/>) which



students could use to self-control the rotation of the models, and this became a main teaching source for the 3D teaching method (Figure 3.20 and 3.21).

### **3.8.1 Development of 3D Printed Models**

3D printing is a type of rapid prototyping and is described as an additive manufacturing process. Additive manufacturing means that sequential layers of material are laid down on top of one another to build up a complete model (Figure 3.16). In contrast to subtractive manufacturing, where the model is created, and parts removed afterward to create the final product. CT and MRI scans can be used to create a digital input to 3D printers in the form of STL files (Stereolithography). The STL files created from CT or MRI scans are then modified by computer programs before printing, allowing endless possibilities regarding adding structures to the 3D models, scaling the model, digitally sculpting the model, or slicing the model in any plane. This capability was utilised to make the 3D printed models for this research.



*Figure 3.16 Photograph of 3D printer printing canine shoulder, stifle and hip models.*

The main areas of teaching on the AB1 course included in this research were the head, forelimb, hindlimb, and pharynx. It was decided that 3D printed models would be created of the main limb joints including the shoulder, elbow, hip, and stifle with the ligaments/tendons printed in a flexible material and the bones in a solid material. A model of the pharynx was to be included, but due to the limitations of the Author's skills, a 3D computer model was created but not printed.

The following equipment and software/programs were used to create the 3D printed models:

- Siemens<sup>3</sup> Volume Zoom, 4-slice helical CT unit.

---

<sup>3</sup> Siemens AG, Werner-von-Siemens-Straße 1, 80333 Munich, Germany.

- OsiriX<sup>4</sup>- computer software that converts CT scans to STL files.
- Blender (Roosendaal, 2002)- software used for creating animated films, visual effects, art, 3D printed models, interactive 3D applications and video games.
- Objet 260 Connex<sup>5</sup> 3D printer- the 'base' materials used were VeroWhite Plus and TangoPlus. Digital blends of these two materials gave a range of flexibility options.

An initial attempt to create the 3D printed models used CT scans from patients in the R(D)SVS small animal hospital. However, it was found that the clinical scans contained too much anatomical information that was not solely isolated to the bones, ligaments, and tendons of the joints. To overcome this, a set of medium-sized dog bones from the student bone box collection were selected, and scans of the individual bones were created and converted to STL files using Osirix.

A free computer program called Blender (Roosendaal, 2002), designed for creating computer games, was used to edit, sculpt, and modify the STL files into desirable models ready for 3D printing. The method was complex and is summarised below.

1. Tidying the meshes of the bones

The structure of the bones (mesh) imported into Blender is made up of millions of polygons each with n number of vertices, edges, and faces depending on their shape.

---

<sup>4</sup> Pixmeo SARL, 266 Rue de Bernex, CH-1233 Bernex, Switzerland.

<sup>5</sup> Stratasys, 9600 West 76th Street Eden Prairie, MN 55344 United States.

Although detailed CT scans of the bones were taken some detail is lost when the STL file is created, and this means there are holes in the mesh.

## 2. Articulating the bones correctly

Articulating the bones involved a great deal of manipulation and trial and error. The bones required for each joint were aligned into their articular positions; an exact anatomical fit was difficult to achieve, and therefore a compromise with the best fit for articulation was used.

## 3. Making the ligaments

This process took the longest and involved much trial and error as the ligaments were designed from scratch to be printed in a flexible material, meaning they could not touch the bones as the models were 3D printed (additive manufacturing with two materials).

## 4. Finishing touches ready for 3D printing

Finishing touches were then made to the models in order to be 3D printed.

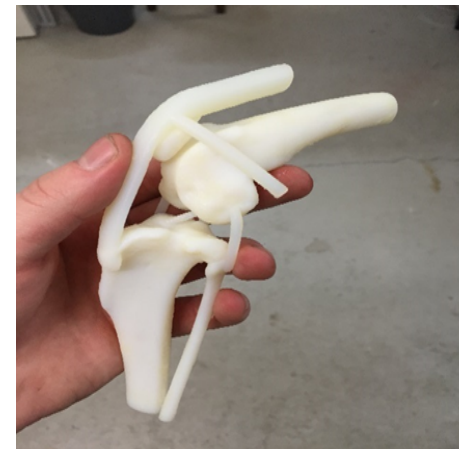
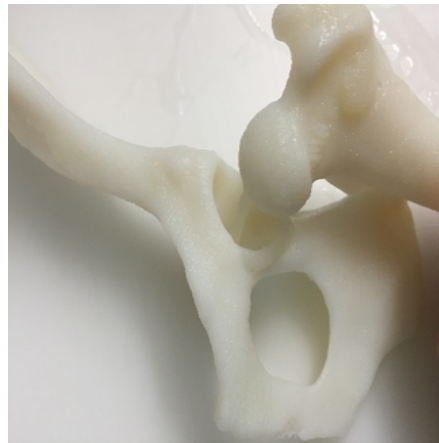
The models were printed in collaboration with the Edinburgh College of Art where the 3D printer was located.

The stifle joint was the first model to be constructed as it is the most complex joint, and it was felt that this presented the greatest potential to learn the art of computer modelling. The first 3D print was of the stifle joint at half scale with ligaments/tendons at flexibility grade 2 (the 3D printer has a flexibility scale ranging from 1-8) (Figure 3.17). This model was found to be too flimsy; so following

consultation with orthopaedic surgeons at the R(D)SVS small animal hospital, number 6 was selected to ensure a true likeness to real ligaments as well as ensuring robustness. The final 3D printed models were made to scale (Figure 3.18), and three prints of each joint were produced.



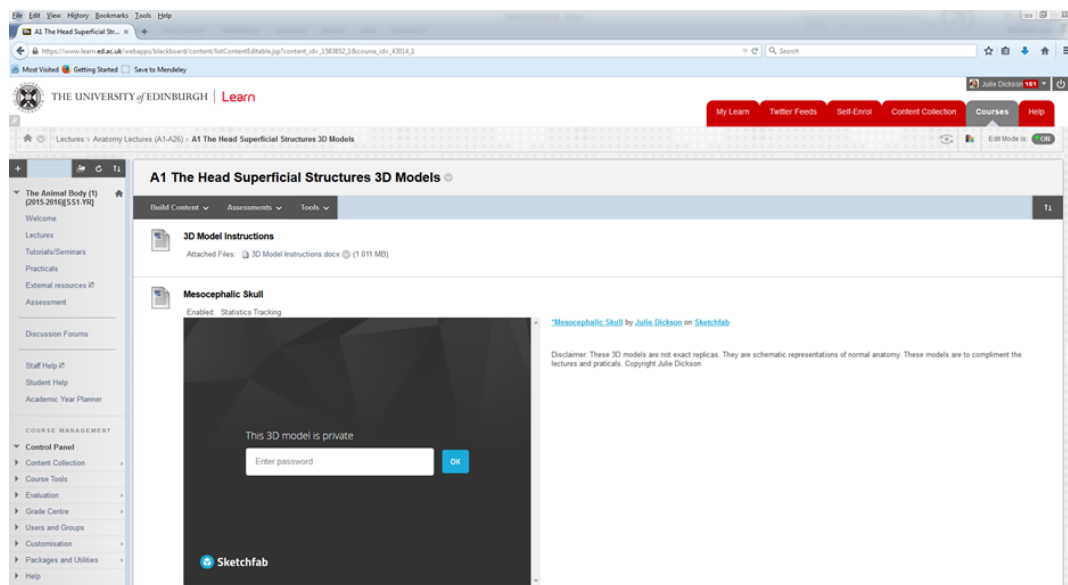
*Figure 3.17 Half scale 3D printed stifle model of option 2 flexibility. Photograph courtesy of Richard Collins.*



***Figure 3.18 Photographs of the shoulder (top left), elbow (top right), hip (bottom left) and stifle (bottom right) 3D printed joint models. Photographs courtesy of Richard Collins.***

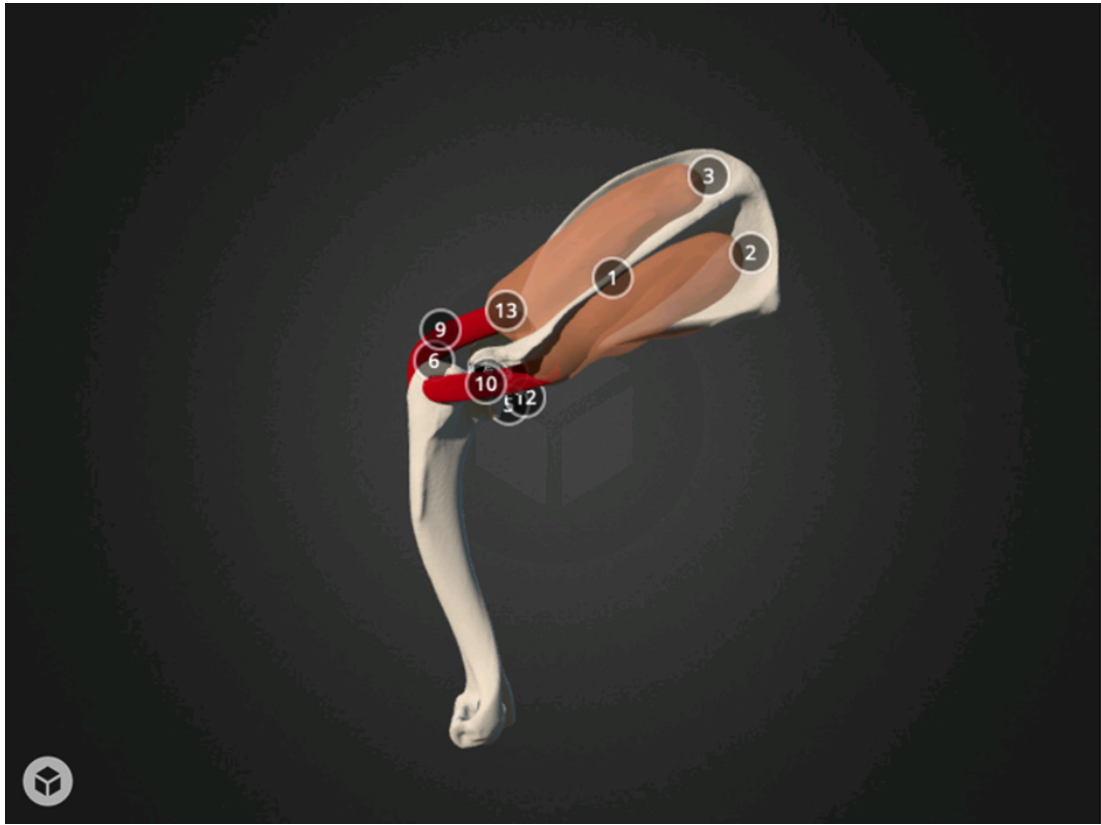
## 3.8.2 Development of 3D Computer Models

The 3D models created on Blender were uploaded and viewed on a website called Sketchfab (Denoyel, Pinson & Passet, n.d.). To access this website users must register and create a profile. Following discussion with the company about the needs of the project, an upgrade was obtained, which allows more flexibility for viewing the 3D models. The Sketchfab models were embedded on the VLE where lectures, practicals, discussions, and courses can be accessed by students (Figure 3.19). The students had access to the 3D models from the morning of the relevant lecture. All the 3D models were password protected and not publicly available.



**Figure 3.19** Screenshot of Sketchfab viewer embedded in LERN.

The 3D computer models included more anatomical features than was possible to include within the 3D printed models, due to the complexities of printing structures such as the menisci of the stifle joint, and the supraspinatus, infraspinatus and subscapularis muscles of the shoulder joint (Figure 3.20) moreover, also financial constraints.



*Figure 3.20 Shoulder joint model showing extra features not on 3D printed model.*

Because Sketchfab provides a way of viewing structures in 3D, it was used to create 3D computer models of all the 2D diagrams drawn in the 2D teaching method (Table 3.5), to make the teaching as spatially orientated as possible, and also distinct from the 2D teaching method (see figure 3.24 for comparison of the 2D to the 3D teaching methods).



3D Computer Models Created and Viewable on Sketchfab	
(* = also 3D printed)	
1. Mesocephalic skull	21. Short digital muscles
2. Brachycephalic skull	22. Innervation of forelimb
3. Muscles of facial expression	23. Suprascapular nerve
4. Superficial structures of the head	24. Axillary nerve
5. Muscles of mastication	25. Radial nerve
6. Hyoid apparatus	26. Musculocutaneous nerve
7. Extrinsic muscles of the tongue	27. Median nerve
8. Sagittal section of dog head	28. Ulnar nerve
9. Hinge joint	29. Skeleton of canine hindlimb
10. Condylar joint	30. Hindlimb bony prominences
11. Ellipsoidal and saddle joints	31. Pelvis hipbones
12. Ball and socket joint	32. Hip joint*
13. Pivot and plane joints	33. Stifle joint*
14. Skeleton of forelimb	34. Menisci and meniscal ligaments without femur
15. Forelimb bony prominences	35. Menisci and meniscal ligaments with femur
16. Extrinsic muscles of the trunk	36. Bones and joints of the tarsus
17. The shoulder of the dog*	37. Innervation of the hindlimb
18. Elbow joint of the dog*	38. Sciatic nerve
19. Bones and joints of the carpus	39. Obturator nerve
20. Superficial and deep digital flexor tendons	40. Femoral nerve

**Table 3.5** List of all 3D computer models created. \* = also 3D printed.

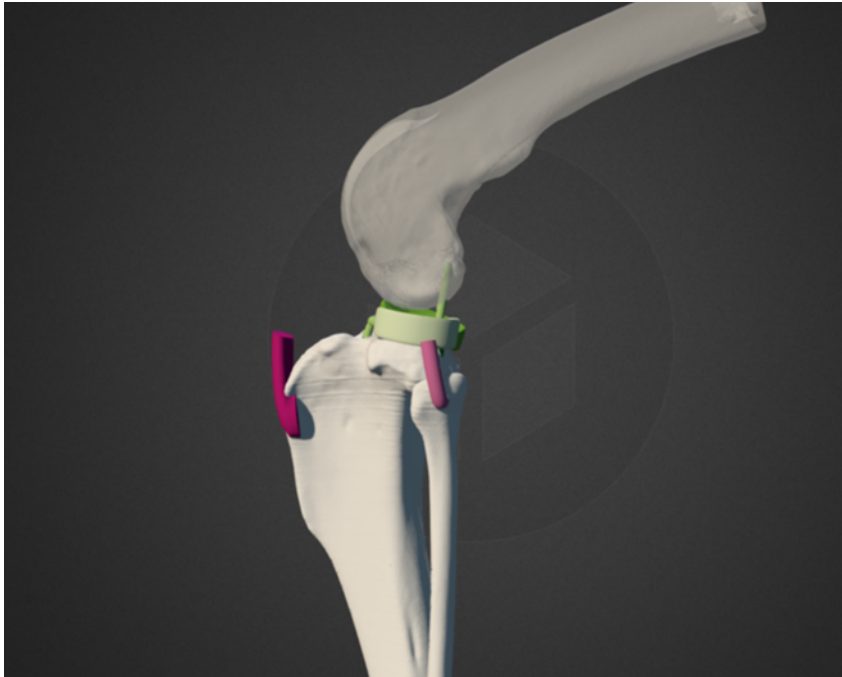
To create the 3D computer models of the manus, pes, and hyoid apparatus, the same process of creating an STL file from a CT scan was used, although the bones of the manus and pes were glued together before scanning to prevent complex articulation of small bones using Blender. CT scans from the R(D)SVS small animal hospital were also obtained for different skull shapes, and to also create a sagittal section of the canine dog head that could be used to support spatial anatomy teaching on the head and pharynx.

Once finished, all 3D computer models were rendered with colour on Blender using the website colour generator colorbrewer (Brewer, Harrower & The Pennsylvania State University, n.d., <http://colorbrewer2.org/>) to ensure the combination of colours used were colour-blind friendly. For the joint models, the bones were kept the same colour as when imported into Blender, as these naturally had a bone-like colour, except for the manus and pes where individual smaller bones were coloured to make identification easier (Figure 3.21), and the ligaments, tendons, or cartilage were also given a colour (Figure 3.22).



*Figure 3.21 Example of colour rendering on small carpal bones.*

Additionally, some of the anatomical structures were made translucent to help explain spatial aspects. Examples include a translucent mandible for the ‘Extrinsic muscles of the tongue’ model and a transparent femur for the ‘Menisci and meniscal ligaments with femur’ (Figure 3.22).



*Figure 3.22 Example of translucent structure in 3D model to aid learning.*

On Sketchfab up to 20 annotations can be added to each 3D computer model, providing students with, for example, the name of salient anatomical structures including a written description of these. The limited number of annotations per model was challenging because some 2D diagrams that the 3D models were created from contained more than 20 annotations, e.g., diagrams for the nerves and bony prominences were heavily annotated. To overcome this, a colour coded key was added to the 3D model and/or the extra annotation was added to another similar model (Figure 3.23). The numerous annotations were particularly difficult for the 2D diagram of nerves as it contained much information. In this case, one 3D model was made which showed all the nerves without annotations, and then additional individual 3D models were made for each nerve including the appropriate annotations for that specific nerve.



*Figure 3.23 Example of key used to identify more structures due to limitations of annotation numbers per model.*

Some structures were missing from the models, such as small sesamoid bones, and so an annotation was put in place where the structure would be located with an explanation stating it was not visible on this model.

The order in which annotations are given can be predetermined, as can the view shown when the annotation is selected. Including, how 'zoomed in or out' the view of the model is when displayed, the angle at which it is shown, and the view created when moving from one annotation to another. All these capabilities were exploited to encourage students to appreciate the spatial nature of the anatomy they were learning, and to ensure the teaching was spatial and not simply displaying 3D representations.

Sketchfab allowed the annotations to be controlled by selecting them on the model or by clicking on buttons at the bottom of the screen. These buttons allowed the students to access:

- A list of the annotations available.
- An option to play the annotations on autopilot in ascending order.
- An option allowing the annotations to be hidden.

Students could self-control the rotation of the model using the computer mouse in the x, y, and z-axes, they could also zoom in and out of the model, and move the position of the model on the screen using the computer mouse.

### **Tracking of 3D Computer Model Usage**

The University's online learning platform LEARN was able to track individual student usage of the 3D computer models. However, during data collection and tallying, it was noted that the total number of views for each separate model in the same lecture was the same. After consulting with relevant technical staff, it was discovered there was a bug within LEARN affecting this function. Therefore these usage data were incorrect and could not be used for any further analysis. The Sketchfab website does, however, track the number of views for each model, (although not at the level of individual student usage). These data were collected three months after the students from cohort UoE Vet 2 sat the end-of-course examinations.

### **3.8.3 Development of 3D Lectures**

The lectures were delivered in a lecture theatre using PowerPoint to support the delivery. The main strategies of the lecture design were:

1. To avoid using 2D diagrams as far as possible.

2. To develop 3D models of the 2D diagrams drawn from the 2D teaching method.
3. Where 2D diagrams were used, to ensure they showed the area under consideration in many views or even in non-standard, obscure, oblique views.
4. To use diagnostic images to help further illustrate the anatomy.

The last lecture used for the 2D teaching method, 'Introduction to ultrasound', was converted to an 'Introduction to Diagnostic Imaging' lecture, and given at the start of the 3D teaching method as part of the introductory series (Figure 3.24). The lecture format was changed in order for the students to understand and interpret the diagnostic images used as part of the 3D spatial teaching method.

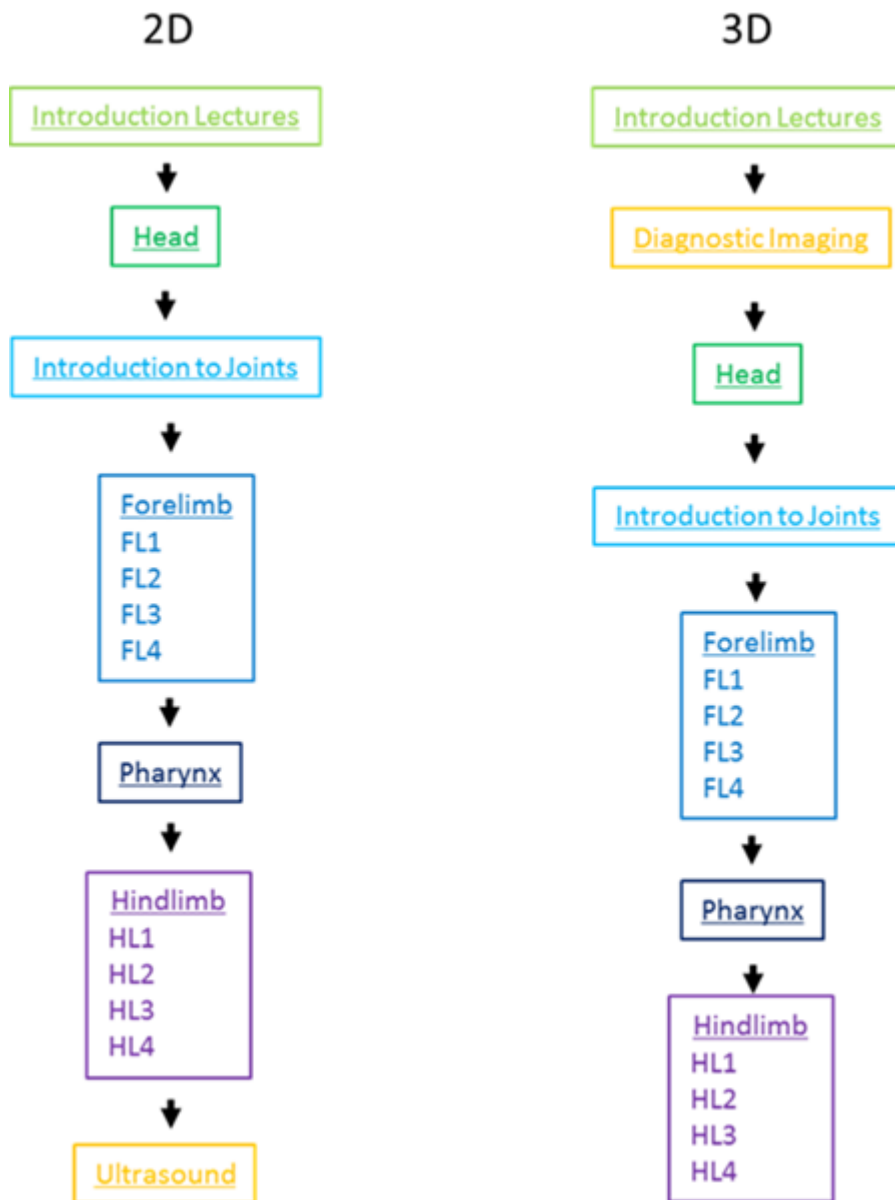


Figure 3.24 Comparison of order of teaching.

The content in the lectures was not altered from the 2D teaching method to the 3D teaching method, and the same notes were used from cohort UoE Vet 1. The learning objectives remained the same. The written notes in the 2D method were typed onto PowerPoint slides for the 3D teaching method. Where required the content was restructured to align with a more spatially oriented approach. For instance, the two lectures on the head for the 2D teaching method explored the content in the following order:

1. Muscles of facial expression
2. Superficial structures of the head
3. Ear
4. Sagittal section of dog head
5. Muscles of mastication
6. Hyoid apparatus
7. Muscles of the tongue

Whereas for the 3D teaching method, the order was changed to represent a superficial to deep approach towards the anatomy:

1. Muscles of facial expression (the 3D model for this was annotated for the ear as well)
2. Superficial structures of the head
3. Muscles of mastication
4. Hyoid apparatus
5. Extrinsic muscles of the tongue
6. Sagittal section of the dog head

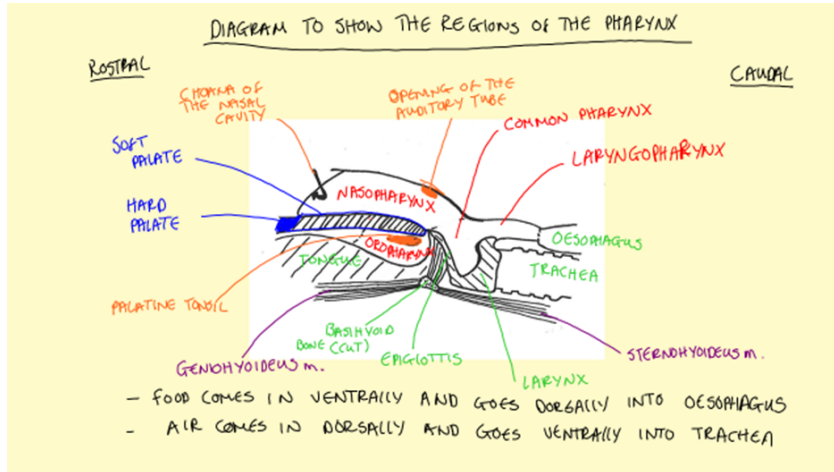


In the lecture, the 3D models were used to rotate the model and display the order of the annotations as they would have been written/drawn in the 2D teaching method. The students were encouraged to bring their own electronic devices to view the 3D models in real time as they were presented in the lecture. The 3D models were made available to the students from the morning of the lecture and were available online 24/7 during the academic year.

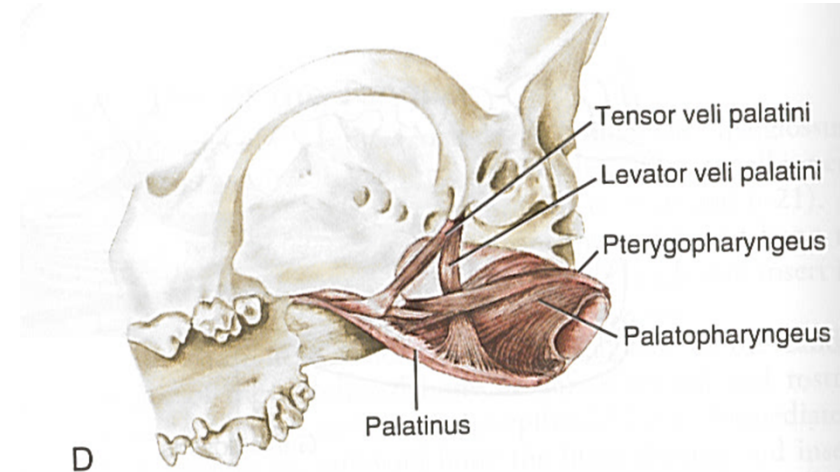
Diagnostic images were used throughout the lectures and replaced the textbook images used in the 2D teaching method along with the 3D computer models. For instance, instead of a schematic image from a textbook of the bones of the carpus, a radiograph of the carpus was used. Videos of CT scans were also played during lectures and MRI images. Radiographs and ultrasound were all used to compliment diagrams and 3D models. Diagnostic images were used due to their potential to promote the use of spatial ability, because of both the understanding of the anatomy and understanding of how the image was produced, which can be spatially demanding.

The 'pharynx' lecture was the only lecture in which a specific 3D model was not constructed. Several attempts were made to construct a specific model, but the intricate nature of the structure proved too challenging. Therefore during this lecture images from textbooks were used to explain the anatomy. These images depicted as many views as possible and a 2D diagram of the pharynx was also drawn (Figure 3.25). The 3D model 'Sagittal section of the dog head' from the Head 2 lecture was also used in the pharynx lecture to help explain the position and location of the pharynx (Figure 3.25C). The author also used photos of a real canine head dissection to explain the pharynx (Figure 3.25D). Notes that had been printed for cohort UoE Vet 1 were made available on LEARN for UoE Vet 2.

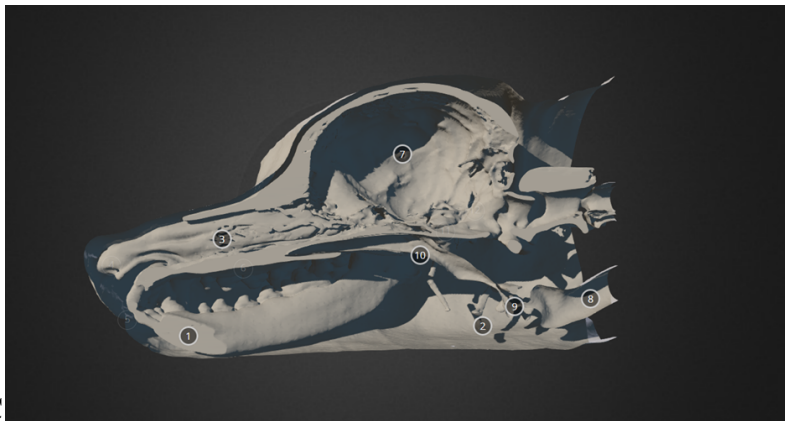
A



B



C



D

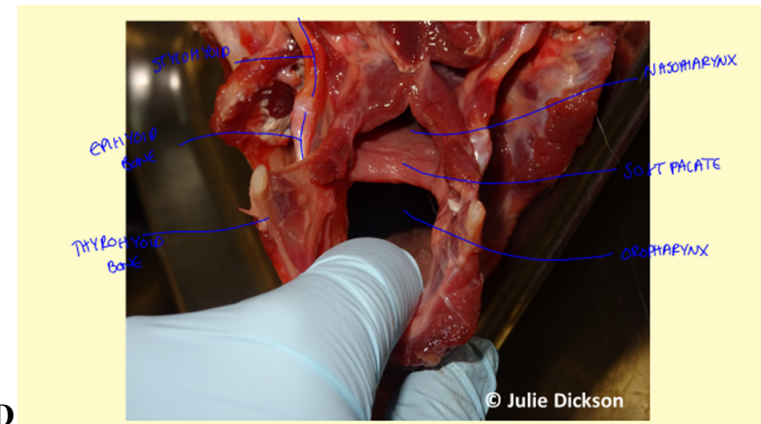


Figure 3.25 Example of 2D image drawn (A), flat 3D images (B and D), and 3D computer model (C) used to help explain the pharynx.

Figures 3.26, 3.27, 3.28, 3.29, 3.30 and 3.31 summarise the diagrams or 3D models/diagnostic images used between the two teaching methods per section of teaching.

<u>Introduction Lectures</u>	<u>Introduction Lectures</u>
<ul style="list-style-type: none"> <li>• Language of Anatomy</li> <li>• Anatomy of the Adult Body</li> <li>• Physiology of the Adult Body</li> </ul>	<ul style="list-style-type: none"> <li>• Language of Anatomy</li> <li>• Anatomy of the Adult Body</li> <li>• Physiology of the Adult Body</li> <li>• Introduction to Diagnostic Imaging</li> </ul>

Figure 3.26 Comparison of introduction lectures. Left is 2D and right is 3D.

<u>Head</u>	<u>Head</u>
<p>Superficial:</p> <ul style="list-style-type: none"> <li>• Mimetic muscles (Lat)</li> <li>• Superficial structures (Lat)</li> </ul> <p>Deep:</p> <ul style="list-style-type: none"> <li>• Sagittal section dog head</li> <li>• Muscles of mastication (Lat)</li> <li>• Hyoid apparatus (Lat)</li> <li>• Extrinsic muscle of the tongue (Lat)</li> </ul>	<p>Superficial:</p> <ul style="list-style-type: none"> <li>• Mesocephalic skull</li> <li>• Brachycephalic skull</li> <li>• Mimetic muscles</li> <li>• Superficial structures</li> <li>• CT video canine head</li> </ul> <p>Deep:</p> <ul style="list-style-type: none"> <li>• Muscles of mastication</li> <li>• CT video muscles mastication</li> <li>• MRI scan muscles mastication</li> <li>• Hyoid apparatus</li> <li>• Extrinsic muscles of the tongue</li> <li>• Sagittal section dog head *</li> <li>• Sagittal section dog head</li> <li>• MRI scan sagittal head</li> <li>• CT video sagittal head</li> </ul>

Figure 3.27 Comparison of head section. \* = also 2D diagram. Left is 2D and right is 3D.

<u>Introduction to Joints</u>	<u>Introduction to Joints</u>
<ul style="list-style-type: none"> <li>• Typical Synovial Joint</li> <li>• Articular Cartilage</li> <li>• Articular Disc</li> <li>• Articular Menisci</li> <li>• Synovial Joints (hinge, condylar, ellipsoidal, ball &amp; socket, saddle)</li> </ul>	<ul style="list-style-type: none"> <li>• Typical Synovial Joint*</li> <li>• Articular Cartilage*</li> <li>• Articular Disc*</li> <li>• Articular Menisci*</li> <li>• Hinge Joint</li> <li>• Condylar Joint</li> <li>• Ellipsoidal &amp; saddle Joint</li> <li>• Ball &amp; Socket Joint</li> <li>• Pivot &amp; Plane Joint</li> </ul>

Figure 3.28 Comparison of introduction to joints section. \* = also 2D diagram. Left is 2D and right is 3D.

<u>Forelimb</u>	<u>Forelimb</u>
<p>FL1:</p> <ul style="list-style-type: none"> <li>• Skeleton</li> <li>• Bony Prominences (Lat)</li> <li>• Bony Prominences (Med)</li> <li>• Extrinsic muscles (Cr)</li> <li>• Shoulder Joint (Lat)</li> </ul> <p>FL2:</p> <ul style="list-style-type: none"> <li>• Elbow Joint (Cd, Med, Cr)</li> <li>• Carpus (Dr)</li> </ul> <p>FL3:</p> <ul style="list-style-type: none"> <li>• Carpus Sagittal</li> <li>• DDFT and SDFT (Pal)</li> <li>• Interosseous muscle (Lat)</li> <li>• Interosseous muscle (Cr)</li> </ul> <p>FL4:</p> <ul style="list-style-type: none"> <li>• Brachial Plexus (schematic)</li> <li>• Innervation (Lat)</li> <li>• Innervation (Med)</li> </ul>	<p>FL1:</p> <ul style="list-style-type: none"> <li>• Skeleton</li> <li>• Bony Prominences</li> <li>• Extrinsic muscles</li> <li>• Shoulder Joint</li> <li>• Radiograph shoulder (Lat)</li> <li>• Radiograph shoulder (CrCd)</li> <li>• CT video shoulder</li> <li>• Elbow Joint</li> <li>• Radiograph elbow (Lat)</li> <li>• Radiograph elbow (CrCd)</li> <li>• CT video elbow</li> </ul> <p>FL2:</p> <ul style="list-style-type: none"> <li>• Carpus Joint</li> <li>• Carpus sagittal*</li> <li>• Radiograph carpus (Lat)</li> <li>• Radiograph carpus (DorPal)</li> <li>• DDFT &amp; SDFT</li> <li>• Interosseous muscle</li> <li>• Brachial Plexus*</li> <li>• All Forelimb nerve</li> <li>• Radial Nerve</li> <li>• Musculocutaneous nerve</li> <li>• Median nerve</li> <li>• Ulnar nerve</li> <li>• Axillary nerve</li> <li>• Suprascapular nerve</li> </ul> <p>FL3:</p> <ul style="list-style-type: none"> <li>• Tutorial 1</li> </ul> <p>FL4:</p> <ul style="list-style-type: none"> <li>• Tutorial 2</li> </ul>

*Figure 3.29 Comparison of forelimb section. \* = also 2D diagram. Left is 2D and right is 3D.*

<u>Pharynx</u>	<u>Pharynx</u>
<ul style="list-style-type: none"> <li>• Regions of Pharynx Sagittal</li> </ul>	<ul style="list-style-type: none"> <li>• Regions of Pharynx Sagittal*</li> <li>• MRI sagittal head</li> <li>• Sagittal Head</li> <li>• Textbook diagrams different orientations</li> </ul>

*Figure 3.30 Comparison of pharynx section. \* = also 2D diagram. Left is 2D and right is 3D.*

Hindlimb	Hindlimb
<p>HL1:</p> <ul style="list-style-type: none"> <li>• Bony Prominences (Lat) (HW)</li> <li>• Bony Prominences (Med) (HW)</li> <li>• Hipbone (Lat)</li> <li>• Hip Joint (Ven)</li> </ul> <p>HL2:</p> <ul style="list-style-type: none"> <li>• Menisci (Dor)</li> <li>• Menisci (Cr)</li> <li>• Tibial Plateau (Dor)</li> <li>• Stifle Joint (Cr)</li> <li>• Stifle Joint (Lat)</li> </ul> <p>HL3:</p> <ul style="list-style-type: none"> <li>• Tarsus (Dor)</li> </ul> <p>HL4:</p> <ul style="list-style-type: none"> <li>• Innervation (Lat)</li> <li>• Innervation (Lat)</li> </ul>	<p>HL1:</p> <ul style="list-style-type: none"> <li>• Skeleton</li> <li>• Bony Prominences</li> <li>• Hipbones</li> <li>• Hip Joint</li> <li>• CT video hips</li> <li>• Stifle joint</li> <li>• Tibial Plateau without femur</li> <li>• Tibial Plateau with translucent femur</li> <li>• MRI stifle scan (Lat)</li> <li>• MRI stifle scan (CrCd)</li> <li>• Menisci (Dor)*</li> </ul> <p>HL2:</p> <ul style="list-style-type: none"> <li>• Tarsus</li> <li>• Radiograph Tarsus (Lat)</li> <li>• Radiograph Tarsus (DP)</li> <li>• All Hindlimb Nerve</li> <li>• Sciatic nerve</li> <li>• Obturator nerve</li> <li>• Femoral nerve</li> </ul> <p>HL3:</p> <ul style="list-style-type: none"> <li>• Tutorial 1</li> </ul> <p>HL4:</p> <ul style="list-style-type: none"> <li>• Tutorial 2</li> </ul>

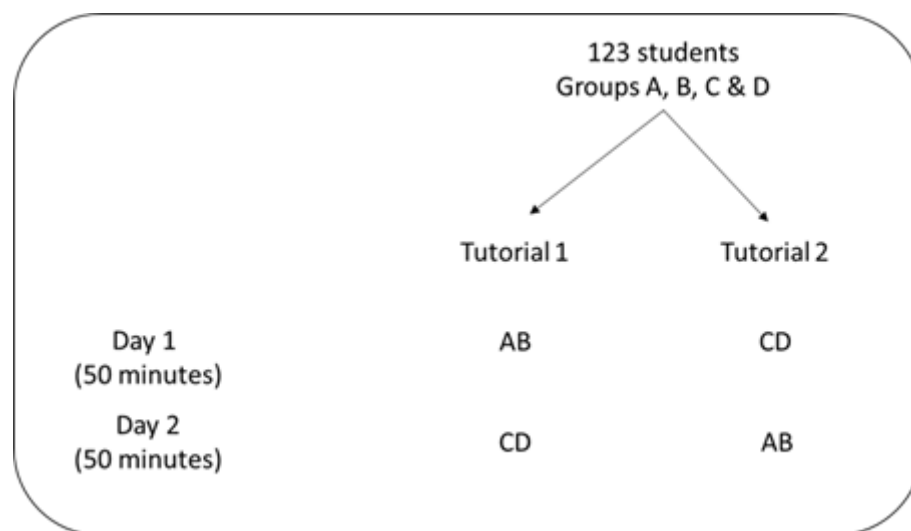
Figure 3.31 Comparison of hindlimb section. \* = also 2D diagram. Left is 2D and right is 3D.

### 3.8.4 Development of Tutorials

To keep the total hours of teaching the same across both cohorts and to incorporate the 3D printed models into teaching, the forelimb and hindlimb lectures were delivered as two lectures to cohort UoE Vet 2 instead of four, with the remaining two lectures converted into tutorials designed as small-group work utilising the 3D printed models. One half of the class at a time had tutor contact, while the other half was in a computer lab working on an independent task. These tutorials lasted the same time as the lectures (50 minutes each).

The aim of these tutorials was for students to engage in manipulating and critiquing the 3D printed models. Twelve 3D models were printed (three of each major joint;

shoulder, elbow, hip, and stifle). The shoulder and elbow models were used for the forelimb tutorial and the hip and stifle models for the hindlimb tutorial. There were 123 students in cohort 2 hence the year was divided in half, and two tutorials were delivered simultaneously, and then students swapped after 50 minutes as shown in Figure 3.32.



*Figure 3.32 Logistics of 3D tutorials.*

Tutorial 1 involved the students answering questions based on the 3D printed models. Tutorial 2 involved the students answering questions based on the 3D computer models.

### **Tutorial 1 (Forelimb 3 and Hindlimb 3)**

Before the start of this tutorial, the students were given ground rules (Figure 3.33).

## Ground Rules

- Everyone should participate
- Everyone should come on time
- Safe space: what we do in here stays in here
- Respect one another (don't talk over one another or dominate)
- Ask as many questions as you like, there are no stupid questions
- Be careful with the models (they cost £100 each)

Password for iPad= 2015

*Figure 3.33 Ground rules for tutorial 1.*

The course administrator randomly allocated the students into nine groups, with six to seven students in each group. These groups were the same for both the forelimb 3 and hindlimb 3 tutorials. The students were asked to work as a team and complete three stations with 15 minutes allocated for each station. The tutorial was designed to utilise other teaching resources such as textbooks, the 3D computer models, and lecture notes (see Appendix 5 for a tutorial handout).

Stations 1 and 2 were of a similar design. They involved a 3D printed model with three to four questions on that model and area of anatomy. The first question was always an identification question with the students identifying salient structures on the 3D printed model. The aim of this was to get the students acquainted and orientated with the model, to set the scene. The second question often involved highlighting a flaw of the 3D printed model, such as outlining the location of a missing structure (i.e., biceps brachii tendon of origin), or asking what was incorrect about the structures on the 3D printed model (e.g., collateral ligaments). The purpose of this was to encourage students to critically review what they observed and to recognise the inaccuracies of the 3D printed models while gaining a spatial

appreciation. The third/fourth question to stations one and two then progressed to a problem-solving clinical scenario related to anatomy knowledge and understanding.

Station 3 involved the students using an articulated canine skeleton limb (either forelimb or hindlimb, depending on the tutorial) and coloured plastic cords to trace the path the different nerves of the limb followed (Figure 3.34). The articulated canine skeleton limb was of an average sized dog skeleton from the R(D)SVS skeleton/bone collection that was due to be cleaned and restored.



*Figure 3.34 Photograph of students using coloured cord to trace nerves on an articulated forelimb skeleton.*

The answers to the tutorials were given during the class once the groups were finished on that station. The answers were also uploaded to LEARN for the students to access later if they wished.

The 3D printed models were also made available in the dissection laboratory for the students to look at during dissection practicals forelimb 1-5 and hindlimb 3-5. However, only two students took up this opportunity.



## **Tutorial 2 (Forelimb 4 and Hindlimb 4)**

Tutorial 2 was designed to be a self-directed session using the 3D computer models. The aim was to encourage students to rotate the models in order to answer the questions. To accomplish this, the annotation facility of Sketchfab was exploited.

To create the questions, duplicates of the 3D computer models were generated with one model for each question. Some of the 3D models had the rendered colour completely removed to prevent the question being answered from memory of the colour. Instead, students had to demonstrate their understanding of the anatomy by answering a total of seven questions (A-G), presented in a random order from the lecture content. The students were given more questions than could be typically managed to ensure sufficient content for the 50-minute tutorial. The students were asked to work in groups of two to encourage teamwork and to manage limited computer availability.

The students had to rotate the 3D model to find the annotation number to the question they were answering, once the number was found the students could answer the question. The student handout (see Appendix 5) for the tutorial included the questions to answer, such as “Annotations 1-10: Identify the structures numbered 1-10” or “Annotation 12: find this annotation and state the classification of the joint shown and state the type of joint within this classification.” The students would need to rotate the model to find the annotation and once found, answer the question on the handout, with the answer then revealed by clicking on the annotation, thus receiving instant feedback on their learning.

For tutorial 2 there was the following classification of questions:

1. Identification of structures (e.g., Annotations 1-10: Identify the structures).
2. Description/explanation questions (e.g., Name and describe the articulation that occurs at annotation 4).
3. Identification of a nerve.

Again, the 3D models were schematic representations of normal anatomy. As in tutorial 1, some questions were formed around the fact that the models were not exactly anatomically correct to ensure the students understood the content. This tutorial was available on LEARN afterward for students to access in their own time.

### **3.8.5 Practical Classes (3D)**

The dissection classes for the 3D teaching method were the same as for the 2D method, except for the ability to look at the 3D printed models from the newly designed tutorials. The models were made available after the tutorials during the forelimb 1-5 and hindlimb 3-5 dissection practicals. Students could also rent out bone boxes from the library in the same way as cohort UoE Vet 1.

## **3.9 Quantitative Data Analysis**

A mixed methods approach, which combined both quantitative and qualitative methods, was used in the research of this thesis. This approach was used to provide a richer analysis by gaining both a numerical value indicative of students' anatomy knowledge and understanding and their spatial ability while collecting the students' views on the teaching and learning experience using focus groups. The methods used for the qualitative analysis are explained in detail at the start of chapter 8 before the

results of the qualitative analysis are presented. This current section explains the quantitative analysis methods before the presentation of results in the next four chapters (chapters 4-7).

All quantitative data analysis was conducted using the software R (R Core Team, 2018). Statistical significance was taken to be at  $p < 0.05$ , with all p-values, rounded to three decimal places (unless near zero then quoted as  $p < 0.001$ ) and test statistics rounded to two decimal places.

### **3.9.1 Descriptive Statistics**

All quantitative analyses began with descriptive statistics of variables and then inferential statistical analysis. The descriptive analysis was performed to gain an initial idea of the difference between variables. A table was then generated to summarise the data quoting the median (M) and range for non-normally distributed data, and the mean ( $\bar{x}$ ) and standard deviation (SD) for normally distributed data. Correlations were also performed between variables of interest to establish any possible relationships. A Spearman rank correlation was used because the majority of parameters were not normally distributed. To compare the difference between two dependent correlations (such as in Chapter 6, section 6.3.1, when comparing the non-spatial MCQ and spatial MCQ question correlations of each spatial ability test to one another), a dependent correlation comparison was performed (Revelle, 2018) (R 'psych' package 'r.test').

### **3.9.2 Univariable Analysis**

Univariable analysis was performed using t-tests, and Wilcoxon rank sum/signed rank tests for within cohort comparisons such as comparison of genders (male and female) scores on the spatial ability tests (see section 4.2.2). A 2-proportion test or

Fisher's exact test was used to compare UoE Vet 1 and UoE Vet 2 scores for each question of the spatial MCQ (see section 6.3.4).

### **3.9.3 Multivariable Analysis**

The inferential analysis was primarily conducted by multivariable analysis using Generalised Linear Models (GLM). The GLMs allowed for differences in categorical and continuous explanatory variables such as gender, handedness, spatial ability score, or cohort to be accounted for. GLMs were simplified by stepwise reduction of variables and interactions by the highest p-value, with the final model including only significant explanatory variables. The variable 'cohort' (UoE Vet 1, UoE Vet 2, UoB Vet or UoE Psych) was the main variable of interest in GLMs involving all four cohorts and was not removed during model simplification.

After model simplification, residual diagnostic plots were checked to ensure goodness-of-fit. If there was overdispersion in any of the GLMs (identified by the residual deviance being greater than the residual degrees of freedom (Crawley, 2012:p.580), then a Quasibinomial or Quasipoisson error family was used to ensure goodness-of-fit. Overdispersion indicates that there is extra variation in the response variable that could have occurred because of some unmeasured factor, and the overdispersion is adjusted for by using a Quasipoisson/binomial error family instead to fit an appropriate dispersion parameter.

A Poisson error family ( $GLM_P$ ) was used for the response variables involving the spatial ability test scores (CRT and MRT only) or the spatial MCQ score because these variables were count data, which the Poisson distribution is based on. Previous studies investigating spatial ability have divided participants into high and low spatial ability for analysis, which categorises data into two categories. This dichotomising of data loses the fine granular detail of using a complete data set and

can increase Type I errors (Senn, 2003; Altman, 2006). This study does not do this by using the whole data set of spatial ability test scores.

A Binomial error family ( $GLM_B$ ) was used for the response variable involving exam parameters (in-course, oral, spot, interpretation, short, MCQ, or total exam) this was because the exam scores were calculated as a proportion to account for differences in the number of exam questions between cohorts UoE Vet 1 and UoE Vet 2, as outlined in Table 3.4. A  $GLM_B$  was also used to provide the GLM with not only the number of correct answers but also the number of incorrect answers to calculate an appropriate a model as possible. Additionally, a  $GLM_B$  was used for the exam parameters because it bounds the response variable by 0 and 1, i.e., there are no exam scores greater than 100% and no negative scores <0%. A  $GLM_B$  was also used for the spatial ability test SDT because this test had a dramatically left shifted distribution (Figure 3.35).

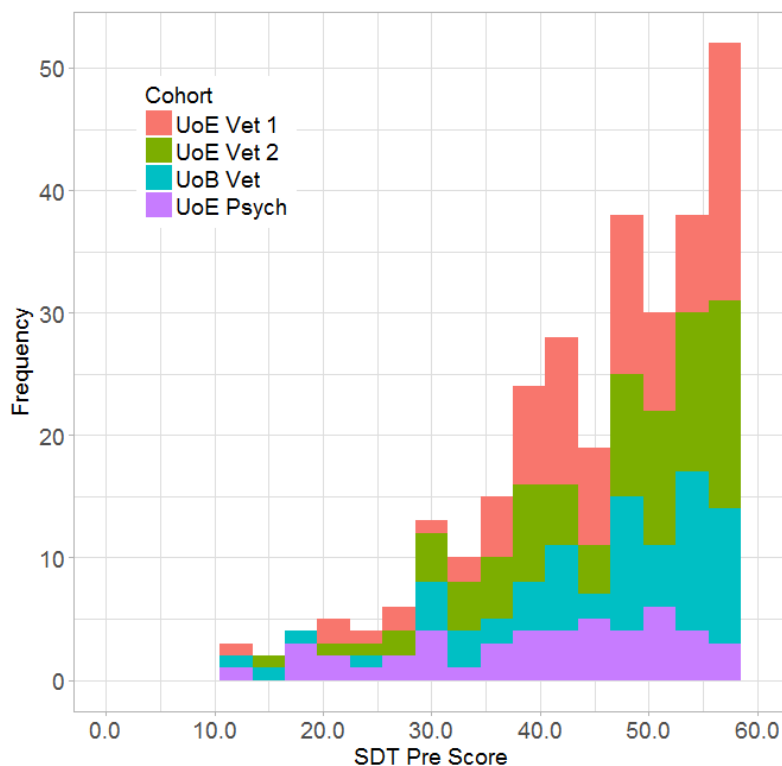


Figure 3.35 Histogram showing skewed distribution of SDT Pre for each cohort.

A  $GLM_B$  model was used for the SDT to take into account the skewed distribution by approximating to a binomial distribution involving proportions, by providing the GLM with the correct and incorrect SDT scores. Thus, the model could calculate the proportion of participants that got a given SDT score and the probability of getting that proportion of answers correct, as no distribution could be used to approximate the probability of the skewed raw SDT scores. Therefore the SDT score will be a proportion throughout the remainder of this thesis. Chapters 4-7 each include a table summarising the specific GLM analyses performed for the study presented in that chapter.

An explanation of the specific quantitative analysis conducted in each chapter is included under the chapter section 'Data Analysis'.



# Chapter 4 Quantitative Analysis of Baseline Spatial Ability and Demographic Parameters

## 4.1 Introduction to Chapter 4

Chapters 1 and 2 discussed the literature on spatial ability and anatomy learning. From this literature it was identified the Mental Rotation Test (Vandenberg & Kuse, 1978; Peters *et al.*, 1995) has commonly been used in anatomy education research to measure spatial ability (Lufler *et al.*, 2012; Langlois *et al.*, 2009; Nguyen *et al.*, 2013; Cui *et al.*, 2016; Gutierrez *et al.*, 2017; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999b, 1999a). The study presented in this chapter used the Mental Rotation Test (MRT, Peters *et al.*, 1995), and two further spatial ability tests - the Card Rotation Test (CRT, Ekstrom *et al.*, 1976) and the Surface Development Test (SDT, Ekstrom *et al.*, 1976) to measure the spatial ability of veterinary students learning anatomy. The aim of using the CRT was to further investigate the spatial ability sub-category of mental rotation (section 2.3.3) in two-dimensions (2D), as the MRT measures mental rotation in three-dimensions (3D). The SDT measures the spatial visualisation sub-category of spatial ability - an area less frequently measured in anatomy education research and is included in this study with the aim of investigating this sub-category for students learning anatomy.

Studies on spatial ability and anatomy learning have primarily involved medical students at one academic institution. These studies have primarily investigated the links between spatial ability and anatomy knowledge (Rochford, 1985; Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a), and the use of 3D technology to teach anatomy along with the relationship to a students' spatial ability (Nguyen *et al.*, 2013; Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001; Tan *et al.*, 2012; Berney *et al.*, 2015; Keedy *et al.*, 2011). However, so far limited studies are researching these areas



in veterinary students (Provo-Klimek, 2002; Preece *et al.*, 2013; Gutierrez *et al.*, 2017). For instance, it is unknown whether the spatial ability of students differs between cohorts of veterinary students (i.e., how stable a trait spatial ability generally is between cohorts of students entering veterinary school). Any differences could be important given the potential link between spatial ability and anatomy learning (Lufler *et al.*, 2012; Garg, Norman & Sperotable, 2001; Rochford, 1985; Sweeney, Hayes & Chiavaroli, 2014; Vorstenbosch *et al.*, 2013a). This chapter aims to explore this using the CRT, MRT, and SDT spatial ability tests.

Within the anatomy education literature comparisons of spatial ability have been made between novice, intermediate, and expert anatomists (Fernandez, Dror & Smith, 2011), but as yet the spatial ability between cohorts of students of the same or different academic institution have not been investigated. To explore this specifically for veterinary students, this chapter presents the quantitative analyses comparing the baseline spatial ability (Pre scores) within and between each of the four cohorts of this study (University of Edinburgh Veterinary Students 1 (UoE Vet 1), University of Edinburgh Veterinary Students 2 (UoE Vet 2), University of Bristol Veterinary Students (UoB Vet), and University of Edinburgh Psychology Students (UoE Psych) for each of the three spatial ability tests measuring two sub-categories (CRT, MRT, and SDT).

The demographic parameters collected of the participants in this study included gender, handedness, and age. Gender differences are known between males and females on tasks of spatial ability (section 2.3.4), with males, in general, performing better on tasks of mental rotation and in particular 3D mental rotation (Masters & Sanders, 1993; Reilly & Neumann, 2013). However smaller effect sizes between genders have been found for tasks such as the paper folding test (classified under the spatial visualisation sub-category) and the Hidden Figures Test (Voyer, Voyer & Bryden, 1995). This chapter includes analysis of any gender differences within the four cohorts of this study for each of the three spatial ability tests to compare with

other published studies. In addition, data on the age and handedness of the students in each of the four cohorts and its relation to spatial ability were also analysed as has been carried out in several other studies (Burnett, Lane & Dratt, 1982; Somers *et al.*, 2015; Li, Zhu & Nuttall, 2003).

The overall aim of this study was to establish whether spatial ability differs between cohorts of veterinary students from two academic institutions and to compare this to a cohort of non-veterinary university students. This baseline analysis will also inform the subsequent analyses comparing the 2D and 3D spatial teaching methods (chapter 5) and student scores on a spatial MCQ (chapter 6).

## **4.2 Data Analysis**

### **4.2.1 Descriptive Statistics**

Shapiro-Wilk normality tests were performed and histograms plotted for the continuous parameters CRT Pre, MRT Pre, SDT Pre, and age to check for normality. To compare the distributions of each cohort for the three spatial ability tests, Kolmogorov-Smirnov tests were conducted for all possible pairwise combinations, and these were examined separately by gender (Female and Male) and by handedness (Left and Right).

During input of the MRT Pre test results, it was noted that a total of 30 test sheets for the UoE Vet cohorts had printed in grey rather than black and white (Cohort UoE Vet 1 = 22 and Cohort UoE Vet 2 = 8). Therefore the students with grey MRT test sheets were identified on the MRT Pre histograms by a separate colour to ensure they

were not clustered in the distribution (Figure 4.1) and could be included in the analysis.

Descriptive statistical analysis was performed including the mean or median with accompanying SD or range for each spatial ability test pre score and age. Spearman rank correlations were performed for each pairwise combination of spatial ability test scores for each of the four cohorts and all four cohorts combined into one dataset.

#### **4.2.2 Within Cohort Comparison**

To compare the mean spatial ability test score between genders (Male vs. Female) and between handedness (Left vs. Right) a two-sample t-test was used for CRT Pre for UoE Vet 2 and UoE Psych, and for MRT Pre for UoE Vet2, UoB Vet, and UoE Psych. A Wilcoxon rank sum test was used to compare the spatial ability between genders (Male vs. Female) and between handedness (Left vs. Right) for CRT Pre for UoE Vet 1 and UoB Vet, for MRT Pre for UoE Vet 1 and all four cohorts for all SDT Pre comparisons.

To analyse the effect of age within each cohort separately a generalised linear model (GLM) analysis was performed with the response variable spatial ability test (e.g., CRT Pre, MRT Pre or SDT Pre) and the explanatory variable *age*. This analysis was conducted for each cohort separately for each spatial ability test. A  $GLM_P$  was performed for CRT Pre and MRT Pre because these spatial ability scores were in the form of count data, and a  $GLM_B$  was performed for SDT Pre to inference to a binomial distribution (see section 3.9.3).

### 4.2.3 Between Cohort Comparison

Separate GLMs ( $GLM_P$  for CRT Pre and MRT Pre, and  $GLM_B$  for SDT Pre see section 3.9.3) were performed, with the response variable spatial ability test (CRT Pre/MRT Pre/SDT Pre) and the categorical explanatory variables cohort, gender, handedness, and the continuous explanatory parameter age (Table 4.1).

Parameter	Analysis	Response variable	Explanatory variables	
			Categorical	Continuous
CRT Pre	$GLM_P$	CRT Pre	Cohort, Gender, Handedness	Age
MRT Pre	$GLM_P$	MRT Pre	Cohort, Gender, Handedness	Age
SDT Pre	$GLM_B$	SDT Pre	Cohort, Gender, Handedness	Age
Age	$GLM_P$	CRT Pre	N/A	Age
	$GLM_P$	MRT Pre	N/A	Age
	$GLM_B$	SDT Pre	N/A	Age

**Table 4.1 Summary of GLM analysis of baseline parameters.**

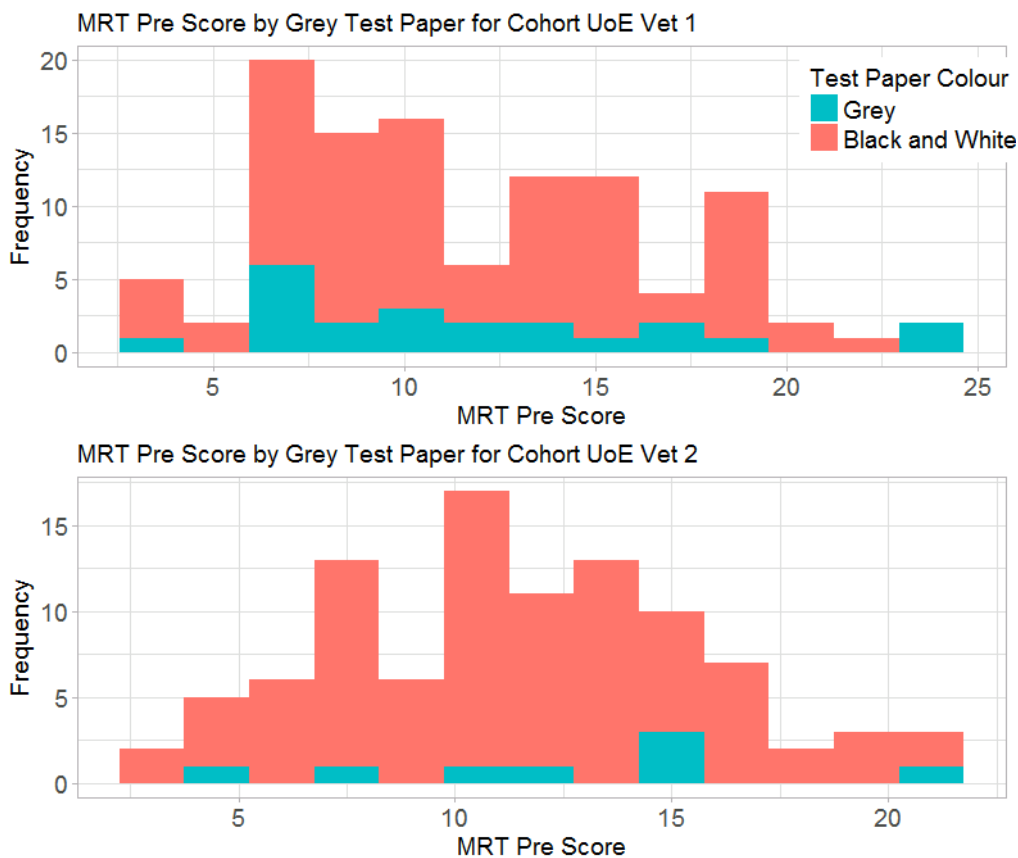
The explanatory variable cohort involved 4 levels (UoE Vet 1, UoE Vet 2, UoB Vet, and UoE Psych) and was not removed during stepwise simplification of the model. Analysis of deviance table was calculated on the final simplified GLMs to assess whether there was an overall cohort effect. If cohort had an overall significant effect

( $p < 0.05$ ) then a post-hoc Tukey was performed to determine which pair-wise combination of the four cohorts were significantly different.

## 4.3 Results

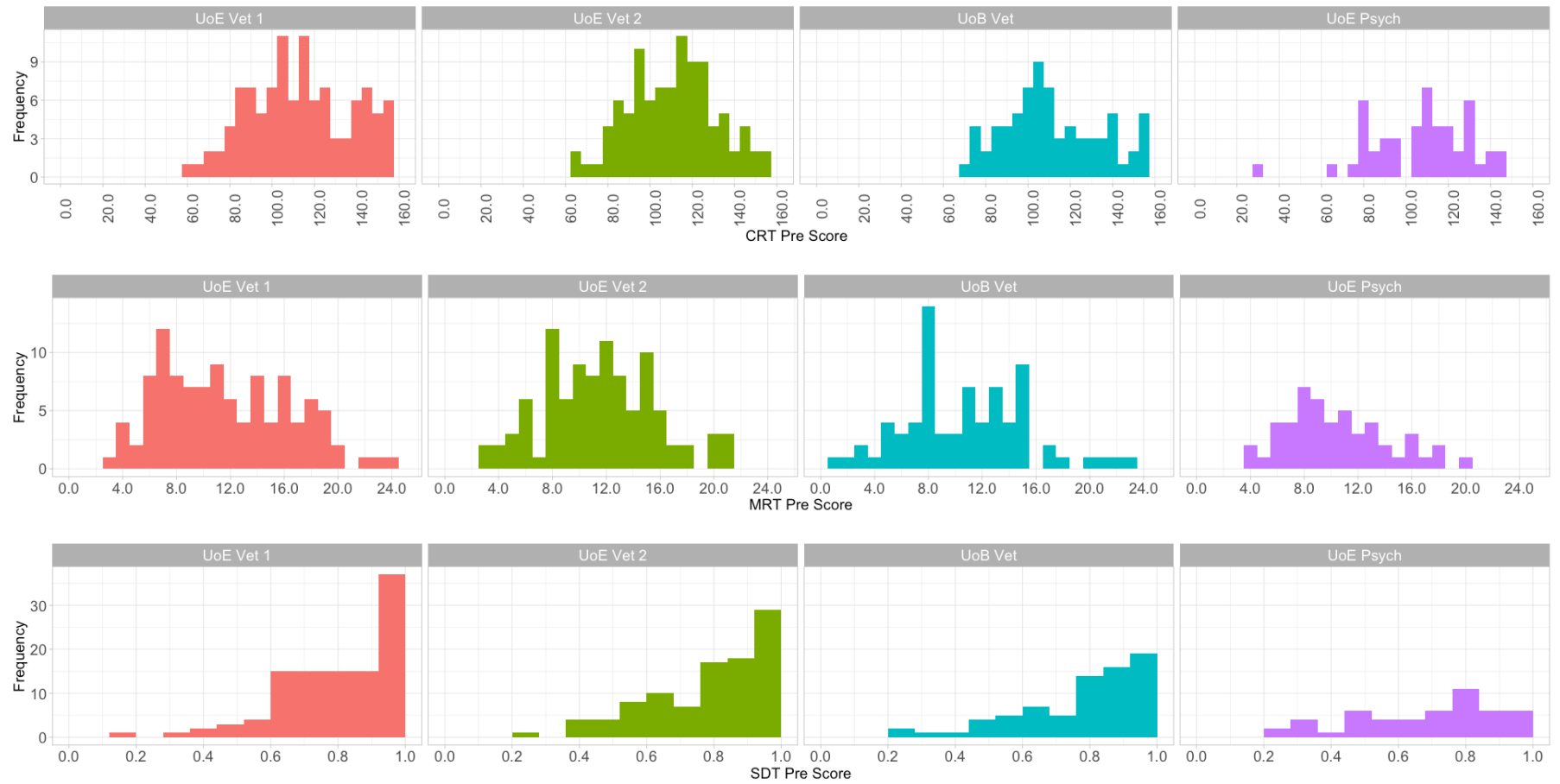
### 4.3.1 Descriptive Statistics

The participants with grey printed MRT Pre test papers in cohorts UoE Vet 1 and 2 were not clustered within the whole cohort distribution (Figure 4.1), suggesting the grey printing did not adversely affect their performance. Therefore these participants were included in the study.



*Figure 4.1 Distribution of participant scores with grey printed MRT test papers compared to rest of cohort.*

The majority of the parameters were not normally distributed ( $p < 0.05$ ) with SDT Pre showing a markedly left shifted distribution particularly for the veterinary student cohorts suggesting this spatial ability test was easier. Figure 4.2 shows the distributions of each spatial ability test for each of the four cohorts.



*Figure 4.2 Comparison of histograms for each cohort by spatial ability test.*

The CRT Pre distribution for each cohort is slightly left-skewed, although all cohorts look to have a similar distribution, the UoE Psych distribution has more participants with lower scores. Kolmogorov-Smirnov tests comparing the CRT Pre distributions of each cohort to one another confirmed the distributions were not statistically different ( $p > 0.05$ ) (Table 4.2).

CRT Pre	Cohort UoE Vet 2	Cohort UoB Vet	Cohort UoE Psych
Cohort UoE Vet 1	D = 0.14 p = 0.291	D = 0.09 p = 0.866	D = 0.16 p = 0.351
Cohort UoE Vet 2		D = 0.11 p = 0.717	D = 0.11 p = 0.847
Cohort UoB Vet			D = 0.14 p = 0.573

**Table 4.2 Matrix comparing CRT Pre distributions for all possible pairwise combinations of cohorts by Kolmogorov-Smirnov test.**

Additionally when the cohorts were divided by gender (Male and Female) for the CRT Pre distribution there was no statistically significant difference comparing the female or the male distributions between each cohort ( $p > 0.05$ , Table 4.3).

Cohort	Gender n (%)		Handedness n (%)		Female (Left) n	Female (Right) n	Male (Left) n	Male (Right) n
	Female	Male	Left	Right				
UoE Vet 1	93 (86)	15 (14)	14 (13)	94 (87)	11	82	3	12
UoE Vet 2	84 (86)	14 (14)	10 (10)	88 (90)	8	76	2	12
UoB Vet	60 (81)	14 (19)	13 (18)	61 (82)	11	49	2	12
UoE Psych	37 (74)	13 (22)	5 (10)	45 (90)	4	33	1	12

**Table 4.3 Number and percentage of male, female, left and right handed participants in each cohort, F = female, M = male, L = left, R = right.**



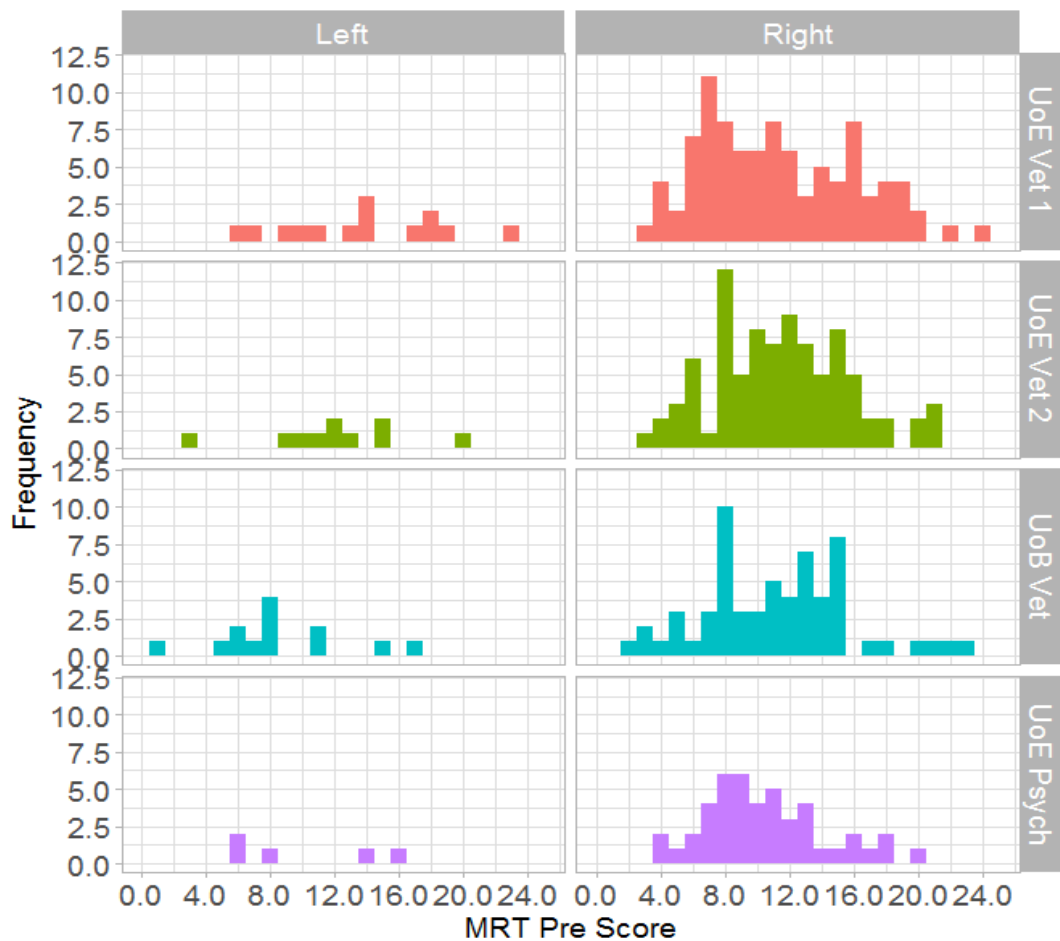
Comparisons between handedness (Left and Right) identified no differences between left-handed students of each cohort and the right-handed students of each cohort. Thus identifying no statistically significant differences of the baseline CRT Pre score distributions between each cohort.

The MRT Pre score distribution of each cohort showed a slight right skew for all four cohorts possibly suggesting this spatial ability test was more challenging (Figure 4.2). Kolgomorov-Smirnov test comparison of the distributions of each cohort identified no differences between each cohort (Table 4.4).

MRT Pre	Cohort UoE Vet 2	Cohort UoB Vet	Cohort UoE Psych
Cohort UoE Vet 1	D = 0.11 p = 0.597	D = 0.16 p = 0.185	D = 0.17 p = 0.274
Cohort UoE Vet 2		D = 0.14 p = 0.380	D = 0.16 p = 0.365
Cohort UoB Vet			D = 0.08 p = 0.983

**Table 4.4 Matrix comparing MRT Pre distributions for all possible pairwise combinations of cohorts by Kolgomorov-Smirnov test.**

Division of this comparison by gender (Males and Females) identified no statistically significant differences when comparing the male distribution of MRT Pre scores across the four cohorts or when comparing females. However, when divided by handedness statistically significant differences were noted with the left-handed UoB Vet cohort scoring less on the MRT Pre compared to UoE Vet 1 and UoE Vet 2 (D = 0.55, p = 0.034, and D = 0.59, p = 0.038 respectively), but not compared to UoE Psych (Figure 4.3). This finding could be because the left-handed participants of UoB Vet found the Mental Rotation Test more challenging.



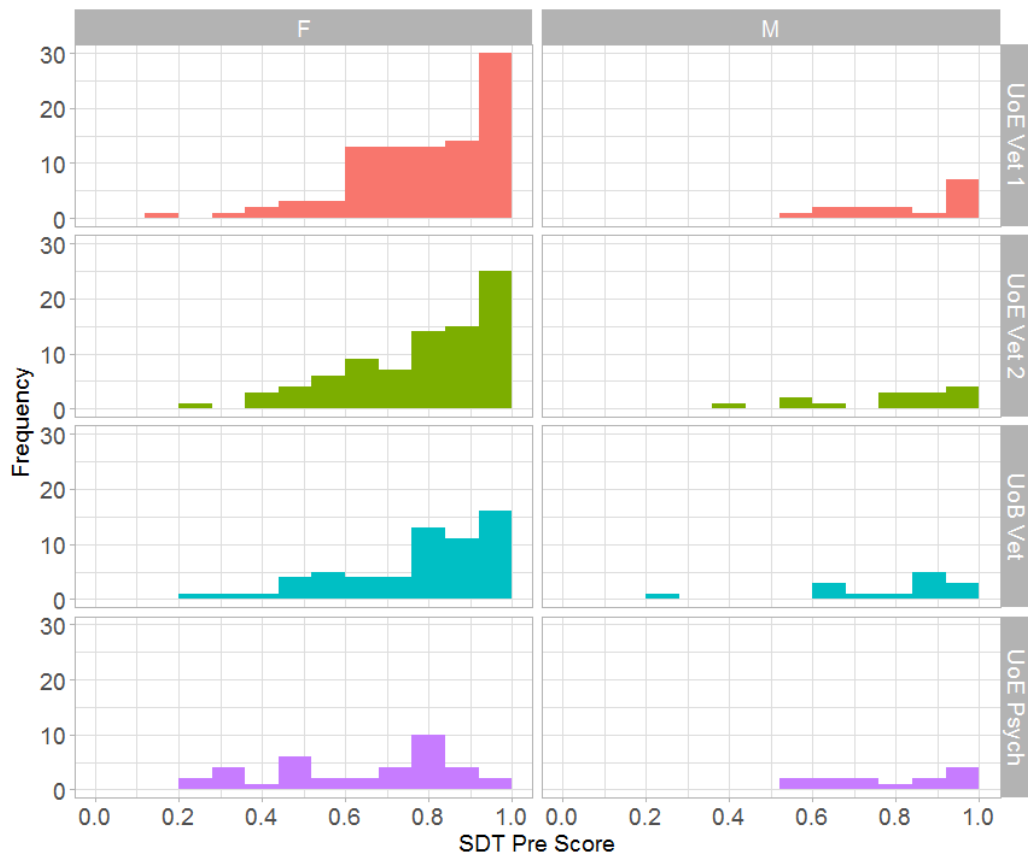
*Figure 4.3 MRT Pre score distribution divided by handedness.*

The distribution of the SDT Pre score showed a markedly left skewed distribution for UoE Vet 1, UoE Vet 2, and UoB Vet possibly indicating this test was easier for these cohorts (Figure 4.2), whereas UoE Psych exhibited a less dramatically skewed distribution. Kolmogorov-Smirnov test comparisons confirmed the UoE Psych distribution was statistically significantly different to the remaining three veterinary cohorts. No statistically significant differences were found between the three veterinary cohorts (Table 4.5).

SDT Pre	Cohort UoE Vet 2	Cohort UoB Vet	Cohort UoE Psych
Cohort UoE Vet 1	D = 0.08 p = 0.898	D = 0.09 p = 0.879	D = 0.24 p = 0.037
Cohort UoE Vet 2		D = 0.05 p = 1.00	D = 0.27 p = 0.017
Cohort UoB Vet			D = 0.28 p = 0.022

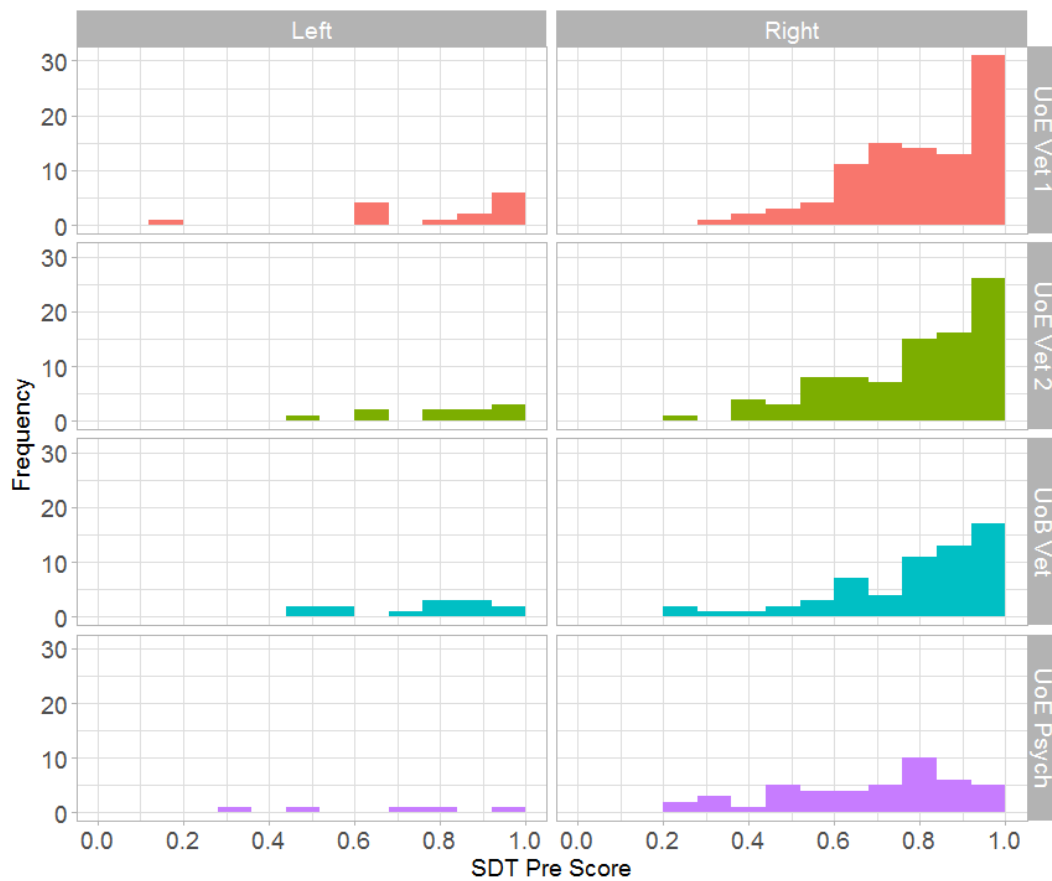
**Table 4.5 Matrix comparing SDT Pre distributions for all possible pairwise combinations of cohorts by Kolmogorov-Smirnov test.**

To ascertain whether the differences of UoE Psych were due to a particular gender or handedness, Kolmogorov-Smirnov comparisons when divided by gender identified statistically significant differences in the distribution of female UoE Psych students compared to the females of the remaining three cohorts ( $p < 0.05$ ) (Figure 4.4). No statistically significant differences were found when comparing the distribution of SDT scores for males between participants.



**Figure 4.4** SDT Pre score distribution divided by gender, *F* = female, *M* = male.

No statistically significant differences were noted for the SDT Pre distribution when comparing the three veterinary cohorts when divided by gender. When the SDT Pre distributions were divided by handedness and compared, UoE Psych right-handed participants scored less well on the SDT Pre compared to UoE Vet 2 right-handed participants ( $D = 0.28$ ,  $p = 0.022$ ). No other statistically significant differences were noted between the cohorts SDT Pre distributions when divided by handedness (Figure 4.5).



**Figure 4.5** SDT Pre score distribution divided by handedness.

The baseline SDT Pre distribution comparisons identified a dramatic left skew and this spatial ability test is likely to be exhibiting a ceiling effect although, UoE Psych do not exhibit a dramatic skew as the remaining three veterinary cohorts possibly suggesting the UoE Psych cohort found this test more challenging than the veterinary cohorts. The difference of the UoE Psych cohort SDT Pre distribution could be due to the female students scoring lower.

Correlations between each spatial ability Pre-test scores (CRT, MRT, and SDT) when all four cohorts were combined into one cohort were moderate in size and

statistically significant (Table 4.6) - confirming the positive manifold effect (Spearman, 1904) as discussed in section 2.2.1.

Cohort Full	MRT Pre	SDT Pre
CRT Pre	$r_s = 0.53, p < 0.001$	$r_s = 0.47, p < 0.001$
MRT Pre		$r_s = 0.52, p < 0.001$

**Table 4.6 Matrix correlation of SA tests for all four cohorts combined.**

Correlations between each spatial ability test Pre score within each cohort were moderate in size ( $r_s < |0.58|$ ) and statistically significant (Tables 4.7, 4.8, 4.9, & 4.10). Confirming each of the three spatial ability tests were testing the same trait as per the positive manifold effect.

Cohort UoE Vet 1	MRT Pre	SDT Pre
CRT Pre	$r_s = 0.56, p < 0.001$	$r_s = 0.48, p < 0.001$
MRT Pre		$r_s = 0.58, p < 0.001$

**Table 4.7 Matrix correlation of spatial ability tests for cohort UoE Vet 1.**

Cohort UoE Vet 2	MRT Pre	SDT Pre
CRT Pre	$r_s = 0.48, p < 0.001$	$r_s = 0.44, p < 0.001$
MRT Pre		$r_s = 0.53, p < 0.001$

**Table 4.8 Matrix correlation of spatial ability tests for cohort UoE Vet 2.**

Cohort UoB Vet	MRT Pre	SDT Pre
CRT Pre	$r_s = 0.55, p < 0.001$	$r_s = 0.47, p < 0.001$
MRT Pre		$r_s = 0.51, p < 0.001$

*Table 4.9 Matrix correlation of spatial ability tests for cohort UoB Vet.*

Cohort UoE Psych	MRT Pre	SDT Pre
CRT Pre	$r_s = 0.52, p < 0.001$	$r_s = 0.52, p < 0.001$
MRT Pre		$r_s = 0.30, p = 0.035$

*Table 4.10 Matrix correlation of spatial ability tests for cohort UoE Psych.*

### 4.3.2 Within Cohort Comparison

Statistically, significant gender differences were identified for UoE Psych with males scoring higher. The other three cohorts did not display any statistically significant gender differences for the CRT Pre spatial ability test scores measuring the sub-category of 2D mental rotation (Figure 4.6 and Table 4.11).

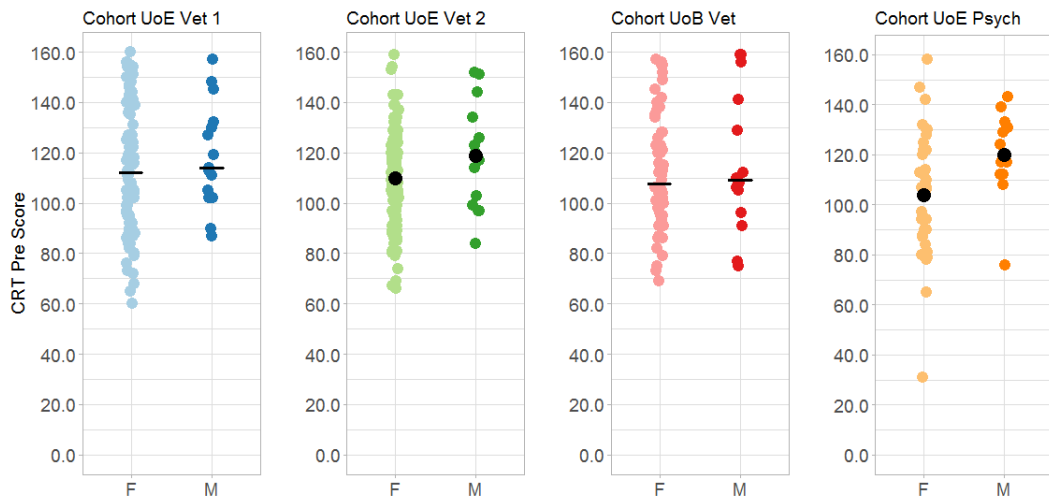


Figure 4.6 CRT Pre score by gender for each cohort, bar = median, dot = mean (points jittered).

CRT Pre by Gender	Female ( $\bar{x}/M$ )	Male ( $\bar{x}/M$ )	Test-Statistics	df	CI/Range	Difference in Location	p-value
Cohort UoE Vet 1	M = 112	M = 114	W = 605	-	Range: 60 – 160	-6.0	= 0.414
Cohort UoE vet 2	$\bar{x}$ = 109.93	$\bar{x}$ = 118.64	t = -1.42	17.38	CI: -21.63 – 4.2	-	= 0.173
Cohort UoB Vet	M = 107.5	M = 109	W = 380	-	Range: 69 – 159	-3.0	= 0.586
Cohort UoE Psych	$\bar{x}$ = 103.92	$\bar{x}$ = 120.08	t = -2.57	31.12	CI: -28.97 – -3.35	-	= 0.015

Table 4.11 Comparison of Gender (female vs male) CRT Pre scores for each cohort.

Gender differences, with males scoring higher, were statistically significantly noted for the MRT Pre scores testing the spatial ability sub-category mental rotation in 3D (Figure 4.7 and Table 4.12). Only Cohort UoB Vet did not exhibit a statistically significant gender effect with the MRT Pre.



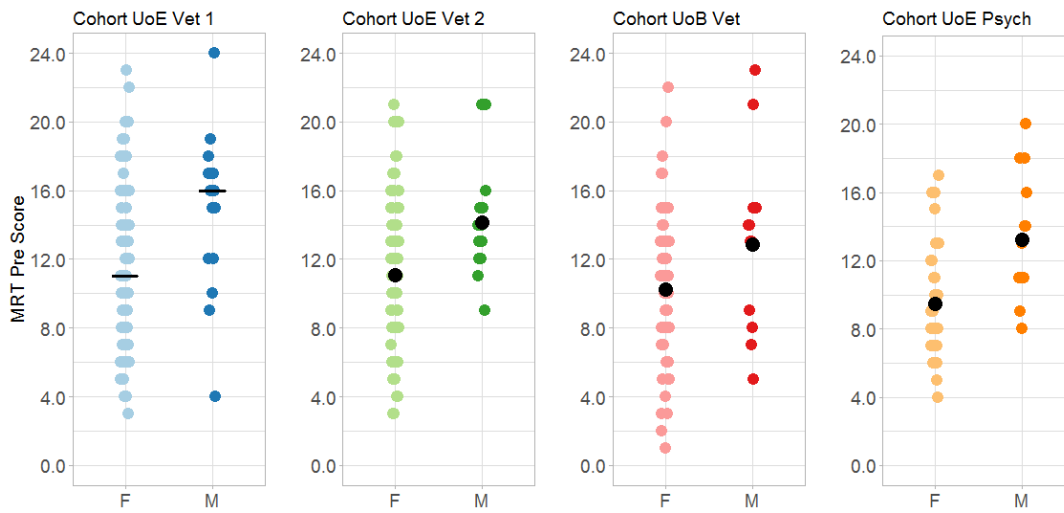


Figure 4.7 MRT Pre score by gender for each cohort, bar = median, dot = mean (points jittered).

MRT Pre by Gender	Female ( $\bar{x}$ /M)	Male ( $\bar{x}$ /M)	Test-Statistics	df	CI/Range	Difference in location	p-value
Cohort UoE Vet 1	M = 11	M = 16	W = 401.5	-	Range: 3 – 24	-4.0	= 0.009
Cohort UoE Vet 2	$\bar{x}$ = 11.08	$\bar{x}$ = 14.14	t = -3.0	20.03	CI: -5.19 – -0.93	-	= 0.007
Cohort UoB Vet	$\bar{x}$ = 10.22	$\bar{x}$ = 12.86	t = -1.78	17.66	CI: -5.77 – 0.49	-	= 0.093
Cohort UoE Psych	$\bar{x}$ = 9.46	$\bar{x}$ = 13.23	t = -3.14	18.72	CI: -6.29 – -1.25	-	= 0.005

Table 4.12 Comparison of Gender (female vs male) MRT Pre scores for each cohort.

Only cohort UoE Psych displayed statistically significant gender differences for the sub-category of spatial visualisation as measured by the SDT (Figure 4.8 and Table 4.13).

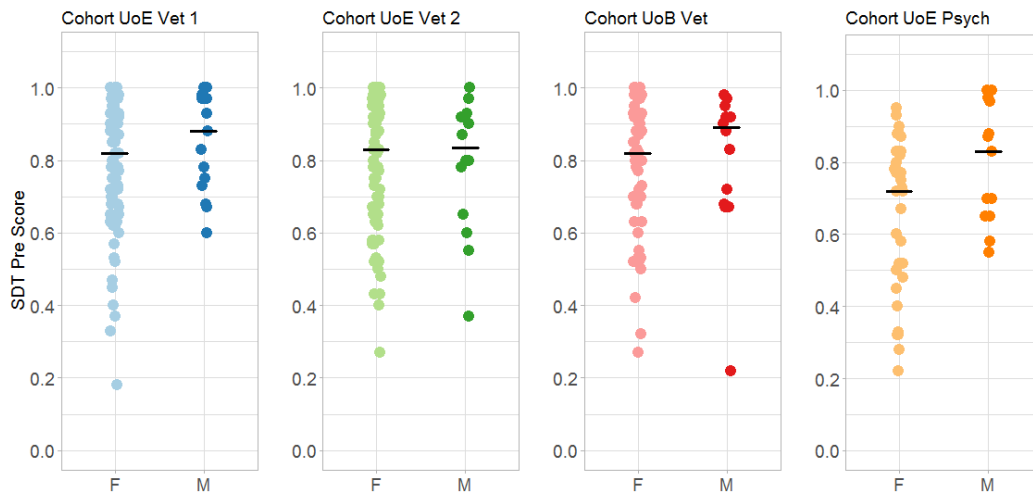


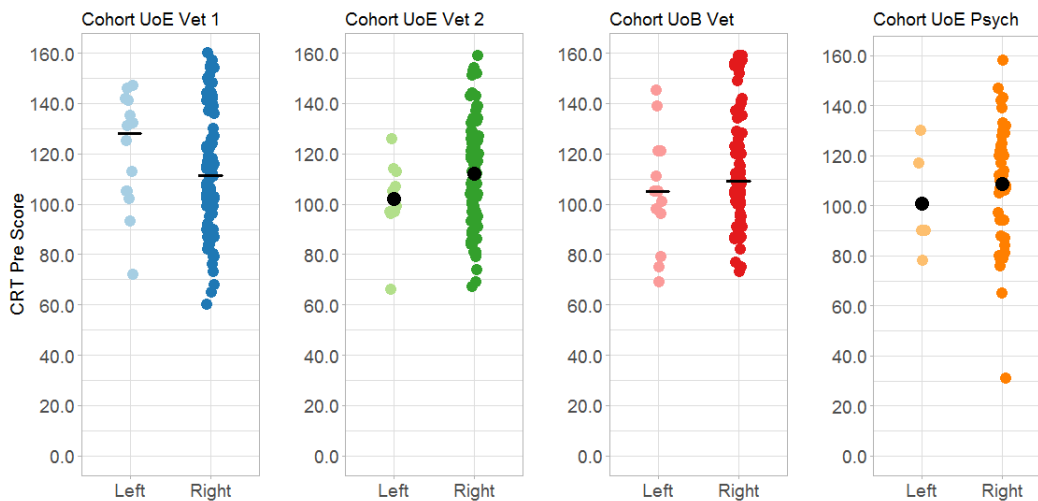
Figure 4.8 SDT Pre score by gender for each cohort, bar = median, dot = mean (points jittered).

SDT Pre by Gender	Female ( $\bar{x}$ /M)	Male ( $\bar{x}$ /M)	Test-Statistics	Range	Difference in location	p-value
Cohort UoE Vet 1	M = 0.82	M = 0.88	W = 568.5	0.18 – 1.0	-0.04	= 0.253
Cohort UoE Vet 2	M = 0.83	M = 0.84	W = 597	0.27 – 1.0	$2.93 \times 10^{-6}$	= 0.931
Cohort UoB Vet	M = 0.82	M = 0.89	W = 384.5	0.22 – 1.0	-0.02	= 0.629
Cohort UoE Psych	M = 0.72	M = 0.83	W = 148.5	0.22 – 1.0	-0.14	= 0.043

Table 4.13 Comparison of Gender (female vs male) SDT Pre scores for each cohort.

To summarise the gender differences, cohort UoE Psych (Table 4.11, 4.12, and 4.13) showed a statistically significant gender effect for all of the three spatial ability tests (CRT Pre  $t_{31.12} = -2.57$ ,  $p = 0.015$ ; MRT Pre  $t_{18.72} = -3.14$ ,  $p = 0.005$ ; and SDT Pre  $W = 148.5$ ,  $p = 0.043$ ). The remaining three cohorts did not show any gender differences for CRT Pre and SDT Pre (Table 4.11 and 4.13), however, gender differences were identified with the MRT Pre scores with males scoring higher except for UoB Vet.

Comparison of left and right-handed student scores on the CRT Pre identified no statistically significant differences within each cohort (Table 4.14, Figure 4.9).

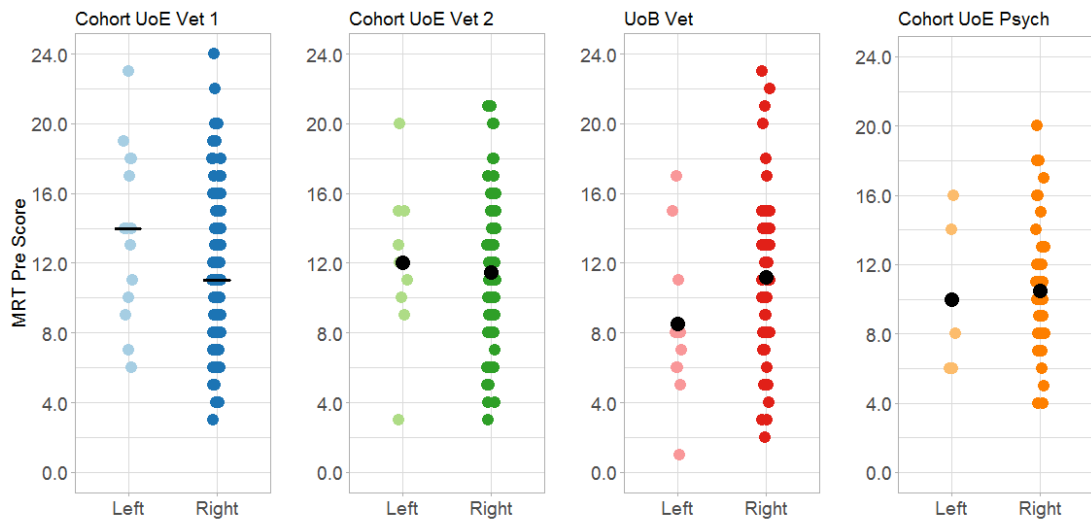


**Figure 4.9** Comparison of handedness for CRT Pre for each cohort, bar = median, dot = mean (points jittered).

CRT Pre by Hand	Left ( $\bar{x}$ /M)	Right ( $\bar{x}$ /M)	Test-Statistics	df	CI/Range	Difference in location	p-value
Cohort UoE Vet 1	M = 128	M = 111.5	W = 778.5	-	Range: 60 – 160	9.0	= 0.272
Cohort UoE vet 2	$\bar{x}$ = 102	$\bar{x}$ = 112.22	t = -1.86	13	CI: -22.11 – 1.68	-	= 0.086
Cohort UoB Vet	M = 105	M = 109	W = 315	-	Range: 69 – 159	-8.0	= 0.250
Cohort UoE Psych	$\bar{x}$ = 101	$\bar{x}$ = 108.91	t = -0.77	5.22	CI: -34.14 – 18.32	-	= 0.477

**Table 4.14** Comparison of handedness (left vs right) scores for CRT Pre for each cohort.

Handedness differences were not statistically significant for the MRT spatial ability test Pre scores (Figure 4.10 and Table 4.15).

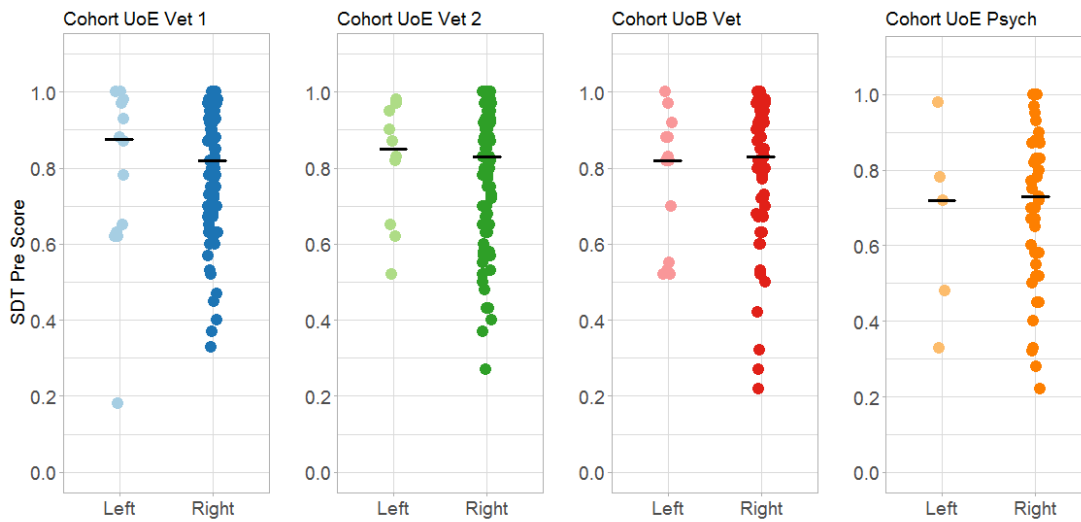


**Figure 4.10 Comparison of handedness for MRT Pre scores for each cohort, bar = median, dot = mean (points jittered).**

MRT Pre by Hand	Left (x/M)	Right (x/M)	Test-Statistics	df	CI/Range	Difference in location	p-value
Cohort UoE Vet 1	M = 14	M = 11	W = 840.5	-	Range: 3 – 24	3.0	= 0.095
Cohort UoE Vet 2	$\bar{x} = 12$	$\bar{x} = 11.47$	t = 0.36	10.89	CI: -2.72 – 3.79	-	= 0.724
Cohort UoB Vet	$\bar{x} = 8.54$	$\bar{x} = 11.18$	t = -2.03	18.69	CI: -5.37 – 0.09	-	= 0.057
Cohort UoE Psych	$\bar{x} = 10$	$\bar{x} = 10.49$	t = -0.23	4.59	CI: -6.22 – 5.25	-	= 0.832

**Table 4.15 Comparison of handedness (left vs. right) scores for MRT Pre for each cohort.**

Statistically significant differences between left and right-handed participants were not identified for the SDT spatial ability test Pre scores measuring the spatial visualisation sub-category (Figure 4.11 and Table 4.16).

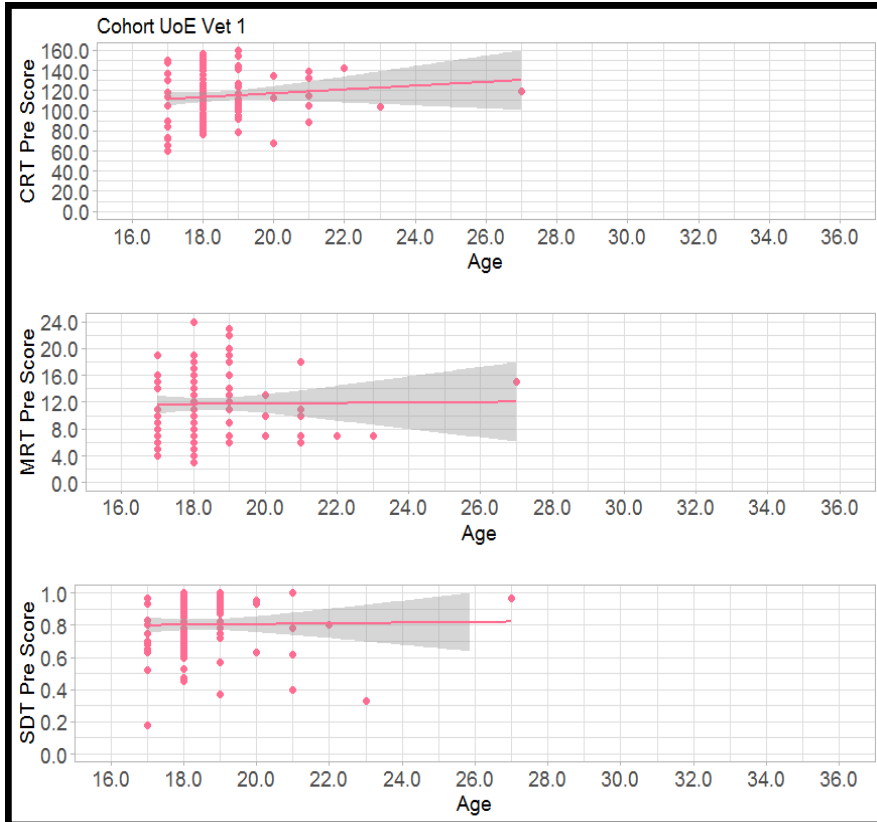


**Figure 4.11** Comparison of handedness for SDT Pre scores for each cohort, bar = median, dot = mean (points jittered).

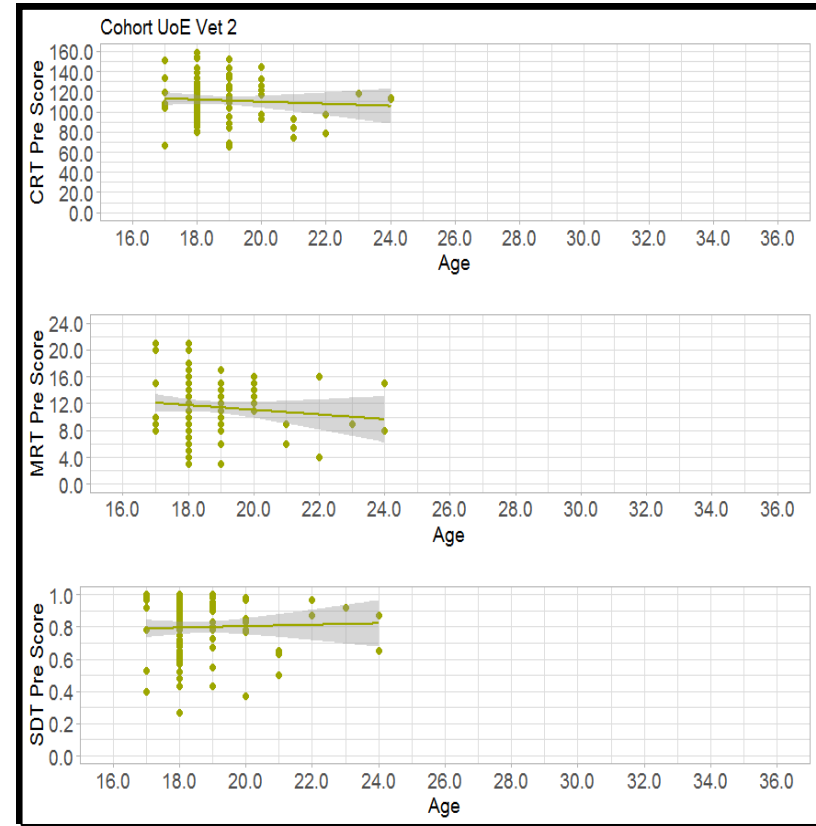
SDT Pre by Hand	Left ( $\bar{x}$ /M)	Right ( $\bar{x}$ /M)	Test-Statistics	Range	Difference in location	p-value
Cohort UoE Vet 1	M = 0.88	M = 0.82	W = 698.5	0.18 – 1.0	0.01	= 0.714
Cohort UoE vet 2	M = 0.85	M = 0.83	W = 448.5	0.27 – 1.0	$3.49 \times 10^{-5}$	= 0.925
Cohort UoB Vet	M = 0.82	M = 0.83	W = 353	0.22 – 1.0	-0.03	= 0.541
Cohort UoE Psych	M = 0.72	M = 0.73	W = 104.5	0.22 – 1.0	-0.04	= 0.808

**Table 4.16** Comparison of handedness (left vs. right) scores for SDT Pre for each cohort.

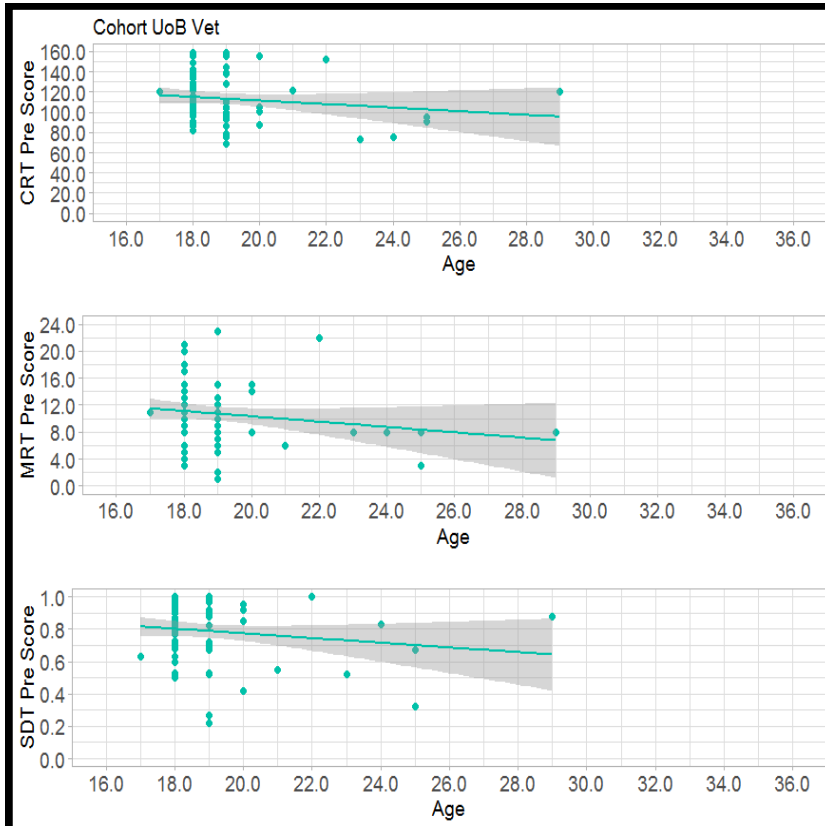
Furthermore, the demographic parameter age was not statistically significant for each cohort and each spatial ability test, as analysed by GLM (Figure 4.12, 4.13, 4.14, and 4.15, and Table 4.17).



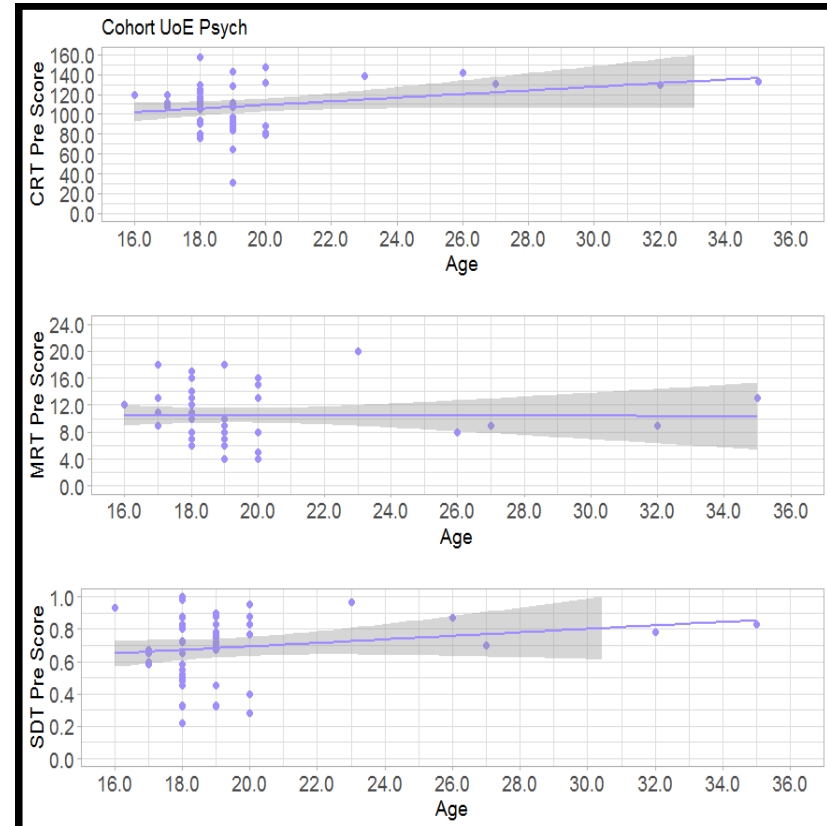
*Figure 4.12 Age by spatial ability test for cohort UoE Vet 1, regression line and 95% confidence interval included.*



*Figure 4.13 Age by spatial ability test for cohort UoE Vet 2, regression line and 95% confidence interval included.*



*Figure 4.14 Age by spatial ability test for cohort UoB Vet, regression line and 95% confidence interval included.*



*Figure 4.15 Age by spatial ability test for cohort UoE Psych, regression line and 95% confidence interval included.*

To summarise the results comparing gender, handedness, and age within each cohort, no statistically significant differences in spatial ability test scores were found between left, and right-handed participants and for different ages. The only demographic parameter to exhibit statistically significant differences was gender, with male participants scoring higher on the MRT as is well documented in the psychology literature (Masters & Sanders, 1993; Reilly & Neumann, 2013). The UoE Psych cohort exhibited a gender difference for all three spatial ability tests.



Parameter	Cohort UoE Vet 1				Cohort UoE Vet 2				Cohort UoB Vet				Cohort UoE Psych			
	$\bar{x}/M$	SD	Range	n	$\bar{x}/M$	SD	Range	n	$\bar{x}/M$	SD	Range	n	$\bar{x}/M$	SD	Range	n
CRT Pre	M = 113	-	60-160	108	$\bar{x} = 111.17$	20.97	-	98	M = 108.5	-	69-159	74	$\bar{x} = 108.1$	24.19	-	50
MRT Pre	M = 11	-	3-24	108	$\bar{x} = 11.52$	4.19	-	98	$\bar{x} = 10.72$	4.62	-	74	$\bar{x} = 10.44$	3.82	-	50
SDT Pre	M = 0.82	-	0.18 – 1.0	108	M = 0.83	-	0.27 – 1.0	98	M = 0.83	-	0.22 – 1.0	74	M = 0.73	-	0.22 – 1.0	50
Age	M = 18	-	17-27	108	M = 18	-	17-24	98	M = 18	-	17-29	74	M = 18	-	16-35	50

*Table 4.17 Mean ( $\bar{x}$ ), median (M), range, sample size (n) and standard deviation (SD) for each continuous parameter.*

### 4.3.3 Between Cohort Comparison

Comparison of cohorts' CRT Pre scores by GLM for the spatial ability sub-category 2D mental rotation identified no statistically significant differences between cohorts (Figure 4.16, Table 4.17 and 4.18). There was no statistically significant difference between all pairwise combinations of the four cohorts' CRT Pre scores ( $p > 0.05$ ).

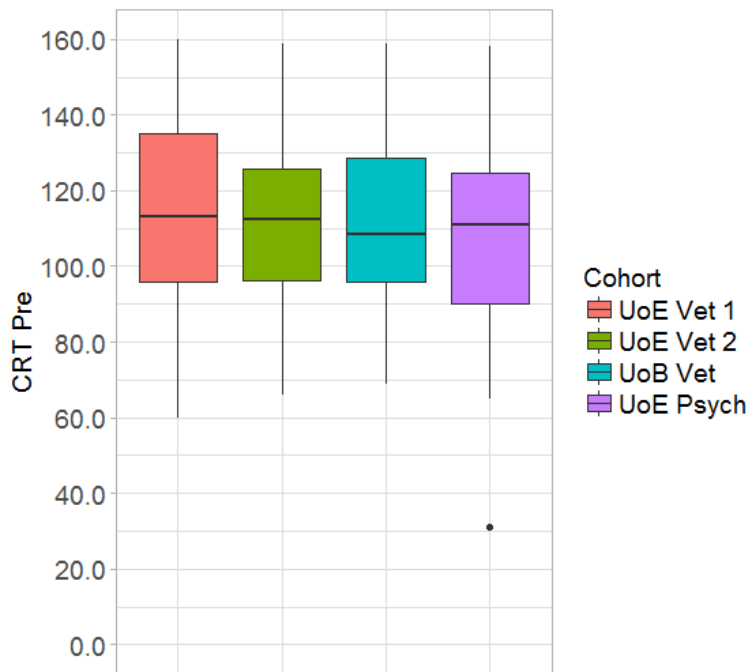
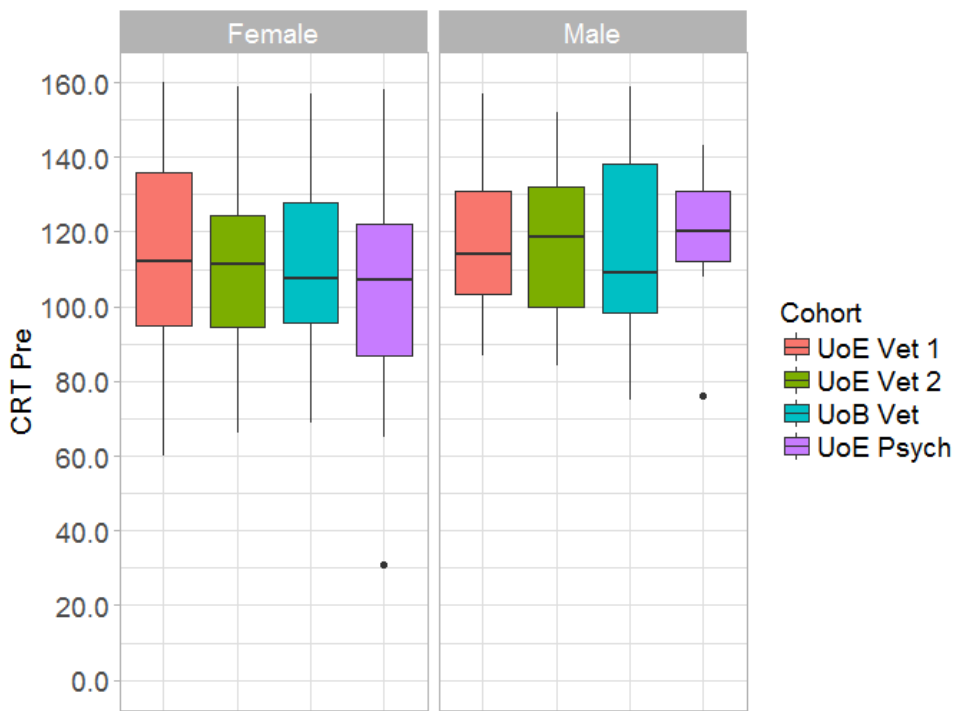


Figure 4.16 CRT Pre score comparison between all four cohorts.

Response Variable	Significant Explanatory Variable	Relative Risk Ratio (RR)	95% CI	p-value	Non-significant Explanatory Variables Removed
CRT Pre	Cohort				Age Handedness
	UoE Vet 1	Ref			
	UoE Vet 2	0.97	0.92 – 1.03	= 0.365	
	UoB Vet	0.99	0.93 – 1.05	= 0.684	
	UoE Psych	0.94	0.88 – 1.01	= 0.082	
	Gender				
	Female	Ref			
	Male	1.07	1.01 – 1.14	= 0.018	

Table 4.18 CRT Pre GLM<sub>P</sub> output after model simplification, ref = reference level.

Gender was shown to be statistically significant for CRT Pre (RR = 1.07,  $p = 0.018$ , adjusted for cohort) with males scoring higher than females (Figure 4.17).



*Figure 4.17 Effect of gender with CRT Pre score for each cohort.*

However, an odds ratio of 1.07 is a small effect size. Handedness and age were both not statistically significant for CRT Pre scores.

Mental Rotation Test (MRT) Pre scores were statistically significantly different between UoE Vet 1 and UoE Psych, with UoE Psych scoring lower (RR = 0.86,  $p = 0.030$ , adjusted for gender differences). There were no statistically significant differences between all combinations of the remaining three veterinary cohorts (Figure 4.18 and Table 4.19 and 4.17).

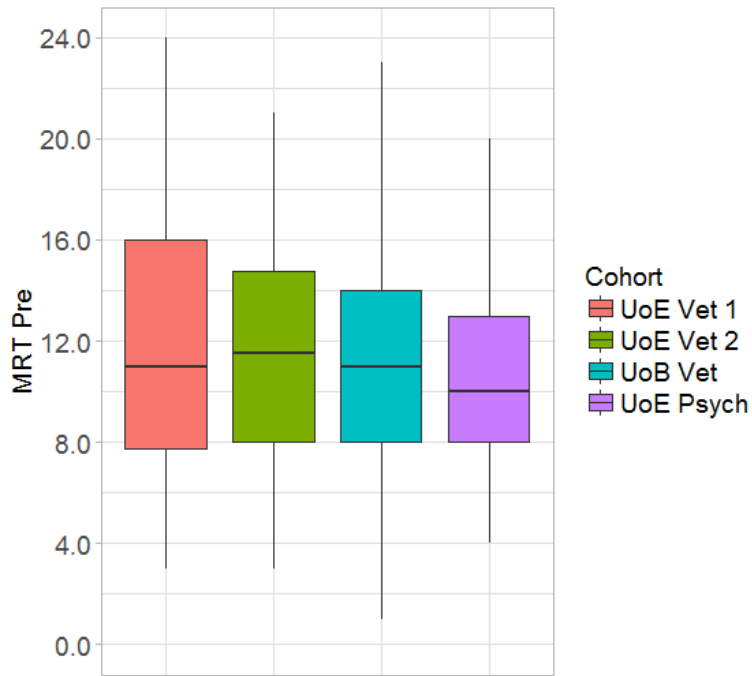
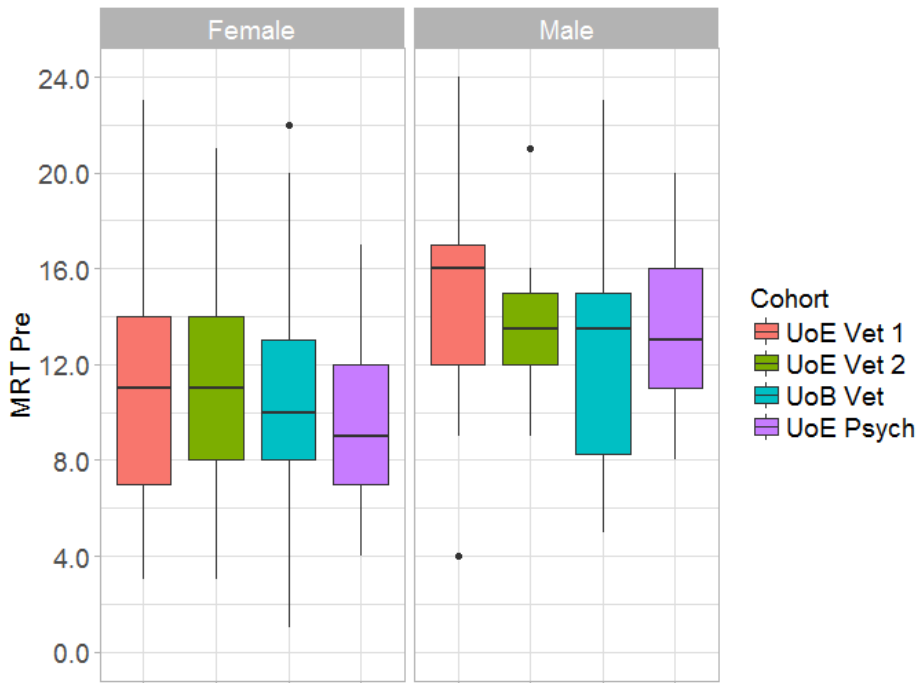


Figure 4.18 MRT Pre score comparison between all four cohorts.

Response Variable	Significant Explanatory Variable	Relative Risk Ratio (RR)	95% CI	p-value	Non-significant explanatory variables removed
MRT Pre	Cohort				Age Handedness
	UoE Vet 1	Ref			
	UoE Vet 2	0.99	0.89 – 1.09	= 0.782	
	UoB Vet	0.90	0.81 – 1.01	= 0.086	
	UoE Psych	0.86	0.76 – 0.98	= 0.030	
	Gender				
	Female	Ref			
	Male	1.31	1.18 – 1.45	< 0.001	

Table 4.19 MRT Pre GLM<sub>P</sub> output after model simplification, ref = reference level.

The explanatory variable gender (RR = 1.31,  $p < 0.001$ , adjusted for cohort) was significant for MRT Pre scores with males scoring higher than females (Figure 4.19 and Table 4.19).

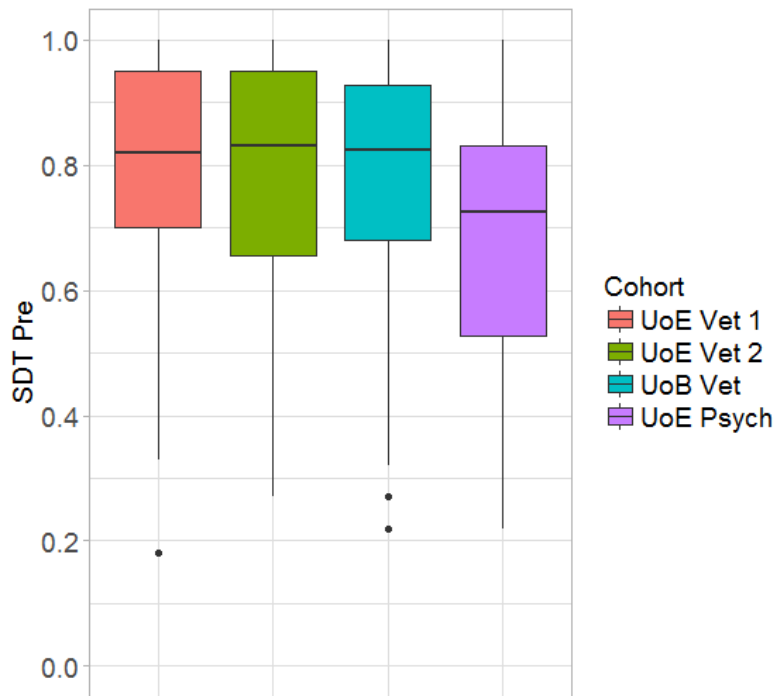


**Figure 4.19** Effect of gender with MRT Pre score for each cohort.

Comparison of cohorts SDT Pre scores for the spatial ability sub-category spatial visualisation, identified statistically significant differences between UoE Psych and each of the three veterinary cohorts (Figure 4.20 and Table 4.20 and 4.17), with UoE Psych scoring lower (UoE Vet 1 OR = 1.85,  $p < 0.001$ ; UoE Vet 2 OR = 1.78,  $p < 0.001$ ; and UoB Vet OR = 1.69,  $p = 0.004$ ). There were no statistically significant differences between all pairwise combinations of the three veterinary cohorts.

Response Variable	Significant Explanatory Variable	Odds Ratio (OR)	95% CI	p-value	Non-significant explanatory variables removed
SDT Pre	Cohort	Ref			Age Handedness Gender
	UoE Psych	1.85	1.32 – 2.58	$< 0.001$	
	UoE Vet 1	1.78	1.27 – 2.50	$< 0.001$	
	UoB Vet	1.69	1.18 – 2.43	$= 0.004$	

**Table 4.20** SDT Pre GLM<sub>B</sub> output after model simplification.



*Figure 4.20 SDT Pre score comparison between all four cohorts.*

The explanatory variables gender, handedness, and age were not statistically significant for the Surface Development Test (SDT) scores.

## 4.4 Discussion

The purpose of the study presented in this chapter was to analyse the baseline spatial ability of veterinary students, using three different spatial ability tests. Specifically, the study aimed to determine if this cognitive ability varies among veterinary students of different cohorts and across two academic institutions. Additionally, a control cohort of psychology students was used to determine if there were any differences between veterinary students and students from another academic discipline.

Chapter 4 of the “Cambridge Handbook of Visuospatial Thinking” entitled “Individual Differences in Spatial Abilities” (Hegarty & Waller, 2005), provides an excellent discussion on the cognitive psychology literature on why performances on spatial ability tests can differ. In this chapter it is explained that to investigate individual differences, psychologists breakdown the cognitive processes used to solve items in a test, to thus identify which processes are used. Psychologists then compare participants’ scores on spatial ability tests and scores on tests designed to isolate the basic processes used, to identify reasons for differences in test scores. Hegarty and Waller (2005) explain that from this ‘componential approach’ it is found that differences in spatial ability tests can depend on the speed of processing, strategies, working memory, and mental imagery. The findings from the studies presented by Hegarty and Waller’s (2005) chapter on individual differences will be discussed in relation to the findings of the study presented in this chapter.

Studies investigating individual differences in spatial ability tests have identified that individuals’ speed of processing is an important source of variation in test scores (Hegarty & Waller, 2005; Mumaw *et al.*, 1984; Lohman, 1986). Individuals with a faster processing speed perform better than those with slower processing speed. Simpler spatial ability tests, for example involving 2D rotations as with the CRT in the current chapters study, are identified as being easily answered by most individuals (Hegarty & Waller, 2005). Thus, when a speed limit is placed on these tests, differences in individuals’ test scores are identified, and this reflects individuals’ differences in speed of processing.

However, with more complex tests involving multi-step rotations and manipulations, such as 3D mental rotation and spatial visualisation tests (e.g., the MRT and SDT used in the current study of this chapter), these tests cannot be solved easily by all individuals (Hegarty & Waller, 2005). This holds even when long liberal time limits are given to complete tests (Lohman, 1986). Hegarty and Waller (2005) explain that Lohman (1986) identified this by graphing the speed of completing tasks against

accuracy, for individuals with low and high spatial ability as measured by the Shepard and Metzler Mental Rotation Test (1971). It was found both curves asymptote with low spatial ability individuals plateauing at a lower level than high spatial ability individuals. These results indicate that there is a maximum level of accuracy obtained that did not increase with longer time limits, and this was lower for low spatial ability individuals.

In the study presented in this chapter, the Card Rotation Test (CRT) can be categorised as measuring 2D mental rotation (Linn & Petersen, 1985). There are currently limited studies linking performance on the CRT to anatomy education (see section 1.4). In this study, it was shown the scores on the CRT were not statistically different across veterinary students from two academic institutions, or when comparing veterinary students to psychology students. The lack of differences in the CRT Pre scores could be due to the simpler nature of the CRT compared to other tests of spatial ability as discussed. Due to the simplicity of the test, any differences in baseline spatial ability may not be large enough to be observed unless more stringent time limits are implemented.

In contrast, the MRT (Mental Rotation Test) and SDT (Surface Development Test) can be categorised as involving multi-step rotations and manipulations, and these tests were found to have statistically significant differences between the four cohorts. In particular, the SDT scores were found to be lower for the UoE Psychology cohort compared to all three veterinary cohorts, but no differences were found when comparing the three veterinary cohorts only. While for the MRT only the UoE Vet 1 and UoE Psych cohorts exhibited differences in baseline scores. Differences in the students' speed of processing could explain these observed differences as discussed above.



Individual differences in working memory are another source of variation in test scores, particularly for tests of spatial visualisation (Hegarty & Waller, 2005). Studies have found the quality of spatial representations mentally created by individuals with low and high spatial ability differs, and the quality of these images can differ after mental transformation (Lohman, 1988; Just & Carpenter, 1985; Mumaw & Pellegrino, 1984). Low spatial individuals are less able to maintain mental spatial representations after the image has been transformed, while high spatial individuals exhibit better storage and processing skills of spatial representations (Hegarty & Waller, 2005). For instance, using the Cube Comparison test (participants decide whether two cubes with letters on each face are the same or different to one another), low spatial individuals tend to 'lose' information about the letter on the sides of the cubes once they have been rotated 'out of view' (Hegarty & Waller, 2005; Just & Carpenter, 1985). Thus, for this study, differences in the veterinary and psychology students' scores on the SDT could be due to differences in working memory along with their speed of processing.

The SDT in the study presented in this chapter exhibited the most consistent differences with the UoE Psych students scoring lower compared to all three veterinary cohorts. The SDT is categorised as a test of spatial visualisation, a test that involves multi-step complex manipulations. Tests of this category can, therefore, be solved by a variety of strategies, some of which will be more efficient than others (Hegarty & Waller, 2005). In the psychology literature on individual differences, this contribution of strategies has been identified as a source of variation in tests of spatial visualisation.

Hegarty and Waller (2005) explain that Just and Carpenter (1985) investigated this with the Cube Comparison Test. The main strategies identified from this study were a mental rotation strategy, a perspective-taking strategy (where the individual imagines changing their perspective of the cube), and a strategy comparing the orientation-free descriptions of the cubes (i.e., comparing the orientation of the

letters on the cube faces). Another strategy identified for tests of spatial visualisation is the use of an analytic strategy (Kyllonen, Lohman & Snow, 1984). To explain this further Hegarty and Waller use the paper folding test as an example (participants in this test are to identify out of five options which one correctly shows the hole punch pattern of a piece of folded paper punched with holes). They explain that if the paper were folded in half, then the pattern of the punched holes would be symmetrical. This analytic strategy is often best used for tests of spatial visualisation, and this finding has been used as a reason for why tests of this category load highly onto the general intelligence factor  $g$  (Lohman, 1988 and as discussed in section 2.4).

Differences in the students'  $g$  factors could, therefore, account for differences in the SDT scores. This theory would hence predict the psychology student cohort would have a lower general intelligence compared to the veterinary students because of the lower scores on the SDT, while the veterinary students' scores exhibited a ceiling effect with the majority of students scoring high. In the study presented in this chapter, no measure of general intelligence ( $g$ ) was obtained, such as the use of Raven's Matrices to obtain a light indication of intelligence as was done in the study by Hegarty et al. (2009) with dentistry students. Differences in general intelligence could account for the differences in the scores on the SDT.

Another possible explanation for differences in the SDT scores in relation to  $g$  could be due to each degree attracting different applicants. The entry requirements for a veterinary degree in the UK include obtaining high grades at high school on first sitting, suitable industry work experience at a range of organisations (e.g. farms, stables, kennels, veterinary practice both large animal and small animal), and an entrance interview often involving multiple circuits examining a range of skills and abilities. The numbers of places available on a veterinary course are also limited due to the specialty of the subject, which therefore adds a competitive element to obtaining a place.

In comparison, entrance into a psychology degree in the UK requires high grades at high school without the added requirement of work experience or an entrance interview, and places are less competitive due to larger cohort sizes (the numbers when this current study was conducted were roughly 120 veterinary students and 400 psychology students). Therefore veterinary degrees could attract, and possibly accept, more homogenous populations of applicants, while psychology degrees could attract a wider heterogeneous population of applicants providing extra sources of variation between the populations. Furthermore, the sample of psychology students in this study was 50 out of a possible 400, and may not truly be representative of this cohort of students.

The finding in this study of no difference between the three veterinary cohorts spatial ability scores indicates the CRT, MRT, and SDT are consistent at measuring this ability in this population of students. Professional examinations across veterinary schools in the UK can vary between institutions – each institution designs and conducts their examinations although this is regulated by accrediting bodies (Rhind & Baillie, 2018; Baillie, Warman & Rhind, 2014; Pooley & Wapenaar, 2018). The identification from this study that spatial ability as measured by the same three tests was stable across two academic institutions could be used to provide a uniform measure of assessment across veterinary students in the UK. However, with the discovery that spatial ability is a malleable trait (Uttal *et al.*, 2013, section 2.3.6) and can potentially change across a professional career with skill acquisition (Keehner *et al.*, 2006, section 1.3.2), the use of these tests should be used to assess and inform educational practices and not for entrance to veterinary schools or degree qualifications.

The analytic approach to solving spatial problems was used in the design of the 3D spatial teaching method of this thesis, with the aim of encouraging the students to think analytically in relation to anatomy as well as rotationally. The next chapter

presents the statistical analysis comparing the novel 3D spatial teaching method to a 2D traditional teaching method.



# Chapter 5 Comparison of a 2D and 3D Spatial Anatomy Teaching Method

## 5.1 Introduction to Chapter 5

In the previous chapter the baseline spatial ability scores on three spatial ability tests (Card Rotation Test (Ekstrom *et al.*, 1976), Mental Rotation Test (Peters *et al.*, 1995), and Surface Development Test (Ekstrom *et al.*, 1976)) were presented and compared between four cohorts of undergraduate university students. The conclusions of chapter 4 were the baseline spatial ability scores of veterinary students on the Card Rotation Test (CRT), Mental Rotation Test (MRT), and Surface Development Test (SDT) was not different across cohorts of students from two veterinary academic institutions or the same veterinary academic institution. However, veterinary students scores on the SDT were statistically significantly higher than those of first-year psychology students (control cohort).

This current chapter expands on the research investigating the links between spatial ability and anatomy teaching (section 1.3.2 and 1.4) by presenting the quantitative analyses investigating the relationship between spatial ability and anatomy teaching methods. Two cohorts of students are presented in this chapter, University of Edinburgh Veterinary Students 1 (UoE Vet 1) and University of Edinburgh Veterinary Students 2 (UoE Vet 2). The two cohorts of students presented in this chapter each received a different teaching method. UoE Vet 1 received a two-dimensional (2D) teaching method (section 3.7) and UoE Vet 2 a three-dimensional (3D) spatial teaching method (section 3.8).

Previous studies have investigated the use of teaching methods involving 3D technology to improve anatomy knowledge and understanding (Garg *et al.*, 1999b,

1999a; Garg, Norman & Sperotable, 2001; Berney *et al.*, 2015; Nguyen, Nelson & Wilson, 2012; Smith *et al.*, 2018; Plumley *et al.*, 2013; Cui *et al.*, 2016; Küçük, Kapakin & Göktaş, 2016; Preece *et al.*, 2013), with most studies additionally investigating the relationship to students' spatial ability (Berney *et al.*, 2015; Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001; Nguyen, Nelson & Wilson, 2012; Tan *et al.*, 2012; Hoyek *et al.*, 2009; Cui *et al.*, 2016).

Varying 3D technologies have been implemented in anatomy teaching methods such as 3D computer models, 3D printed models, diagnostic imaging, augmented reality, and stereoscopic displays. All of which have focused on presenting anatomical material in different dimensional perspectives (Garg *et al.*, 1999b; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Küçük, Kapakin & Göktaş, 2016; Plumley *et al.*, 2013; Smith *et al.*, 2018). However, spatial ability is thought of as having two aspects (Lohman, 1996, and section 2.4 of this thesis); the first is the direct 3D manipulation of structures (rotating, cutting, slicing, floor-plan view, change in perspective) and involves looking at objects from a different dimensional perspective, and is often how spatial ability is first perceived. The second is a more analytical aspect; spatial ability can also be used abstractly to look at problems from different possibilities and involve analytical, problem solving, creative cognitive processes (Lohman, 1996; Hegarty & Waller, 2005; Lohman, 1988).

What exactly constitutes a 3D teaching method has not yet been defined or confirmed. Is it the inclusion of 3D teaching resources to the learning experience as a means to show different perspectives, or is it something more immersive? For example, encouraging the students to think in an analytical spatial capacity every time they approach the topic of anatomy. By designing learning resources to not only present the material in a different 3D perspective as has previously been implemented but also designed to support the spatial analytical thinking of students. Often individual learners require different explanations to understand a concept, and

in anatomy, these different explanations can take the form of a different 3D perspective, which involves the first aspect of spatial ability (e.g. “imagine viewing the dog in dorsal recumbency”). Explanations of anatomy can also involve the second aspect of spatial ability, by thinking abstractedly or analytically about an anatomy problem (e.g., appreciating when learning the tributaries of the caudal vena cava the cranial and caudal mesenteric veins drain into the hepatic portal vein). These different approaches to anatomy explanations will use spatial ability to achieve a new way of thinking, and in doing so could enhance a student’s ability to think spatially about anatomy (both three-dimensionally and in a problem-solving capacity).

The research presented in this chapter takes a different approach to 3D teaching by immersing students in a spatial environment that not only incorporated learner-controlled 3D computer models and 3D prints of anatomy, but also delivered this teaching across a gross anatomy curriculum involving diagnostic images, spatially orientated tutorials, and spatial analytical explanations of anatomy. The overall aim of this approach was to guide and encourage the students to think spatially about anatomy, in both a 3D perspective and involving spatial analytical cognitive processes.

Studies have primarily researched the educational value of 3D technology interventions outwith the core curriculum as extra optional tutorials (Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001; Berney *et al.*, 2015; Tan *et al.*, 2012; Keedy *et al.*, 2011). This approach has primarily been employed for ethical reasons, as a new intervention may be disadvantageous or advantageous to learning. This has resulted in varied contact time with the teaching intervention (ranging roughly between 3 minutes to 5 hours). Learners need time to develop and grow while they navigate new and challenging material in anatomy. Providing 3D resources during a one-off tutorial may not be enough to push through difficult learning concepts (Meyer & Land, 2003; Cousin, 2006). It is also unknown exactly



how long the 3D teaching intervention needs to be to improve anatomy knowledge (or even spatial ability), and also how quickly, and by how much, each learner will improve (section 1.3.4). So far there are no anatomy education studies or literature that answer or discuss these questions.

A study by Peterson and Mlynarczyk (2016) incorporated 3D material across a veterinary curriculum and compared this to a traditional veterinary curriculum with no 3D material. The researchers' 3D material aimed to move in such a way as to show more of the 3D nature of objects, and this was achieved using a variety of resources (such as rotatable CT and MRI scans, 3D virtual dissector programs, and video tutorials). The 3D teaching resources were incorporated into the most challenging sections of the course as identified by student questionnaires and included: the central nervous system, cranial nerves, skull, head and neck, muscles, brachial plexus, heart, upper respiratory tract, and pelvis. The authors found the 3D augmented traditional teaching method significantly improved anatomy knowledge ( $p < 0.001$ ), and this was particularly so for cadaveric questions ( $p < 0.001$ ). Unfortunately, no spatial ability tests were used in this study to investigate the links to this cognitive ability. The study presented in this thesis investigates the incorporation of a 3D spatial teaching method into a veterinary anatomy course and uses three spatial ability tests to measure the relationship to spatial ability.

End-of-course examination scores were used to compare the efficacy of the novel 3D spatial teaching method of this study to students who received a traditional 2D anatomy teaching method. This was to assess whether a novel 3D spatial teaching method improved anatomy knowledge and understanding as measured by end-of-course examinations (section 3.6.1). The end-of-course examinations in this study incorporate a range of assessment types including MCQ, short answer, interpretation (interp), spot, oral, and in-course assessments involving dissection workbooks. Additionally, this chapter investigates the possible links between performance on anatomy examinations and spatial ability by analysing whether spatial ability as

measured by the Card Rotation Test (CRT), Mental Rotation Test (MRT), and Surface Development Test (SDT) was related to anatomy examination scores.

This chapter addresses the following research questions:

1. Does teaching anatomy using a traditional 2D method improve students' anatomy knowledge and understanding?
2. Does anatomy teaching incorporating diagnostic imaging/ 3D images and 3D printing improve students' anatomy knowledge and understanding?
3. Is spatial ability a predictor of success in anatomy examinations?

## **5.2 Data Analysis**

### **5.2.1 Descriptive Statistics**

Cohorts University of Edinburgh Veterinary Students 1 (UoE Vet 1) and University of Edinburgh Veterinary Students 2 (UoE Vet 2) were included in the analysis of this chapter. UoE Vet 1 received the 2D teaching method and UoE Vet 2 the 3D spatial teaching method. Shapiro-Wilk normality tests were performed and histograms plotted for the continuous parameters in-course, oral, short, interpretation (interp), MCQ, spot, and total exam as a proportion of total maximum score. Additionally, the Animal Body 2 (AB2) exam result was included as a means to account for academic differences between the two cohorts and was checked for normality. All examination parameters (in-course, oral, short, interpretation (interp), MCQ, spot, and total exam) were a proportion to account for differences in examination questions between the two cohorts (section 3.6.1, Table 3.4).

Descriptive statistical analysis was performed including the mean or median with accompanying SD or range for each continuous parameter. Spearman rank correlations were performed between each exam parameter and each spatial ability test score (Card Rotation Test (CRT), Mental Rotation Test (MRT), and the Surface Development Test (SDT)) to check for any associations for each cohort.

## 5.2.2 Between Teaching Method Comparison

Univariable analysis by a two-sample t-test for the exam parameter short questions and Wilcoxon-rank sum tests for in-course, oral, interp, spot, and total exam were performed to compare scores between teaching methods/cohorts. To further compare the exam results between each cohort/teaching method separate  $GLM_B$  analyses (section 3.9.3) were performed, one for each exam parameter. For each  $GLM_B$  the response variable was 'exam score' as a proportion of total maximum score (e.g., for in-course, oral, short, interp, MCQ, spot, and total exam) with the continuous explanatory variables age, AB2 score, CRT Pre, MRT Pre, and SDT Pre, and the categorical explanatory variables teaching method/cohort (2D and 3D), gender (Male and Female), and handedness (Left and Right). The interactions CRT Pre\*Cohort, MRT Pre\*Cohort, and SDT Pre\*Cohort were added to each model to assess further whether there were differences between the two teaching methods/cohorts and spatial ability sub-category to exam parameter.

The separate  $GLM_B$  analyses included spatial ability as a continuous explanatory parameter to investigate whether the spatial ability Pre scores for each test (CRT, MRT, and SDT) were predictive of examination performance for each exam parameter. AB2 score was identified as a confounding variable, therefore to account for the extra influence of AB2 score all Odds Ratios (OR) for this chapter are adjusted for AB2 score by not removing AB2 score during model simplification. Table 5.1 summarises the  $GLM_B$  analyses conducted in this chapter.

Exam Parameter	Response Variable (as a proportion)	Explanatory Variables		Interactions
		Continuous	Categorical	
In-Course	Exam Parameter “In-Course”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
Oral	Exam Parameter “Oral”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
Short	Exam Parameter “Short”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
Interp	Exam Parameter “Interp”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
MCQ	Exam Parameter “MCQ”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
Spot	Exam Parameter “Spot”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
Total Exam	Exam Parameter “Total Exam”	AB2, Age, CRT Pre, MRT Pre, SDT Pre	Teaching method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort

*Table 5.1 Summary of GLM<sub>B</sub> analysis conducted for chapter 5.*

## 5.3 Results

### 5.3.1 Descriptive Statistics

All AB1 exam parameters correlated statistically significantly and positively with one another for 2D cohort (UoE Vet 1) ( $r_s = |0.95|$ ,  $p < 0.05$ ) and 3D cohort (UoE Vet 2) ( $r_s = |0.95|$ ,  $p < 0.05$ ), except for the correlations in-course  $\alpha$  oral and in-course  $\alpha$  MCQ for 2D (UoE Vet 1). Additionally, AB2 correlated positively and statistically significantly with all exam parameters for both cohorts identifying this variable as a confounding variable (Table 5.2).

CRT Pre and MRT Pre both measure the sub-category of mental rotation, in 2D and 3D respectively, and were not statistically significantly related to any of the exam parameters. Whereas SDT Pre, measuring spatial visualisation, had a low statistically significant correlation with oral, interp, spot, and total exam for 2D cohort (UoE Vet 1) ( $r_s = |0.24|$ ,  $p < 0.05$ ), and with interp and total exam for 3D cohort (UoE Vet 2) ( $r_s = |0.24|$ ,  $p < 0.05$ ). However many students had similarly high marks for the SDT Pre which may be falsely contributing to these significant relationships (Table 5.2).

	In- Course	Oral	Short	Interp	MCQ	Spot	Total Exam	AB2	CRT Pre	MRT Pre	SDT Pre
In-Course	<b>In-course</b>	$r_s = 0.14$ $p = 0.144$	$r_s = 0.30$ $p = 0.002$	$r_s = 0.40$ $p < 0.001$	$r_s = 0.11$ $p = 0.254$	$r_s = 0.30$ $p = 0.002$	$r_s = 0.35$ $p < 0.001$	$r_s = 0.21$ $p = 0.028$	$r_s = -0.01$ $p = 0.942$	$r_s = -0.06$ $p = 0.514$	$r_s = 0.11$ $p = 0.244$
Oral	$r_s = 0.36$ $p < 0.001$	<b>Oral</b>	$r_s = 0.54$ $p < 0.001$	$r_s = 0.36$ $p < 0.001$	$r_s = 0.26$ $p = 0.007$	$r_s = 0.52$ $p < 0.001$	$r_s = 0.55$ $p < 0.001$	$r_s = 0.50$ $p < 0.001$	$r_s = 0.10$ $p = 0.285$	$r_s = 0.06$ $p = 0.552$	$r_s = 0.19$ $p = 0.044$
Short	$r_s = 0.41$ $p < 0.001$	$r_s = 0.53$ $p < 0.001$	<b>Short</b>	$r_s = 0.47$ $p < 0.001$	$r_s = 0.35$ $p < 0.001$	$r_s = 0.70$ $p < 0.001$	$r_s = 0.87$ $p < 0.001$	$r_s = 0.56$ $p < 0.001$	$r_s = 0.10$ $p = 0.283$	$r_s = 0.07$ $p = 0.480$	$r_s = 0.16$ $p = 0.099$
Interp	$r_s = 0.25$ $p = 0.013$	$r_s = 0.31$ $p = 0.002$	$r_s = 0.47$ $p < 0.001$	<b>Interp</b>	$r_s = 0.27$ $p = 0.005$	$r_s = 0.60$ $p < 0.001$	$r_s = 0.65$ $p < 0.001$	$r_s = 0.44$ $p < 0.001$	$r_s = 0.0$ $p = 0.990$	$r_s = 0.07$ $p = 0.470$	$r_s = 0.20$ $p = 0.040$
MCQ	$r_s = 0.26$ $p = 0.011$	$r_s = 0.39$ $p < 0.001$	$r_s = 0.59$ $p < 0.001$	$r_s = 0.51$ $p < 0.001$	<b>MCQ</b>	$r_s = 0.50$ $p < 0.001$	$r_s = 0.51$ $p < 0.001$	$r_s = 0.34$ $p < 0.001$	$r_s = 0.01$ $p = 0.881$	$r_s = -0.09$ $p = 0.367$	$r_s = 0.04$ $p = 0.703$
Spot	$r_s = 0.41$ $p < 0.001$	$r_s = 0.56$ $p < 0.001$	$r_s = 0.64$ $p < 0.001$	$r_s = 0.55$ $p < 0.001$	$r_s = 0.64$ $p < 0.001$	<b>Spot</b>	$r_s = 0.95$ $p < 0.001$	$r_s = 0.55$ $p < 0.001$	$r_s = 0.12$ $p = 0.216$	$r_s = 0.16$ $p = 0.096$	$r_s = 0.24$ $p = 0.014$
Total Exam	$r_s = 0.42$ $p < 0.001$	$r_s = 0.57$ $p < 0.001$	$r_s = 0.81$ $p < 0.001$	$r_s = 0.66$ $p < 0.001$	$r_s = 0.72$ $p < 0.001$	$r_s = 0.95$ $p < 0.001$	<b>Total Exam</b>	$r_s = 0.62$ $p < 0.001$	$r_s = 0.12$ $p = 0.232$	$r_s = 0.13$ $p = 0.197$	$r_s = 0.24$ $p = 0.012$
AB2	$r_s = 0.49$ $p < 0.001$	$r_s = 0.54$ $p < 0.001$	$r_s = 0.58$ $p < 0.001$	$r_s = 0.47$ $p < 0.001$	$r_s = 0.47$ $p < 0.001$	$r_s = 0.65$ $p < 0.001$	$r_s = 0.68$ $p < 0.001$	<b>AB2</b>	$r_s = 0.15$ $p = 0.115$	$r_s = -0.01$ $p = 0.929$	$r_s = 0.13$ $p = 0.186$
CRT Pre	$r_s = 0.16$ $p = 0.126$	$r_s = 0.16$ $p = 0.113$	$r_s = 0.08$ $p = 0.408$	$r_s = 0.16$ $p = 0.124$	$r_s = 0.04$ $p = 0.690$	$r_s = 0.10$ $p = 0.309$	$r_s = 0.12$ $p = 0.235$	$r_s = 0.07$ $p = 0.468$	<b>CRT Pre</b>	$r_s = 0.56$ $p < 0.001$	$r_s = 0.48$ $p < 0.001$
MRT Pre	$r_s = 0.04$ $p = 0.695$	$r_s = 0.16$ $p = 0.112$	$r_s = -0.01$ $p = 0.914$	$r_s = 0.04$ $p = 0.700$	$r_s = 0.06$ $p = 0.557$	$r_s = 0.10$ $p = 0.314$	$r_s = 0.08$ $p = 0.450$	$r_s = 0.06$ $p = 0.557$	$r_s = 0.48$ $p < 0.001$	<b>MRT Pre</b>	$r_s = 0.58$ $p < 0.001$
SDT Pre	$r_s = 0.17$ $p = 0.101$	$r_s = 0.15$ $p = 0.138$	$r_s = 0.12$ $p = 0.257$	$r_s = 0.24$ $p = 0.016$	$r_s = 0.17$ $p = 0.096$	$r_s = 0.20$ $p = 0.053$	$r_s = 0.20$ $p = 0.043$	$r_s = 0.11$ $p = 0.275$	$r_s = 0.44$ $p < 0.001$	$r_s = 0.53$ $p < 0.001$	<b>SDT Pre</b>

Table 5.2 Correlation matrix for exam parameters and spatial ability tests. The top half of the matrix above the grey diagonal shows 2D (UoE Vet 1) correlations and below the diagonal 3D (UoE Vet 2) correlations. Yellow box = significant at  $p < 0.05$ .

### 5.3.2 Between Teaching Method Comparison

Univariable analysis identified statistically significant differences between 2D (UoE Vet 1) and 3D (UoE Vet 2) for in-course ( $W = 6798$ ,  $p < 0.001$ ), oral ( $W = 3812.5$ , 95%:  $-0.1 - -0.05$ ,  $p < 0.001$ ), interp ( $W = 6377.5$ , 95%:  $2.65 \times 10^{-5} - 8.0 \times 10^{-2}$ ,  $p = 0.011$ ), MCQ ( $W = 4226$ , 95%:  $-1.3 \times 10^{-1} - -8.31 \times 10^{-6}$ ,  $p = 0.012$ ), and short questions ( $t_{200.46} = -2.96$ , 95%:  $-0.11 - -0.02$ ,  $p = 0.003$ ) (Figure 5.1 and Table 5.3).

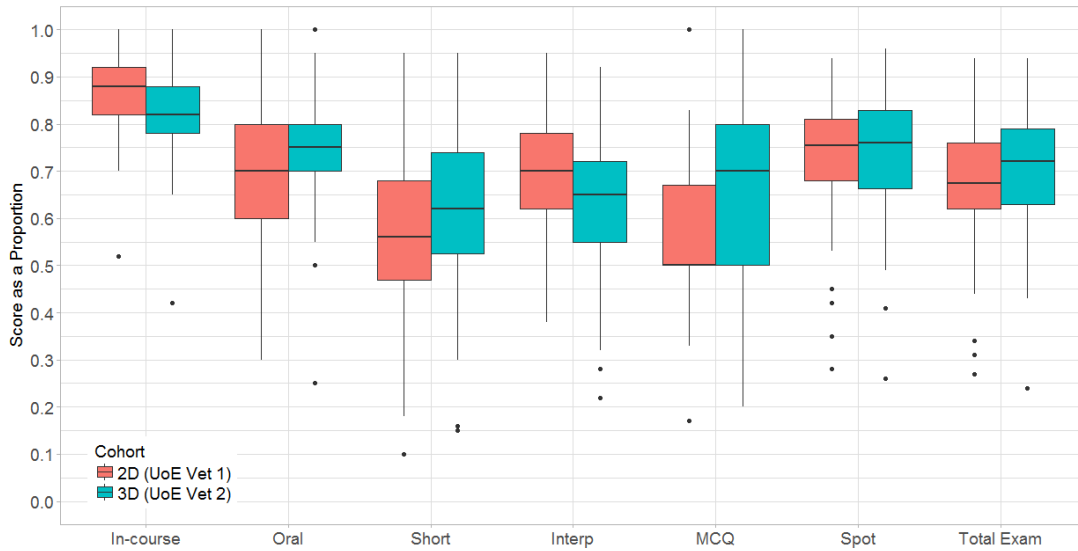
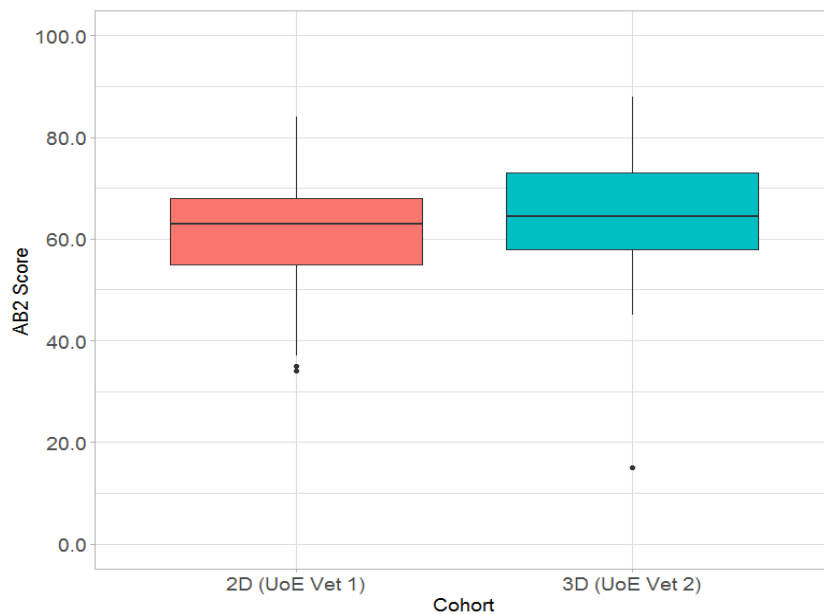


Figure 5.1 Comparison of exam parameter between teaching method/cohorts.

Parameter	2D (UoE Vet 1)					3D (UoE Vet 2)				
	$\bar{x}/M$	SD	Range	n	Max Score	$\bar{x}/M$	SD	Range	n	Max Score
In-Course	M = 0.88	-	0.52 – 1.0	108	20	M = 0.82	-	0.42 – 1.0	98	20
Oral	M = 0.70	-	0.3 – 1.0	108	10	M = 0.75	-	0.25 – 1.0	98	10
Short	$\bar{x} = 0.56$	0.16	-	108	60	$\bar{x} = 0.62$	0.16	-	98	40
Interp	M = 0.70	-	0.38 – 0.95	108	20	M = 0.65	-	0.22 – 0.92	98	20
MCQ	M = 0.50	-	0.17 – 1.0	108	6	M = 0.70	-	0.2 – 1.0	98	10
Spot	M = 0.76	-	0.28 – 0.94	108	130	M = 0.76	-	0.26 – 0.96	98	130
Total Exam	M = 0.68	-	0.27 – 0.94	108	216	M = 0.72	-	0.24 – 0.94	98	200
AB2	$\bar{x} = 61.75$	10.24	-	108	100	M = 64.5	-	15.0 – 88.0	98	100

**Table 5.3 Mean ( $\bar{x}$ ), median (M), range, and standard deviation (SD) for each exam parameter as a proportion and AB2 score.**

No statistically significant differences were identified for spot ( $p = 0.639$ ), total exam ( $p = 0.102$ ), and AB2 score comparisons ( $p = 0.053$ , Figure 5.2).



**Figure 5.2 Comparison of AB2 score between teaching method/cohort.**



For the separate  $GLM_B$  analyses (Table 5.4) the explanatory variables age, gender, and handedness were not statistically significant ( $p > 0.05$ ) for all exam parameters (in-course, oral, short, interp, MCQ, spot, and total exam).

Response Variable	Significant Explanatory Variables	OR	95% CI	p-value	Non-significant Explanatory Variables Removed
In-course	AB2	1.02	1.01 – 1.03	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, CRT Pre, Gender, Handedness, MRT Pre, Age
	SDT Pre	1.56	1.01 – 2.40	= 0.046	
	Cohort	Ref			
	2D (UoE Vet 1)	1.44	1.24 – 1.69	< 0.001	
Oral	AB2	1.04	1.03 – 1.05	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, CRT Pre, SDT Pre, Gender, MRT Pre, Handedness, Age
	Cohort	Ref			
	3D (UoE Vet 2)	1.26	1.07 – 1.47	= 0.005	
Short	AB2	1.04	1.03 – 1.05	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, CRT Pre, MRT Pre, Age, SDT Pre, Gender, Handedness
	Cohort	Ref			
	3D (UoE Vet 2)	1.18	1.01 – 1.37	= 0.040	
Interpretation (Interp)	AB2	1.03	1.02 – 1.03	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, Handedness, MRT Pre, Age, CRT Pre, Gender
	SDT Pre	1.94	1.29 – 2.90	= 0.002	
	Cohort	Ref			
	2D (UoE Vet 1)	1.35	1.17 – 1.56	< 0.001	
MCQ	AB2	1.04	1.03 – 1.05	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, Handedness, CRT Pre, Gender, MRT Pre, Age, SDT Pre, Cohort

Response Variable	Significant Explanatory Variables	OR	95% CI	p-value	Non-significant Explanatory Variables Removed
Spot	AB2	1.04	1.03 – 1.04	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, Age, MRT Pre, Handedness, Gender, Cohort, CRT Pre
	SDT Pre	1.58	1.11 – 2.23	= 0.011	
Total Exam	AB2	1.04	1.03 – 1.04	< 0.001	CRT Pre*Cohort, MRT Pre*Cohort, SDT Pre*Cohort, MRT Pre, Handedness, Age, Cohort, Gender, CRT Pre
	SDT Pre	1.49	1.10 – 2.02	= 0.011	

**Table 5.4 Output of GLM<sub>B</sub> analyses after model simplification by highest p-value. Ref = reference level for categorical explanatory variable.**

AB2 score was statistically significant ( $p < 0.001$ ) for all exam parameters (in-course, oral, short, interpretation, MCQ, spot, and total exam), which could be interpreted as performing well in one exam will mean performing well in another and so further confirms AB2 as a confounding variable. Therefore all Odds Ratios (OR) for this chapter are adjusted for AB2 score.

For in-course score, the 2D teaching method cohort performed statistically significantly better than the 3D teaching method cohort (OR = 1.44,  $p < 0.001$ ) (Table 5.4 and 5.3), although this is a small effect size with a difference between  $M = 0.88$  (2D) and  $M = 0.82$  (3D). The two spatial ability tests selected to test mental rotation in 2D and 3D were not statistically significantly predictive of the in-course score. However, the spatial ability sub-category of spatial visualisation as measured by the Surface Development Test (SDT) was statistically significantly predictive of the in-course score (OR = 1.56,  $p = 0.046$ ). Meaning for each 1% increase in SDT score the odds of scoring each 1% higher in the in-course was 1.56.

The 3D teaching method cohort performed statistically significantly higher on the oral exam compared to the 2D teaching method cohort (OR = 1.26,  $p = 0.005$ ), with a difference between  $M = 0.75$  and  $M = 0.70$  respectively (Table 5.3). An odds ratio of 1.26 is a small effect size and may or may not equate to a large educational impact on learning. Spatial ability as measured by the three tests (CRT, MRT, and SDT) was not statistically significantly predictive of the oral score. The in-course oral examination involved a composite mark of cell biology, histology, and gross anatomy so although included in this study, and students who had experienced the 3D spatial teaching method exhibited higher scores, this part of the assessment was not solely assessing gross anatomy knowledge. Therefore, the other components of this assessment, histology and cell biology, could compensate for lower scores in gross anatomy.

For the short exam parameter spatial ability as measured by the three tests was not statistically predictive of scores on short answer questions. The 3D teaching method cohort performed statistically significantly higher on the short questions compared to the 2D teaching method cohort (OR = 1.18,  $p = 0.040$ ).

For the interpretation score the 2D teaching method cohort performed statistically significantly higher compared to the 3D spatial teaching method cohort (OR = 1.35,  $p < 0.001$ ), with a difference of  $M = 0.70$  and  $M = 0.65$  for 2D and 3D respectively. The Surface Development Test (SDT) was statistically significantly predictive of interpretation score (OR = 1.94,  $p = 0.002$ ), while the spatial ability tests of mental rotation (CRT and MRT) were not statistically predictive.

Neither the 2D nor 3D teaching method was statistically significant for MCQ score, and spatial ability as measured by the Card Rotation Test, Mental Rotation Test, and Surface Development Test was not statistically significantly predictive of MCQ score (Table 5.3 and 5.4).

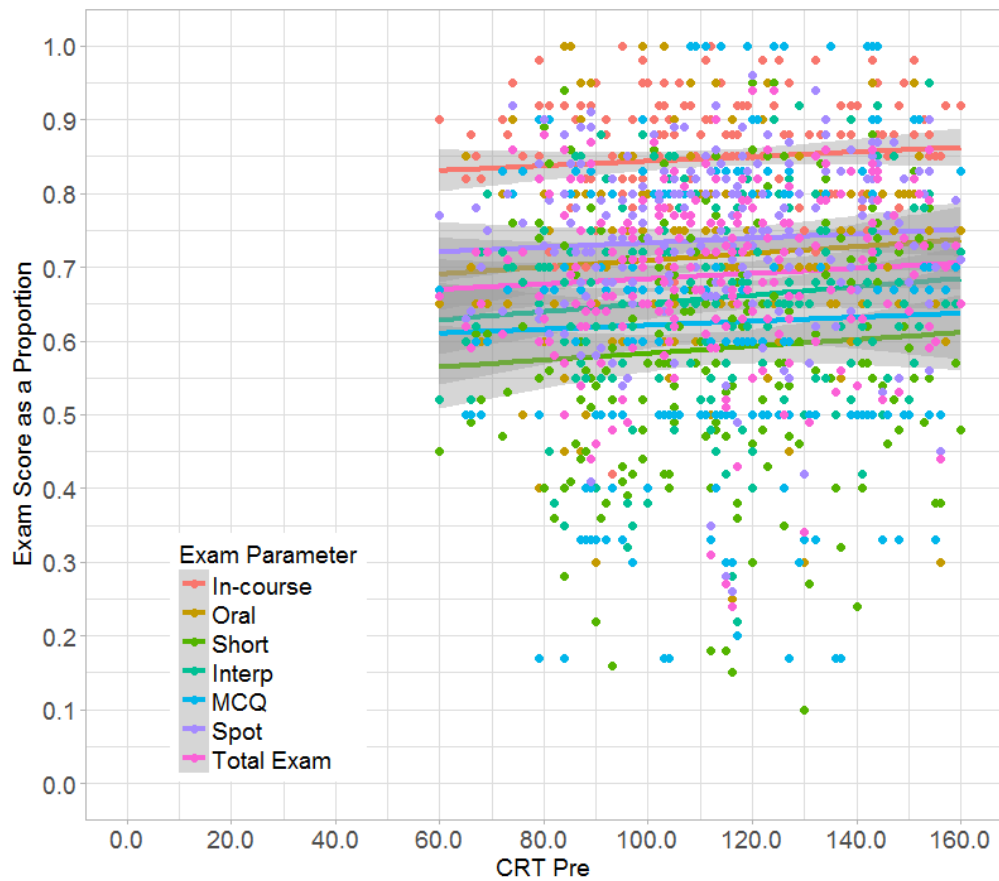
The Surface Development Test was predictive of the spot exam parameter (OR = 1.58,  $p = 0.011$ ) and the total exam parameter (OR = 1.49,  $p = 0.011$ ). Neither the Card Rotation Test nor the Mental Rotation Test was predictive of spot or total exam score. Additionally there was no statistically significant difference between the 2D or 3D teaching method for spot ( $M = 0.76$  vs  $M = 0.76$ , respectively) or total exam scores ( $M = 0.68$  vs  $M = 0.72$ , respectively).

To summarise the results of the  $GLM_B$  analysis of each exam parameter. The demographic parameters age, gender, and handedness were not statistically significant for all exam parameters. The 2D teaching method cohort performed statistically significantly higher on the in-course and interpretation, while the 3D teaching method cohort performed statistically significantly higher for the oral and short (Figure 5.1). There was no statistically significant difference in exam parameter scores between teaching methods for MCQ, spot or total exam scores.

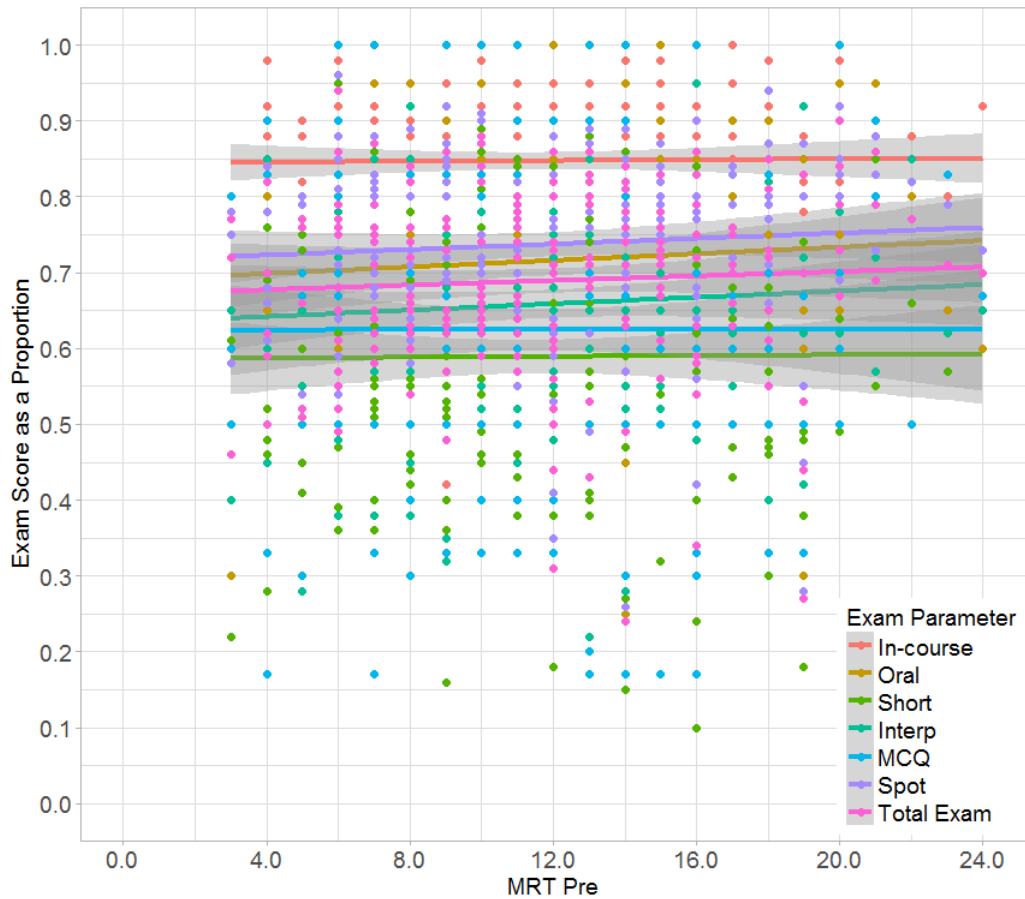
The odds ratio represents the effect size of the difference between the cohorts and is the ratio of the odds of exposure, to the odds of no exposure. Thus an odds ratio of 1 indicates no difference. Odds ratios greater than 1 indicate the exposure is associated with a higher odds of the outcome, while an odds ratio less than 1 indicates the exposure is associated with a lower odds of the outcome.

All of the odds ratios for the difference between the 2D and 3D teaching methods presented above statistically represent a small effect size (Table 5.4). Comparisons of the difference in median (or mean for short questions) do not show a large difference although this was shown to be statistically significant. The statistically significant small effect sizes may or may not have a large educational impact on the students' learning (see discussion section 5.4 below).

The spatial ability sub-category of mental rotation as measured by the CRT and MRT was not statistically significantly predictive for any of the exam parameters (Figure 5.3 and 5.4).

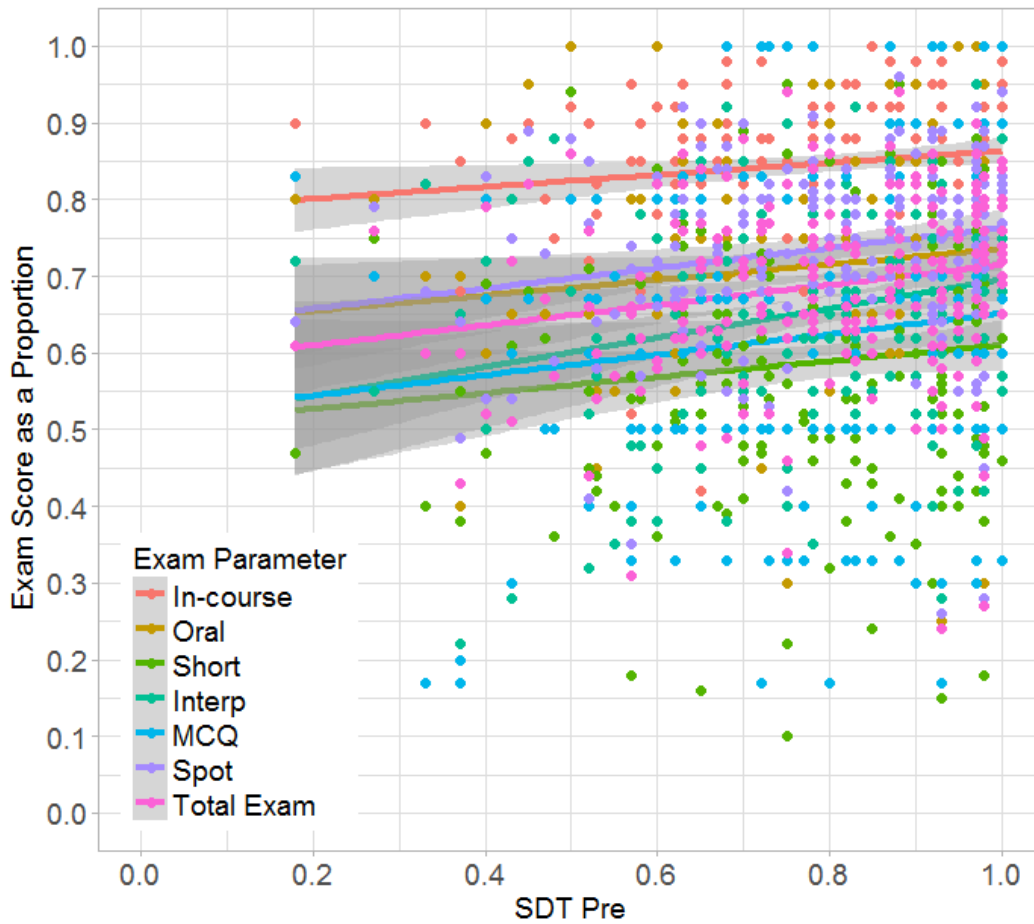


*Figure 5.3 Exam parameter score as a function of CRT Pre score, regression line with 95% confidence interval included.*



*Figure 5.4 Exam parameter score as a function of MRT Pre score, regression line with 95% confidence interval included.*

The SDT was statistically significant for predicting in-course, interpretation, spot, and total exam. However, the effect sizes for SDT are small in size and the majority of students scored highly for the SDT indicating there is less variance in scores, meaning any large differences in SDT scores would more likely be statistically significant. Therefore, this needs to be appreciated when interpreting the SDT effect size.



*Figure 5.5 Exam parameter score as a function of SDT Pre score, regression line with 95% confidence interval included.*

## 5.4 Discussion

The purpose of the study presented in this chapter was first to determine whether a 3D spatial teaching method improved end-of-course anatomy examination scores compared to a 2D teaching method. The second aim of this study was to determine whether spatial ability as measured by three tests (CRT, MRT, and SDT) was related to end-of-course anatomy examination scores.

### **5.4.1 Does teaching anatomy using a traditional 2D method or 3D spatial method improve students' anatomy knowledge and understanding?**

The main aim of the 3D spatial teaching method in this study was to teach anatomy spatially. Previous 3D anatomy teaching methods have utilised 3D technology such as 3D computer models, 3D printed models, augmented reality, and diagnostic imaging techniques (Garg *et al.*, 1999b; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Küçük, Kapakin & Göktaş, 2016; Plumley *et al.*, 2013; Smith *et al.*, 2018). It would be expected given the 3D nature of anatomy (section 1.3.1) a 3D teaching approach would improve anatomy knowledge and understanding, and this theory has been confirmed by a meta-analysis that found 3D teaching methods to improve anatomy knowledge (Yamine & Violato, 2015).

In the study of this chapter, a 3D spatial teaching method compared to a 2D teaching method identified no statistically significant differences in examination scores for MCQs, 'spot' type questions, and the total examination score (composed of the MCQ, interpretation, short, and spot scores combined). The main source of 3D teaching in this study was the use of learner-controlled 3D computer models, with a smaller component of 3D printed resources and other techniques outlined below. Other studies have also found no difference between the use of 3D computer models compared to 2D teaching methods on anatomy examination scores (Berney *et al.*, 2015; Tan *et al.*, 2012; Garg *et al.*, 1999b). An hypothesised reason for this is the 3D teaching methods implemented were not truly 3D (Tan *et al.*, 2012).

The 3D spatial teaching method implemented in this study employed learner-controlled 3D computer models, 3D printed models, diagnostic images, and tutorials



designed with the aim of presenting anatomy in 3D and spatially, and is arguably a 3D method. Another relevant factor to consider when exploring differences in results and experiences using 3D and 2D teaching methods is cognitive load theory (Ayres & Paas, 2012; van Merriënboer & Ayres, 2005). Cognitive load theory research is grounded in psychology research on memory - particularly working and long-term memory along with their interactions. Cognitive load specifically refers to working memory load, i.e. how much information the brain can handle at once.

Poor instructional design (such as complicated 3D technology) or dealing with complex material (such as anatomy) can increase the working memory load meaning there are fewer cognitive resources to be devoted to learning. The cognitive load of a learner can be divided into three parts: intrinsic cognitive load (the inherent nature of the learning task), extraneous cognitive load (the way in which the task(s) are presented), and germane cognitive load (the cognitive resources that are relevant to that task) (van Merriënboer & Ayres, 2005). Therefore, if the 3D technology or spatial teaching methods are not intuitive, the extraneous cognitive load will increase. Learners will then use the majority of valuable working memory purely on working out how to use the technology rather than for learning. Interestingly, Nguyen, Nelson, and Wilson (2012:p.106) state their 3D computer model manipulation "... was not intuitive, and as such it is possible that merely operating it produced additional cognitive demands on interactive participants." The authors of the study found a 3D computer model of a non-anatomical object (a cube) improved anatomy knowledge for low spatial ability participants (Nguyen, Nelson & Wilson, 2012). Thus it may be worth considering whether the use of non-anatomical objects may be helpful in supporting the development of spatial ability and in turn improving anatomical understanding (Hoyek *et al.*, 2009).

Cognitive mediated interfaces such as using a mouse to manipulate a 3D computer model are thought to increase extraneous cognitive load. In contrast, perceptually mediated interfaces using systems that are more 'automatic' to humans (such as

using your hand to manipulate an object), are thought to reduce extraneous cognitive load. Barrett and Hegarty (2016) investigated Aptitude-Treatment Interactions (APIs), i.e. the match between an individual's aptitude and the treatment they received i.e. the location of the hand-held device to manipulate. The authors found co-locating the manipulation device with the visual display was better for performing tasks quicker. Their study showed that a perceptual-mediated setup was better for students with low spatial ability and made no difference to those with high spatial ability.

So far limited studies are researching the cognitive load associated with the use of 3D anatomical education techniques. This present study did not explore cognitive load and would be an interesting aspect to explore in future research studies. It should be noted however that students experiencing the 3D spatial teaching method in this study did not identify difficulty with model use (as identified through focus group discussions presented in chapter 8). Küçük, Kapakin, and Göktaş (2016), when researching the use of augmented reality in the context of anatomy learning, subjectively investigated the cognitive load of students using AR technology using a 9-point Likert scale questionnaire. They found the AR group to have a lower cognitive load than the control non-AR group.

Another reason for lack of improvement of examination results for the 3D spatial teaching method implemented in this chapter could be because of a limitation in the study design. To explain, the AB1 gross anatomy course included fourteen sections in total (head, joints, forelimb, neck, pharynx, body wall, vertebrae, thorax, abdomen, hindlimb, pelvis, ultrasound, nervous system, and lymphatics), all of which were examined. For the 3D spatial teaching method, only six of these sections were taught with this method and the remaining eight sections, taught by other lecturers, were taught traditionally in 2D. In contrast, for the 2D teaching method, all fourteen sections were taught 2D. Meaning, on a comparison of the two teaching methods, all fourteen sections were taught with a 2D teaching method, while six out

of the fourteen sections were taught with a 3D spatial teaching method. It is possible, therefore, the 3D spatial method may have had a greater impact (at least regarding examination results) if it had been used across the entire fourteen sections of course content. To overcome this, analysis specifically comparing the exam questions involving only the six sections could be compared to investigate further whether the 3D spatial method improved anatomy knowledge specifically for these sections.

The 3D spatial teaching method described in this chapter, in addition to using 3D technology, also took a new approach by aiming to ‘tap’ into the analytical strategy and problem-solving aspect of spatial ability (Lohman, 1996 and section 2.4 of this thesis). In comparison, previous studies have focussed on teaching methods involving 3D representations of anatomical structures as previously discussed (Tan *et al.*, 2012; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999b, 1999a; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Smith *et al.*, 2018). The examination questions in the end-of-course examination that are aimed at assessing problem-solving and integration of knowledge are the interpretation questions (AB1 Exam Board Chair, personal communication). It may, therefore, be expected the students who experienced the 3D spatial teaching method would exhibit significantly higher scores for these questions.

However, in this study, it was found the 2D teaching method improved anatomy knowledge for interpretation questions. A possible reason for this could be because the 3D spatial teaching method was not potent enough as was given across only six sections, and/or because of differences in academic ability between the two cohorts. However, comparison of AB2 examination results, that were from a completely different unrelated part of the course were found to have no statistically significant differences between cohorts ( $p = 0.053$ ), suggesting a similar academic ability. Additionally, AB2 score was included as a covariate in the statistical analysis as a way to account for academic differences. Another way of accounting for possible academic differences in future research would be to include other cognitive ability

tests to establish and compare the general intelligence of different cohorts (Hegarty *et al.*, 2009; Gutierrez *et al.*, 2017). This could also investigate the link that different cognitive abilities could have to the learning of anatomy.

In this study, the students experiencing the 3D spatial teaching method exhibited statistically significantly higher scores for short answer questions. For the R(D)SVS the short answer questions assess core knowledge based directly on the lecture learning outcomes. Core knowledge of a subject is critical for long term memory along with the integration of knowledge to problem solve (Miller, 1990; Bloom, 1984), and with the 3D spatial teaching method providing evidence for improvement in short answer scores, this teaching method provides the initial promising potential to improve core knowledge.

Another source of variance in this study is the difficulty of the end-of-course examinations. This study was conducted over two academic years, and the examination questions received by each cohort were different. As with all Higher Education Institutions, the R(D)SVS examinations are critiqued and peer reviewed by internal and external examiners to ensure standardisation across academic years. Nevertheless, there will inevitably be variability between examination papers. The Hofstee standard-setting approach is used at the R(D)SVS to address this to a certain extent but only with the parameters of 48-52% as the pass mark.

In the psychology education literature, the inclusion of 3D technology that uses visual and verbal representations is defined by Mayer as ‘multimedia learning’ (Mayer, 2001). Multimedia learning is described as learning from words (spoken or printed), and pictures which can either be static (photographs, figures, graphs, and drawings) or dynamic (animations or videos).

Mayer provides strong evidence from peer-reviewed research linking multimedia learning to human cognitive theory (i.e., cognitive processes and abilities). Through this Mayer has proposed eight guiding principles of multimedia learning based on cognitive theory; multimedia principle, contiguity principle, coherence principle, modality principle, redundancy principle, personalisation principle, interactivity principle, and signalling principle (Mayer, 2002).

Interestingly Mayer's (2002) explanation of the cognitive processes learners use to understand new concepts, utilising multimedia learning, includes visuospatial thinking or spatial ability. Mayer explains this ability is used to integrate and mentally manipulate the two inputs (visual and verbal) to thus construct knowledge and learn. Other cognitive processes and abilities are used such as memory, both working, and long term.

While investigating the contiguity principle theory (i.e., where deeper learning is achieved when both the verbal and visual representations are delivered together) Mayer and Sims (1994) also researched the relationship to an individual's spatial ability. They hypothesised two effects: the ability-as-compensator effect and the ability-as-enhancer effect. The ability-as-compensator effect hypothesises that when poor instructions are given to learners' of high spatial ability they will compensate for this by having the ability to maintain an image in working memory, while problem-solving and constructing knowledge. In contrast, low spatial ability learners' may struggle to maintain 3D images in working memory while deciphering poor instructions at the same time. The ability-as-enhancer effect hypothesises when instructions are good, learners' spatial ability will enhance learning, particularly for high spatial ability learners' as less cognitive resources will be required for deciphering instructions.

In Mayer and Sims study, the contiguity principle (Mayer, 2002) was confirmed and was dependent on the spatial ability of the students providing strong evidence for the ability-as-enhancer hypothesis. They demonstrated, using two controlled experiments, that when visual and verbal presentations were given simultaneously (i.e., good instructions) compared to successively (i.e., poor instructions), learners of high spatial ability had higher scores than low spatial ability learners on transfer of knowledge questions (i.e., application of knowledge questions).

However, dichotomising of data into high and low spatial ability can give false positive results (Senn, 2003; Altman, 2006). So, when the authors analysed the data in a non-dichotomised way, spatial ability was found to be significant for one experiment but not for the other. Additionally, the authors found no interaction between spatial ability and presentation of visual and verbal material when data were not dichotomised, meaning spatial ability was not a significant factor. Investigations into the effect of the contiguity principle in relation to anatomy teaching methods could help explain the use of such resources for teaching.

#### **5.4.2 Is spatial ability a predictor of success in anatomy examinations?**

The second aim of the study presented in this chapter was to determine the relationship of spatial ability to anatomy examinations to understand the relationship between spatial ability and anatomy. The study presented in this chapter found spatial ability, measured by the sub-categories of 2D mental rotation (CRT) and 3D mental rotation (MRT), to be unrelated to anatomy knowledge as measured by traditional end-of-course and in-course examinations. Spatial ability as measured for the sub-category of spatial visualisation (SDT) was significant for four exam parameters including in-course workbooks, interpretation, spot, and the total exam.

The mental rotation test (MRT) has been used in the vast majority of anatomy education research on spatial ability (Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999b, 1999a; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Cui *et al.*, 2016; Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a; Langlois *et al.*, 2009). This spatial ability test has been shown to be related to anatomy examination scores and as a predictor of performance (Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a), with one study finding high spatial ability students to be 2.2 times more likely to score above 90% on anatomy examinations (Lufler *et al.*, 2012). However, other studies have shown no statistical relationship between scores on the MRT and anatomy examination scores (Hegarty *et al.*, 2009; Hoyek *et al.*, 2009). The study reported in this chapter found no statistically significant relationship between scores on the MRT and anatomy examination scores, or on another spatial ability test on the sub-category of mental rotation (Card Rotation Test).

In this study, spatial ability test scores were not categorised into low or high spatial ability or any other division as has been implemented in some previous studies (Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a; Nguyen, Nelson & Wilson, 2012). One study which categorised students into four groups of spatial ability, found statistically significant relationships between scores on the MRT and anatomy examination scores (Lufler *et al.*, 2012). Categorising spatial ability scores increases the chances of finding a statistically significant difference (Altman, 2006; Senn, 2003). Similarly, a study comparing two sets of scores on the MRT between medical students and educational science students found the medical student group to have the largest improvement in scores. These results led the researchers to infer that spatial ability was used to learn anatomy (Vorstenbosch *et al.*, 2013a). However, how much of the improvement was due to the practice of the test, i.e., the 'retest effect', is unknown as no means to control for this were accounted for (Goldberg *et al.*, 2015; McCaffrey & Westervelt, 1995; Lezak *et al.*, 2012:chap.5).

In the study presented in this chapter, the spatial ability sub-category of spatial visualisation (as measured by the Surface Development Test) was statistically significantly predictive of examination scores for in-course, interpretation, spot, and the total examination. Tests of spatial visualisation have previously not been investigated in relation to anatomy knowledge and understanding (section 1.4). Tests of spatial visualisation load highly onto the general intelligence factor *g* (Lohman, 1988 and as discussed in section 2.4). Therefore the finding of spatial visualisation significantly related to anatomy examination scores may indicate the use of general intelligence to answer anatomy examination questions, which would be expected.

Spatial ability is one of the numerous correlated sub-factors of general intelligence (Spearman, 1904; Carroll, 1993). Therefore, it is difficult, or impossible, to measure just that one single ability despite best efforts. Similarly, assessments are usually designed to assess at different levels of Blooms taxonomy in order to ensure not just requisite knowledge but an ability to utilise and interact with that knowledge and problem solve appropriately to the context. Thus, sitting examinations uses a range of cognitive abilities meaning spatial ability could be related to any examination from any subject, not specifically anatomy (as supported by the finding in this study that a test of spatial visualisation which, loads highly onto *g*, was significantly related to examination scores). Therefore, other cognitive abilities could be more prominently utilised when sitting examinations, such as verbal comprehension, memory, or processing-speed as also mentioned in the above section (section 5.4.1).

To isolate whether spatial ability specifically is used, future research studies should be designed to investigate whether spatial ability is involved during the learning process as opposed to answering anatomy examination questions. This could be achieved by using the componential approach used by psychologists to breakdown and isolate the cognitive processes used while learning anatomy (Hegarty & Waller, 2005). Furthermore, research investigating the relationship between anatomy learning and other cognitive abilities could further identify what cognitive abilities



are used to learn anatomy and teaching methods could be aimed at utilising these abilities.

Despite studies finding a relationship between spatial ability and anatomy learning, and studies finding no relationship (Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a; Hegarty *et al.*, 2009; Hoyek *et al.*, 2009), there is general agreement amongst anatomy educators of the importance of spatial ability when learning anatomy (Berney *et al.*, 2015; Nguyen *et al.*, 2013; Lufler *et al.*, 2012; Keedy *et al.*, 2011). Along with discussions on the importance of spatial ability to anatomy learning in peer-reviewed publications, the qualitative aspect of the relationship has been reported in qualitative studies (Nguyen *et al.*, 2013; Preece *et al.*, 2013, and chapter 8 of this thesis).

The next chapter follows on from these results to evaluate a spatial MCQ designed to assess anatomy knowledge both non-spatially and spatially by comparing to three spatial ability test scores. The study also explores whether teaching anatomy with a 2D or a 3D emphasis improves scores on the spatially designed MCQ.

# Chapter 6 Assessing Non-spatial and Spatial Anatomy Knowledge and Understanding

## 6.1 Introduction to Chapter 6

Leading on from chapter 5 which presented and discussed the effect of a two-dimensional (2D) or a three-dimensional (3D) spatial teaching method on anatomy knowledge and understanding, chapter 6 focusses on spatial anatomy assessments. Many studies investigating 3D anatomy teaching resources have additionally designed some form of spatial anatomy assessment as a means to investigate whether spatial anatomy knowledge improved as a result of the teaching approach (Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001; Keedy *et al.*, 2011; Schubert, Schnabel & Winkelmann, 2009; Guillot *et al.*, 2006; Berney *et al.*, 2015; Tan *et al.*, 2012; Provo, Lamar & Newby, 2002; Rochford, 1985; Hegarty *et al.*, 2009; Langlois *et al.*, 2017) (see also section 1.3.5).

Chapter 6 explores spatial multiple choice question (MCQ) assessments by evaluating student performances on a new spatial MCQ test (Appendix 7). The spatial MCQ was designed as part of this research as a means to test spatial anatomy knowledge and understanding (henceforth referred to as the “spatial MCQ”). This chapter expands on previous studies and the area of spatial anatomy assessment by specifically looking at the design of an MCQ spatial anatomy assessment, as to date the relationship of non-spatial questions to spatial ability tests is unknown for MCQs (Langlois *et al.*, 2017).

The study presented in this chapter aims to address the following:

- Whether a spatial MCQ can be designed,
- Whether teaching via the 2D and 3D methods described earlier influences students' performance on the spatial MCQ,
- Whether performance on the spatial MCQ can predict students' end-of-course anatomy examination scores.

An additional aim was to explore the validity and reliability of the spatial MCQ by post-examination item analysis. Within the literature, on spatial anatomy knowledge, there are discussions on whether anatomical knowledge can be categorised into spatial or non-spatial knowledge (Yammine & Violato, 2015; Nguyen, Nelson & Wilson, 2012). The difficulty in the distinction between the two categories could be due to the 3D nature of anatomy, and how learners could think and problem-solve spatial anatomy knowledge. For instance, students may access the spatial anatomy knowledge by memory, while others may problem solve using spatial ability and intelligence to 'figure it out', or a combination of these approaches may be used (as discussed in section 1.3.5). This complicated relationship could explain why spatial anatomy assessments are challenging to design.

Various approaches have been used to design MCQs to examine spatial anatomy knowledge, such as, cross-sections (of teeth, and of the aorta, trachea and oesophagus) (Hegarty *et al.*, 2009; Nguyen, Nelson & Wilson, 2012), using rotated views of carpal bones (Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001), changing the format of a traditional 'spot' exam (Schubert, Schnabel & Winkelmann, 2009), the use of Likert scale responses to ascertain whether radiographic anatomy experts considered a question to be spatial or not (Keedy *et al.*, 2011), and the use of a cognitive task analysis to examine the thought processes used when answering spatial anatomy questions (Berney *et al.*, 2015).

Studies have also used a range of approaches to attempt to design spatial questions, e.g. rotation of complete structures, identification of features on rotated structures, asking what level a cross-section image was taken at, and to identify the order of movement for a structure (Berney *et al.*, 2015; Hegarty *et al.*, 2009; Nguyen, Nelson & Wilson, 2012; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999b, 1999a).

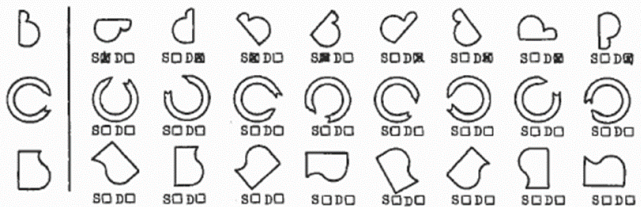
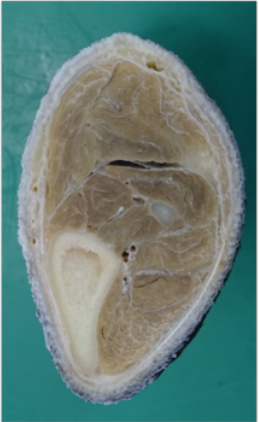
Studies have primarily used the Mental Rotation Test (MRT) to assess if questions are spatially demanding. These studies have attempted to identify if there is a relationship between scores on the MRT and scores on the spatial anatomy questions (Berney *et al.*, 2015; Garg *et al.*, 1999b, 1999a; Garg, Norman & Sperotable, 2001; Guillot *et al.*, 2006; Hegarty *et al.*, 2009; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Tan *et al.*, 2012). So far MCQs examining non-spatial anatomy knowledge have had limited design, and their relationship to tests of spatial ability is unclear (Langlois *et al.*, 2017). The study presented in this chapter provides an analysis of student scores on a novel spatial MCQ assessment designed to examine anatomy knowledge non-spatially and spatially. The approach taken for the design of the novel spatial MCQ in this study makes specific reference to the Card Rotation Test, the Mental Rotation Test, and the Surface Development Test designs.

## **6.2 Method of Spatial MCQ Design**

In order to combine the assessment of spatial ability and anatomy an MCQ assessment to include both non-spatial factual anatomy questions and questions with a spatial focus were designed. The rationale was to generate a measure of spatial ability in the specific context of veterinary anatomy. This MCQ replaced the third in-course anatomy workbook which, had been required for cohorts before this research study.

As noted by other authors (Nguyen, Nelson & Wilson, 2012; Yammine & Violato, 2015) developing the Spatial MCQ test was challenging as determining which anatomy questions are specifically spatially demanding can be open to debate. This is because anatomy is inherently a spatial subject as extensively discussed in the Literature Review of chapter 1.

Initially, when creating questions, the author considered non-spatial questions first, for example, “What is the action of the quadriceps femoris muscle?” However, on reflection, it was felt that this type of question could involve spatial ability. For the example mentioned above, imagining the object (muscle) and then rotating or moving the object could be required to answer the question (figuring out the action). To overcome this, the author analysed the lecture content to be assessed (hindlimb, pelvis, and ultrasound) and selected anatomy facts that were covered in lectures, for example, “What type of joint is the hip joint?” The author felt this type of question relied on recalling information of the different synovial joint types more than spatially processing information. The spatial anatomy questions were developed by the author and were developed based on the spatial ability tests used in this study. Four types of spatial anatomy questions were devised; 1. Identification of the anatomical aspect of a cross-section, 2. Identification of the anatomical aspect of a whole bone, 3. Identification of a soft tissue structure on cross-section, and 4. Identification of ‘what level’ a cross section was taken at. Examples of each question type are shown in Table 6.1 along with the corresponding spatial ability tests their design was based.

Spatial ability test	Example of Spatial ability test question	Example of equivalent Spatial MCQ
Card Rotation Test (CRT)	 <p>Copyright © 1976 Educational Testing Service.  www.ets.org. Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.</p>	<p>The photo below is a transverse section through a canine hindlimb. Which side is medial?</p> <p>A B C D</p> 

Spatial ability test	Example of Spatial ability test question	Example of equivalent Spatial MCQ
<p>Mental Rotation Test (MRT)</p>	<div data-bbox="577 470 1211 592" style="text-align: center;"> </div> <p data-bbox="577 635 1151 663">Copyright permission granted by Peters et al. (1995).</p>	<p data-bbox="1272 395 1749 416">Identify the leg and orientation shown in the photograph below.</p> <ul data-bbox="1317 424 1518 512" style="list-style-type: none"> <li>a) Left leg, caudal view</li> <li>b) Left leg, cranial view</li> <li>c) Right leg, caudal view</li> <li>d) Right leg, cranial view</li> </ul> <div data-bbox="1547 419 1787 1070" style="text-align: center;"> </div> <p data-bbox="1261 1086 1738 1102">The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.</p> <ul data-bbox="1294 1107 1496 1166" style="list-style-type: none"> <li>a) Cranialis tibialis</li> <li>b) Deep digital flexor, lateral head</li> <li>c) Deep digital flexor, medial head</li> <li>d) Vastus intermedeus</li> </ul> <div data-bbox="1263 1171 1608 1394" style="text-align: center;"> </div>

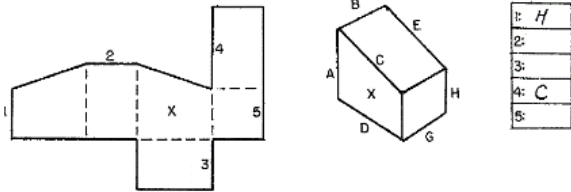
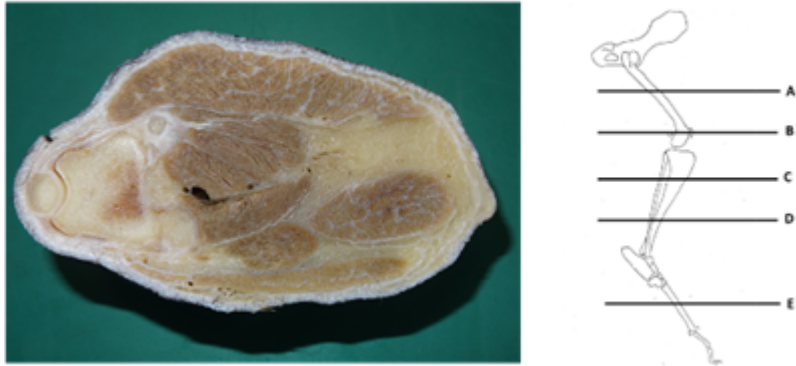
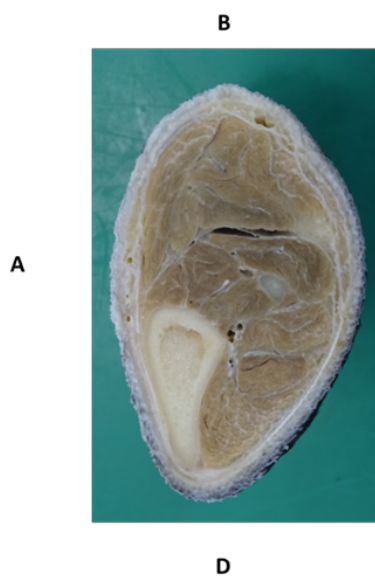
Spatial ability test	Example of Spatial ability test question	Example of equivalent Spatial MCQ
Surface Development Test (SDT)	 <p>Copyright © 1976 Educational Testing Service.  www.ets.org. Practice questions from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.</p>	<p>At which level of the limb is this cross section taken at?</p> 

Table 6.1 Example spatial MCQ questions based on each spatial ability test.

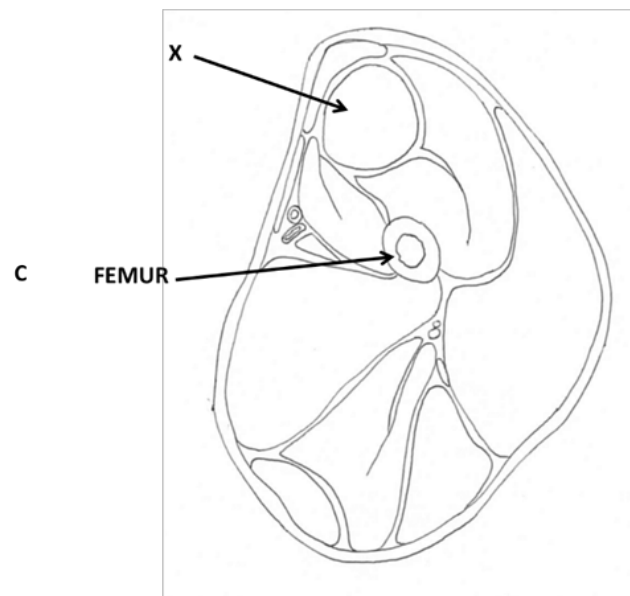


For initial quality assurance purposes before post-examination item analysis, the author and supervisors of this research met to discuss the author's design of the spatial anatomy questions, the developed questions were then critiqued by a group of senior anatomists and educationalists in the R(D)SVS. This group met on three separate occasions to critique the questions regarding anatomy content, to consider whether the questions were fair, and to ensure the questions adhered to school MCQ writing guidelines.

The MCQ contained photographs of specimens (Figure 6.1) and diagrammatic representations of anatomy as seen in textbooks (Figure 6.2).



*Figure 6.1 Example of spatial MCQ containing real photograph of canine hindlimb cross section.*



*Figure 6.2 Example spatial MCQ containing diagrammatic representation of canine hindlimb cross section.*

The MCQ critique group decided that if the question was asking the students to identify a muscle or structure, then a diagrammatic representation should be used as

this reduced ambiguity; whereas if the question was not asking for identification of a specific structure then a photograph of the real structure could be used.

Formalin-fixed and frozen right and left canine hindlimbs (estimated to be from a Staffordshire Bull Terrier breed) were cut by a band saw to provide the desired cross-sections. These were then photographed with a Sony Cyber-shot DSC HX60 Camera. The cross-sections were taken at points along the limb to match two cross-sectional diagrams in one of the recommended course textbooks (Evans and de Lahunta (2010) *Guide to the Dissection of the Dog*. Seventh edition. Elsevier, Missouri), with additional sections taken at salient points along the limb.

The Spatial MCQ was designed with a total of thirty questions, equally divided between non-spatial ( $n = 15$ ) and spatial anatomy questions ( $n = 15$ ). A total time of 60 minutes was set for the exam (i.e., 2 minutes per question). Originally, the spatial MCQ was intended to be computer-based however it was decided the MCQ exam should be given on paper, as the spatial ability tests were paper-based. Students requiring extra time because of academic adjustments received this as normal.

For Cohort University of Edinburgh Veterinary Students 1 (UoE Vet 1) who received the 2D teaching method, it was noticed after the exam that the diagram in question 29 was incorrect (the arrow had changed position at printing). This question was therefore excluded from the statistical analysis. Question 29 was corrected for the exam for Cohort University of Edinburgh Veterinary Students 2 (UoE Vet 2) who received the 3D spatial teaching method but was excluded from the statistical analysis for this thesis. The Spatial MCQ was included in the students' final in-course assessment mark meaning that students could have the same exam driven incentive to perform well in the Spatial MCQ as with any other in-course assessment.

## 6.3 Data Analysis

### 6.3.1 Descriptive Statistics

Shapiro-Wilk normality tests were performed and histograms plotted for the non-spatial questions (nSNQ,  $n = 15$ ) and the spatial questions (SNQ,  $n = 14$ ). The ‘difference-in’ score ( $\Delta_{\text{nSNQ-SNQ}}$ ), the nSNQ minus SNQ, was calculated. The ‘difference-in’ score was calculated as a proportion because the total number of nSNQ was different to the total number of SNQ, as question 29 was removed from the spatial MCQ due to a change in question format at printing. The nSNQ, SNQ, and  $\Delta_{\text{nSNQ-SNQ}}$  scores were checked for normality for each teaching method/cohort.

Descriptive statistical analysis was performed including the mean (or median) with SD (or range) and correlations. Spearman Rank correlations were performed between each category of MCQ question (nSNQ or SNQ) and each of the three spatial ability tests (CRT Pre, MRT Pre, and SDT Pre). To compare the correlations between the nSNQ and SNQ to assess if the correlations were different, a dependent correlation comparison was performed (i.e., to compare the CRT Pre  $\alpha$  nSNQ correlation to the CRT Pre  $\alpha$  SNQ correlation a dependent correlation comparison was performed). This comparison of correlations was performed for each spatial ability test (CRT, MRT, and SDT) and each cohort (section 3.9.1).

### 6.3.2 Between Teaching Method Comparison on Spatial MCQ

Initially, SNQ and nSNQ variables were treated separately in the statistical analysis with two separate  $GLM_p$  analyses (section 3.9.3). The first  $GLM_p$  model had ‘nSNQ’ as the response variable and a second model had ‘SNQ’ as the response variable. Each model had the categorical explanatory variables teaching method/cohort, gender, and handedness, and the continuous explanatory variables, age, AB2 score,

and spatial ability (CRT, MRT, and SDT). Additionally, the interactions CRT Pre\*Cohort, MRT Pre\*Cohort, and SDT Pre\*Cohort were added to the model. The explanatory variable AB2 score was identified as a confounding variable and was not removed during stepwise simplification. These separate  $GLM_p$  analyses were conducted to compare if there was a difference between the two teaching methods (2D or 3D spatial) for the nSNQ and SNQ scores. Table 6.2 summarises the two  $GLM_p$  models.

Spatial MCQ Parameter	Analysis	Response variable	Explanatory variable		Interactions
			Continuous	Categorical	
nSNQ	$GLM_p$	'nSNQ'	Age, CRT Pre, MRT Pre, SDT Pre, AB2	Teaching Method/Cohort, Gender, Handedness,	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort
SNQ	$GLM_p$	'SNQ'	Age, CRT Pre, MRT Pre, SDT Pre, AB2	Teaching Method/Cohort, Gender, Handedness	CRT Pre*Cohort MRT Pre*Cohort SDT Pre*Cohort

**Table 6.2 Summary of  $GLM_p$  analyses for between cohort comparisons on spatial MCQ in chapter 6.**

To test whether there was an interaction between teaching method/cohort and MCQ score, the  $\Delta_{nSNQ-SNQ}$  score was compared to zero by a 1-sample Wilcoxon Signed Rank test for each teaching method, and compared between the two teaching methods by a Wilcoxon Rank Sum test. A positive  $\Delta_{nSNQ-SNQ}$  score indicates the number of nSNQ answered correctly was higher than the SNQ, whereas a negative  $\Delta_{nSNQ-SNQ}$  score indicates the number of nSNQ answered was lower than the SNQ (Figure 6.3).

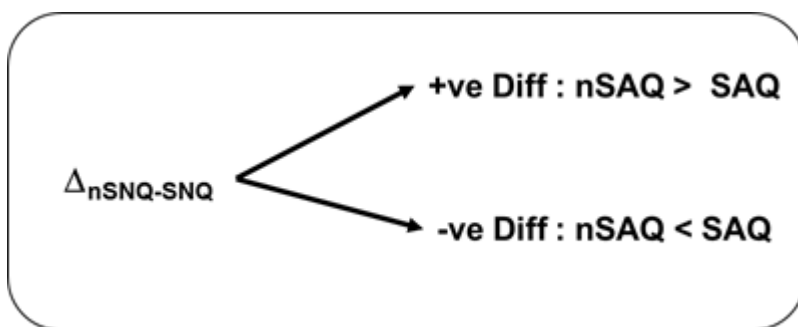


Figure 6.3 Diagrammatic representation of Diff score.

### 6.3.3 Spatial MCQ Score and Relationship to End-of-Course Examination Scores

To investigate whether a student’s score on the spatial MCQ was predictive of a student’s anatomy end-of-course examination result (as a proportion of the total score), separate  $GLM_B$  analyses (section 3.9.3) were performed with the response variable ‘Examination Parameter’ (either in-course, oral, short, interpretation, MCQ, spot, or total exam). In these  $GLM_B$  models the continuous explanatory variables ‘Total Spatial MCQ Score’ (SMS), age, CRT Pre, MRT Pre, SDT Pre score, AB2 score, and the categorical explanatory variables teaching method/cohort (2D and 3D), gender (male and female), and handedness (left and right) were added. The interactions SMS\*CRT Pre, SMS\*MRT Pre, SMS\*SDT Pre, and SMS\*Cohort were also added to each model (Table 6.3).

Exam Parameter	Response Variable	Explanatory Variables		Interaction
		Continuous	Categorical	
In-course	Exam Parameter ‘In-Course’	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort

Exam Parameter	Response Variable	Explanatory Variables		Intercation
		Continuous	Categorical	
Oral	Exam Parameter 'Oral'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort
Short	Exam Parameter 'Short'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort
Interpretation (Interp)	Exam Parameter 'Interp'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort
MCQ	Exam Parameter 'MCQ'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort
Spot	Exam Parameter 'Spot'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort
Total Exam	Exam Parameter 'Total Exam'	SMS, CRT Pre, MRT Pre, SDT Pre, Age, AB2	Gender, Handedness, Teaching Method/ Cohort,	SMS*CRT Pre SMS*MRT Pre SMS*SDT Pre SMS*Cohort

**Table 6.3 Summary of GLM<sub>B</sub> analyses with spatial MCQ score as a predictor of end-of-course examination parameter.**

### **6.3.4 Post-examination Item Analysis of Spatial MCQ**

To obtain post-examination item analysis on the reliability of the spatial MCQ assessment as a whole, the analysis output from the software Speedwell (Speedwell, n.d.) was used. The reliability estimate (Kuder-Richardson 20) and item discrimination index (point bi-serial) from the Speedwell analysis were used. The Kuder-Richardson 20 (KR-20) provides an estimate of the overall reliability of the test. The KR-20 measures how well the individual items/questions are functioning together to measure the same underlying construct, i.e. a reliability estimate, and should ideally be above 0.70 (Tavakol & Dennick, 2011).

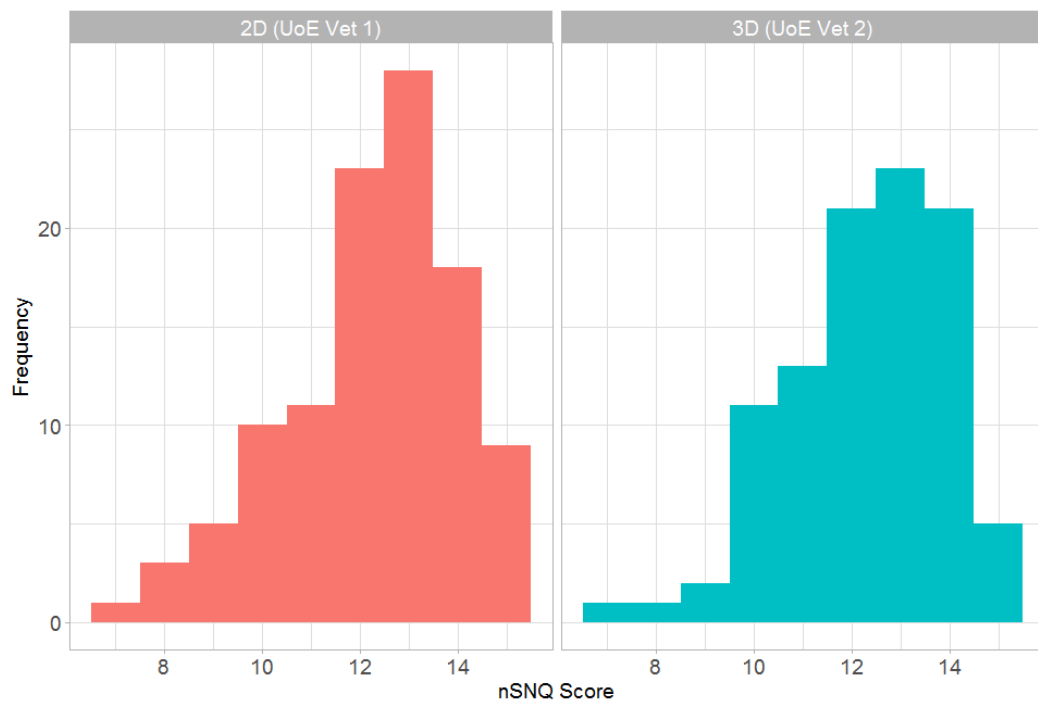
The point bi-serial index was calculated to examine item-discrimination; this is the correlation between students' scores on the question and their score on the total exam. The point bi-serial helps to determine if the students who performed well overall on the test performed well on a question. A negative value indicates that students who performed poorly overall on the test performed well on the question and often indicates the presence of an item flaw. A higher point bi-serial indicates good discrimination between stronger and poorer students (Tavakol & Dennick, 2011).

Furthermore to identify if there were particular questions that, were answered differently between the two teaching methods/cohorts a 2 proportion test (Fisher's exact test was used for question 1 only, due to small sample sizes for the incorrect answers in the contingency table) was performed on the proportion of correct answers for each question. The post-examination analysis included all students who sat the spatial MCQ but did not necessarily sit the spatial ability tests.

## 6.4 Results

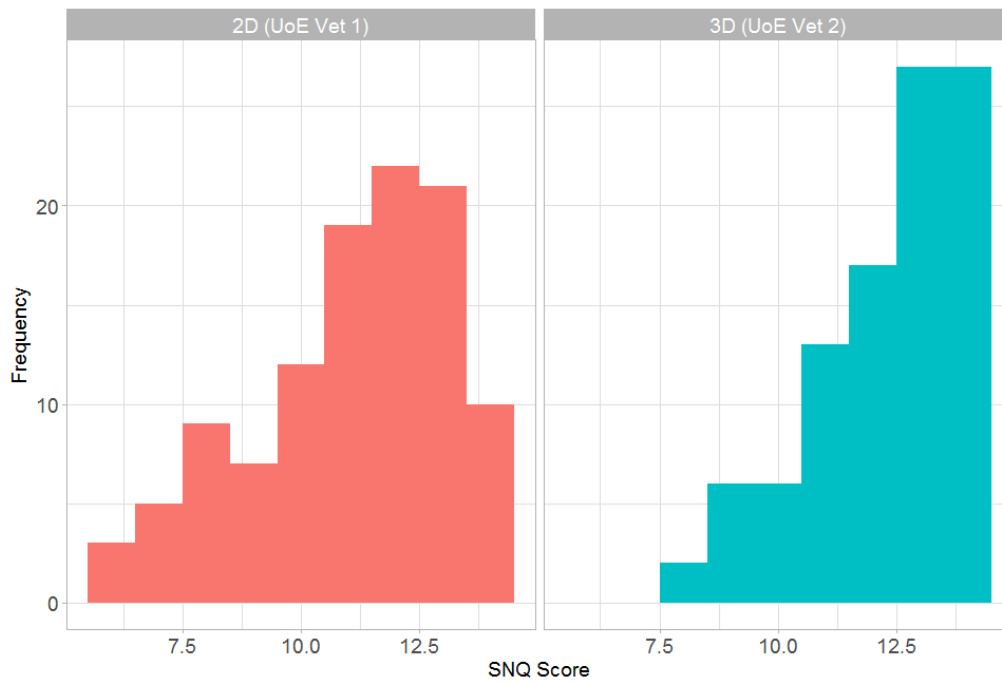
### 6.4.1 Descriptive Analysis

For both cohorts, the parameters nSNQ and SNQ were not normally distributed with a left skewed distribution, which could be interpreted as the nSNQ and SNQ were both easily answered (Figures 6.4 and 6.5).



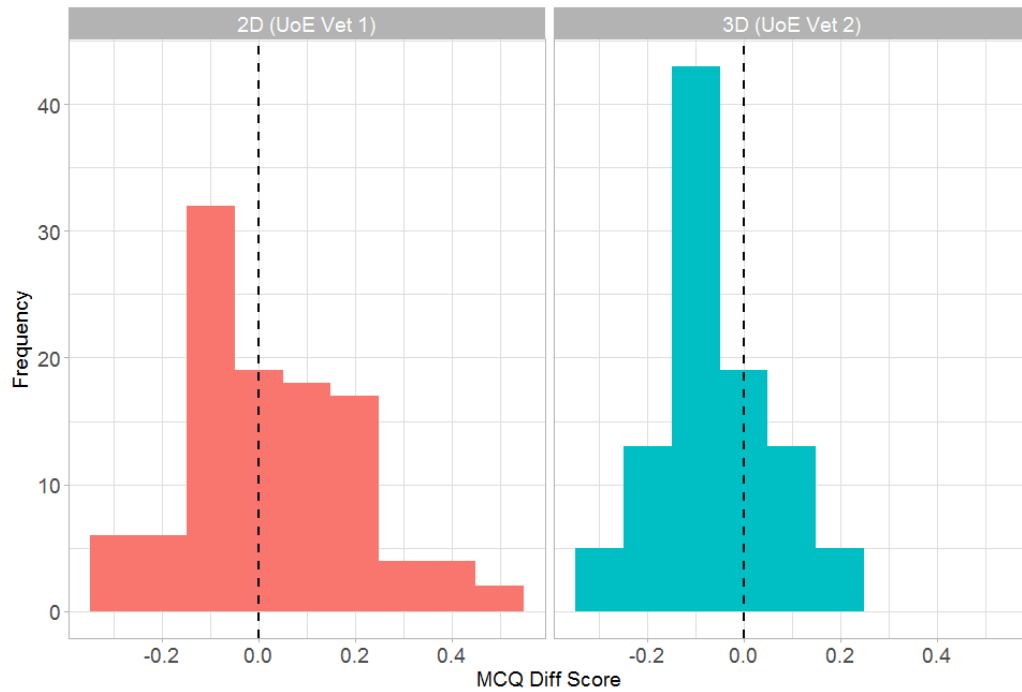
*Figure 6.4 Histograms of nSNQ score for cohort UoE Vet 1 and UoE Vet 2.*





**Figure 6.5. Histograms for SNQ for cohort UoE Vet 1 and UoE Vet 2.**

The  $\Delta_{nSNQ-SNQ}$  scores were not normally distributed for UoE Vet 1 with a slight right-skewed distribution, whereas the distribution of the  $\Delta_{nSNQ-SNQ}$  scores for UoE Vet 2 were normally distributed (Figure 6.6).



**Figure 6.6** Histograms of spatial MCQ Diff scores for UoE Vet 1 and UoE Vet 2.

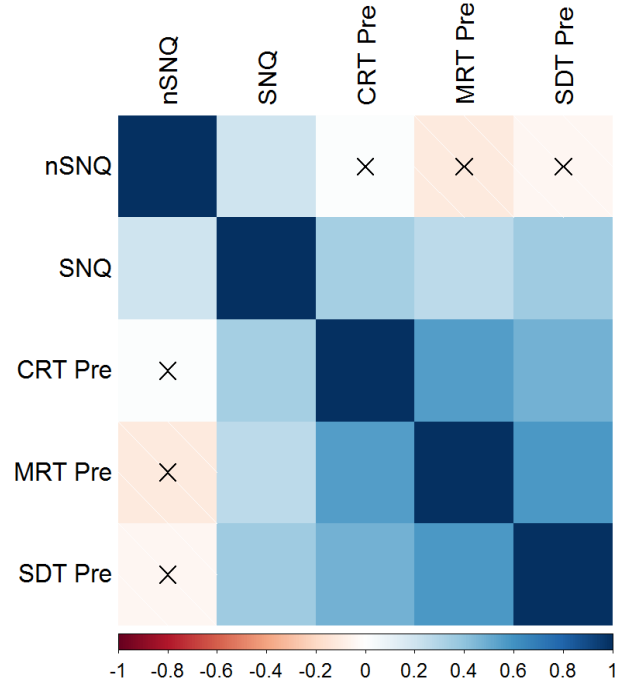
Correlations of the nSNQ to each spatial ability test identified no statistically significant correlations (Table 6.4). SNQ had a low but statistically significant correlation to all three spatial ability tests for 2D (UoE Vet 1) ( $r_s = |0.36|$ ,  $p < 0.05$ ) and two of the three spatial ability tests for 3D (UoE Vet 2) ( $r_s = |0.26|$ ,  $p < 0.05$ ), as the MRT did not correlate statistically significantly with the SNQ for 3D (UoE Vet 2) (Table 6.4).

	nSNQ	SNQ	CRT Pre	MRT Pre	SDT Pre
nSNQ		$r_s = 0.21$ $p = 0.032$	$r_s = 0.02$ $p = 0.877$	$r_s = -0.11$ $p = 0.249$	$r_s = -0.04$ $p = 0.647$
SNQ	$r_s = 0.44$ $p < 0.001$		$r_s = 0.34$ $p < 0.001$	$r_s = 0.27$ $p = 0.005$	$r_s = 0.36$ $p < 0.001$
CRT Pre	$r_s = -0.01$ $p = 0.945$	$r_s = 0.26$ $p = 0.011$		$r_s = 0.56$ $p < 0.001$	$r_s = 0.48$ $p < 0.001$
MRT Pre	$r_s = -0.01$ $p = 0.928$	$r_s = 0.14$ $p = 0.184$	$r_s = 0.48$ $p < 0.001$		$r_s = 0.58$ $p < 0.001$
SDT Pre	$r_s = 0.17$ $p = 0.097$	$r_s = 0.24$ $p = 0.017$	$r_s = 0.44$ $p < 0.001$	$r_s = 0.53$ $p < 0.001$	

**Table 6.4 Correlation matrix of nSNQ, SNQ, CRT Pre, MRT Pre, and SDT Pre for UoE Vet 1 above the diagonal line, and for UoE Vet 2 below the diagonal line. Yellow = significant at  $p < 0.05$ .**

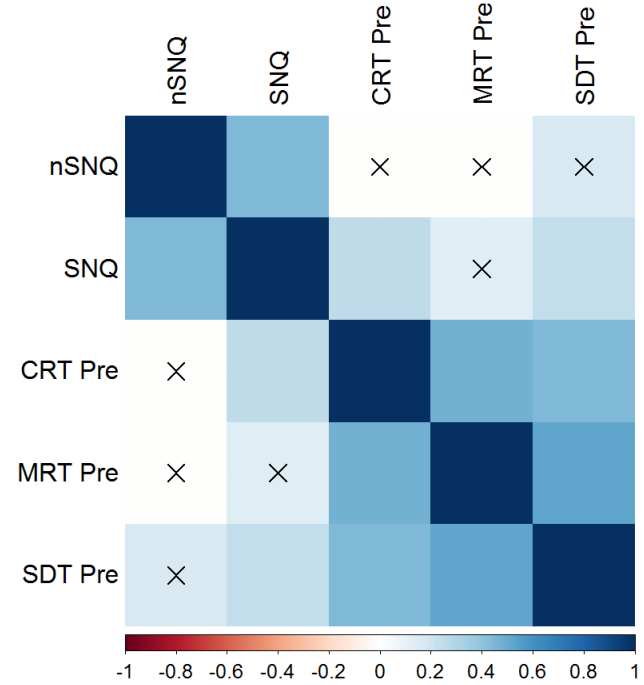
The nSNQ and SNQ correlated statistically significantly and positively with one another for each cohort (2D:  $r_s = 0.21$   $p = 0.032$ , 3D:  $r_s = 0.44$   $p < 0.001$ ). AB2 score correlated statistically significantly with the nSNQ and the SNQ for both teaching methods/cohorts and was identified as a confounding variable. Figures 6.7 and 6.8 graphically show the differences in the correlation matrices of each cohort.

**Cohort UoE Vet 1 Matrix Correlation Graph**



*Figure 6.7 Correlation matrix graph, the colour of the tile indicates the size of the correlation, x = not statistically significant.*

**Cohort UoE Vet 2 Matrix Correlation Graph**



*Figure 6.8 Correlation matrix graph, the colour of the tile indicates the size of the correlation, x= not statistically significant.*

The dependent correlation comparisons were all found to be statistically significant for each spatial ability test for 2D (UoE Vet 1) (CRT Pre  $t = 2.77$   $p = 0.007$ , MRT Pre  $t = 3.27$   $p = 0.002$ , SDT Pre  $t = 3.51$   $p < 0.001$ ) (Table 6.5).

Cohort UoE Vet 1	nSNQ correlation	SNQ correlation	nSNQ $\alpha$ SNQ correlation	Test statistic	p-value
CRT Pre	0.02	0.34	0.21	$t = 2.77$	$p = 0.007$
MRT Pre	-0.11	0.27	0.21	$t = 3.27$	$p = 0.002$
SDT Pre	-0.04	0.36	0.21	$t = 3.51$	$p < 0.001$

**Table 6.5** *Dependent correlation comparisons for 2D (UoE Vet 1).*

However, 3D (UoE Vet 2) only had a statistically significant difference between the nSNQ and SNQ correlations to CRT Pre ( $t = 2.6$ ,  $p < 0.01$ ) and not for MRT Pre ( $t = 1.4$ ,  $p < 0.160$ ) or SDT Pre ( $t = 0.66$ ,  $p < 0.510$ ) (Table 6.6).

Cohort UoE Vet 2	nSNQ $\alpha$ SA correlation	SNQ $\alpha$ SA correlation	nSNQ $\alpha$ SNQ correlation	Test statistic	p-value
CRT Pre	-0.01	0.26	0.44	$t = 2.6$	$p = 0.011$
MRT Pre	-0.01	0.14	0.44	$t = 1.4$	$p = 0.160$
SDT Pre	0.17	0.24	0.44	$t = 0.66$	$p = 0.510$

**Table 6.6** *Dependent correlation comparisons for 3D (UoE Vet 2).*

## 6.4.2 Between Teaching Method Comparison on Spatial MCQ

For the  $GLM_p$  analysis comparing only the nSNQ scores, of 3D (UoE Vet 2) and 2D (UoE Vet 1) cohorts were not statistically significantly different from one another ( $p > 0.05$ ) (Figure 6.9, and Table 6.7 and 6.8). AB2 score was statistically significant with a very small effect size ( $RR = 1.0042$ ).

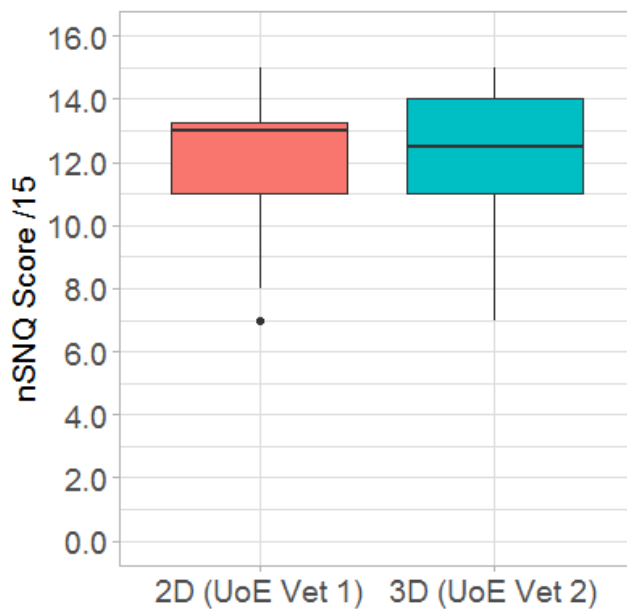


Figure 6.9 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) nSNQ scores.

Parameter	Cohort UoE Vet 1				Cohort UoE Vet 2			
	$\bar{x}/M$	Range	n	Max Score	$\bar{x}/M$	Range /SD	n	Max Score
nSNQ	M = 13.0	7 – 15	108	15	M = 12.5	7 – 15	98	15
SNQ	M = 11.0	6 – 14	108	14	M = 13.0	8 – 14	98	14
Diff	M = 0.01	-0.33 – 0.5	108	-	$\bar{x} = -0.06$	0.12	98	-

Table 6.7 Median ( $M$ ), range,  $SD$  = standard deviation, max score = maximum score, and  $n$  = number of students for nSNQ, SNQ and Diff scores on the spatial MCQ.

Response Variable	Significant Explanatory Variable	RR	95% CI	p-value	Non-significant Explanatory Variables Removed
nSNQ	AB2	1.00	1.00 – 1.01	= 0.021	Age, CRT Pre, MRT Pre, SDT Pre, Teaching Method/Cohort, Gender, Handedness, SA*Cohort
SNQ	Cohort	Ref			Age, MRT Pre, SDT Pre, Teaching Method/Cohort, Gender, Handedness, SA*Cohort
	2D (UoE Vet 1)	1.11	1.02 – 1.20	= 0.013	
	3D (UoE Vet 2)				
	CRT Pre	1.00	1.00 – 1.00	= 0.035	
	AB2	1.00	1.00 – 1.01	= 0.077	

Table 6.8 Summary of GLM<sub>p</sub> output for nSNQ and SNQ, ref = reference level.

GLM<sub>p</sub> analysis comparing the SNQ scores between the two teaching methods showed a statistically significant difference (RR = 1.11, p = 0.013 adjusted for CRT Pre and AB2) with the 3D spatial teaching method (UoE Vet 2) scoring higher (Figure 6.10 and Table 6.8).

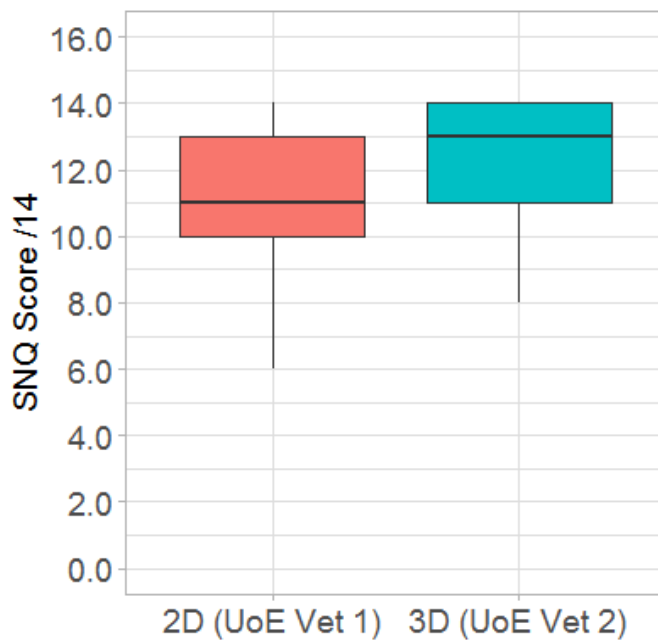
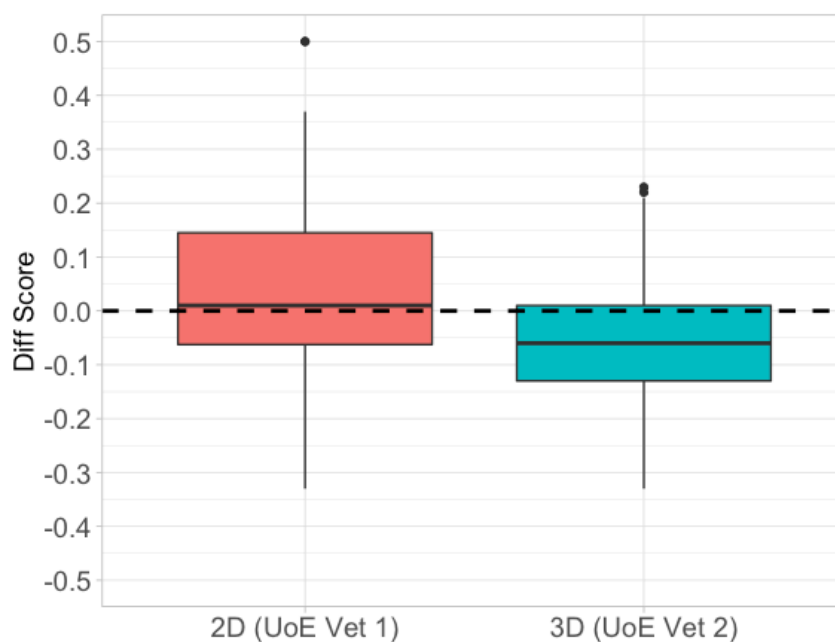


Figure 6.10 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) SNQ scores.

CRT Pre score was also statistically significant for SNQ score. However the effect size for the difference was very small and therefore unlikely to be of large educational significance. Furthermore, the confounding variable AB2 score was not statistically significant for SNQ scores, however, remained in the GLM model to account for differences in academic ability between the two cohorts.

Comparison of the Diff scores to zero by a Wilcoxon Signed Rank test for each cohort identified no statistically significant difference for 2D (UoE Vet 1) ( $p > 0.125$ ) and a statistically significant difference for 3D (UoE Vet 2) ( $p < 0.001$ ) with a negative Diff ( $\Delta_{nSNQ-SNQ}$ ) value (Figure 6.11 and Table 6.7).



*Figure 6.11 Comparison of 2D (UoE Vet 1) and 3D (UoE Vet 2) Diff scores.*

The negative Diff score ( $\Delta_{nSNQ-SNQ}$ ) for 3D (UoE Vet 2) indicates the SNQ score was higher than the nSNQ score and there was no difference for the 2D group, suggesting the 2D group performed similarly on both sets of questions while the 3D group scored higher on the SNQ possibly due to the teaching method.



Comparison of the Diff scores between cohorts by Wilcoxon Rank Sum test identified a statistically significant difference with the 2D group (UoE Vet 1) exhibiting a higher Diff score ( $W = 7027$ , 95% CI: 0.05 – 0.13,  $p < 0.001$ ). Further confirming the 3D group scored higher on the SNQ compared to the nSNQ.

### 6.4.3 Spatial MCQ Score and Relationship to End-of-Course Examination Score

$GLM_B$  analysis of total spatial MCQ score (SMS) identified SMS to be a statistically significant predictor of oral, short, interpretation, MCQ, spot, and total exam parameters. However, the effect sizes for these relationships are small and therefore could be of little educational significance (Table 6.9).

Response Variable	Significant Explanatory Variables	OR	95% CI	p-value	Non-significant Explanatory Variables Removed
In-course	AB2	1.02	1.01 – 1.03	< 0.001	Age, gender, handedness, CRT Pre, MRT Pre, SMS, Cohort*SMS, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS
	SDT Pre	1.56	1.01 – 2.40	= 0.046	
	Cohort 3D (UoE Vet 2) 2D (UoE Vet 1)	Ref 1.44	1.24 – 1.69	< 0.001	
Oral	AB2	1.04	1.03 – 1.04	< 0.001	Gender, handedness, age, Cohort*SMS, MRT Pre, SDT Pre, MRT Pre*SMS, SDT Pre*SMS
	CRT Pre	0.96	0.94 – 0.99	= 0.009	
	Cohort 2D (UoE Vet 1) 3D (UoE Vet 2)	Ref 1.24	1.06 – 1.46	= 0.008	
	SMS	0.86	0.75 – 0.98	= 0.025	
	CRT Pre*SMS <sup>a</sup>	1.00	1.00 – 1.00	= 0.008	
Short	AB2	1.04	1.03 – 1.04	< 0.001	Age, gender, handedness, Cohort, CRT Pre, MRT Pre, SDT Pre, CRT Pre*SMS, MRT Pre*Cohort, SDT Pre*Cohort, Cohort*SMS
	SMS	1.06	1.03 – 1.09	< 0.001	

Response Variable	Significant Explanatory Variables	OR	95% CI	p-value	Non-significant Explanatory Variables Removed
Interpretation (Interp)	SMS	1.07	1.05 – 1.10	< 0.001	Age, gender, handedness, CRT Pre, MRT Pre, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS, Cohort*SMS
	SDT Pre	1.61	1.09 – 2.37	= 0.017	
	Cohort 3D (UoE Vet 2) 2D (UoE Vet 1)	Ref 1.46	1.27 – 1.67	< 0.001	
	AB2	1.02	1.01 – 1.03	< 0.001	
MCQ	SMS	1.08	1.04 – 1.12	< 0.001	Age, gender, handedness, CRT Pre, MRT Pre, SDT Pre, Cohort, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS, Cohort*SMS
	AB2	1.03	1.02 – 1.04	< 0.001	
Spot	SMS	1.08	1.06 – 1.10	< 0.001	Age, gender, handedness, MRT Pre, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS, Cohort*SMS
	Cohort 3D (UoE Vet 2) 2D (UoE Vet 1)	Ref 1.17	1.05 – 1.31	= 0.006	
	SDT Pre	1.51	1.06 – 2.15	= 0.022	
	CRT Pre	1.00	0.99 – 1.00	= 0.033	
	AB2	1.03	1.03 – 1.04	< 0.001	
Total Exam	AB2	1.03	1.03 – 1.04	< 0.001	Age, gender, handedness, CRT Pre, MRT Pre, SDT Pre, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS
	Cohort 3D (UoE Vet 2) 2D (UoE Vet 1)	Ref 2.50	1.13 – 5.53	= 0.025	
	SMS	1.09	1.06 – 1.12	< 0.001	
	Cohort*SMS	0.96	0.93 – 1.00	= 0.034	

**Table 6.9 Summary of GLM<sub>B</sub> output for spatial MCQ score predicting exam parameter score. a = extremely small values for OR and CI, ref = reference level.**

The explanatory variables age, gender (Figure 6.12), and handedness (Figure 6.13) were not statistically significant.

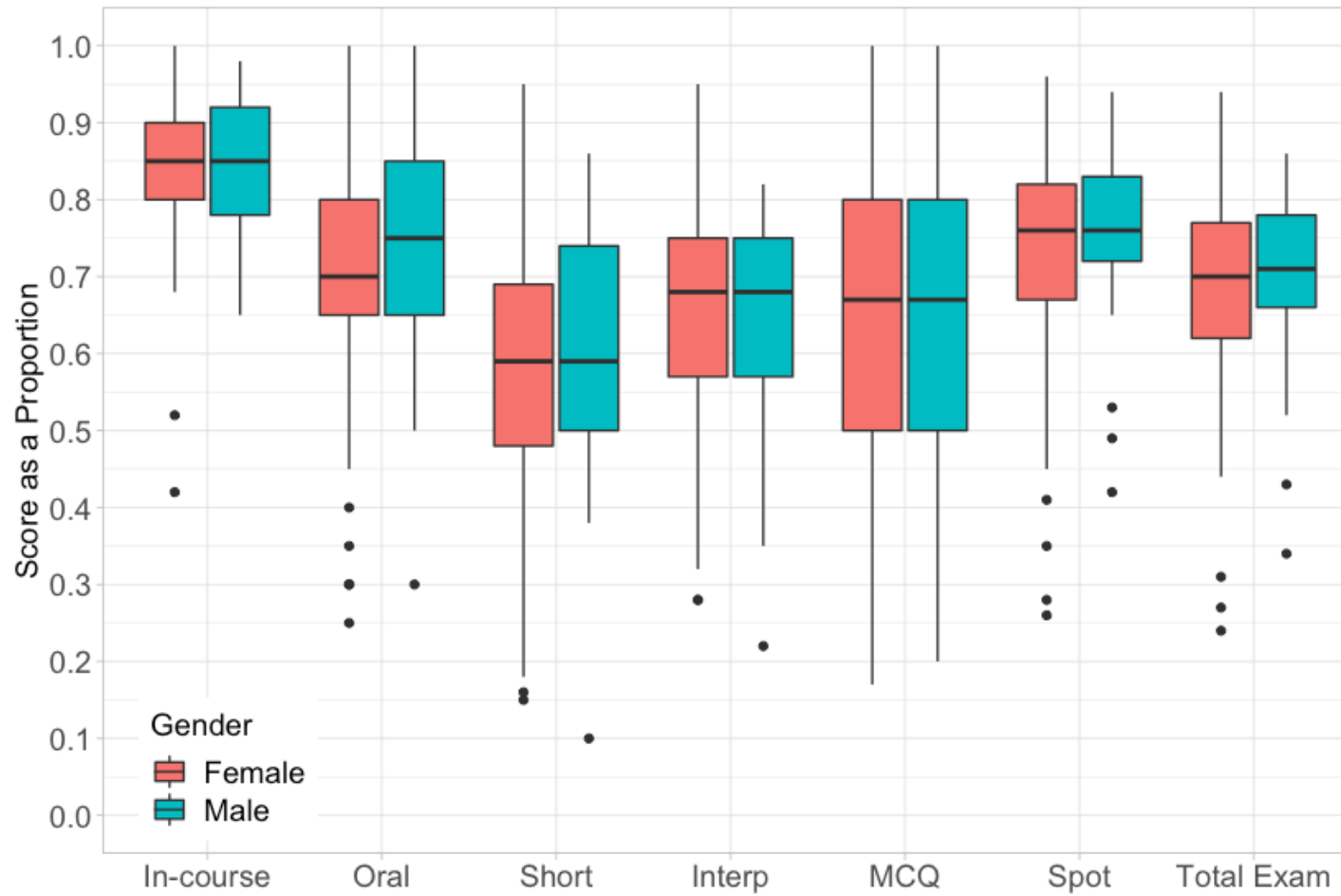


Figure 6.12 Gender vs examination parameter score.

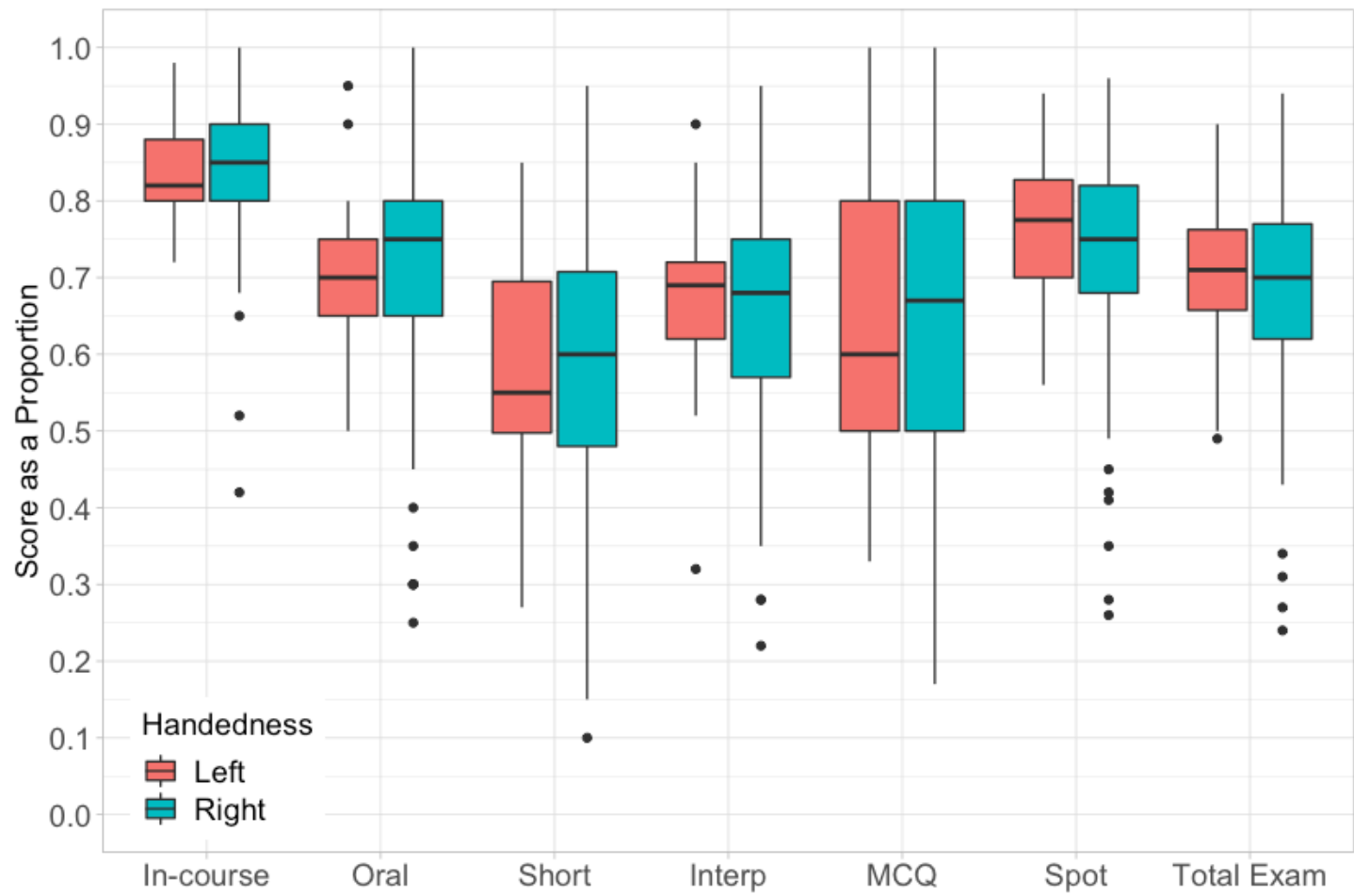


Figure 6.13 Handedness vs examination parameter score.

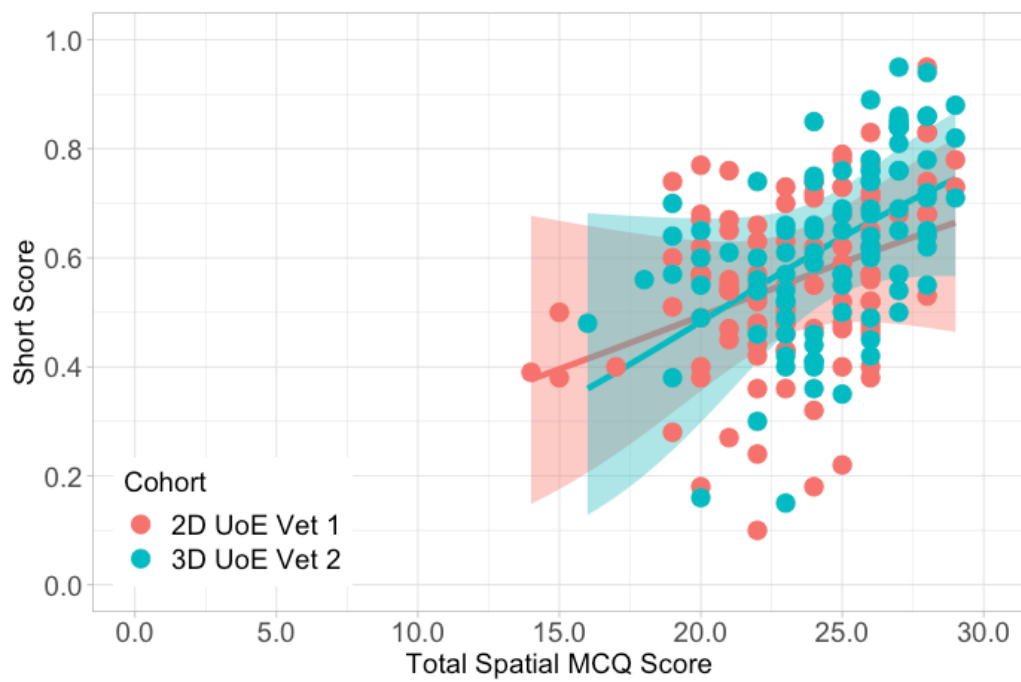
For the response variable oral the CRT Pre\*SMS interaction was statistically significant, however although significant exhibited a negligible effect size (OR = 1.0016) and was therefore removed from the final model and also the predictors SMS and CRT Pre were removed because these predictors became not statistically significant after removal of the negligible interaction. Table 6.10 shows the updated regression output after removal of the CRT Pre\*SMS interaction, and CRT Pre and SMS predictors.

Response Variable	Significant Explanatory Variables	OR	95% CI	p-value	Non-significant Explanatory Variables Removed
Oral	AB2	1.04	1.03 – 1.05	< 0.001	Age, gender, handedness, CRT Pre, MRT Pre, SDT Pre, SMS, Cohort*SMS, CRT Pre*SMS, MRT Pre*SMS, SDT Pre*SMS
	Cohort 2D (UoE Vet 1) 3D (UoE Vet 2)	Ref 1.26	1.07 – 1.47	= 0.005	

*Table 6.10 Updated summary of GLM<sub>B</sub> output for spatial MCQ score predicting oral score.*

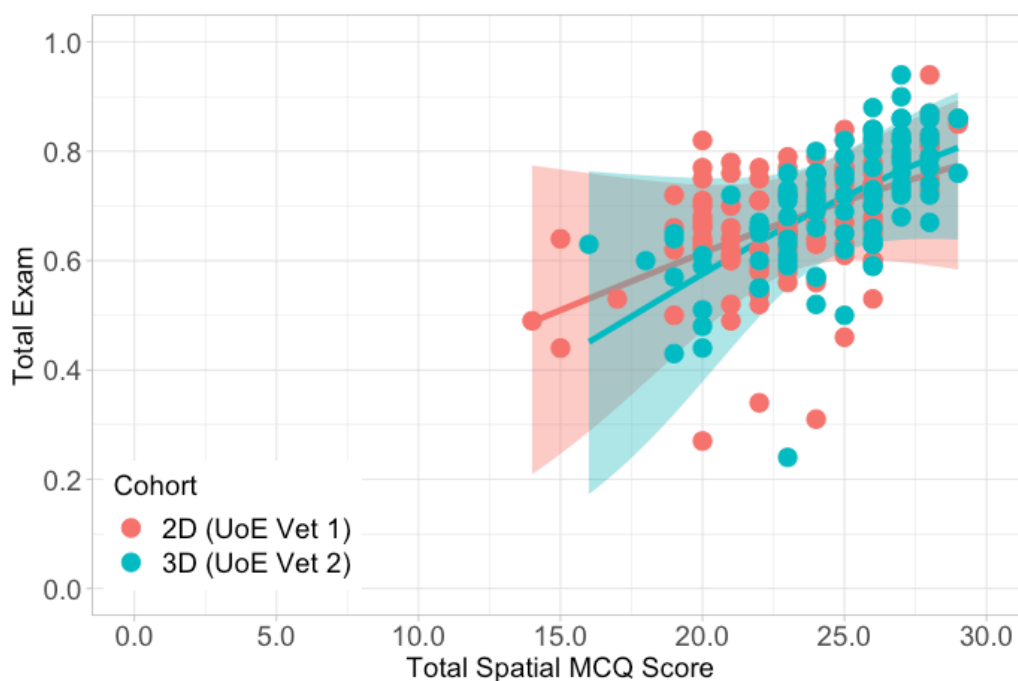
The spatial ability test scores for SDT Pre were statistically significant for three of the exam parameters. Specifically, SDT Pre was statistically significant for in-course (OR = 1.56, p = 0.046 adjusted for cohort and AB2), interpretation (OR = 1.61, p = 0.017, adjusted for cohort, SMS, AB2), and spot (OR = 1.51, p = 0.022, adjusted for cohort, SMS, CRT Pre, and AB2). Meaning with each 1% mark increase in SDT Pre score the odds of scoring each 1% higher in the in-course, interpretation, and spot exams were 1.56, 1.61, 1.51 times greater respectively. Additionally for the spot exam, CRT Pre was statistically significant (OR = 1.00, p = 0.033).

Spatial MCQ score (SMS) was a statistically significant predictor for short answer questions (OR = 1.06,  $p < 0.001$  adjusted for AB2 score) (Figure 6.14).



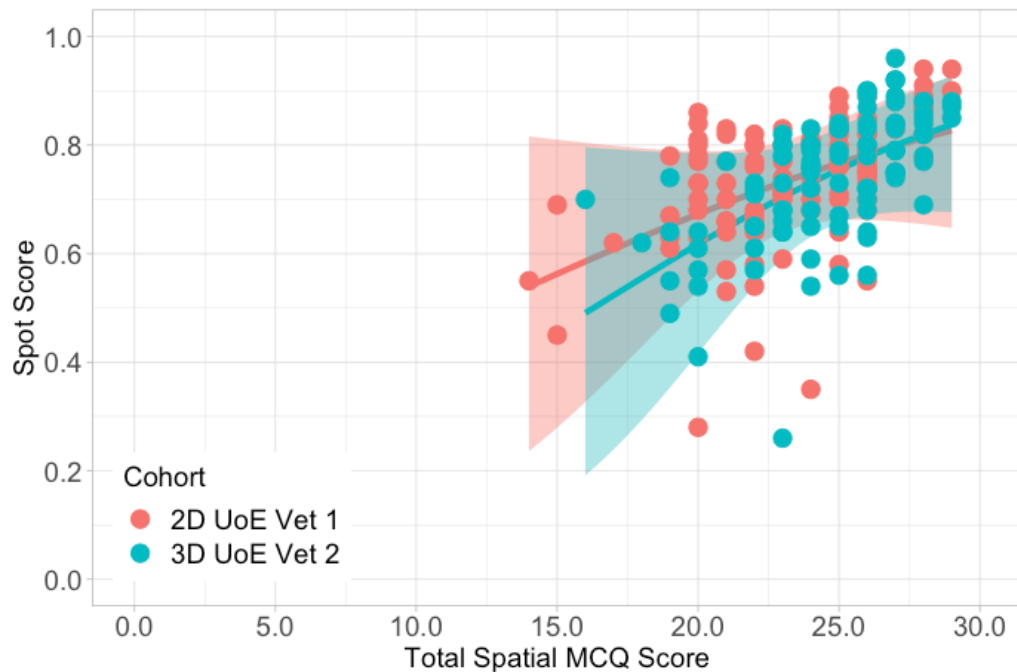
**Figure 6.14** Total spatial MCQ score as a predictor of short answer questions, regression line with 95% confidence interval included.

Total exam was also statistically significantly predicted by SMS (OR = 1.09,  $p < 0.001$  adjusted for cohort, AB2, and cohort\*SMS). The 2D teaching method cohort scored statistically significantly higher than the 3D teaching method cohort (OR = 2.50,  $p = 0.025$  adjusted for SMS, AB2, and cohort\*SMS). A significant interaction between teaching method and spatial MCQ score was found (OR = 0.96,  $p = 0.034$ ). This interaction means that the relationship between the total exam score and spatial MCQ score is different for whether the students received the 2D or 3D spatial teaching method. The significant interaction demonstrates (Figure 6.15) that as spatial MCQ score increased total exam score increased with an odds ratio of 1.09. However the students receiving the 3D spatial teaching method have significantly higher scores than those receiving the 2D method. While the opposite effect was found with decreasing spatial MCQ scores; total exam score decreased with decreasing spatial MCQ scores, and this was lower for students receiving the 3D spatial teaching method than those receiving the 2D method (Figure 6.15).



**Figure 6.15** Total spatial MCQ score as a predictor of total exam score, regression line with 95% confidence interval included.

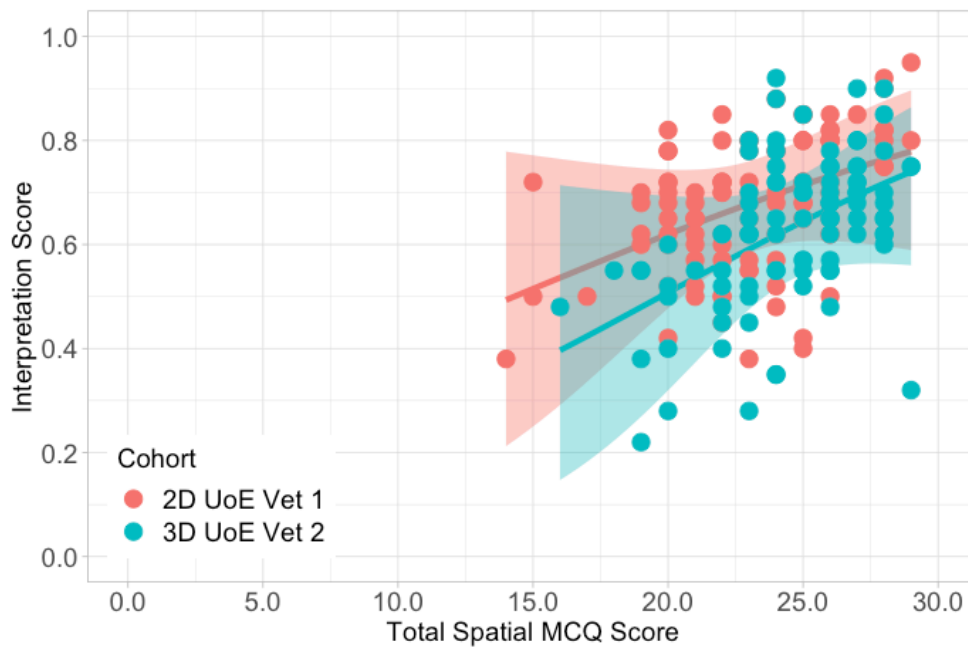
The spatial MCQ (SMS) score was statistically significantly predictive of students' spot exam scores (OR = 1.08,  $p < 0.001$  adjusted for cohort, SDT Pre, CRT Pre, and AB2). The 2D teaching method cohort scored statistically significantly higher than the 3D spatial cohort on the spot questions (OR = 1.17,  $p = 0.006$  adjusted for SMS, SDT Pre, CRT Pre, and AB2) (Figure 6.16).



**Figure 6.16** Total spatial MCQ score as a predictor of spot exam score, regression line with 95% confidence interval included.

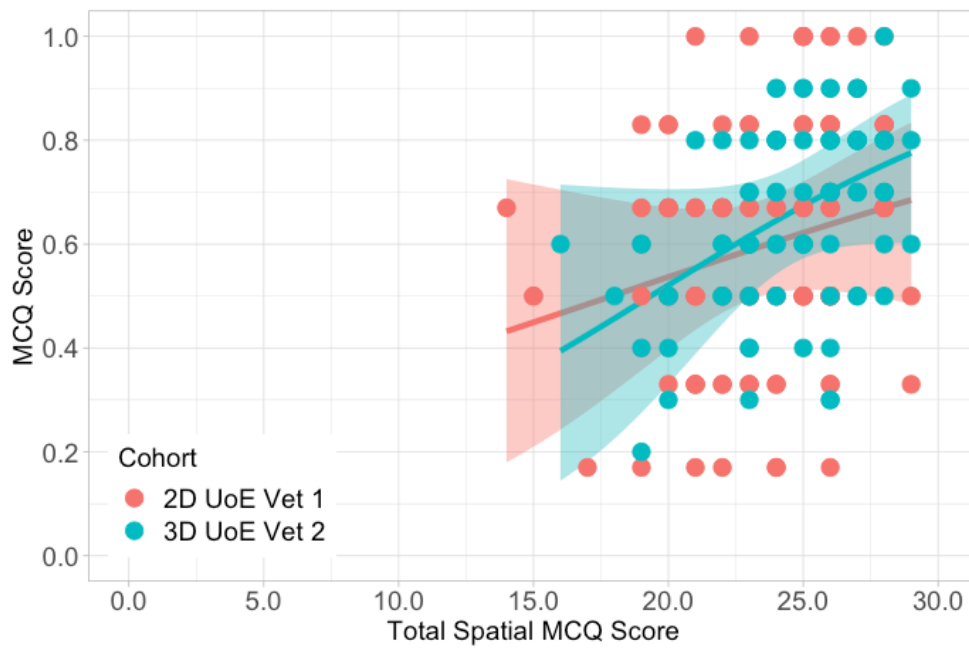


Interpretation scores were also statistically significantly predicted by SMS (OR = 1.07,  $p < 0.001$  adjusted for SDT Pre, cohort, and AB2). The 2D teaching method cohort scored significantly higher on the interpretation questions than the 3D spatial teaching method, as was discussed in chapter 5 (Figure 6.17).



*Figure 6.17 Total spatial MCQ score as a predictor of interpretation exam score, regression line with 95% confidence interval included.*

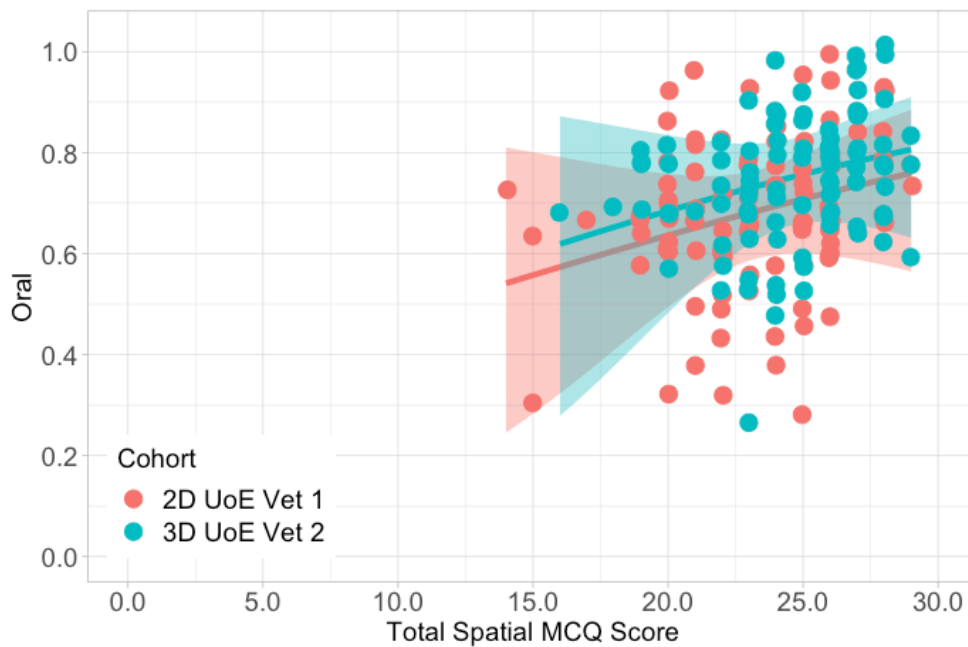
Spatial MCQ score (SMS) was a statistically significant predictor for end-of-course MCQ scores (OR = 1.08,  $p < 0.001$  adjusted for AB2) (Figure 6.18).



*Figure 6.18 Total spatial MCQ score as a predictor of MCQ exam score, regression line with 95% confidence interval included.*

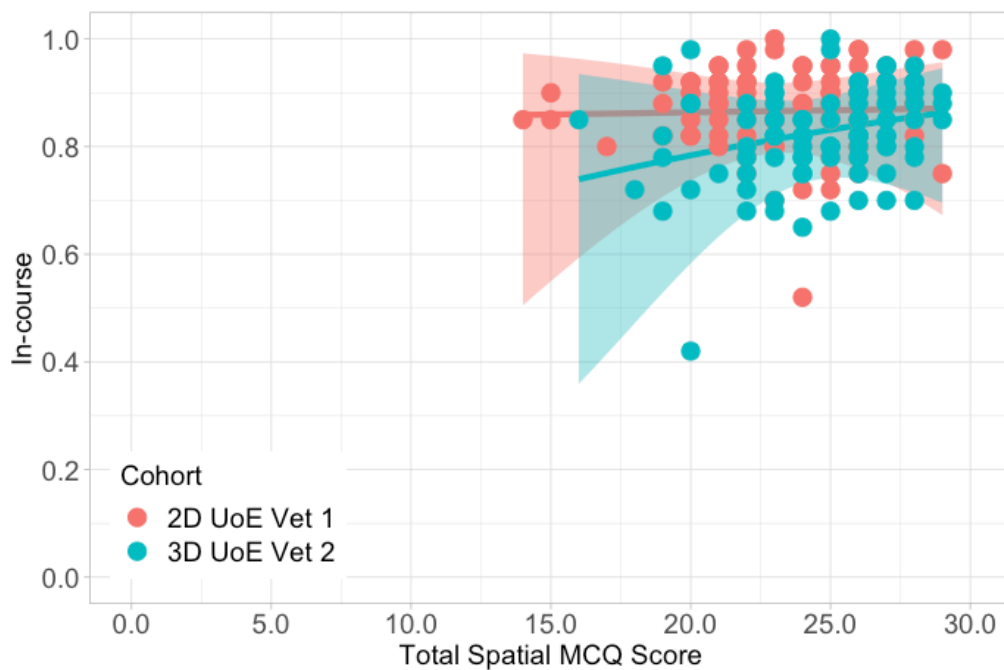
The above exam parameters with a statistically significant spatial MCQ predictor mean with each 1 mark increase in spatial MCQ score the odds of scoring 1% higher is 1.06 for short answer score, 1.09 for total exam score, 1.08 for spot score, 1.08 for MCQ score, and 1.07 times greater for the interpretation score.

Spatial MCQ score was not statistically significantly predictive for oral scores (Table 6.10). AB2 score was statistically significantly predictive of oral scores (OR = 1.04,  $p < 0.001$ , Table 10), and the 3D teaching method cohort scored statistically significantly higher than the 2D teaching method (OR = 1.26,  $p = 0.005$ , Table 10).



**Figure 6.19** Total spatial MCQ score as a predictor of oral score, regression line with 95% confidence interval included.

Spatial MCQ score was not a statistically significant predictor of in-course anatomy workbooks ( $p > 0.05$ ). The 2D teaching method cohort scored statistically significantly higher on the in-course assessment than the 3D spatial teaching method (Figure 6.20).



*Figure 6.20 Total spatial MCQ score as a predictor of in-course score, regression line with 95% confidence interval included.*

Similar to chapter 5, the odds ratio represents the effect size of any statistically significant explanatory variables and is the ratio of the odds of exposure, to the odds of no exposure. Therefore an odds ratio of 1 indicates no difference. Odds ratios greater than 1 indicate the exposure is associated with a higher odds of outcome, while an odds ratio less than 1 indicates the exposure is associated with a lower odds of outcome. Therefore the statistically significant odds ratios identified in this study are small and could represent a small educational impact. A further understanding of the educational impact from these quantitative studies will be investigated and discussed with the qualitative data collected for this thesis (see chapter 8).

#### 6.4.4 Post-examination Analysis of Spatial MCQ

The KR-20 for 2D (UoE Vet 1) and 3D (UoE Vet 2) on the MCQ was 0.63 and 0.71 respectively (Table 6.11).

Cohort	KR-20	Mean	SD	SEM	n
UoE Vet 1	0.63	23.2	3.07	1.86	119
UoE Vet 2	0.71	24.2	3.24	1.75	121

*Table 6.11 KR-20, mean, standard deviation (SD), standard error of measurement (SEM) and total number of students (n) of the spatial MCQ for each cohort.*

The KR-20 should ideally be above 0.70, and so the reliability of this spatial MCQ is on, or slightly below, what is expected however for a test with relatively few items this would be acceptable reliability. There were differences in the mean, SD, and SEM for each cohort, with 2D (UoE Vet 1) exhibiting a lower mean, lower SD and higher SEM. This indicates 2D (UoE Vet 1) on average scored slightly lower on the spatial MCQ but with little variance in scores, and had a slightly higher error inherent in an individual's test score (Table 6.11).

To analyse whether the stronger students answered the spatial MCQ questions consistently correct compared to the weaker students the item-discrimination index was calculated (point bi-serial). To further analyse the spatial MCQ at the question level in relation to the 3D spatial teaching method, the proportion of students answering each question correctly (i.e., item difficulty) was compared between the two teaching methods (Table 6.12).

Question	% correct (item-difficulty index)		2-proportion test/ Fisher's exact test	Point Bi-serial (item-discrimination index)	
	2D (UoE Vet 1)	3D (UoE Vet 2)		2D (UoE Vet 1)	3D (UoE Vet 2)
1	94.1	97.5	OR = 0.4, 95%: 0.07 – 1.84, p = 0.214	0.201	0.108
2	100	100	n/a	0.0	0.0
3	100	100	n/a	0.0	0.0
4	66.4	74.4	$X^2 = 1.84$ , 95%: -0.20 – 0.04, p = 0.175	0.314	0.531
5	66.4	77.7	$X^2 = 3.81$ , 95%: -0.23 – 0.0, p = 0.051	0.326	0.318
6	98.3	91.7	$X_2 = 5.47$ , 95%: 0.01 – 0.12, p = 0.019	0.264	0.156
7	99.2	99.2	n/a	-0.025	0.005
8	79.0	84.3	$X^2 = 1.13$ , 95%: -0.15 – 0.04, p = 0.288	0.447	0.376
9	25.2	17.4	$X^2 = 2.21$ , 95%: -0.02 – 0.18, p = 0.137	0.169	0.111
10	64.7	57.9	$X^2 = 1.19$ , 95%: -0.05 – 0.19, p = 0.276	0.381	0.482
11	68.1	79.3	$X_2 = 3.93$ , 95%: -0.22 – 0.0, p = 0.047	0.381	0.306
12	85.7	81.8	$X^2 = 0.67$ , 95%: -0.05 – 0.13, p = 0.413	0.377	0.431
13	80.7	82.6	$X^2 = 0.16$ , 95%: -0.12 – 0.08, p = 0.693	0.397	0.262
14	96.6	92.6	$X^2 = 1.95$ , 95%: -0.02 – 0.1, p = 0.163	0.224	0.445
15	92.4	84.3	$X^2 = 3.87$ , 95%: 0.0 – 0.16, p = 0.050	0.151	0.171
16	90.8	90.1	$X^2 = 0.03$ , 95%: -0.07 – 0.08, p = 0.859	0.246	0.309

Question	% correct (item-difficulty index)		2-proportion test/ Fisher's exact test	Point Bi-serial (item-discrimination index)	
	2D (UoE Vet 1)	3D (UoE Vet 2)		2D (UoE Vet 1)	3D (UoE Vet 2)
17	89.9	92.6	$X^2= 0.53, 95\%: -0.1 - 0.05, p = 0.468$	0.265	0.357
18	69.7	79.3	$X^2= 2.91, 95\%: -0.21 - 0.01, p = 0.088$	0.426	0.439
19	89.9	88.4	$X^2= 0.14, 95\%: -0.06 - 0.09, p = 0.711$	0.338	0.428
20	89.9	92.6	$X^2= 0.53, 95\%: -0.1 - 0.05, p = 0.468$	0.338	0.572
21	91.6	86.8	$X^2= 1.44, 95\%: -0.03 - 0.13, p = 0.230$	0.166	0.256
22	47.1	66.9	$X_2= 9.7, 95\%: -0.32 - -0.08, p = 0.002$	0.479	0.474
23	81.5	89.3	$X^2= 2.89, 95\%: -0.17 - 0.01, p = 0.089$	0.409	0.250
24	89.9	91.7	$X^2= 0.24, 95\%: -0.09 - 0.05, p = 0.625$	0.329	0.519
25	84.0	89.3	$X^2= 1.42, 95\%: -0.14 - 0.03, p = 0.234$	0.332	0.400
26	44.5	76.9	$X_2= 26.31, 95\%: -0.44 - -0.21, p < 0.001$	0.313	0.291
27	83.2	89.3	$X^2= 1.86, 95\%: -0.15 - 0.03, p = 0.173$	0.261	0.275
28	52.1	70.2	$X_2= 8.32, 95\%: -0.30 - -0.06, p = 0.004$	0.261	0.472
29	QUESTION REMOVED				
30	96.6	93.4	$X^2= 1.33, 95\%: -0.02 - 0.09, p = 0.248$	0.239	0.344

**Table 6.12** Item-difficulty index, item-discrimination index and 2 proportion test result for each spatial MCQ question for each cohort. Questions 1 – 15 are non-spatial and questions 15 – 30 are spatial.

The item-difficulty index ranged from 0.252 – 1 for 2D (UoE Vet 1) and 0.174 – 1 for 3D (UoE Vet 2). For 2D (UoE Vet 1), questions 1, 2, 3, 6, 7, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 24, 25, 27, and 30 had an index greater than 0.80 (Table 6.12). This was the same for 3D (UoE Vet 2) with the addition of question 8. Only question 9 had an index less than 0.30 for both cohorts.

The point bi-serial coefficient ranged between -0.025 – 0.479 for 2D (UoE Vet 1) and 0 – 0.572 for 3D (UoE Vet 2) (Table 6.12). Three questions (2, 3, and 7) had a zero, or near zero, point bi-serial with question 7 exhibiting a negative value for 2D (UoE Vet 1) indicating this question should be reviewed.

Comparison of the proportion of students answering correctly between cohorts identified statistically significant differences for question 6 ( $X^2 = 5.47$ ,  $p = 0.019$ ) with 2D (UoE Vet 1) scoring higher. While the 3D spatial teaching method cohort (UoE Vet 2) scored statistically significantly higher for question 11 ( $X^2 = 3.93$ ,  $p = 0.047$ ), question 22 ( $X^2 = 9.7$ ,  $p = 0.002$ ), question 26 ( $X^2 = 26.31$ ,  $p < 0.001$ ), and question 28 ( $X^2 = 8.32$ ,  $p = 0.004$ ).

## 6.5 Discussion

### 6.5.1 Design of a spatial MCQ

The first aim of the study presented in this chapter was to design an MCQ to assess anatomy knowledge and understanding in both a spatial format and a non-spatial format (i.e., a 'spatial MCQ'). The design of the MCQ was based on the question formats of three spatial ability tests (Card Rotation Test (Ekstrom *et al.*, 1976), Mental Rotation Test (Peters *et al.*, 1995), and Surface Development Test (Ekstrom *et al.*, 1976). To statistically analyse whether a spatial MCQ is successfully designed,



the relationship of the spatial MCQ scores to the spatial ability test scores are analysed, most commonly by correlation (Hegarty *et al.*, 2009; Nguyen, Nelson & Wilson, 2012; Keedy *et al.*, 2011; Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a; Tan *et al.*, 2012; Berney *et al.*, 2015).

This current study found a statistically significant distinction between the non-spatial questions and the spatial questions, confirmed by correlation and the dependent correlation comparisons (Table 6.5, 6.6, and 6.7). This finding confirms anatomy knowledge can be assessed by both formats using MCQs (Langlois *et al.*, 2017). Furthermore, the finding of AB2 score as not statistically significant in the  $GLM_p$  model for SNQ further confirms this finding. It would be expected that if a student performs well in one assessment, they will more than likely perform well in another, as was identified in chapter 5 with AB2 score significantly predicting the end-of-course examination parameters. This relationship was identified for AB2 and nSNQ although exhibited a small effect size. However, a non-significant relationship was identified for SNQ suggesting these questions are measuring a different trait (or cognitive abilities) to the AB2 exam. This finding provides further evidence of the split design of the spatial MCQ.

Many studies have found significant relationships with tests of spatial ability and spatial anatomy MCQs (Nguyen, Nelson & Wilson, 2012; Hegarty *et al.*, 2009; Garg *et al.*, 1999a; Garg, Norman & Sperotable, 2001). However, not all studies have included non-spatial questions to make comparisons between spatial ability tests and non-spatial questions, or to make comparisons between non-spatial questions and spatial questions (studies which have included non-spatial questions are Keedy *et al.*, 2011; Tan *et al.*, 2012). Successfully designed spatial questions have included questioning on the identification of structures on cross-section, asking 'what level' a cross-section was taken at, identification of features on rotated bones, questions requiring students to state the correct order of movement of a structure, and questions involving intersecting pins through the superficial skin or bones (Berney *et al.*, 2015;

Garg *et al.*, 1999a; Garg, Norman & Sperotable, 2001; Nguyen, Nelson & Wilson, 2012; Hegarty *et al.*, 2009). The study of this chapter replicated the format of three well validated and robust spatial ability tests to design a spatial anatomy MCQ (Ekstrom *et al.*, 1976; Peters *et al.*, 1995) and primarily used anatomical cross-sections. Since the spatial anatomy MCQs were designed based on the format of well-validated spatial ability tests, this finding suggests the spatial anatomy questions are ‘tapping’ into the spatial ability factor. Future research investigating the relationship of scores on the spatial MCQ to other tests of spatial ability would further validate this theory.

There are many important factors to consider when designing exam questions that ensure the quality and validity of the questions. MCQ questions should be written to align with the principles and theory underpinning their design such as; are the distractors plausible, no ambiguity to the question, and do the questions require higher order thinking rather than simple recall of facts (Bloom, 1984; Case & Swanson, 2002). Thus, differences in general question writing technique may also account for the differences observed across anatomy education study findings (Keedy *et al.*, 2011; Tan *et al.*, 2012; Garg, Norman & Sperotable, 2001; Hegarty *et al.*, 2009; Nguyen, Nelson & Wilson, 2012).

### **6.5.2 Post-examination Analysis of the Spatial MCQ**

The reliability of the novel spatial MCQ was at an acceptable level as measured by KR-20. Two questions (2 and 3) had a zero item difficulty and zero point bi-serial coefficients, which were explained by all students answering correctly. As these questions were examining core knowledge, they were not removed from the test. Question 7 was a similar situation with a close to zero point bi-serial coefficient and 99.2% of students answered correctly for each cohort. Eighteen of the 29 questions (62%) had a greater than 0.80 item difficulty (percentage correct) suggesting these questions were easier. These questions being easier is also supported by the low to

moderate point bi-serial coefficients, as high coefficients indicate discriminating questions which tend to be more challenging and thus have a lower item difficulty. Only question 9 had an item difficulty less than 0.30 (and point bi-serial  $<0.25$ ) indicating this question was a difficult question yet despite this, it did not discriminate as well as some of the other questions for both cohorts. This question was specifically asking about the cruciate ligaments and since both cohorts found this question difficult perhaps this identifies the material as a difficult topic for students or an area requiring more emphasis during teaching. [The difficulty with question 9 was also identified in the subsequent cohort of first-year veterinary students (academic year 16/17), which were taught by a combination of a 2D and 3D approach – data not shown].

Only questions 6, 11, 22, 26, and 28 (Figure 6.21, 6.22, 6.23, 6.24, and 6.25) were identified as having a statistically significant difference in the proportion of students correctly answering when comparing the two cohorts by the 2-proportion test. Of these five questions, UoE Vet 1 (2D) exhibited a higher proportion of students answering correctly for question 6 (asking about the structure of the menisci see Figure 6.21 below). The higher proportion could be because there was no isolated 3D model of the menisci; the menisci were shown in a 3D model of the tibial plateau primarily showing the meniscal ligaments. UoE Vet 2 (3D) exhibited a higher proportion of students answering questions 11, 22, 26, and 28 correctly. Of these questions, question 11 is the only non-spatial anatomy question (nSNQ) and was asking about the fabellae bones (Figure 6.22). This difference between cohorts may be due to the fact the fabellae were better explained through the use of 3D models, but there is no good evidence for this. Questions 22, 26, and 28 were spatial anatomy questions (SNQ) (Figures 6.23, 6.24, 6.25) and show dramatic differences between the two cohorts (with differences ranging between 18% - 32%). The differences could be because the spatial 3D teaching did improve anatomy knowledge. Chapter 8 explores this possibility qualitatively through focus groups with student views of the teaching methods and the spatial MCQ.

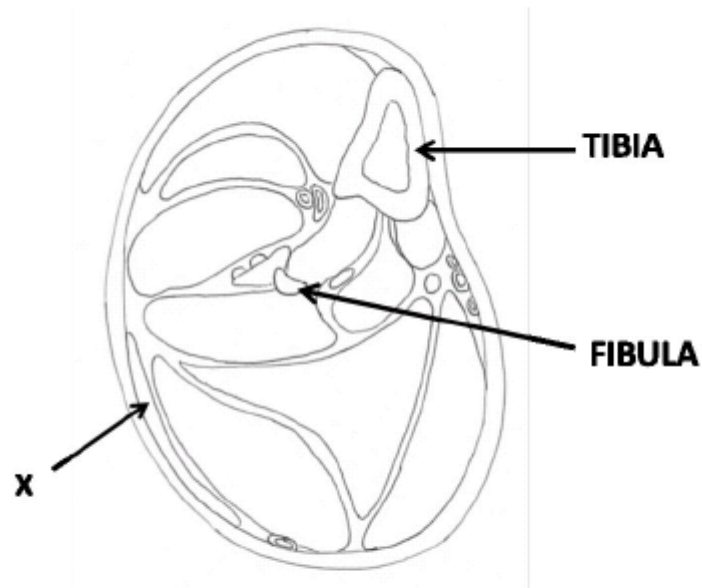
6. What is the structure of the menisci?
- a. Circular shaped, fibrocartilaginous wedges
  - b. Circular shaped, ligamentous wedges
  - c. Crescent shaped, fibrocartilaginous wedges
  - d. Crescent shaped, ligamentous wedges

*Figure 6.21 Question 6 from the spatial MCQ*

11. The fabellae are found in the muscle bellies of which muscle?
- a. Deep digital flexor
  - b. Gastrocnemius
  - c. Quadriceps femoris
  - d. Semitendinosus

*Figure 6.22 Question 11 from the spatial MCQ.*

22. The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.



- a. Biceps femoris
- b. Cranialis tibialis
- c. Semitendinosus
- d. Tensor fasciae latae

*Figure 6.23 Question 22 from the spatial MCQ.*

26. In relation to this X-ray of a normal stifle, which of the following statements could be correct?

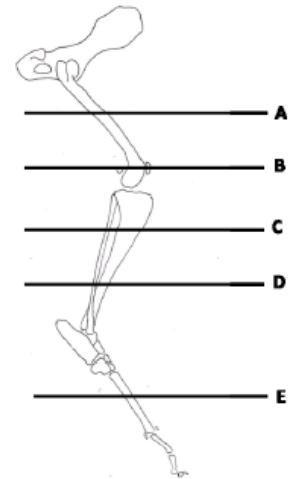
1. Right leg, cranial view
2. Right leg, caudal view
3. Left leg, cranial view
4. Left leg, caudal view



- a. 1 and 2 could be correct
- b. 2 and 3 could be correct
- c. 1 and 3 could be correct
- d. 2 and 4 could be correct
- e. 1 and 4 could be correct

*Figure 6.24 Question 26 from the spatial MCQ.*

28. At which level of the limb is this cross section taken at



*Figure 6.25 Question 28 from the spatial MCQ.*

Another possible reason for the difference between cohorts on specific questions could be because of the use of images in the question. As previously mentioned, Vorstenbosch et al. (2013b) found that the use of an anatomical image can influence the difficulty of the question making it harder or easier. To further this research, Vorstenbosch et al. (2014) used the think-aloud technique with seventeen medical students answering extended match questions of two types; either with an image or without an image. The authors concluded that the students used different cognitive processes to answer each question type. Questions without an image seemed to test the students' mental image of the anatomy, while questions with an image tested the students' interpretation of visual information. While spatial ability is likely to be involved in these processes; it was not a specific aim of the study.

Mental images can be defined as a representation in a person's mind of a structure in the real world. In the study of anatomy, students and experts could create mental

images of anatomical structures in order to memorise anatomical structures regardless of whether non-spatial or spatial. By creating these mental models learners will have taken cognitive steps in their construction and stored them in memory, ready for recall. Therefore, if a question includes an image how is it known whether the student is answering the question based on recall of a previously created and memorised mental image, or due to applying spatial ability to figure out the answer? If the anatomy topic under question is deemed 3D and so therefore defined as ‘spatial anatomy knowledge’ this does not mean the question could only be answered by using spatial ability (section 1.3.5, p28).

### **6.5.3 Between Teaching Method Comparison on Spatial MCQ**

No difference between the non-spatial anatomy question (nSNQ) score between cohorts was identified. However, UoE Vet 2 (3D) scored statistically significantly higher on the spatial anatomy questions (SNQ). This difference in SNQ score could be due to the 3D spatial teaching method improving, enhancing or prompting students to think more spatially about anatomy. Furthermore, comparison of ‘diff’ spatial MCQ scores (i.e., nSNQ minus SNQ) also confirmed this finding by identifying UoE Vet 2 (3D) exhibiting a negative ‘diff’ score meaning students scored higher on the SNQ than the nSNQ.

So far previous anatomy education studies comparing 2D and 3D teaching methods have identified no statistically significant differences between the two teaching methods on spatially designed MCQs (Keedy *et al.*, 2011; Berney *et al.*, 2015; Tan *et al.*, 2012). The 3D teaching method employed by these studies involved computer-controlled 3D computer models. This is interesting because the 3D computer models used in this chapter’s study, which found the 3D teaching group to score higher on the spatial anatomy questions, utilised learner-controlled 3D computer models. This

finding further confirms that learner control of the computer models is an important factor (Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999a).

However, one main consideration when comparing the performance on the spatial MCQ by the two cohorts is the possibility that UoE Vet 1 advised UoE Vet 2 on the details of the assessment. This communication means UoE Vet 2 could have specifically prepared for the type of questions, e.g., revising cross sections, whereas UoE Vet 1 was completely naïve to the assessment because the assessment had never been administered before this cohort. This possibility was confirmed while conducting focus groups with UoE Vet 2 students:

“...quite a few of the 2nd years were like ‘look at transverse sections of the leg.’”

Prior knowledge of the assessment content may account for the differences observed such as higher scores for UoE Vet 2. To overcome this, different questions could be used for UoE Vet 2, although the questions in each section (nSNQ and SNQ) were designed similarly and therefore would only change question content and not prevent discussions on the format of the questions (McCaffrey & Westervelt, 1995; Lezak *et al.*, 2012). Alternatively UoE Vet 2 could have performed better on the examination because they had better anatomy knowledge due to the 3D spatial teaching method.

#### **6.5.4 Spatial MCQ Score and Relationship to End-of-Course Examination Scores**

So far anatomy education research studies have compared scores on spatial MCQs to spatial ability tests score, but not to end-of-course anatomy examinations (Berney *et al.*, 2015; Garg *et al.*, 1999a; Garg, Norman & Sperotable, 2001; Nguyen, Nelson & Wilson, 2012; Hegarty *et al.*, 2009). In this study, the spatial MCQ was analysed in



relation to end-of-course anatomy examinations. It was found the spatial MCQ predicted end-of-course examination scores except for scores on an in-course assessment involving dissection workbooks and oral scores.

The spatial MCQ is an assessment itself, and so it would be expected one assessment should predict another, i.e., comparing like with like, as was observed with AB2 score and the end-of-course anatomy examinations in chapter 5. However, this was not observed for the spatial MCQ and AB2 score. Specifically, the AB2 score did not predict the spatial anatomy questions (SNQ) but did predict the non-spatial anatomy questions (nSNQ). This finding suggests the nSNQ were similar to the AB2 questions, while the SNQ were different to the AB2 questions. Thus, highlighting the dual role of the spatial MCQ (to assess anatomy non-spatially and spatially).

Furthermore, the finding of AB2 score predicting nSNQ and the end-of-course examination parameters suggests these questions are similar in nature, while SNQ is not. Therefore perhaps the use of end-of-course anatomy examinations is not a reliable indicator to investigate whether a relationship exists between anatomy and spatial ability. This theory is supported by the significant relationships found by other studies using other spatial anatomy assessments such as practical assessments, 3D synthesis from 2D views, drawing of views, and cross-sections (Langlois *et al.*, 2017; Rochford, 1985; Provo, Lamar & Newby, 1998; Lufler *et al.*, 2012).

To conclude this chapter, the design aims of the spatial MCQ to assess anatomy knowledge non-spatially/factually and spatially was confirmed. The significant correlations of the SNQs to the spatial ability tests and the insignificant correlations to the nSNQ, along with the significant dependent correlation comparisons statistically confirm this design aim. Therefore suggesting that a spatial MCQ assessment can be designed. The size of the correlations are low, and the dependent correlation comparisons were not significant for all three spatial ability tests for UoE

Vet 2 indicating future research to investigate and refine the design of the spatial MCQ.

The results of the post-examination analysis identified potential questions that could be reviewed to help improve the SNQs, and this may help to improve the correlations. The spatial MCQ could be used to identify struggling students as it significantly predicted the end-of-course examination scores. Overall the initial design of the spatial MCQ was acceptable with good reliability, and only one question (question 9) was identified to have a low item difficulty and point bi-serial coefficient. Specifically, on comparison of item difficulty, four questions had significantly higher scores when students received a 3D spatial teaching method. Furthermore, the students receiving the 3D spatial teaching method performed better on the SNQ section of the spatial MCQ compared to a cohort of students taught with a 2D teaching method. This difference could be because of the teaching method. However there was communication between the two cohorts on the nature of the assessment, therefore possibly rendering UoE Vet 2 not naïve. Chapter 8 investigates this further by exploring thematically analysed student views on the spatial MCQ and the two teaching methods.

The next chapter investigates whether spatial ability improved with the learning of veterinary anatomy.



# Chapter 7 Comparison of Changes in Spatial Ability Following Two Different Anatomy Teaching Methods

## 7.1 Introduction to Chapter 7

In the previous chapter, a spatial Multiple Choice Question (MCQ) assessment was designed, and student performance analysed. This chapter concludes the quantitative analysis of spatial ability and anatomy learning in veterinary students by investigating whether spatial ability improved with the learning of anatomy. Chapter 2 (section 2.3.5) previously discussed how spatial ability tests are prone to the practice/retest effect, where improvements in performance on a given spatial ability test can be due to the practice of the test, in contrast to any true improvement. To incorporate this potential practice effect research study designs can include a control group i.e., a group of participants that do not receive the intervention but still take the spatial ability tests twice, and at the same time interval as the intervention group (Uttal *et al.*, 2013; Goldberg *et al.*, 2015; Lezak *et al.*, 2012; McCaffrey & Westervelt, 1995).

In a meta-analysis conducted by Uttal *et al.* (2013) investigating the malleability of spatial ability, the inclusion criteria of studies included the presence of an adequate control group. The researchers explain and discuss that there are three possible research designs for measuring cognitive ability improvement, each with different advantages. The first research design, called the 'within-subjects-only' incorporates a pre-test, post-test design on one single group with no control group. The second, called the 'between-subjects' design is where an intervention group and a control group are compared on post-test scores. Lastly, the third design, called 'mixed', is when pre-test and post-test measures are taken for both an intervention group and a control group. Each of these designs can differ regarding the implications for any

improvement that the control group if included, may demonstrate. For instance, the within-subjects-only design has no control group and so doesn't take account of any possible practice-effect. The between-subject design has a control group although, since performance is measured only once, it is inferior to the mixed design, which allows effect sizes to be calculated independently for the intervention group and the control group, along with investigating the effect size between the two groups. Therefore the mixed study design is regarded as a gold standard approach for calculating true differences. Uttal et al. only included studies with the mixed design into their meta-analysis. The mixed study design is the approach used for the study presented in this chapter.

There are several other approaches to incorporating the practice effect other than including a control group, such as, administering the test three times and measuring the difference between the first and third tests, as the biggest practice effect is observed between the first and second tests (McCaffrey & Westervelt, 1995). Another approach is to use an alternate version of the same test, but this does not account for the fact that the similarity of the test questions may have an impact (Lezak *et al.*, 2012). This last approach has been used with the Mental Rotation Test (MRT) for an anatomy education study by Cui et al. (2016), where the MRT-B version was used at the post-test (consisting of the same questions as the MRT-A but in a different order). In Cui et al.'s study, a control group was also incorporated into the study design, although the researchers do not state whether this consisted of students studying anatomy or not. Additionally, the performance of the control group was not statistically compared to the 2D or 3D teaching intervention groups, and therefore the practice effect was not accounted for. In Cui et al.'s study, no significant difference was found between the Pre score and the Post scores, and MRT scores were not significantly related to anatomy knowledge as assessed by a fifteen question test on function and structure, spatial relationship questions, and questions based on images of the 3D models used in the study.

Recognition of the practice effect concerning cognitive ability tests has often not been incorporated into anatomy education study designs, such as Rochford (1985), Hegarty (2009), Lufler et al. (2012), Vorstenbosch et al. (2013a), and Gutierrez et al. (2017). This omission has possibly led to the false conclusion that spatial ability improves with the learning of anatomy (Lufler *et al.*, 2012; Vorstenbosch *et al.*, 2013a; Gutierrez *et al.*, 2017). Furthermore, the addition of dichotomising participants into high or low spatial ability further confounds a practice effect. Dichotomising further confounds because dividing participants into extremes and comparing these will increase the likelihood of finding a statistically significant difference (Altman, 2006; Senn, 2003), and combined with a practice effect could exaggerate the difference.

A recent example of a study falling into the ‘trap’ of not acknowledging the practice effect is a veterinary anatomy education study by Gutierrez et al. (2017). In this study first-year veterinary medical students were given three cognitive ability tests to measure spatial ability (Guay’s visualisation of views test and the Mental Rotation Test) and visual reasoning ability (Raven’s Advanced Progressive Matrices Test-short form). The tests were given at the start and end of a thirty-two week integrated anatomy curriculum with no control group or consideration of previous anatomy learning. The study hypothesis was: do spatial visualisation, and visual reasoning ability improve due to exposure to an integrated curriculum? The authors found that students’ scores on each of the three cognitive ability tests statistically significantly improved as would be expected given the practice effect.

There is one anatomical study which has incorporated a control group that participated in the MRT at the same time interval as the intervention groups and incorporated the control group into the statistical analysis to account for the practice effect (Hoyek *et al.*, 2009). In this study, the control group did not learn anatomy and the two intervention groups were enrolled in a functional anatomy learning module. One of the intervention groups additionally received specific training aimed to

improve mental rotation (MR) (therefore creating a total of three groups; control group, non-MR training anatomy group, and an MR training anatomy group).

The training sessions did not involve the MRT or the use of anatomical structures and a total of 12 x 20-minute sessions were given. During these training sessions, the non-MR training anatomy group and control group were engaged in physical activity. The researchers found a statistically significant group effect ( $F = 12.6, p < 0.001$ ) and individual two-sample t-tests found that the MR training anatomy group scored statistically significantly greater than the non-MR anatomy training group ( $t = 4.14, p < 0.001$ ) and the control group ( $t = 4.03, p < 0.001$ ). No statistically significant differences in the MRT scores were found between the non-MR training group and the control group ( $t = 0.3, p > 0.05$ ). However, a post-hoc Tukey test to compare the three groups would have been a more reliable statistical method as this adjusts p-values for multiple testing.

Hoyek et al.'s study (2009) also incorporated an MCQ anatomy test with both spatial and non-spatial (factual) questions, given to both the MR anatomy group and the non-MR anatomy group. Statistical analysis of covariance, with MRT Pre score as a covariate, identified the MR anatomy group did not perform statistically significantly better than the non-MR anatomy group on the spatial questions ( $F = 0.02, p = 0.07$ ), or the non-spatial questions ( $F = 0.02, p = 0.80$ ).

The study presented in this chapter incorporates the practice effect by using the mixed study design described earlier. The study design included a control group of first-year psychology students, and three intervention groups of first-year veterinary students, with each group completing three spatial ability tests (Card Rotation Test (CRT), Mental Rotation Test (MRT), and the Surface Development Test (SDT)). Each of the three spatial ability tests was administered twice, with the same time interval between pre and post-tests for all four cohorts. A generalised linear model

(GLM) statistical approach was taken to incorporate all four cohorts into the one analysis, and a post-hoc Tukey performed to identify which comparison of cohorts improved in spatial ability.

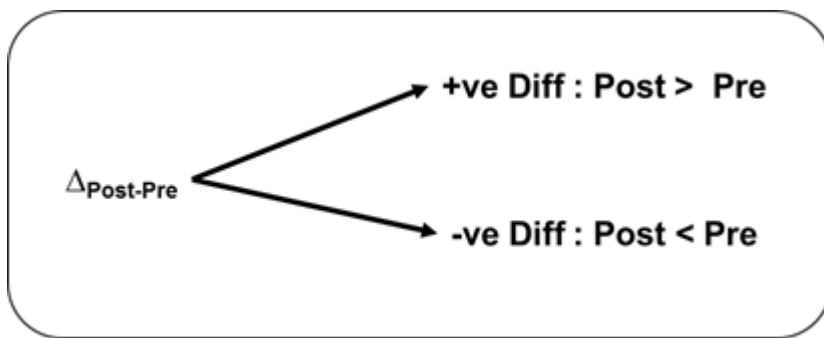
This study aimed to investigate whether learning gross anatomy improves spatial ability in veterinary students across two different sub-categories of spatial ability (Mental Rotation and Spatial Visualisation). Furthermore, to investigate whether teaching 3D and spatially had a larger effect on spatial ability improvement compared to a 2D anatomy teaching method, comparisons were made between University of Edinburgh Veterinary Students 1 (UoE Vet 1) and University of Edinburgh Veterinary Students 2 (UoE Vet 2), who received a 2D and 3D spatial teaching method respectively.

## **7.2 Data Analyses**

### **7.2.1 Descriptive Statistics**

Shapiro-Wilk normality tests were performed and histograms plotted for the continuous parameters pre, post, and diff for each spatial ability test (CRT, MRT, and SDT) and age to check for normality. The spatial ability difference (diff) score was calculated by subtracting the pre-test score from the post-test score ( $\Delta_{\text{Post-Pre}}$ ). If the diff score was positive, this indicated the post score was higher, while a negative diff score indicated a higher pre score (Figure 7.1).





*Figure 7.1 Diagrammatic representation of Diff score.*

Descriptive statistical analysis was performed including the mean or median with accompanying SD or range for each continuous parameter. Spearman Rank correlations were also performed between each spatial ability test score to check for any associations for each cohort.

## **7.2.2 Baseline Comparisons**

The number of participants that sat both the pre and post spatial ability tests are summarised in Figures 3.5, 3.6, 3.7, and 3.8. The number of participants who sat both tests was less than the number of participants that sat the pre-test only, therefore, the baseline parameters (CRT Pre, MRT Pre, and SDT Pre) were re-analysed to identify any possible baseline differences between the four cohorts before comparison of post scores. Additionally, due to the grey printing of some of the MRT test papers, the distribution of grey printed MRT test papers for UoE Vet 1, and UoE Vet 2 were plotted to ensure no disadvantages.

Pairwise Kolmogorov-Smirnov tests were initially performed between each cohort to check for differences between the distributions, with the further comparison by gender and handedness. Individual Generalised Linear Model (GLM, section 3.9.3)

analyses were then performed with the response variable ‘pre’ (CRT Pre, MRT Pre or SDT Pre) and the categorical explanatory variables cohort, gender, and handedness, and the continuous explanatory variable age (Table 7.1).

Parameter	Analysis	Response Variable	Explanatory Variables	
			Categorical	Continuous
CRT Pre	$GLM_P$	CRT Pre score	Cohort, Gender, Handedness	Age
MRT Pre	$GLM_P$	MRT Pre score	Cohort, Gender, Handedness	Age
SDT Pre	$GLM_B$	SDT Pre score	Cohort, Gender, Handedness	Age

**Table 7.1 Summary of  $GLM_{P/B}$  analysis comparing baseline Pre-test score for each spatial ability test.**

If ‘cohort’ was statistically significant and because this explanatory factor had 4 levels (UoE Vet 1, UoE Vet 2, UoB Vet and UoE Psych) an analysis of deviance table was performed on the GLM. The analysis of deviance table was performed to test overall if there was any statistically significant effect of cohort, and a post-hoc Tukey performed if statistically significant to determine which pairwise combinations of cohorts were statistically significantly different to one another.

### **7.2.3 Comparison of Post Spatial Ability Test Scores**

To compare the post-test scores between each cohort separate GLMs were performed, one for each spatial ability test. For each GLM the response variable was ‘post’ e.g., CRT Post, MRT Post, and SDT Post, and the categorical explanatory variables were cohort, gender, and handedness, and the continuous explanatory variables were pre-test score and age. The explanatory variable pre-test score was identified as a confounding variable, and so all risk/odds ratios stated are adjusted for

pre score. Additionally, the GLMs were re-leveled with UoE Psych to be the comparison cohort because this cohort was the control group, i.e. for the GLM each cohort was compared to the control group. A post-hoc Tukey test was performed to compare all pairwise combinations of cohort comparisons if analysis of deviance of the GLM identified cohort as a significant variable. Table 7.2 summaries the GLM analyses conducted to compare post-test scores.

Parameter	Analysis	Response Variable	Explanatory Variables	
			Categorical	Continuous
CRT Post	$GLM_P$	CRT Post score	Cohort, Gender, Handedness	CRT Pre (confounding), Age,
MRT Post	$GLM_P$	MRT Post score	Cohort, Gender, Handedness	MRT Pre (confounding), Age,
SDT Post	$GLM_B$	SDT Post score	Cohort, Gender, Handedness	SDT Pre (confounding), Age,

**Table 7.2 Summary of  $GLM_{P/B}$  analysis comparing Post test scores for each spatial ability test.**

## 7.3 Results

### 7.3.1 Baseline Comparisons

Comparison of the distributions of the spatial ability tests between the three veterinary cohorts identified no statistically significant differences, including when divided by gender or handedness ( $p > 0.05$ ) (Figures 7.2 – 7.5). Tables 7.3 and 7.4 show the mean/median, SD/range, and numbers (and percentages) of participants for each parameter.

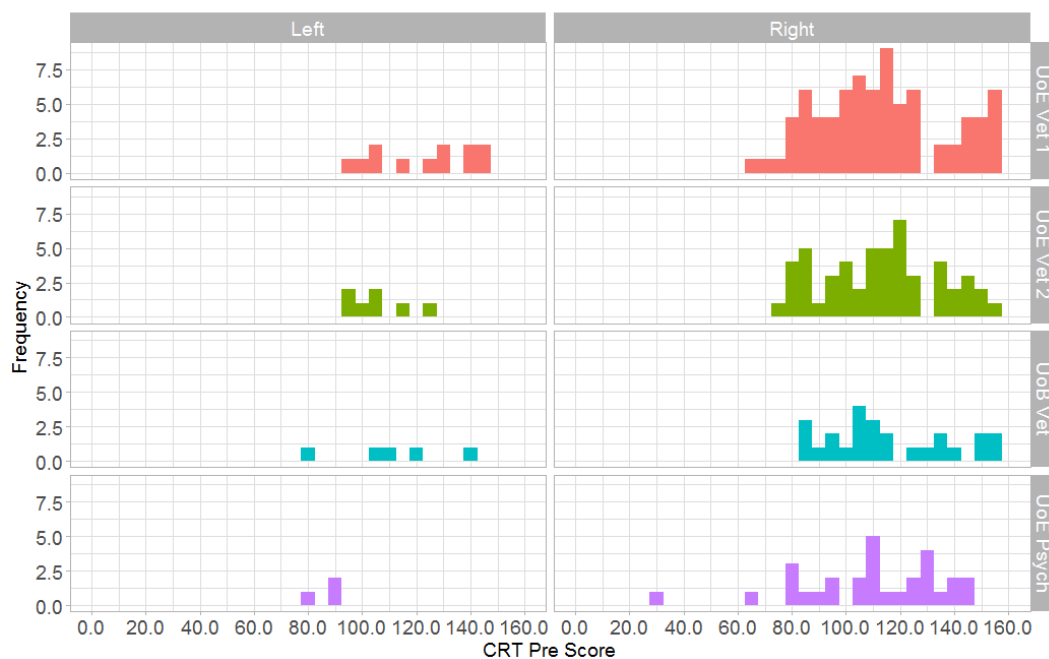
Cohort	Gender n (%)		Handedness n (%)	
	Female	Male	Left	Right
UoE Vet 1	79 (86.8)	12 (13.2)	12 (13.2)	79 (86.8)
UoE Vet 2	52 (86.7)	8 (13.3)	7 (11.7)	53 (88.3)
UoB Vet	27 (84.4)	5 (15.6)	5 (15.6)	27 (84.4)
UoE Psych	25 (75.8)	8 (24.2)	3 (9.1)	30 (90.9)

***Table 7.3 Number and percentage of female, male, left and right-handed participants in each cohort, with Pre and Post test scores.***

Parameter	Max Score	Cohort UoE Vet 1 (n = 91)			Cohort UoE Vet 2 (n = 60)			Cohort UoB Vet (n = 32)			Cohort UoE Psych (n = 33)		
		$\bar{x}/M$	SD	Range	$\bar{x}/M$	SD	Range	$\bar{x}/M$	SD	Range	$\bar{x}/M$	SD	Range
CRT Pre	160	M= 113	-	65 – 160	$\bar{x}$ = 112.4	21.24	-	$\bar{x}$ = 117.8	23.83	-	$\bar{x}$ = 109.1	27.22	-
MRT Pre	24	M= 12	-	4 – 24	$\bar{x}$ = 11.63	4.37	-	$\bar{x}$ = 11.75	5.27	-	$\bar{x}$ = 10.52	3.70	-
SDT Pre	60	M= 0.87	-	0.37 – 1.0	M= 0.88	-	0.4 – 1.0	M= 0.86	-	25 – 60	$\bar{x}$ = 0.68	0.21	-
CRT Post	160	M= 137	-	83 – 160	M= 141	-	74 – 160	M= 140	-	82 – 160	M= 137	-	40 – 160
MRT Post	24	M= 16	-	4 – 24	M= 16	-	0 – 23	$\bar{x}$ = 14.91	5.09	-	$\bar{x}$ = 14.12	4.67	-
SDT Post	60	M= 0.95	-	0.53 – 1.0	M= 0.97	-	39 – 60	M= 0.95	-	0.65 – 1.0	M= 0.83	-	0.38 – 1.0
CRT Diff	-	$\bar{x}$ = 21.55	16.45	-	M= 24.5	-	-41 – 56	$\bar{x}$ = 23.25	15.41	-	$\bar{x}$ = 19.12	16.79	-
MRT Diff	-	$\bar{x}$ = 3.71	3.32	-	M= 4	-	-15 – 10	$\bar{x}$ = 3.16	3.09	-	$\bar{x}$ = 3.61	3.28	-
SDT Diff	-	M= 0.08	-	-0.09 – 0.37	M= 0.08	-	-1 – 34	M= 0.07	-	-0.02 – 0.57	$\bar{x}$ = 0.12	0.14	-
Age	-	M= 18	-	17 – 27	M= 18	-	17 – 24	M= 18	-	18 – 22	M= 19	-	17 – 35

**Table 7.4** Number (n), mean ( $\bar{x}$ ), median (M), standard deviation (SD) and range of Pre, Post and Diff test scores for each spatial ability test for each cohort.

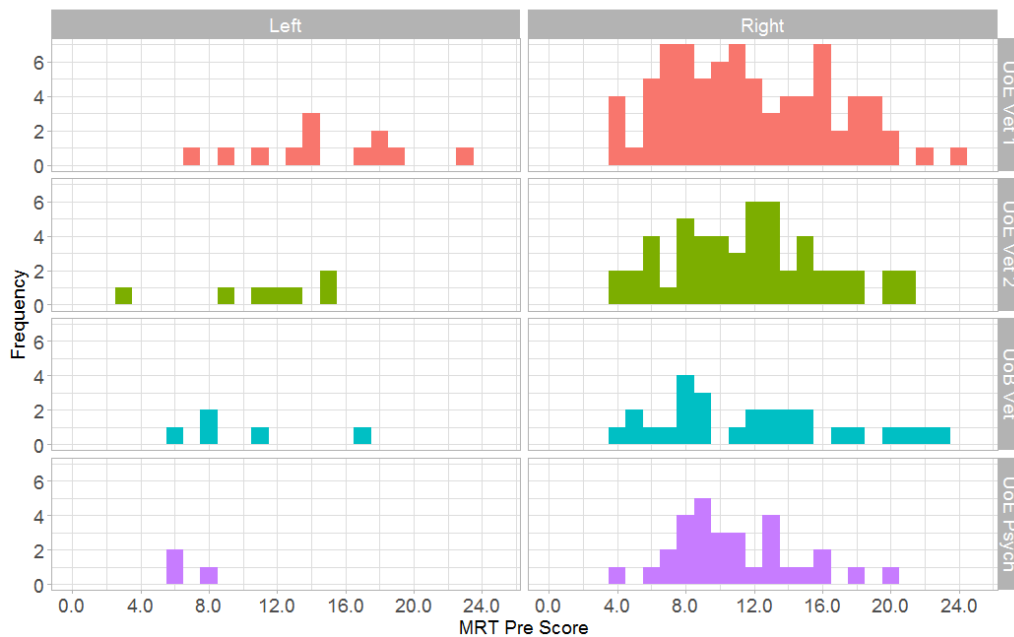
However, statistically significant differences were identified when comparing the UoE psych cohort to the three veterinary cohorts ( $p < 0.05$ ). When comparing the overall distributions of each spatial ability test between the UoE Psych cohort and the three veterinary cohorts, without dividing by gender or handedness, no statistically significant differences were found for CRT Pre ( $p > 0.680$ ). However, when the distribution of the CRT Pre was divided by handedness (left and right), a statistically significant difference was found between the left-handed UoE Psych students, which was more right-skewed, and UoE Vet 1 ( $D = 1, p = 0.016$ ) and UoE Vet 2 ( $D = 1, p = 0.023$ ) (Figure 7.2). No statistically significant differences were identified when divided by gender or for right-handed students ( $p > 0.216$ ).



**Figure 7.2 CRT Pre distribution per cohort divided by handedness.**

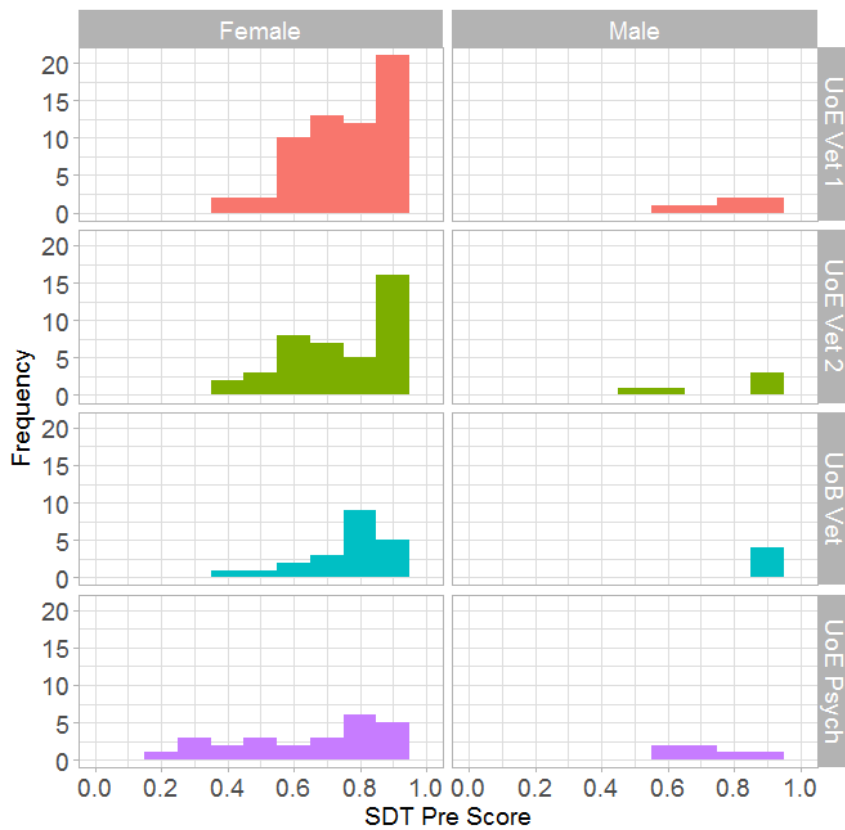
Comparison of the overall distribution of MRT Pre scores between UoE Psych and each of the three veterinary cohorts was not statistically significantly different ( $p > 0.172$ ), however when divided by handedness the UoE Psych left-handed students distribution was more right-skewed than UoE Vet 1 ( $D = 0.92, p = 0.035$ ) (Figure

7.3). There were no statistically significant differences when divided by gender or for right-handed students ( $p > 0.05$ ).



**Figure 7.3 Comparison of MRT Pre distributions divided by handedness.**

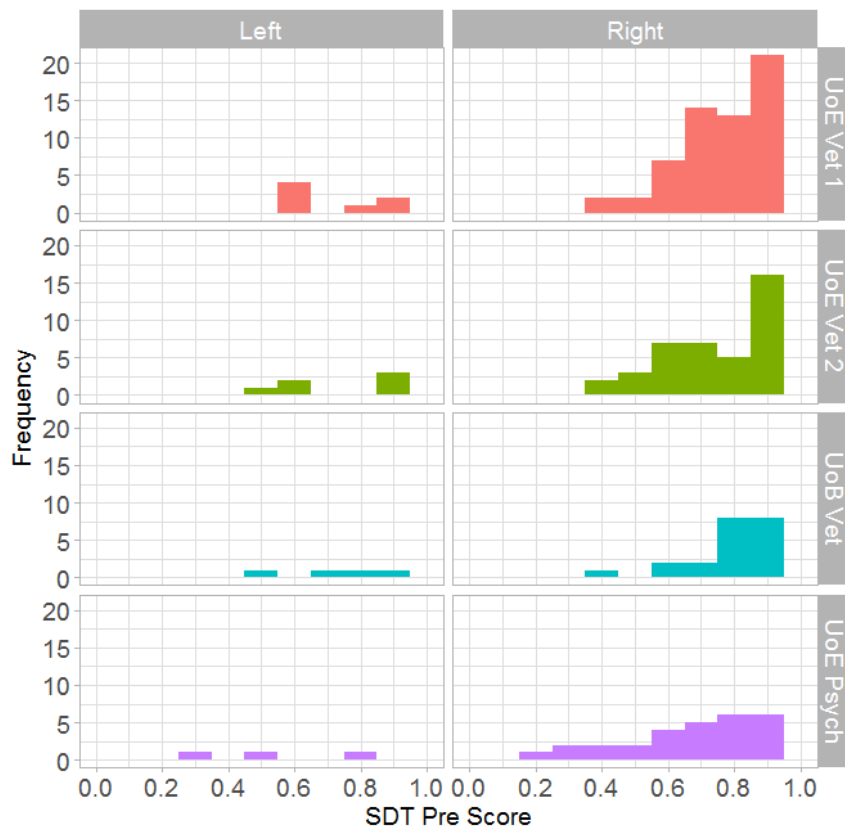
The overall SDT pre-distribution of the UoE Psych cohort was statistically significantly different to each of the three veterinary cohorts ( $p < 0.05$ ). This difference in the SDT pre-distribution was apparent when divided by gender, with the female UoE Psych participants' distribution more right shifted compared to UoE Vet 1 ( $D = 0.36$ ,  $p = 0.013$ ) and UoE Vet 2 ( $D = 0.36$ ,  $p = 0.023$ ), but not for UoB Vet ( $D = 0.37$ ,  $p = 0.055$ ) (Figure 7.4). No statistically significant differences were identified when comparing the male distributions for SDT Pre ( $p > 0.061$ ).



**Figure 7.4 Comparison of SDT Pre distribution divided by gender.**

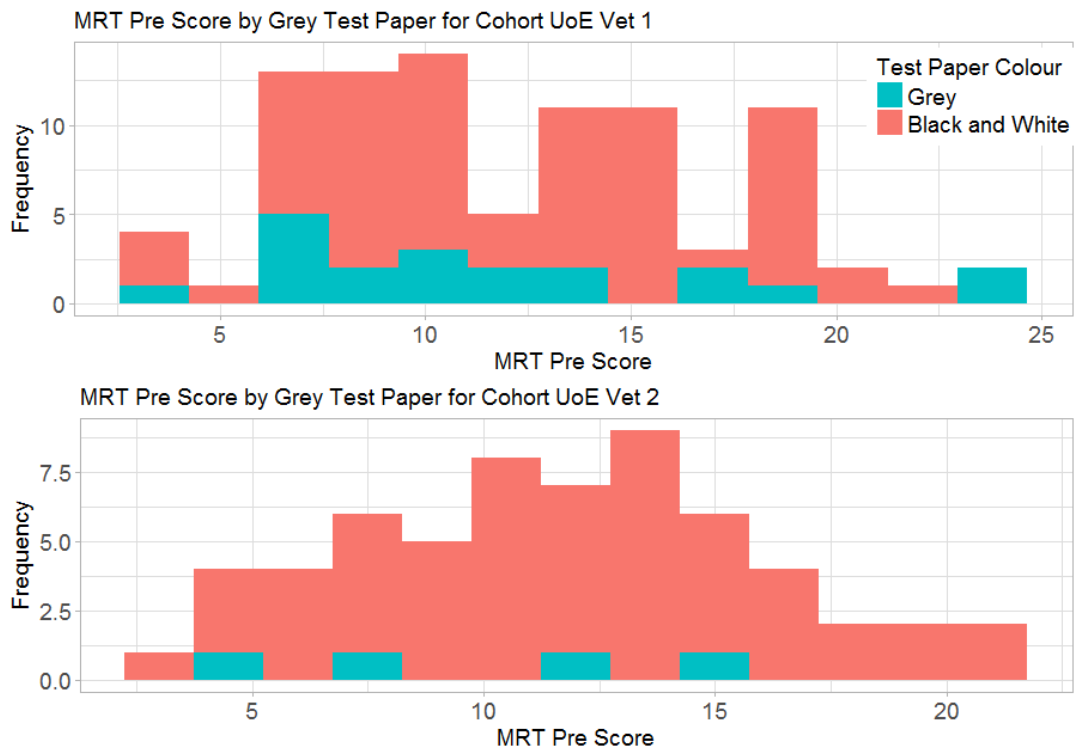
The right handed UoE Psych students SDT Pre distribution was more right skewed compared to UoE Vet 1 ( $D = 0.32$ ,  $p = 0.025$ ) and UoE Vet 2 ( $D = 0.34$ ,  $p\text{-value} = 0.025$ ) (Figure 7.5). No statistical differences were identified for UoB Vet students for handedness ( $p > 0.062$ ).





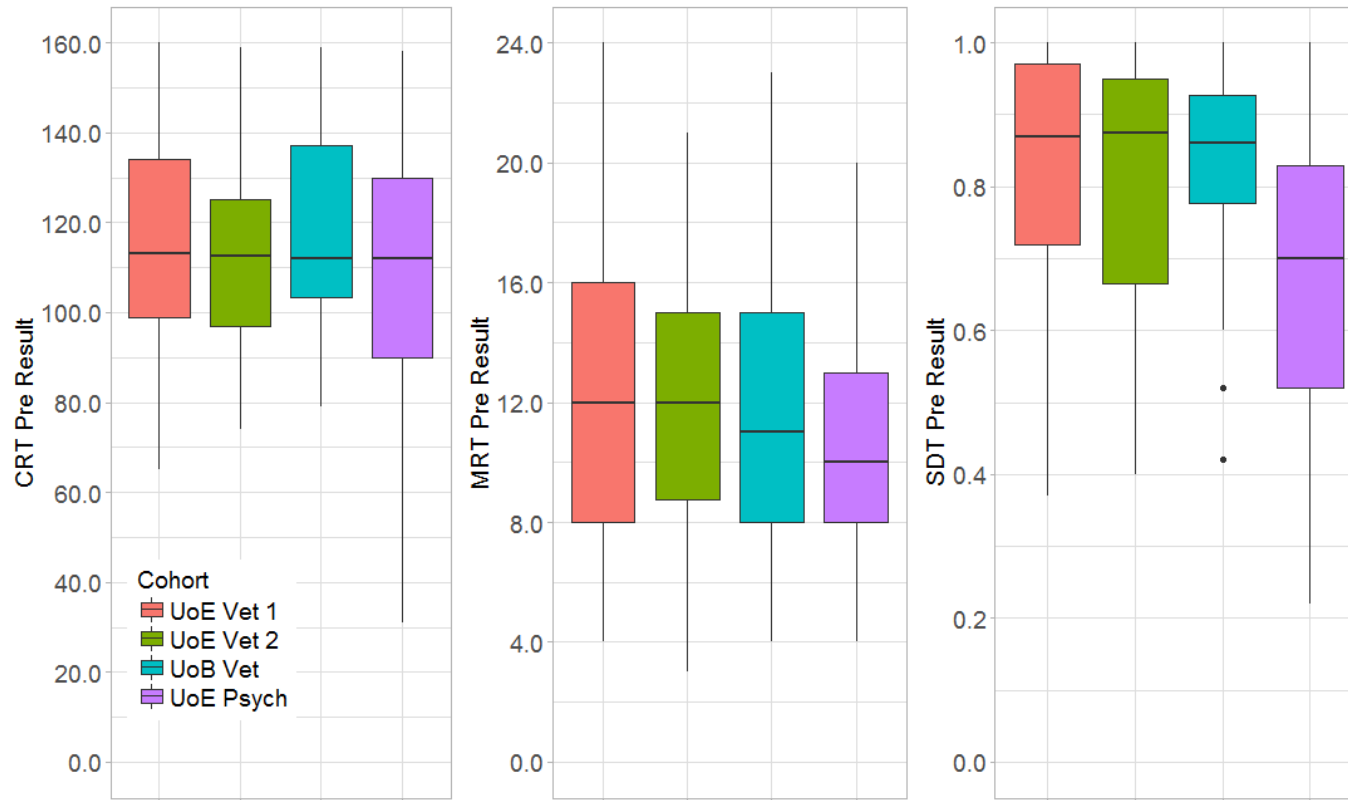
*Figure 7.5 Comparison of SDT Pre distributions divided by handedness.*

The participants in cohorts UoE Vet 1 and UoE Vet 2 with grey printed MRT test papers were randomly distributed, and so included in the analysis (Figure 7.6).



*Figure 7.6 Distribution of grey printed MRT Pre test papers for UoE Vet and UoE Vet 2 for chapter 7.*

Figure 7.7 shows the comparison of each spatial ability test Pre score across each cohort.



*Figure 7.7 Comparison of spatial ability Pre scores.*

The three GLM analyses comparing the spatial ability Pre scores for each spatial ability test identified no statistically significant differences between cohorts for CRT Pre (Table 7.5 and Figure 7.7).

Response Variable	Significant Explanatory Variables	Relative Risk ratio (RR)	95% CI	p-value	Non-significant explanatory variables removed
CRT Pre	Cohort	Ref			Handedness, Age
	UoE Vet 1	0.98	0.91 – 1.04	= 0.482	
	UoE Vet 2	1.02	0.94 – 1.11	= 0.623	
	UoB Vet	0.94	0.86 – 1.02	= 0.117	
	UoE Psych				
	Gender	Ref			
	Female	1.12	1.04 – 1.20	= 0.004	
	Male				

**Table 7.5 GLMp analysis for CRT Pre cohort comparison.**

When comparing MRT Pre scores across cohorts, UoE Psych scored significantly lower for MRT Pre compared to UoE Vet 1 (RR = 0.83, p = 0.024 adjusted for gender, Table 7.6 and Figure 7.7).

Response Variable	Significant Explanatory Variables	Relative Risk Ratio (RR)	95% CI	p-value	Non-significant explanatory variables removed
MRT Pre	Cohort	Ref			Handedness, Age
	UoE Vet 1	0.95	0.84 – 1.08	= 0.445	
	UoE Vet 2	0.96	0.82 – 1.11	= 0.560	
	UoB Vet	0.83	0.71 – 0.97	= 0.024	
	UoE Psych				
	Gender	Ref			
	Female	1.33	1.17 – 1.51	< 0.001	
	Male				

**Table 7.6 GLMp analysis for MRT Pre cohort comparison.**

Comparison of SDT Pre scores between cohorts identified UoE Psych scoring statistically significantly lower than UoE Vet 1 (OR = 0.41,  $p < 0.001$  adjusted for gender, Table 7.7 and Figure 7.7).

Response Variable	Significant Explanatory Variables	Odds Ratio (OR)	95% CI	p-value	Non-significant explanatory variables removed
SDT Pre	Cohort	Ref			Handedness, Age
	UoE Vet 1	0.89	0.63 – 1.26	= 0.507	
	UoB Vet	1.00	0.65 – 1.58	= 0.990	
	UoE Psych	0.41	0.28 – 0.60	< 0.001	
Gender	Female	Ref			
	Male	1.84	1.21 – 2.90	= 0.007	

**Table 7.7 GLM<sub>B</sub> analysis for SDT Pre cohort comparison.**

Only the explanatory variable gender was significant for each spatial ability test: CRT Pre (RR = 1.12,  $p = 0.004$  adjusted for cohort), MRT Pre (RR = 1.33,  $p < 0.001$  adjusted for cohort), and SDT Pre (OR= 1.84,  $p = 0.007$  adjusted for cohort), with males scoring higher as expected (section 2.3.4) (Tables 7.5, 7.6, and 7.7).

Analysis of deviance on each GLM model to identify differences between pairwise combinations of cohorts was only significant for SDT Pre, with post-hoc Tukey analysis identifying the pairwise combinations involving UoE Psych to be statistically significantly different ( $p < 0.01$ ), with UoE Psych scoring lower (Figure 7.7).

These differences in the baseline comparisons of each cohort will be accounted for in the statistical analysis comparing Post scores and on the interpretation of this analysis.

### **7.3.2 Correlations**

Tables 7.8, 7.9, 7.10, and 7.11 show the correlation matrices of each cohort for the pre, post, and diff scores for each spatial ability test.

Cohort UoE Vet 1	MRT Pre	SDT Pre	CRT Post	MRT Post	SDT Post	CRT Diff	MRT Diff	SDT Diff
CRT Pre	$r_s = 0.58$ $p < 0.001$	$r_s = 0.49$ $p < 0.001$	$r_s = 0.72$ $p < 0.001$	$r_s = 0.57$ $p < 0.001$	$r_s = 0.37$ $p < 0.001$	$r_s = -0.64$ $p < 0.001$	$r_s = -0.06$ $p = 0.566$	$r_s = -0.43$ $p < 0.001$
MRT Pre		$r_s = 0.58$ $p < 0.001$	$r_s = 0.54$ $p < 0.001$	$r_s = 0.77$ $p < 0.001$	$r_s = 0.44$ $p < 0.001$	$r_s = -0.28$ $p = 0.008$	$r_s = -0.38$ $p < 0.001$	$r_s = -0.54$ $p < 0.001$
SDT Pre			$r_s = 0.48$ $p < 0.001$	$r_s = 0.58$ $p < 0.001$	$r_s = 0.78$ $p < 0.001$	$r_s = -0.25$ $p = 0.018$	$r_s = -0.05$ $p = 0.639$	$r_s = -0.86$ $p < 0.001$
CRT Post				$r_s = 0.52$ $p < 0.001$	$r_s = 0.44$ $p < 0.001$	$r_s = 0.02$ $p = 0.856$	$r_s = -0.05$ $p = 0.654$	$r_s = -0.37$ $p < 0.001$
MRT Post					$r_s = 0.50$ $p < 0.001$	$r_s = -0.29$ $p = 0.006$	$r_s = 0.26$ $p = 0.013$	$r_s = -0.48$ $p < 0.001$
SDT Post						$r_s = -0.12$ $p = 0.263$	$r_s = 0.08$ $p = 0.462$	$r_s = -0.39$ $p < 0.001$
CRT Diff							$r_s = 0.04$ $p = 0.695$	$r_s = 0.26$ $p = 0.012$
MRT Diff								$r_s = 0.12$ $p = 0.260$

**Table 7.8 Correlation matrix of each spatial ability test Pre, Post, and Diff score for cohort UoE Vet 1. Yellow = statistically significant.**

Cohort UoE Vet 2	MRT Pre	SDT Pre	CRT Post	MRT Post	SDT Post	CRT Diff	MRT Diff	SDT Diff
CRT Pre	$r_s = 0.47$ $p < 0.001$	$r_s = 0.42$ $p = 0.001$	$r_s = 0.76$ $p < 0.001$	$r_s = 0.30$ $p = 0.022$	$r_s = 0.43$ $p = 0.001$	$r_s = -0.43$ $p = 0.001$	$r_s = -0.26$ $p = 0.044$	$r_s = -0.29$ $p = 0.027$
MRT Pre		$r_s = 0.47$ $p < 0.001$	$r_s = 0.55$ $p < 0.001$	$r_s = 0.71$ $p < 0.001$	$r_s = 0.43$ $p = 0.001$	$r_s = 0.03$ $p = 0.797$	$r_s = -0.32$ $p = 0.012$	$r_s = -0.38$ $p = 0.003$
SDT Pre			$r_s = 0.55$ $p < 0.001$	$r_s = 0.43$ $p = 0.001$	$r_s = 0.76$ $p < 0.001$	$r_s = 0.02$ $p = 0.860$	$r_s = 0.0$ $p = 0.973$	$r_s = -0.88$ $p < 0.001$
CRT Post				$r_s = 0.37$ $p = 0.003$	$r_s = 0.51$ $p < 0.001$	$r_s = 0.16$ $p = 0.214$	$r_s = -0.22$ $p = 0.086$	$r_s = -0.41$ $p = 0.002$
MRT Post					$r_s = 0.43$ $p = 0.001$	$r_s = 0.0$ $p = 0.974$	$r_s = 0.34$ $p = 0.007$	$r_s = -0.34$ $p = 0.007$
SDT Post						$r_s = 0.01$ $p = 0.948$	$r_s = 0.01$ $p = 0.961$	$r_s = -0.44$ $p < 0.001$
CRT Diff							$r_s = 0.06$ $p = 0.670$	$r_s = -0.02$ $p = 0.908$
MRT Diff								$r_s = -0.01$ $p = 0.930$

**Table 7.9 Correlation matrix of each spatial ability test Pre, Post, and Diff for cohort UoE Vet 2. Yellow = statistically significant.**



Cohort UoB Vet	MRT Pre	SDT Pre	CRT Post	MRT Post	SDT Post	CRT Diff	MRT Diff	SDT Diff
CRT Pre	$r_s = 0.55$ $p = 0.001$	$r_s = 0.43$ $p = 0.014$	$r_s = 0.82$ $p < 0.001$	$r_s = 0.55$ $p = 0.001$	$r_s = 0.36$ $p = 0.043$	$r_s = -0.60$ $p < 0.001$	$r_s = -0.17$ $p = 0.361$	$r_s = -0.29$ $p = 0.104$
MRT Pre		$r_s = 0.51$ $p = 0.003$	$r_s = 0.44$ $p = 0.012$	$r_s = 0.85$ $p < 0.001$	$r_s = 0.47$ $p = 0.007$	$r_s = -0.44$ $p = 0.011$	$r_s = -0.35$ $p = 0.048$	$r_s = -0.28$ $p = 0.116$
SDT Pre			$r_s = 0.44$ $p = 0.011$	$r_s = 0.54$ $p = 0.001$	$r_s = 0.68$ $p < 0.001$	$r_s = -0.26$ $p = 0.150$	$r_s = 0.03$ $p = 0.867$	$r_s = -0.73$ $p < 0.001$
CRT Post				$r_s = 0.49$ $p = 0.005$	$r_s = 0.44$ $p = 0.011$	$r_s = -0.18$ $p = 0.337$	$r_s = -0.03$ $p = 0.876$	$r_s = -0.25$ $p = 0.165$
MRT Post					$r_s = 0.44$ $p = 0.012$	$r_s = -0.36$ $p = 0.045$	$r_s = 0.15$ $p = 0.406$	$r_s = -0.32$ $p = 0.078$
SDT Post						$r_s = -0.13$ $p = 0.479$	$r_s = -0.11$ $p = 0.546$	$r_s = -0.10$ $p = 0.587$
CRT Diff							$r_s = 0.31$ $p = 0.080$	$r_s = 0.24$ $p = 0.185$
MRT Diff								$r_s = -0.05$ $p = 0.790$

**Table 7.10 Correlation matrix of each spatial ability test Pre, Post, and Diff score for cohort UoB Vet. Yellow = statistically significant.**

Cohort UoE Psych	MRT Pre	SDT Pre	CRT Post	MRT Post	SDT Post	CRT Diff	MRT Diff	SDT Diff
CRT Pre	$r_s = 0.44$ p = 0.01	$r_s = 0.56$ p = 0.001	$r_s = 0.76$ p < 0.001	$r_s = 0.45$ p = 0.008	$r_s = 0.67$ p < 0.001	$r_s = -0.29$ p = 0.105	$r_s = 0.13$ p = 0.483	$r_s = -0.03$ p = 0.881
MRT Pre		$r_s = 0.30$ p = 0.086	$r_s = 0.44$ p = 0.010	$r_s = 0.64$ p < 0.001	$r_s = 0.57$ p = 0.001	$r_s = 0.04$ p = 0.807	$r_s = -0.18$ p = 0.313	$r_s = 0.28$ p = 0.189
SDT Pre			$r_s = 0.48$ p = 0.005	$r_s = 0.53$ p = 0.002	$r_s = 0.76$ p < 0.001	$r_s = -0.12$ p = 0.498	$r_s = 0.30$ p = 0.088	$r_s = -0.54$ p = 0.001
CRT Post				$r_s = 0.44$ p = 0.01	$r_s = 0.61$ p < 0.001	$r_s = 0.34$ p = 0.050	$r_s = 0.14$ p = 0.454	$r_s = 0.10$ p = 0.569
MRT Post					$r_s = 0.52$ p = 0.002	$r_s = -0.07$ p = 0.707	$r_s = 0.61$ p < 0.001	$r_s = -0.08$ p = 0.661
SDT Post						$r_s = -0.02$ p = 0.908	$r_s = 0.10$ p = 0.592	$r_s = 0.08$ p = 0.645
CRT Diff							$r_s = -0.09$ p = 0.618	$r_s = 0.28$ p = 0.113
MRT Diff								$r_s = -0.29$ p = 0.106

**Table 7.11 Correlation matrix for each spatial ability test Pre, Post, and Diff score for cohort UoE Psych. Yellow = statistically significant.**

Each of the correlations between the pre and post scores were positive, significant ( $p < 0.05$ ), and moderate to high for all four cohorts (pre:  $r_s = |0.58|$  and post:  $r_s = |0.61|$ ), as would be expected according to the positive manifold effect (section 2.2.1) (Tables 7.8, 7.9, 7.10, and 7.11).

The positive correlations indicate that all three spatial ability tests are assessing the domain of spatial ability. The positive correlation identifies that as pre score increases, post score increases which could be due to the practice effect and/or because of the learning of anatomy. The GLM analyses presented in the next section of this chapter will statistically analyse whether the learning of anatomy improved spatial ability by comparing each veterinary cohort to the control cohort (psychology students) to account for the practice effect.

The correlations of the diff score to the pre score for each spatial ability test (CRT, MRT, or SDT) were negative and statistically significant for each cohort ( $r_s = |-0.88|$ ,  $p < 0.05$ ), except for UoE Psych, where only the spatial ability test SDT demonstrated a statistically significant correlation. A negative correlation between diff score and pre score broadly means as diff score increases (i.e., the post score was higher than the pre score) the pre score decreases, indicating improvement in spatial ability. This improvement could be due to the practice effect or due to the learning of anatomy, which the GLM analysis in the next section will investigate.

The correlation relationship between diff score and post score was also analysed to investigate spatial ability improvement. For this relationship to broadly demonstrate improvement, a positive correlation would be expected. A positive relationship would demonstrate as post score increases diff score increases (i.e., post score was higher than pre score).

The correlations between post score and diff score were statistically significant for cohort UoE Vet 1 for MRT and SDT (Table 7.8), with MRT showing a positive relationship and SDT a negative relationship. Cohort UoE Vet 2 also showed a statistically significant positive relationship for MRT and a negative relationship for SDT (Table 7.9). Cohort UoB Vet demonstrated no statistically significant relationship for all of the three spatial ability test correlations between post and diff scores (Table 7.10) while UoE Psych demonstrated statistically significant positive correlation for MRT (Table 7.11).

### 7.3.3 Comparison of Post Spatial Ability Test Scores

Table 7.9 summarises the separate  $GLM_{P/B}$  outputs for the comparison of each veterinary cohort (UoE Vet 1, UoE Vet 2, and UoB Vet) to the control cohort (UoE Psych) for each spatial ability test.

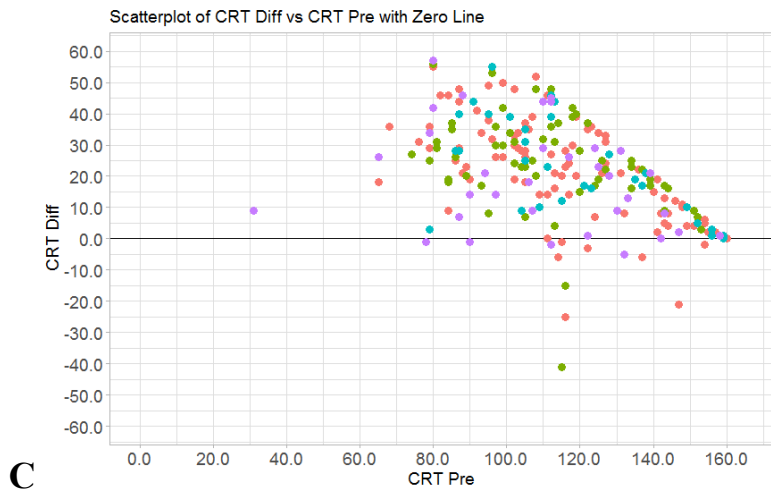
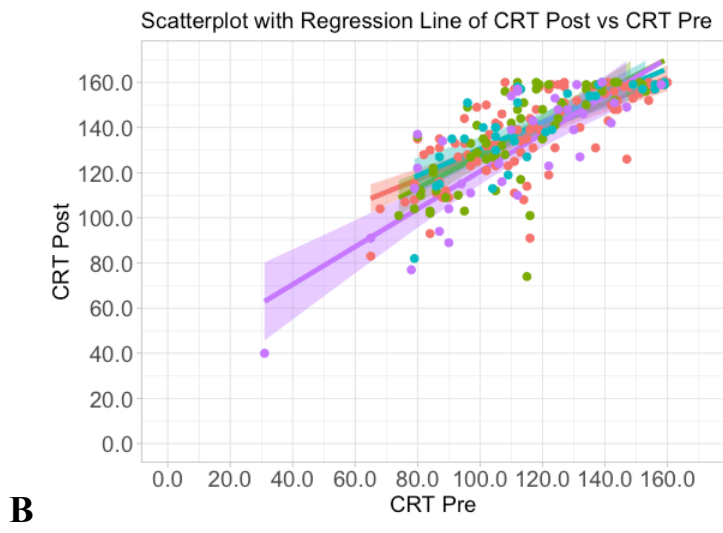
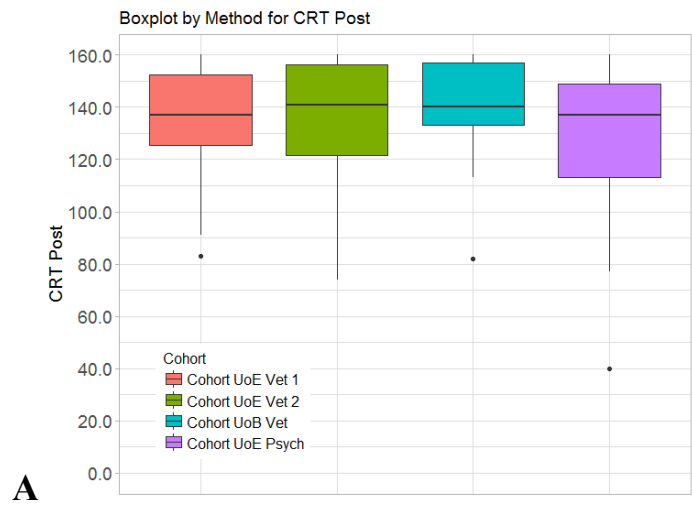
Response Variable	Significant Explanatory Variables	RR/OR	95% CI	p-value	Non-significant Explanatory Variables Removed
CRT Post (GLM <sub>P</sub> )	CRT Pre	1.0	1.00 – 1.01	< 0.001	Gender Handedness Age
	Cohort UoE Psych	Ref			
	UoE Vet 1	1.04	0.99 – 1.08	= 0.111	
	UoE Vet 2	1.05	1.00 – 1.10	= 0.049	
	UoB Vet	1.06	1.00 – 1.12	= 0.047	
MRT Post (GLM <sub>P</sub> )	MRT Pre	1.05	1.04 – 1.05	< 0.001	Handedness Age
	Cohort UoE Psych	Ref			
	UoE Vet 1	1.05	0.94 – 1.17	= 0.390	
	UoE Vet 2	1.03	0.92 – 1.15	= 0.602	
	UoB Vet	0.99	0.87 – 1.13	= 0.908	
	Gender Female Male	Ref 1.11	1.01 – 1.21	= 0.028	

Response Variable	Significant Explanatory Variable	RR/OR	95% CI	p-value	Non-significant Explanatory Variables Removed
SDT Post (GLM <sub>B</sub> )	SDT Pre	81.91	44.24 – 153.07	< 0.001	Handedness Age
	Cohort UoE Psych	Ref			
	UoE Vet 1	1.46	1.08 – 1.97	= 0.014	
	UoE Vet 2	1.86	1.34 – 2.58	< 0.001	
	UoB Vet	1.43	0.98 – 2.12	= 0.070	
Gender Female	Ref				
Male	1.69	1.15 – 2.56	= 0.011		

**Table 7.12 Summary of GLM<sub>P/B</sub> output for each spatial ability test Post score as the response variable.**

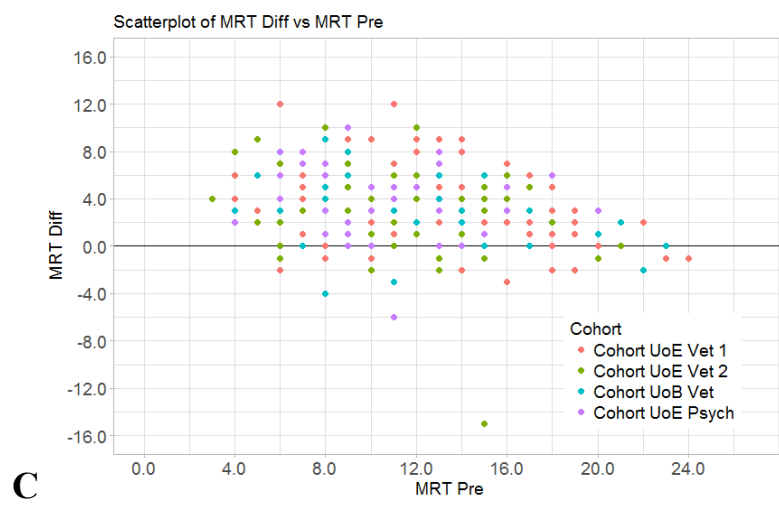
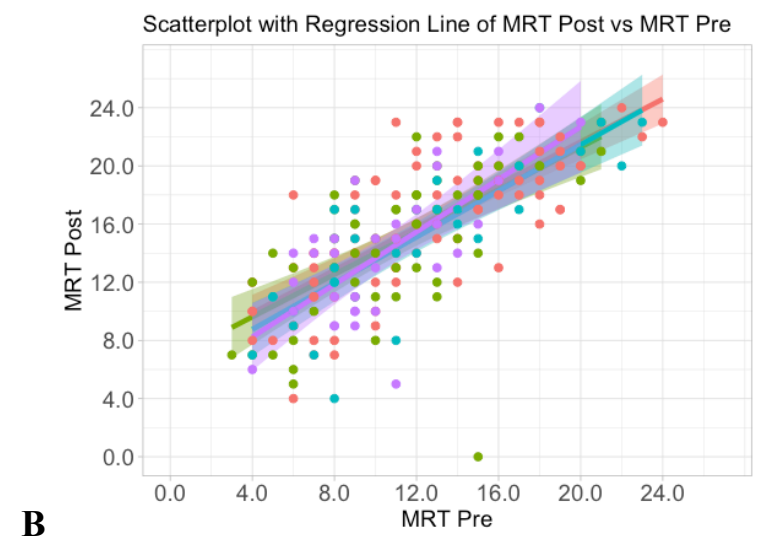
Pre score was shown to be statistically significant for all three spatial ability test GLMs meaning if a student scored well on the pre-test they scored well on the post-test. Additionally, the pre and post scores are highly and statistically significantly correlated to one another (Table 7.8, 7.9, 7.10 and, 7.11,  $p < 0.05$ ), and so pre-test score was a confounding variable. Thus, pre score for all three spatial ability tests remained in the GLM as a covariate. Furthermore, the analysis above comparing the baseline pre score for each test showed no differences between the four cohorts for CRT, and showed the control cohort (UoE Psych) to score lower on the MRT to UoE Vet 1 and on the SDT to all three veterinary cohorts (UoE Vet 1, UoE Vet 2, and UoB Vet). To account for these variations in pre score, pre score was included in each GLM when comparing post scores.

Gender, age, and handedness were not statistically significant for CRT Post scores (Table 7.12). Cohort UoE Vet 2 were 1.05 and Cohort UoB Vet 1.06 times as likely of scoring higher on the CRT Post compared to students who do not learn anatomy when adjusted for CRT Pre score ( $RR = 1.05$ ,  $p = 0.049$ , and,  $RR = 1.06$ ,  $p = 0.047$ , respectively). Cohort UoE Vet 1 CRT Post scores were not significantly different to students who do not learn anatomy ( $RR = 1.04$ ,  $p = 0.111$ ), and post-hoc Tukey analysis revealed no significant differences between pairwise combinations of cohorts for CRT Post (Figure 7.8 A and B).



**Figure 7.8** *A = Comparison of CRT Post score for each cohort. B = CRT Post score vs CRT Pre score for each cohort. C= CRT Diff score vs CRT Pre score for each cohort. The key for all three graphs is as figure A.*

For the MRT Post scores, no statistically significant differences between students who learned anatomy to students who did not learn anatomy when adjusted for MRT Pre score and gender were found (UoE Vet 1: RR = 1.05,  $p = 0.390$ , UoE Vet 2: RR = 1.03,  $p = 0.602$ , and, UoB Vet: RR = 0.99,  $p = 0.908$ ). Male students were 1.11 times as likely to score higher on the MRT Post compared to female students when adjusted for MRT Pre score and cohort (RR = 1.11,  $p = 0.028$ ). This gender effect was also demonstrated in chapter 4 and is in line with the psychology literature on gender differences for tests of mental rotation (section 2.3.4). Post-hoc Tukey analysis revealed no statistically significant differences between pairwise combinations of cohorts for MRT Post ( $p > 0.05$ ) (Figure 7.9).



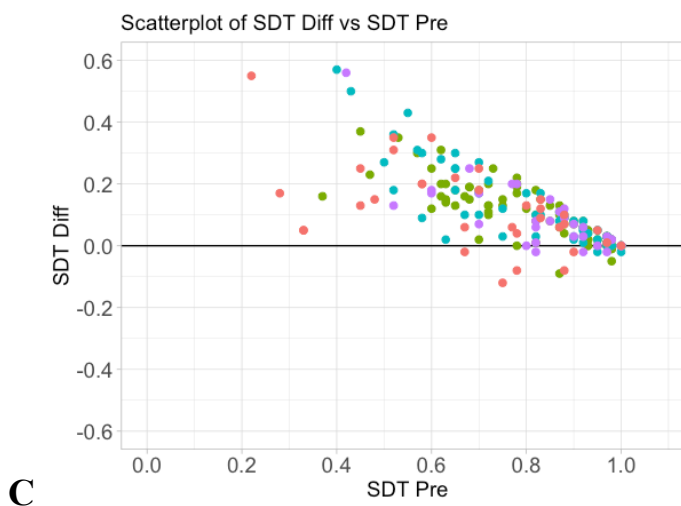
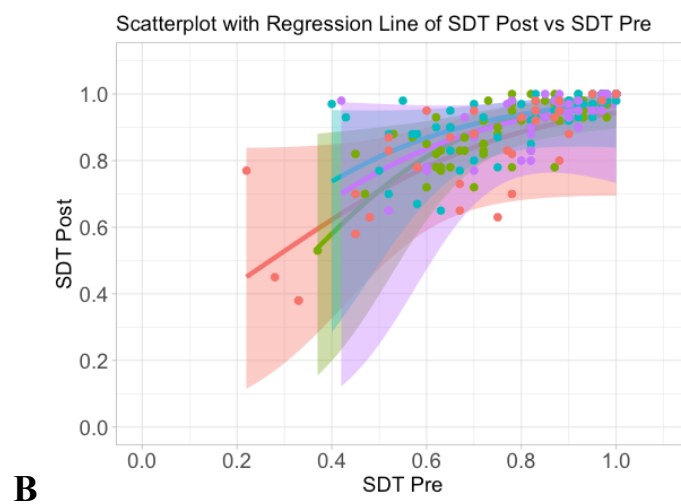
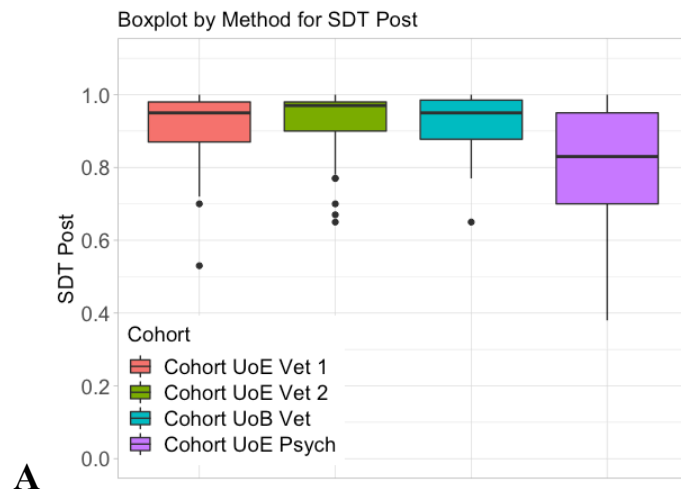
**Figure 7.9** A= Comparison of MRT Post score for each cohort. B= MRT Post score vs MRT Pre score for each cohort. C= MRT Diff score vs CRT Pre score for each cohort. The key for all three graphs is as figure C



These statistical findings demonstrate the 3D mental rotation sub-category of spatial ability did not improve with the learning of anatomy, regardless of whether this learning had a 2D or 3D spatial emphasis in nature.

For each 1% increase in the SDT Post test score, UoE Vet 1 were 1.46 odds and UoE Vet 2 were 1.86 odds of scoring higher than students who do not learn anatomy (OR = 1.46,  $p = 0.014$ ; OR = 1.86,  $p < 0.001$  respectively, adjusted for SDT Pre score and gender). While the difference between UoB Vet students and students who do not learn anatomy was not statistically significant (OR = 1.43,  $p = 0.070$  adjusted for SDT Pre score and gender). Male students had a higher odds (OR 1.69,  $p = 0.011$ ) of scoring higher on the SDT Post compared to female students when adjusted for SDT Pre score and cohort. Post-hoc Tukey analysis identified the same statistical difference between UoE Vet 2 and UoE Psych (estimate = 0.62, SE = 0.17,  $z = 3.74$ ,  $p < 0.001$ ) with UoE Vet 2 scoring higher (Figure 7.10).

These findings show that students learning anatomy score higher on second sitting of the SDT compared to students who do not learn anatomy. However, there was no statistically significant difference in spatial ability improvement between a 2D and a 3D/spatial teaching method.



**Figure 7.10** *A= Comparison of SDT Post score for each cohort. B= SDT Post score vs SDT Pre score for each cohort. C= SDT Diff score vs SdT Pre score for each cohort. The key for all three graphs is as figure A.*

## 7.4 Discussion

### 7.4.1 Does spatial ability improve with the learning of anatomy?

This chapter develops the growing anatomy education research investigating whether spatial ability is improved by anatomy learning (Rochford, 1985; Vorstenbosch *et al.*, 2013a; Gutierrez *et al.*, 2017; Hegarty *et al.*, 2009). The study presented in this chapter is one of the first to research this effect in veterinary students and account for the practice effect by incorporating a control group who do not learn anatomy (Uttal *et al.*, 2013; Lezak *et al.*, 2012:chap.5; McCaffrey & Westervelt, 1995; Goldberg *et al.*, 2015). Furthermore, the study presented in this chapter also takes a novel approach by investigating two different sub-categories of spatial ability as measured by three spatial ability tests (Card Rotation Test (Ekstrom *et al.*, 1976), Mental Rotation Test (Peters *et al.*, 1995), and Surface Development Test (Ekstrom *et al.*, 1976)). In this study, spatial ability was shown to improve for the CRT (Card Rotation Test) and SDT (Surface Development Test), but not for the MRT (Mental Rotation Test). To account for the practice/re-test effect a control cohort of psychology students were incorporated into the study design and regression analysis conducted to account for a practice effect.

A practice or re-test effect was observed for each cohort with 75 - 97% of students exhibiting a positive Diff score across the four cohorts (Table 7.13). The proportion of students exhibiting a negative diff score ranged between 0 – 15% and a zero diff score between 2 – 18% (Table 7.13).

Cohort	SA Test	Diff score		
		-ve Count (%)	0 Count (%)	+ve Count (%)
UoE Vet 1	CRT	7(8)	2(2)	82(90)
	MRT	10(11)	6(7)	75(82)
	SDT	5(5)	16(18)	70(77)
UoE Vet 2	CRT	2(3)	1(2)	57(95)
	MRT	7(12)	3(5)	50(83)
	SDT	2(3)	7(12)	51(85)
UoB Vet	CRT	0(0)	1(3)	31(97)
	MRT	3(9)	4(13)	25(78)
	SDT	3(9)	5(16)	24(75)
UoE Psych	CRT	4(12)	1(3)	28(85)
	MRT	1(3)	4(12)	28(85)
	SDT	5(15)	1(3)	27(82)

**Table 7.13** Count and proportion of students with a negative, zero or positive diff score for each SA test.

The regression analysis results showed the CRT measuring the sub-category of mental rotation in 2D improved for two cohorts of students who learn anatomy (UoE Vet 2 and UoB Vet). The CRT score improvement could be because of the prominent daily use of the skill of mental rotation of anatomical objects in two dimensions when learning anatomy (section 1.3.1). The students receiving the 3D spatial teaching method were shown to have a significant increase in post scores while the students receiving the 2D teaching method did not compared to students who do not learn anatomy. This finding could be because the 3D spatial teaching method improved the students' ability to think spatially and to problem-solve. However, direct comparison of the 2D teaching method to the 3D spatial method by post-hoc Tukey analysis found no statistical difference between the two methods post test scores. Furthermore, the effect size of the increase for CRT is small ( $RR = 1.05$ , equivalent to scoring 105% compared to the control group) and could possibly also be explained by the socioeconomic variance of the control cohort to the intervention cohorts (see section 7.4.3 below).

The Surface Development Test (SDT) scores also demonstrated spatial ability improvement (UoE Vet 1 and UoE Vet 2). However, the distributions of the pre scores of the SDT for all three veterinary cohorts were skewed to the left, particularly for the three veterinary cohorts and less for the control cohort of psychology students. The left skewed distribution of scores for the veterinary cohorts scores indicated a ceiling effect meaning any improvement in their post score could be minimal and possibly reflect the same score as their pre score. This ceiling effect was not shown for the psychology students, and so although there was an improvement in spatial ability scores, this could likely reflect that the Surface Development Test was less challenging for the veterinary students than the psychology students, because the veterinary student scores' exhibited a ceiling effect.

Interestingly for the MRT, there was no statistically significant improvement in spatial ability, as adjusted for pre score when compared to a cohort of students who do not learn anatomy. The reason for no improvement could be because this test was challenging. The MRT distribution has often shown a positive skew meaning most scores are within the lower end of the distribution, suggesting this spatial ability test was harder. Students also subjectively voiced this after the spatial ability test administration (verbal communication identified during focus groups). If learning gross anatomy improves spatial ability and the MRT is the more challenging test, then the biggest difference between pre and post score might be expected with this test, however, this was not found in this study. For the MRT, male students were 1.11 times more likely to score higher on the post-test than female students (when adjusted for pre score) identifying the male students were better at 3D mental rotation supporting the literature on gender differences (Masters & Sanders, 1993; Reilly & Neumann, 2013).

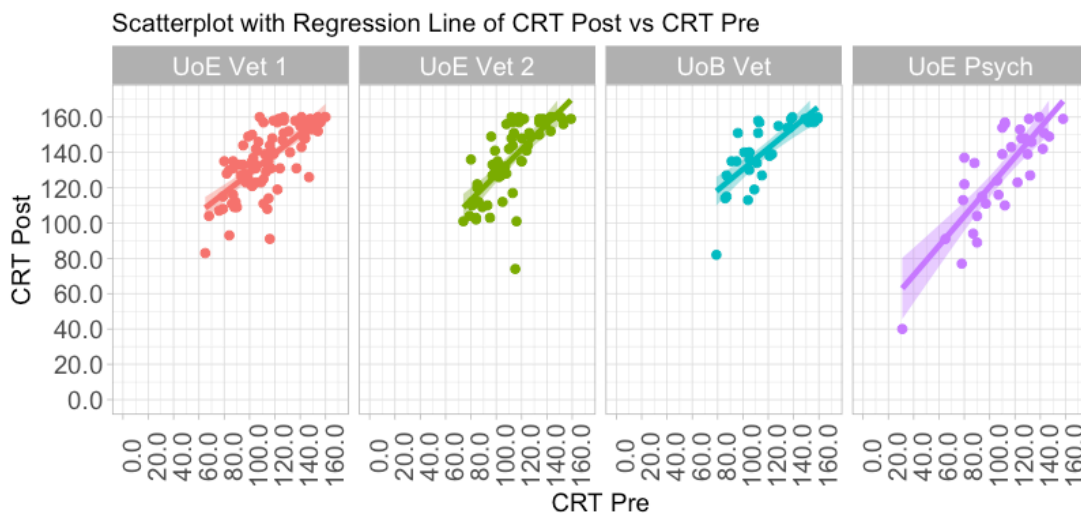
In this study correlations were also conducted to investigate the relationship between spatial ability improvement and anatomy learning. The correlations broadly indicated spatial ability improvement. However, the correlation between post and diff scores

for SDT was negatively correlated for UoE Vet 1 and UoE Vet 2 (Tables 7.8 and 7.9). A negative correlation indicates no improvement in spatial ability. However, the SDT scores for participants from these two cohorts exhibited a ceiling effect for the pre-test and the post-test. Thus, the diff score, which would be the difference between a high pre score and a high post score for this spatial ability test, would be low indicating little to no improvement in spatial ability. Therefore, negative correlations for cohorts UoE Vet 1 and UoE Vet 2 are in keeping with the ceiling effect exhibited by the SDT scores. A negative correlation between post and diff scores broadly indicates no improvement in scores. e.g., as post score increases diff score decreases (i.e., pre score is higher or the same as post score). The correlation between the diff and post scores for UoB Vet and UoE Psych for the SDT were not statistically significant (Table 7.10 and 7.11).

An important factor to consider and discuss with the correlation analyses above (Tables 7.8 – 7.11) is the assumption of consistent gain in spatial ability across students, i.e. the correlations only broadly explain the relationship between the pre, post, and diff scores and assume any gain in spatial ability is consistent across students. However, consistent gain across each cohort and each student is unlikely to be the case, for example, students with a low baseline spatial ability may demonstrate a larger improvement while students with a higher baseline spatial ability may demonstrate a small to moderate improvement. Table 7.13 also supports this by demonstrating not all students spatial ability improved.

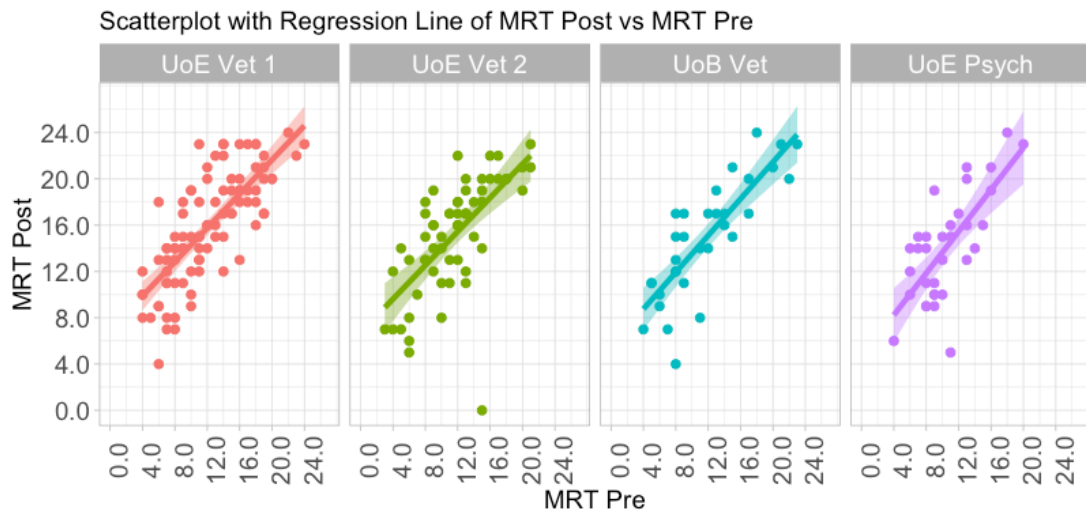
The correlation analysis does not indicate the size of improvement for low and high baseline spatial abilities and the regression analysis did not categorise into these two binary forms (section 3.9.3). However, the gradient of the regression analyses (Figures 7.11, 7.12, 7.13 below) give an indication of the size of improvement for each cohort. The steeper the gradient of the regression slope the larger the improvement in spatial ability.

Comparison of the regression slopes of each cohort for the Card Rotation Test (Figure 7.11) suggest the participants with a lower baseline spatial ability for cohort UoE Psych had a larger improvement in spatial ability compared to the other three cohorts. While comparison of the size of spatial ability improvement for the high baseline spatial ability students suggests a similar improvement in spatial ability. Overall, this suggests the students with a lower baseline CRT spatial ability exhibited a larger improvement for cohort UoE Psych.



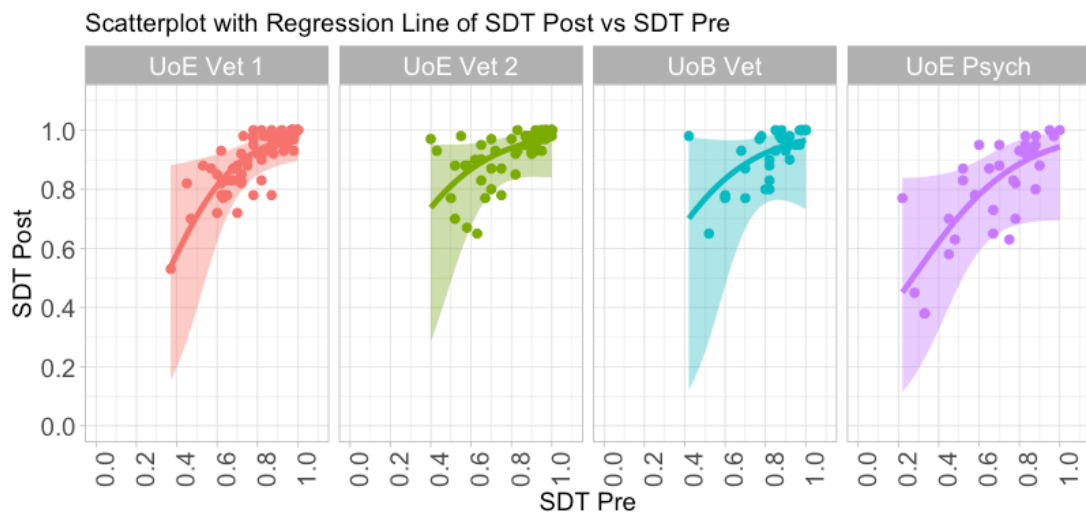
*Figure 7.11 Regression slopes for CRT Post vs CRT Pre per cohort..*

Figure 7.12 shows the regression slopes for each cohort for the Mental Rotation Test. Comparison of the slopes shows similar gradients across the slopes suggesting near equal improvement in spatial ability across all four cohorts and for lower and higher baseline spatial ability scores.



**Figure 7.12** Regression slopes for MRT Post vs MRT Pre per cohort.

Comparison of spatial ability improvement for the Surface Development Test (Figure 7.13) suggests the students with a lower baseline spatial ability for UoE Vet 1 and UoE Psych exhibited a larger improvement, while UoE Vet 2 and UoB Vet displayed a similar improvement. For the students with a higher baseline spatial ability any improvement exhibited appears to be similar across the four cohorts (Figure 7.13).



**Figure 7.13** Regression slopes for SDT Post vs SDT Pre per cohort.



For the process of anatomy learning to improve spatial ability, there must be a positive relationship between the two. To explore this hypothesis, many studies have used examination results as a means to measure learning and have compared this to scores on spatial ability tests (Rochford, 1985; Garg, Norman & Sperotable, 2001; Guillot *et al.*, 2006; Hoyek *et al.*, 2009; Keedy *et al.*, 2011; Lufner *et al.*, 2012; Provo, Lamar & Newby, 2002; Hegarty *et al.*, 2009; Tan *et al.*, 2012; Berney *et al.*, 2015). In chapter 5 of this thesis, anatomy examination scores were compared to scores on spatial ability tests, and it was found the examination scores were not statistically significantly correlated to the CRT and the MRT but did correlate for the SDT (Table 5.2). This finding was also echoed in the GLM models for each examination parameter (Table 5.4).

Other anatomy education studies have shown a positive relationship between examination scores and spatial ability test scores (see systematic review by Langlois *et al.*, 2017). However, the use of examination scores as a measure of learning is a source of variation within itself as the reliability and validity of examinations can vary (e.g., Smith *et al.*, 2018). To investigate this further, the componential approach used by psychologists could be implemented as discussed in chapter 5 (Hegarty & Waller, 2005), or future studies investigating spatial ability and anatomy learning could use another measure of learning rather than examination scores. Recent discussions on this topic within higher education have been focussed around students' learning gain, and currently there are many projects investigating this by bodies such as the Office for Students, Higher Education Academy (now AdvanceHE), the Higher Education Funding Council for England, and the corporation called RAND ('Research ANd Development').

The concept of learning gain developed due to an interest in measuring how much students learn in higher education and has brought about healthy debate (McGrath *et al.*, 2015; Evans, Kandiko Howson & Forsythe, 2018). A report on learning gain by the RAND Corporation (2015:p.xi) defines learning gain "as the difference in student

performance between two stages of their studies, as a variant of the concept of ‘value added’ commonly used in school performance tables, or simply as ‘learning’.” The concept of learning gain is a new field and methods for measuring learning gain are currently being evaluated and debated due to a lack of consensus on an exact definition of what learning gain is (Evans, Kandiko Howson & Forsythe, 2018). Learning gain is perhaps an area that once more developed will help to explore the relationship between anatomy learning and spatial ability further.

To summarise the findings on whether spatial ability improved with the learning of anatomy, this study found this to be true for the CRT and the SDT. However, this finding was not consistent for all three veterinary cohorts. The effect sizes of the improvement were small and could be due to reasons other than the learning of anatomy (see section 7.4.3 below). The next chapter investigates this further by analysing the qualitative views of students on the link between spatial ability and anatomy learning.

#### **7.4.2 Is there a difference between a 2D or 3D spatial teaching method on the improvement of spatial ability?**

This study also assessed whether teaching anatomy with a 2D or a 3D spatial teaching method would improve spatial ability the most. It was expected the 3D spatial teaching method would have the largest effect on improving spatial ability because it was designed to improve the spatial thinking of students, and not only present anatomical material in different dimensional perspectives, as has been done in previous studies (Garg *et al.*, 1999b; Keedy *et al.*, 2011; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Küçük, Kapakin & Göktaş, 2016; Plumley *et al.*, 2013; Smith *et al.*, 2018).

For the CRT, post-hoc Tukey analysis comparing all three veterinary cohorts post-test scores identified no statistically significant differences. Suggesting spatial ability improvement was not significantly different across the veterinary cohorts receiving different teaching methods. However, the 2D teaching method post-test scores did not statistically significantly differ to the control cohort of students who did not learn anatomy while the 3D spatial teaching method post scores did ( $RR = 1.05$ ,  $p = 0.049$ ). These findings demonstrate that the cohort receiving the 2D teaching method (UoE Vet 1) did not score statistically significantly differently on the post-test compared to students who do not learn anatomy ( $RR = 1.04$ ,  $p = 0.111$ ), which could be interpreted as the learning of anatomy in a 2D manner does not improve spatial ability as measured by the CRT. Conversely, the cohort receiving the 3D spatial teaching method scored statistically significantly higher on the post-test to students who do not learn anatomy, and so did a cohort of veterinary students learning anatomy from another academic institution (UoB Vet). Suggesting these two cohorts improved in spatial ability. This could reflect the different curriculum designs of the two institutions, where one teaches a body systems approach from the beginning of the course. It should be noted that although there are significant differences to the control group for UoE Vet 2 and UoB Vet, the effect sizes of these differences are very small and possibly do not reflect a large educational impact.

For the Mental Rotation Test (MRT) post-hoc Tukey analysis, no statistically significant differences were identified between the three veterinary cohorts' post-test scores. Suggesting spatial ability improvement did not vary whether learning anatomy from a 2D or 3D spatial teaching format as measured by the MRT. Furthermore, the MRT post-test scores of anatomy learning veterinary students did not improve statistically significantly to students who do not learn anatomy. These findings of the MRT suggest spatial ability (specifically 3D mental rotation) is not improved by anatomy learning or by different teaching methods.

For the SDT, the 2D and 3D spatial teaching method post scores were both statistically significantly higher compared to students' who do not learn anatomy, and the 3D spatial teaching method exhibited the largest effect size (OR = 1.86 for 3D spatial versus OR = 1.46 for 2D). Suggesting spatial ability improves with the learning of anatomy and the improvement is largest for a 3D spatial teaching method. However, direct comparisons of the three veterinary cohorts' post-test scores identified no statistically significant differences. As previously discussed in chapter 5, six sections out of a total of fourteen were taught with the 3D spatial method. Thus, the effect sizes quoted above may have been larger if the 3D spatial teaching method was delivered across the entire fourteen sections of the gross anatomy course.

In conclusion, the 3D spatial teaching method improved spatial ability more than a 2D teaching method when compared to students who do not learn anatomy as measured by the CRT and SDT. However, this difference was not continued when directly comparing the two teaching methods. The significant effect sizes observed for the 3D spatial teaching method are small and could represent other sources of variation between the two cohorts other than teaching method received. Furthermore, the effect sizes could represent a large or small educational impact. Additional work should be carried out to investigate the potential of the 3D spatial teaching method to improve spatial ability.

### **7.4.3 Study Design**

It is relevant at this stage to discuss some of the limitations to the study presented in this chapter. First, the sample sizes of participants that completed both pre and post-tests were smaller than the total number of participants who sat the pre-test only. These smaller sample sizes should be accounted for when interpreting the statistical analysis of this chapter. To gain an appreciation of the sample size needed to detect

significance a post-hoc sample size calculation was calculated. A sample size of 36 students would be required to detect a small effect size of 0.3 (significance level = 0.05, power = 0.8, degrees of freedom = 5, and a small effect size of 0.3 were set (Cohen, 1988)). Therefore, the sample size of 33 for UoE Psych and 32 for UoB Vet would be too small to detect a small effect size of 0.3 standard deviations (Cohen, 1988).

Another point to consider is that any characteristic biases of the intervention groups and the control group were not accounted for. Randomised controlled trials are considered a 'gold standard' approach to research design because they reduce treatment selection bias (Sullivan, 2011). The study of this chapter was conducted across an academic course, and thus one academic year at a time, therefore a randomised controlled trial could not be implemented. Thus, participants were not randomly assigned to either an intervention or control group, and therefore systematic biases in the populations were not accounted for. Therefore a control group of psychology students, and intervention groups of veterinary students, could have selection biases due to differences in the characteristics of these populations (such as socioeconomic status and academic ability) (Stuart & Rubin, 2008). Stuart and Rubin (2008) in chapter 11 of "Best Practices in Quantitative Methods" explain a way of replicating a randomised experiment for observational studies is to use matching methods. For instance, selecting a sub-sample from the original intervention and control groups that are 'balanced' on possible sources of variance, i.e. the sources of variance are equally distributed in the intervention and the control groups (Stuart & Rubin, 2008). In this study, an attempt to reduce biases in population samples was attempted by initially considering medical students as a control group. However, this was not possible because of the learning of anatomy: it would be difficult to distinguish the causal effect of the medical students' anatomy learning to the veterinary students' anatomy learning. The professional degree of Law was also considered, although was not achieved due to logistic difficulties with degree timetables.

In order to explore the student perspective of the teaching methods, on spatial ability, and the spatial MCQ, the next chapter will present an analysis and discussion of qualitative data from focus groups based on these areas.



# **Chapter 8 Qualitative Analysis of Veterinary Student Experiences of a Spatial MCQ and a 2D or 3D Anatomy Teaching Method**

## **8.1 Introduction**

Chapters four to seven explored the quantitative aspects of this research. This chapter explores the qualitative aspects through thematic analysis of focus group discussions. This analysis aimed to analyse the students' own experiences and views on spatial ability and anatomy learning. This chapter first describes the methods of the thematic analysis. Then the importance of reflexivity in qualitative research is outlined, and the researcher's background and perspectives are described. Lastly, the themes identified from the qualitative analysis are defined and interpreted.

## **8.2 Methods of Qualitative Analysis**

Qualitative data was collected by means of focus groups after The Animal Body 1 (AB1) end-of-course examinations for cohorts University of Edinburgh Veterinary Students 2014/15 (UoE Vet 1) and 2015/16 (UoE Vet 2). Two focus groups, conducted at different dates and times, were organised for each cohort to work around timetables and to recruit as many students as possible. Discussion during the focus groups centered on the in-course spatial MCQ designed for this research, the three spatial ability tests taken by the participants, and the two teaching methods (2D or 3D spatial) (see Appendix 6 for the list and order of questions asked at each focus group). Each focus group was structured in the order of the questions in Appendix 6.



Two staff members were present during the focus groups, the author and a research assistant. The author was present to ensure the questions and discussions were focussed on spatial ability and anatomy teaching. Although, as their teacher, the author's presence could potentially have influenced the students' responses, this was mitigated as far as possible by stating at the start of the focus groups that the students should feel free to speak openly, that the purpose of these discussions was for research only, and that all responses would be anonymised after data collection.

The qualitative data were transcribed by the author verbatim and thematically analysed using the software Nvivo. The thematic analysis coding was checked by the research assistant present during the focus groups to ensure standardisation of coding. All quotes used have been written as intelligible verbatim to optimise comprehension.

### **8.2.1 Thematic Analysis Approach**

Thematic analysis is a flexible approach to analysing qualitative data (Braun & Clarke, 2006) as a variety of different theoretical and epistemological frameworks can be implemented. For instance, themes, or patterns in the data, can be identified by either an inductive (the themes are strongly linked to the data) or theoretical (themes are analyst driven) approach. Different epistemological viewpoints can be taken when handling qualitative data. An essentialist/realist view reports the experiences and realities of the participants, a constructionist view looks at how meanings and experiences relate to society, or a contextualist view combines both essentialist and constructionist views to acknowledge how participants make meanings of their experiences and how society impinges on these meanings (Braun & Clarke, 2006). The themes identified can be at either a semantic level, (i.e., the researcher is not looking for anything different to what a participant has said), or a latent level, (i.e., the researcher looks beyond the semantic meaning to interpret what a participant has said) (Braun & Clarke, 2006).

The data collected from the focus groups for the research presented in this thesis were inductively analysed, to identify semantic themes with an essentialist/realist epistemological view. This deliberately broad approach was taken because, to date, limited studies have conducted thematic analysis on spatial ability and anatomy learning.

The process of coding involved reading through the focus group transcription and highlighting student quotes pertaining to the research questions, then a summary of the quote was written for ease of identification and grouping using Nvivo. Next, the summary comments were handwritten and manually grouped into patterns to identify themes; the themes were then further reviewed (expanded, condensed, or remained the same). The whole process was iterative, with several reviews performed at each stage until saturation was reached with no new themes or patterns identified. This method of coding was performed on data from the first focus group of each cohort. The second focus group's coding was mapped onto the themes identified from the first group. The two cohorts were coded separately to help identify any nuances between them, as each had received a different teaching method (2D or 3D spatial).

### **8.2.2 Importance of Reflexivity**

Braun and Clarke (2006:p.7) discuss the process of thematic analysis and state “[a]n account of themes ‘emerging’ or being ‘discovered’ is a passive account of the process of analysis, and it denies the active role the researcher always plays in identifying patterns/themes, selecting which are of interest, and reporting them to the readers.” They also state, “It is important ... for us to acknowledge our own theoretical positions and values in relation to qualitative research.” This reflexivity of the qualitative researcher has also been identified by others as important and requiring consideration (Langdrige, 2007). It is also acknowledged that different

researchers may have different beliefs and experiences, which may take the research in a different direction, identifying (and possibly emphasising) different themes.

Therefore, it is important for the researcher to be as objective as possible by acknowledging and reflecting on their own beliefs and assumptions as this could impact on the conclusions (Langdrige, 2007:sec.5.4). The following two sections (8.2.3 and 8.2.4) address this by writing in the first person to detail the researcher's background and perspective to ensure transparency and reflexivity.

### **8.2.3 Researcher's Background**

I am a 2011 graduate of the Royal (Dick) School of Veterinary Studies (R(D)SVS). After qualifying in 2011, I worked in a mixed veterinary practice for a year in the north of Scotland and returned to the central belt of Scotland where I worked in small animal veterinary practice for 18 months. I then returned to the R(D)SVS as a Teaching Fellow in Veterinary Anatomy and Physiology in 2013. In the first year of my teaching fellow position, I realised I thought differently about anatomy compared to the majority of students I was teaching. The difference was that I thought of anatomy three-dimensionally (3D) and when I explained anatomy to students with a 3D and spatially orientated explanation, they would often indicate they had suddenly understood something they had found challenging. I also received very positive feedback on my teaching in the end-of-course feedback. Once realising the 3D and spatial explanations helped the students to understand anatomy, I wondered if the incorporation of spatial approaches to the teaching would improve students' anatomy knowledge and understanding. This became the focus of my doctoral research.

## 8.2.4 Researcher's Perspective

As a first-year veterinary student at the R(D)SVS, I remember finding anatomy lectures particularly unhelpful and I often got 'lost' as to where the lecturer was in relation to the 'bigger picture'. In particular, I remember frantically trying to copy the written notes from the overhead projector in anatomy lectures (with my peers often leaning over to look at what each other had written and how to spell it). I also distinctively remember having a lecture on the organs of the abdomen in the dog and how they were all suspended and attached. I remember feeling completely lost in this lecture. I then went home determined to understand and used lots of diagrams from various textbooks and lecture notes. While trying to figure out the diagrams, I thought of the structures in 3D. I started to manipulate the structures in my head, and I then realised I was going to have to think of anatomy in 3D to understand it. Thinking of anatomy in 3D also helped me when I came to learn about diagnostic imaging in my clinical years of study.

Later, when I was a member of staff at the R(D)SVS teaching students anatomy, I believed that thinking about anatomy spatially (using the problem solving spatial aspect of intelligence) and in 3D was important as it ensures the topic is truly grasped and understood. However, I do appreciate that everybody learns differently and at a different pace. I often found I used many approaches to teach the same concept to different students, although I found the most successful teaching approach was trying to encourage the students to think of anatomy spatially and not to rote memorise facts in isolation.

I believe I have a strong affinity to think spatially. At high school, I performed particularly well on subjects that included a spatially demanding element, such as craft and design, mathematics, physics, and physical education. These were also the subjects I was naturally strongest. Additionally, while training to be a veterinary

surgeon I had a difficult time finding a study approach that worked best for me, as I found lectures particularly frustrating. In the third year of my veterinary degree, I decided to tackle this by mind mapping (see Figure 8.1), as a way to take notes during lectures and to study from later. Mind mapping is a spatially demanding task as you must be aware of linking similar concepts on different locations of the map and you also have to navigate where you are on the map; these tasks are performed both visually and cognitively.

My studying became more effective by using mind maps, and I do believe part of this was due to my use of spatial ability to create effective mind maps. I, therefore, have a vested interest for the 3D spatial anatomy teaching method to work, and therefore I need to be aware of the criticisms and limitations of the spatial teaching method to not bias the analysis.

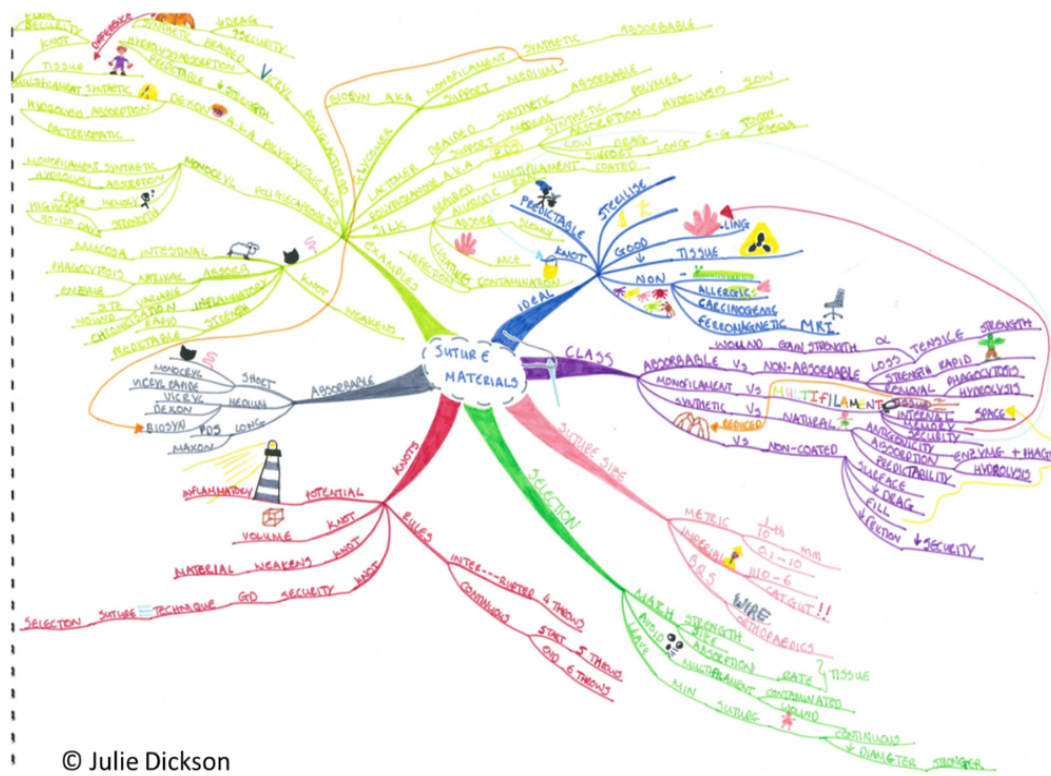
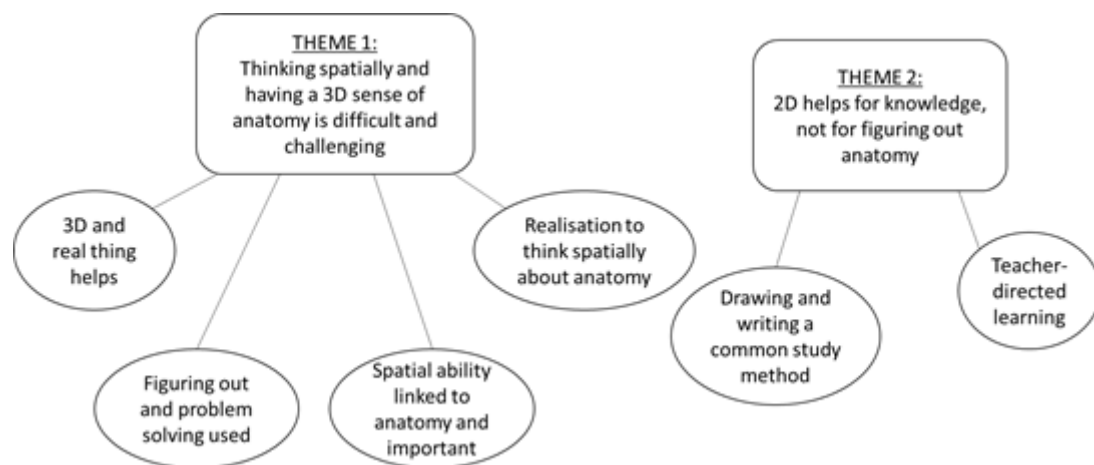


Figure 8.1. Example of a mind map I created from lectures while a veterinary student.

## 8.3 Thematic Analysis of UoE Vet 1 (2D)

### 8.3.1 Main Themes Identified

The main themes and associated sub-themes identified from the thematic analysis of the two focus groups of UoE Vet 1 students (2D teaching method) are depicted in Figure 8.2.



*Figure 8.2. Diagram of the main themes and sub-themes identified from thematic analysis of UoE Vet 1 focus groups.*

The first theme centres on the students' comments throughout the focus groups that 'thinking spatially and having a 3D sense of anatomy is difficult and challenging.' The terms 'difficult' and 'challenging' were selected because these were the terms used by the students. Additionally, although these terms are similar the term, 'challenging' can suggest positive connotations, while 'difficult' does not, and so these terms were selected for this distinction. The second main theme identified was '2D helps for knowledge, but not for figuring out anatomy.' The next two sections explore each theme and associated sub-themes in turn, with interpretations and student quotes to support each theme.

### 8.3.2 Theme 1: Thinking Spatially and having a 3D Sense of Anatomy is Difficult and Challenging

One of the main recurring themes identified throughout the focus groups was the difficulty of thinking of anatomical structures in 3D. This would perhaps seem unsurprising, as anatomy is known to be a challenging subject. For instance, anatomy is often not learned in detail until the higher education stage. Furthermore, as discussed throughout this thesis, anatomical structures are not symmetrical and vary in shape, location, and size according to function (i.e., anatomical structures do not regularly conform to everyday objects). Additionally, there are species variations in veterinary anatomy.

“I mean that just comes down to the structures themselves being a lot more complex in anatomy. You know, rather than just being a few squares you’ve got a very, very complex piece of [...] material to deal with” (student 5)

With this in mind, it is not surprising students could have difficulties when confronted with a complex structure in cross-section, such as the cross-sections used in the spatial MCQ (Appendix 7). Also, the thought of mentally manipulating one of these sections could be daunting for any learner. All of the students in the focus groups voiced that they felt unprepared to answer the spatial anatomy questions from the spatial MCQ. The feeling of unprepared is likely to be because of the 2D teaching method they received, where spatial approaches were not overly emphasised. However, students did draw diagrams in cross-section during lectures and completed a practical on the cross-section imagining tool of ultrasonography.

“[...] like initially when I first saw it (cross-sections) in the exam, [it] was like ‘oh no’” (student 6)

“You’ve only looked at the 2D structures. It’s a lot harder to visualise than when you’re looking at the full structure” (student 8)

“Well I felt unprepared with the transverse sections. I don’t [...] it was just through friends, I think, but we expected there to not be as many transverse pictures as there were. It was purely just transverse” (student 7)

During the learning process students go in and out of understanding, or different liminal states (Meyer & Land, 2003), meaning students can believe they understand a concept but often find they do not when applying the concept to a problem, or students may understand part of a concept but not understand the bigger picture of the concept. The added element of thinking spatially could make this more challenging. For example, in anatomy students need to think of the name of the structure and its function as well as considering the spatial aspects such as location, 3D shape, movement, and relation to other structures.

“I think if I had my dog at university it would have been a lot handier cos I’m trying to imagine it on my dogs, but it’s difficult [...]” (student 6)

The difficulty of thinking spatially about anatomy might have been predicted as students are often ‘uncertain’ or ‘not sure’ of their knowledge if they are still navigating concepts and have unanswered basic questions.

“[...] if I’m looking at it right, it could be this one, but it could be this one, or this one, or this one, and you, like, ‘oh god’” (student 2)



“[...] then the rest of the paper was full of transverse sections and I think a lot of people panicked” (student 8)

The difficulty with thinking spatially about anatomy was also compared to the spatial ability tests themselves, with students commenting that they found the tests easier. So it is not just the manipulating of objects in one’s head that is necessarily challenging on its own, but that it is the specific manipulation of the more complex anatomical structures.

“For me, doing those tests (spatial ability tests) it was more, it was the more consciously I thought about it the harder it was [...] with the bones, I had to think ‘right, this goes here, this goes there’. It’s not just, like, ‘yup, rotate itself, done’ [...]” (student 2)

“I think in the shapes that you were using in the thing (spatial ability tests). I’ve always had the mind to be able to pull it out and rotate it in my mind, but I think with anatomy if, say, you had put transverse sections again of a dog, definitely it would need the anatomy to look at the full thing cos it’s something brand new and fresh to me” (student 8)

However, a student did say that spatial tasks centered on even familiar structures (such as the origami duck used in admission interviews) were also difficult.

“It’s the origami duck all over again” (student 7)

The spatial MCQ designed as part of this project was intended to assess anatomy non-spatially and spatially, and the quantitative analysis of chapter 6 confirms the

split design of the MCQ. Qualitatively, the students voiced the non-spatial questions in the first section of the spatial MCQ were simply factual and the answers to these questions were known with little thought required. Focus group discussions were therefore primarily focused on the spatial questions in the second section and how they were challenging. This knowledge was not factual and had to be synthesised.

“[...] I think the beginning, thought the beginning, was quite logical and if you’d revised you knew it if you didn’t, you didn’t. It felt it was a comfortable way to go in, especially with the second part slightly harder” (student 6)

“No, it was just the transverse questions [...] a few different, there was a few different transverse all at quite similar levels [...] if you got your first one wrong then the rest of them were more likely to be wrong, [...], it was kinda a bit nerve-racking” (student 5)

Discussions during the focus groups also included how the students tackled the difficult spatial aspects of anatomy, and this identified four common sub-themes (Figure 8.1):

1. 3D and ‘real’ thing helps.
2. Figuring out and problem-solving used.
3. Spatial ability linked to anatomy and important.
4. Realisation to think spatially about anatomy.

## Sub-theme 1: 3D and Real Thing Helps

The ‘gold standard’ of anatomy teaching is regarded by many as dissection and the use of real material, e.g. dissected/prosected specimens, bones, plastinated specimens, etc (Theoret, Carmel & Bernier, 2007; Smith *et al.*, 2018). Therefore it is not unexpected the students commonly spoke of the ‘real thing’ helping:

“I’d need to do the anatomy like go into the dissection room or whatever. I wouldn’t necessarily be able to do it just by looking at a piece of paper” (student 9)

“I thought my anatomy revision clicked the most, and it was just before the spot exam, I got out loads of bone boxes and not only did I learn, like, the bone processes but [...] put them on the floor and thought where the muscles go to. That’s when in my head like everything sort of slotted together at that point” (student 6)

The students also voiced how helpful 2D photographs and videos of real specimens were, and interestingly the students did not discriminate against the ‘pseudo-3D’ representation of these compared to the real-life 3D structure. It was the fact that the photographs and videos were showing *real* structures rather than schematic representations that was helpful, as this helped to give a 3D sense of the structures as if they were the next best thing to the real structure.

“[...] when we got transverse section[s] of the actual picture not just a drawing (student is referring to the questions in spatial MCQ with a photograph and not a schematic diagram). I found that easier to do and then that helped me with the other ones” (student 2)

“[...] also even if you show us a 2D picture of the actual, you know there’s a difference between a drawn diagram and then the actual same muscles of something. If a picture like that is shown as well right after you’ve done the diagram that’s also really helpful, which has been done in like a lot of the lectures” (student 9)

“Re-watching lots of videos we took in the dissection classes really helped me [...] the video kinda gave that slightly more 3D as we were describing it, and pulling things out. Instead of just the pictures and the labels” (student 7)

A point to be considered for UoE Vet 1 is that although the use of real structures was identified as helpful, the main source of teaching received by these students was dissection classes involving such real structures. Consequently, the students may be more reliant on what they had the most exposure to; UoE Vet 1 did not have timetabled access to pseudo-real structures such as 3D computer models, or 3D printed models. However some students did say they found the use of a virtual canine website helpful, and the students commonly mentioned how useful 3D models would be.

“But like virtual canine anatomy is really, really good, that is a really good website” (student 5)

“It was great and also it’s really nice (virtual canine anatomy website) in that you can rotate it” (student 8)

“If there was a kind of bone box, but for muscles or something, that would be cool if you could put it together” (student 7)

“It would be so good if you had, like, muscles and then you could insert them into the bones, could build a dog and insert the muscles” (student 2)

Students did say that drawing 2D diagrams was also helpful, although this was mentioned less often, perhaps because this was the main form of teaching during lectures, and so 3D models and the real thing could have ‘novelty value’.

“[...] I learn by that, I write it out, and I draw it, and as soon as I draw it, it seems to be more in my mind. Which that’s why I find anatomy a lot more easier to do instead of cell biology” (student 8)

## **Sub-theme 2: Figuring Out and Problem Solving Used**

The second sub-theme identified that students commonly spoke about ‘figuring out’, ‘deducing’ and ‘problem-solving’ the difficult spatial anatomy when answering the spatial anatomy questions in the MCQ. The figuring out required, supports the overall theme that thinking spatially and having a 3D sense of anatomy is difficult, and this is not simple factual knowledge but requires effort to generate.

“I kinda just worked through it logically, and I just trying to think, and yeh apply your knowledge from what you think (to answer spatial anatomy question)” (student 7)

“[...] getting left and right was kind of ok, but just working out which muscles were like lying on top of each other and stuff and tryin’ a put that into the transverse was a bit hard. But I got them eventually, I think” (student 3)

“[...] when it got to the photos you had to pause and think, and there would be parts were you’d like, right ‘ok, here’s lateral, here’s medial’ like how you say in the diagrams and lectures [...], I actually kinda liked that, the fact that it’s making you think and deducing what it is from your knowledge [...]” (student 8)

“[...] tried to work it out logically but like she said it’s a bit difficult. Ok you can say the quadriceps femoris has 4 heads so let’s, let’s look for one with four divisions [...] so you try to do it logically but it’s a lot harder than that” (student 9)

### **Sub-theme 3: Spatial Ability Linked to Anatomy and Important**

The sub-theme ‘spatial ability linked to anatomy and important’ was identified from students’ explanations of how they ‘figured out’ questions or thought about the anatomy in such a way that spatial ability was used. This theme was also identified when talking about whether the students thought the spatial ability tests had influenced their learning of anatomy or vice versa.

“I definitely did, as I imagined the dog being chopped off and then rotating it in my brain. I would say it did (anatomy improved spatial ability)” (student 7)

“Not consciously, I wasn’t thinking at the time ‘oh I know where this muscle is so that will help me out here’ (to answer spatial ability test). Whether subconsciously because we’ve been using that part of our brains a lot more and that’s improved, like, that could definitely be the case. But like you don’t know that [...], it could definitely. I think it would. I think if you are using that part of your brain a lot more then it’s going to improve” (student 5)

“I imagined what they looked like (muscles) in dissections and match[ed] them to the outlines” (student 6)

“I think it (spatial ability) is very important” (student 8)

“When you’re working as a veterinarian you have to be able to visualise where things are inside the dog. You have to be able to see that, so you know where you’re cutting, where you’re injecting. It’s important” (student 8)

In chapter 2 (section 2.4) evidence was presented supporting the importance of spatial ability to intelligence (i.e., problem-solving). This connection between spatial ability and problem-solving processes was identified during the focus group discussions with some students using their spatial ability to problem-solve, and not purely for the mental manipulation of 3D structures.

“With the pictures you could kinda just take a guess at just the shape. If you imagine just getting a dog cutting it in half, and cutting the leg in half, and then just fixing the shape. Instead of actually going back, referring to your knowledge, finding out all the different muscles instead [...], so there was less kinda back knowledge involved” (student 7)

#### **Sub-theme 4: Realisation to Think Spatially About Anatomy**

Despite the apparent use and importance of spatial ability in anatomy, not all students made the connection initially, and this gap identified the sub-theme ‘realisation to think spatially about anatomy.’

At the start of the focus group, when discussing what the students thought the spatial MCQ was testing, mixed responses were received.

Interviewer: “[...] as the questions got further on in the paper do you feel they were testing your anatomy knowledge or not. What do you feel they were testing?”

Student 2: “yessss (long spoken word) I did”

Interviewer: “As you got on you said the first [section of] question[s] was all the wording, you were quite happy that was quite knowledgeable. If you’d studied you’d be happy. But as we got further on with these, the whole sections, what do you think, what did you feel was being tested? Was it just anatomy, was it something else, was it a combination of things?”

Student 2: “[...] for me it was anatomy but because we’d just done ultrasonography I was like, well, the whole point of that was transverse sections. So it’s probably because of that. But I was like, well, you’re not gonna ultrasound a leg, but you can’t ask us question[s] not on the leg, because that’s [what] the exam [is] on. So I was like, I understand what they are trying to do but I don’t know if getting loads of transverse sections [...] is, I don’t know if it was the right thing”

Interviewer: “so you do feel it was sort of testing your anatomy knowledge (student 2 agrees). Does everybody agree with that, or do you think?”

Student 6: “I felt it was kind of proving to us that, for anatomy, we don’t just need to learn a picture we get in sort of dissection of the dog and label it. It sort of made me realise for my exam I was going to have to know it in a 3D sense.



Rather than just a 2D picture. It kinda made me want to think about everything in 3D when I revised after that more”

Interviewer: “yeh do you guys all agree?”

\*Yeh general agreement\*

The discussion quoted above is about what the students felt the spatial MCQ was examining. One of the student’s felt the test was examining their anatomy knowledge while another student realised they would need to know anatomy in 3D. The rest of the students were asked if they agreed with this realisation about anatomy: there was some agreement, but generally the students were quiet. The quote below provides evidence of a student not making the higher level connection of realising they would need to know anatomy in 3D.

“[...] I was fine with everything except for the transverse sections. I prepared quite a lot for that exam as well. I mean everything else was fine except for that. I didn’t understand why you’d have to be able to identify different structures based on transverse sections in surgery anyway. It was just a bit odd and that was quite a large part of the paper, but apart from that everything that was fine” (student 9)

Shifts in attitudes then changed at the end of the focus group after discussions about the spatial ability tests helping with anatomy or vice versa. At the end of the focus group, the students were asked directly whether they thought anatomy was important and the students strongly agreed that spatial ability was important for anatomy.

Interviewer: “do you actually think or feel that your spatial ability is important for anatomy [...]”

Student 6: “yeh definitely”

Student 5: “oh definitely oh definitely oh definitely 100% it is”

Student 2: “it’s really important”

Student 8: “I think it is very important”

As discussed, these patterns in the responses identified the sub-theme ‘realisation to think spatially about anatomy’ as not all students had made the mental leap of thinking spatially and having a 3D sense of anatomy was important. Furthermore, the sub-theme ‘realisation to think spatially about anatomy’ was also reflected in discussions on specific spatial anatomy questions, with students saying they felt ‘unprepared’ and expressing a feeling of uncertainty about these questions. However, when individual questions were discussed, students often realised they were answerable and not as challenging as first thought.

“That was definitely not as bad as the other ones (about ‘what level’ type of spatial question), it was much, it was, yeh, that was fine” (student 9)

### 8.3.3 Theme 2: 2D Helps for Knowledge, Not for Figuring Out Anatomy

Theme 2 ‘2D helps for knowledge, not for figuring out’ was identified after many comments were made on how 2D helped the students to memorise and know the facts of anatomy. Discussions on the first part of the spatial MCQ, designed to be non-spatial questions, also highlighted this theme, identifying these questions as requiring less problem-solving. It was clear the students did not use any problem-solving for the non-spatial questions.

“The flat pictures are very, very good for just learning off the names. For knowing where your adductor muscle is, you know, just for knowing the names. But when actually comes to their actions their innervations, how they actually are in real life then the transverse sections come in definitely. Because I mean you know it’s handy to the learn names just on the flat picture but actually understand what it does is a lot easier to understand the actions of the biceps if you see it on a transverse section. You can see ‘ok it pulls directly up’ [...]” (student 5)

“[...] thought the beginning was quite logical and if you’d revised you knew it, if you didn’t, you didn’t [...] it was a comfortable way to go in [...]” (student 6)

“I’ve never remembered the beginning at all so it must have gone good” (student 2)

“[...] the beginning part it was just you know, because you’d revised so much it was just quick answers. You knew them immediately when it got to the photos you had to pause and think” (student 8)

However, one student did comment that drawing 2D diagrams helped to figure out the anatomy and this was an active learning process.

“like I think for me it’s more to just remember the names and the kind of rough position rather than actually applying the diagram (drawing a 2D diagram). But I draw it to the diagram and I’m seeing it’s more just applying the knowledge that I’ve got from drawing the diagram. Drawing the diagram makes you actually sit down, makes you actually look at where things are, it makes you read it properly. Cos if you just look at a picture you’re not really taking it in. Whereas if you’ve to draw it out takes a bit of time. You’re kind of thinking about it as you’re doing it as your drawing your saying ‘ok so they’re all over there they’re all the same nerve.’ I think it[s] just making you think rather than applying it directly from your head” (student 5)

### **Sub-theme 1: Drawing and Writing a Common Study Method**

Two sub-themes were identified within theme 2: ‘drawing and writing a common study method’ and ‘teacher-directed learning.’ Students commonly used drawings and writing of notes to study for anatomy, and this was the same teaching method used in the 2D teaching method. Interestingly, although drawing and writing were very common study methods, the students did reflect that this did not necessarily help with deep learning.

“For me I just draw pictures and draw diagrams of the different things straight out of [...] dissection of the dog, and then that just helps me memorise it” (student 5)

“I have to draw and write. I can’t just draw the way she does. I can tell you it’s the yellow muscle but I’ve no idea what nerve that means, so I need to do both definitely” (student 3)

“[...] how deep the muscle goes I think that’s really important as well. You don’t realise it when you see it on the outside, yeh its muscle it covers everything else but that could be much thinner than ones underneath which are much thicker. Or the other way around it could be a really thick one and behind is a really thin one. You don’t really see that in just the pictures that you draw”(student 2)

“With me it’s different. That method with you know drawing it along with the lecturer doesn’t necessary work that much, that well, for me. I need time to just sit and just read it. It doesn’t really go into my head straight away from just writing it down. I really have to sit and examine things” (student 9)

However, as mentioned previously, some students appeared to have made the higher-level connection, realising they wanted to know anatomy in a 3D sense and to accomplish this used real anatomical structures to study from.

“I thought my anatomy revision clicked the most, and it was just before the spot exam. I got out loads of bone boxes and not only did I learn like the bone processes but then where [...] I like put them on the floor and thought where the muscles go to and that’s when in my head like everything sort of slotted together at that point” (student 6)

Throughout the discussions on students’ learning approaches for anatomy, the main method they described was using diagrams and writing. Some students were more reliant on a diagram and using different colours, while others preferred notes. The discussions on this made students reflect that people used different approaches to learn.

## Sub-theme 2: Teacher-Directed Learning

The second sub-theme identified within theme 2 was ‘teacher-directed learning.’ A lot of the students mentioned how they were dependent on the lecturer for drawing diagrams and writing notes within the lecture. The students expressed comfort with this as if it was ‘holding their hands’ and guiding them in a controlled manner through the lecture. Writing notes and drawing diagrams is time-consuming and the slower pace of drawing with the lecturer may have provided the students with a ‘security blanket’; it possibly made them feel they were learning because they were active or because the picture was built up slowly compared to seeing structures labeled all at once.

“During the lectures I thought it was easier when the professors, when they sort of started with an empty picture and then added to it. Instead of just giving us [...] the diagram with all the information on it” (student 1)

“[...] writing the notes with us as well because I write notes in all my lectures but trying to keep up with people who just flick through a powerpoint is just impossible” (student 6)

“I think definitely it was the lectures itself when your when your writing with the lecturer, and as you’re doing the diagrams with them instead of having a diagram put up in front of you draw it and then move on [...]” (student 8)

However, not all students felt this way when copying the notes.

“I thought that when you were handwriting the notes it felt like it was a bit too slow [...] for me it would have been ok if

that was typed. Then you'd just kinda put bullet point after bullet point up. It would have worked for me as well"  
(student 1)

## 8.4 Thematic Analysis of UoE Vet 2 (3D spatial)

This next section discusses the themes of the thematic analysis of the two UoE Vet 2 focus groups. This cohort of students received the 3D spatial teaching method. Despite this, the same two main themes were identified. However, the sub-themes were different, which could reflect the different teaching method received.

### 8.4.1 Main Themes Identified

The two main themes identified from the thematic analysis of UoE Vet 2 focus groups are depicted in Figure 8.3.

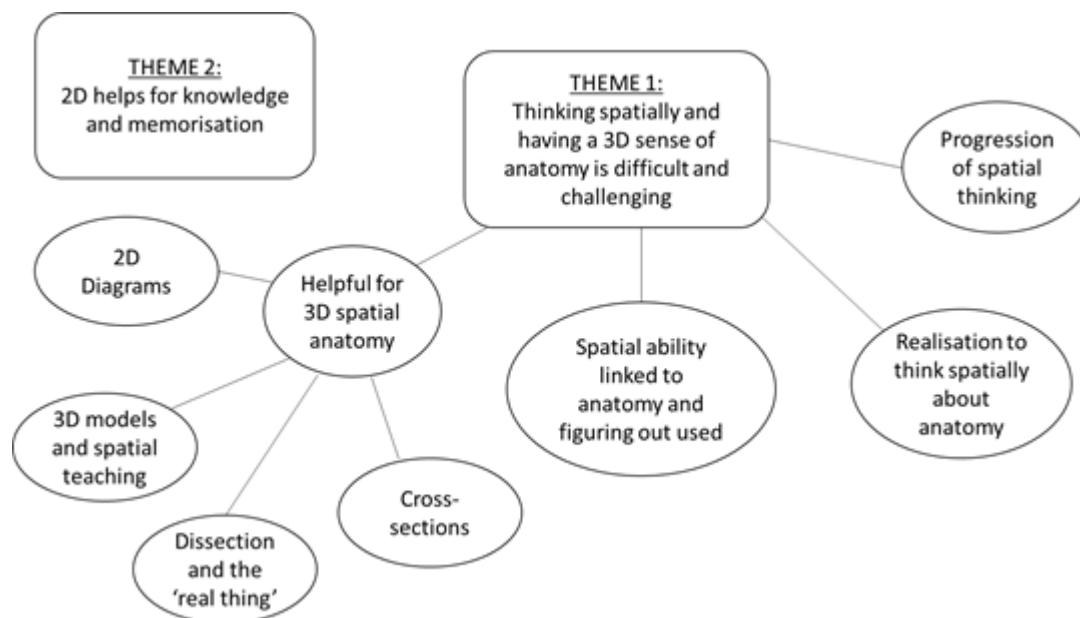


Figure 8.3. Two main themes and associated sub-themes identified for UoE Vet 2.

Theme 1 is the same as for UoE Vet 1: ‘Thinking spatially and having a 3D sense of anatomy is difficult and challenging.’ However, the sub-themes identified were different: ‘helpful for 3D spatial anatomy’ (including four sub-themes), ‘progression of spatial thinking’, ‘spatial ability linked to anatomy and figuring out used’, and ‘realisation to think spatially about anatomy.’ The second main theme identified for UoE Vet 2 was ‘2D helps for knowledge and memorisation.’

#### **8.4.2 Theme 1: Thinking Spatially and having a 3D Sense of Anatomy is Difficult and Challenging**

A similar recurring theme for the students in UoE Vet 2 was the difficulty, despite having received a 3D spatial teaching method, of thinking spatially about anatomy. During the focus groups, it became apparent UoE Vet 2 students were advised by UoE Vet 1 to look at transverse sections for the spatial MCQ. It is uncertain how many students did this, but those that did said they were disconcerted by the transverse cross-sections they saw in textbooks when studying for the MCQ. It could be thought that since the students had prepared for the exam, by studying cross-sections, the students would find the exam easier. However, the students still mentioned how difficult some of the anatomy topics were to learn (such as nerves), and how difficult some of the spatial MCQ questions were. Spatial questions which involved cross-sections of a type not shown in the common veterinary anatomy textbooks were seen as particularly challenging as the students were seeing these specific images for the first time. An additional difficulty was thinking in 2D when learning from 3D.

“I find when I learn from 3D I may not know it when I’m given it in 2D” (student 1)

“[...] it take[s] a lot of effort to actually visualise the thing from 2D to 3D. You need to literally move around and this is what, this is what. Because in 3D if you just move it around,



and you see a lot of views, and in 2D it is really flat you can't go like that, or like that, (rotating). It just looks the same.”  
(student 8)

“Quite a few of the 2nd year were like ‘look at transverse sections of the leg’ and if they hadn’t said that I probably would have got 5 out of 30” (student 13)

“Those transverse section when I first looked at them I was like ‘I am never going to get on with this, this is awful’ ”  
(student 15)

Interestingly, in comparison to UoE Vet 1 (who felt unprepared), the initial responses from UoE Vet 2 in relation to the spatial MCQ were positive, expressing enjoyment, despite the challenge. When asked if there was any teaching that did not help for the MCQ, or if they felt unprepared, there was silence in one focus group. This finding could be reflected in the fact that UoE Vet 2 were not naïve to the MCQ and had studied cross-sections in preparation. Therefore, perhaps memory was used more than spatial ability. These students seemed more comfortable with the MCQ; there was no element of surprise.

“I liked that, I did, we'd named every single one of them didn't we. Cos like me and X did it. How long did it take us to do? About 10 mins? We'd done most of the questions and then obviously went back and checked over it, and then at the back we were labeling every single muscle, and bone, and everything. It was quite nice” (student 14)

However, not all students had looked at transverse sections before the MCQ.

“I must admit I never looked at the transverse section before we got this in the exam, and I had to figure it out” (student 10)

Furthermore, students expressed a difficulty with learning anatomy topics not taught by the spatial 3D teaching method, such as the thorax and abdomen. Arguably the topics of the thorax and abdomen could be deemed more spatially demanding due to the asymmetry and variety of structures located within these 3D body cavities.

“I think it wasn’t as hard with limbs to orientate yourself. But as we say, we were doing abdomen, I struggled with abdomen a lot cos you look at it and your like I can only identify stuff when the dog is in dorsal recumbency, and as soon as something else [...] I should take 10 minutes to try to understand” (student 10)

### **Sub-theme 1: Helpful for 3D Spatial Anatomy**

Discussions in the focus groups encompassed the difficulty of learning anatomy spatially, and in general, this led onto discussions on what teaching could help to learn the spatial nature of anatomy. These recurring discussions and inputs from students on how they studied identified the sub-theme ‘helpful for 3D spatial anatomy.’ Within this sub-theme, four further sub-themes were identified specifically highlighting what students commonly said were helpful and why. The further four sub-themes were ‘2D diagrams’, ‘dissections and the ‘real thing’’, ‘cross-sections’, and ‘3D models and spatial teaching.’

## *2D Diagrams*

One unexpected method that helped students to learn spatial anatomy was 2D diagrams. Students commonly discussed looking at 2D diagrams from textbooks and finding these particularly helpful. Students voiced a preference for looking at 2D diagrams at the start of a topic with progression to 3D methods later.

“I sometimes find though that I might be the opposite way around. I'd rather see a diagram like that and then work out what it is in the 3D. I don't know whether I prefer going 2D to 3D than 3D to 2D” (student 13)

“I think you need the picture first but sometimes I look at it and I'm like hmmm” (student 15)

One of the focus groups mentioned how helpful the diagrams from the thorax lecture were, perhaps because this area of teaching was not taught with the spatial teaching method. Additionally UoE Vet 1 focus groups identified comfort with teacher-directed learning. Perhaps UoE Vet 2 also enjoyed drawing diagrams because they were not challenged to think spatially, which is mentally challenging and were teacher-led. Below is a conversation on the 2D diagrams of the thorax lecture.

Student 12: “but the nice thing with the thorax was because we had the diagrams”

Student 15: “they were amazing”

Student 12: “those diagrams were exactly what was the text, and so you could read the text”

Student 15: “[...] those thorax diagrams are really useful [...]”

Student 12: “in a way the notes were a lot but the diagrams helped”

Student 13: “yeh the diagrams”

For some of the end-of-course examination questions students were required to draw a diagram; they could also opt to draw a diagram to answer any question even if not specifically asked to do so. On this topic, one of the interviewers asked if drawing diagrams was never asked in the exams, and students had to opt for either 2D diagrams or 3D models as a teaching method, which would they prefer? This question was asked in an attempt to find out how exam driven the above responses were because of the necessity to draw diagrams in the exam. The following responses were given:

Student 15: “3D”

Student 14: “Yeh”

Student 15: “all the way”

### *Dissection and the 'Real Thing'*

Similarly to UoE Vet 1 students, UoE Vet 2 discussed how helpful dissections and real structures were for learning anatomy. However, there was a time constraint for finding structures during dissections, and the students felt under pressure with this. Some students mentioned the dissection practical classes helped them to overcome any deficits in other areas of anatomy teaching.

“Yeh I think the dissections are really helpful, but most of the time I found myself more focussed on finding everything on the list rather than taking my time to learn” (student 5)

“Yeh I think you need to see the muscles in dissection before you can do transverse” (student 12)

“[...] I looked at for example the inguinal canal diagram [...] and I looked at that and was like 'I have no idea what she is talking about' [...] then you go into class and you see the real thing and all of a sudden the diagram is perfect. It's a great representation, but until you've seen the real thing it's really hard to understand” (student 15)

In the spatial MCQ assessment, the images of the cross-sections were either of a real photograph or a schematic diagram. UoE Vet 1 (2D) tended to use the questions with a real photograph of a cross-section to answer the equivalent schematic MCQ question. The use of the real photographs could be because dissection of real structures was a significant part of the representation of anatomical structures for this cohort. In contrast, UoE Vet 2 preferred the schematic diagrams to the photographs of real structures, possibly reflecting their main source of teaching was schematic 3D models.

## *Cross-sections*

Cross-sections were identified as ‘helpful for 3D spatial anatomy.’ Cohort UoE Vet 2 studied from cross-sections for the MCQ, and many of them mentioned how helpful the cross-sections were for learning anatomy, particularly for understanding the depth and shape of muscles. One of the students also mentioned that cross-sections helped them to understand the innervation of the muscles when they highlighted specific muscles in colour for a given nerve.

The students gave the impression it would be difficult to look at a cross-section at the outset of learning an area of anatomy; this thought was also echoed when students discussed the use of videos of computed tomography (CT) and magnetic resonance imaging (MRI) scans used in the lectures. The CT and MRI videos were cross-sectional, and students voiced they were difficult to follow during the lecture but became useful at the end of the course. One student suggested a lecture or tutorial on cross-sections and then translating these to CT and MRI images would be helpful.

“I found it easier to visualise the layers of the muscles (using textbooks). Through the transverse sections you can see where the nerves are, so you can sort of visualise which nerves innervates which part of the brachium or the antebrachium. So it was easier for me to study through transverse sections compared to like layers, lateral sections” (student 5)

“I think it’s easier, transverse sections is easier for revision, not from learning itself. If you start from zero then you look at that it would really confuse you. You have to know like a bit of knowledge from like all four lateral, all four views, before you go to transverse section. Or you just confuse yourself” (student 8)

“[...] they’re really useful those transverse section [...] but then the more you kinda learn they were actually quite good at knowing, like, sometimes you see a muscle on say the medial side but you do also see it on the lateral. And you can’t work out why you are seeing it on both and I think those were really useful for that. I don’t know maybe whether that could be incorporated into the lectures as well that would maybe help” (student 15)

“I didn't understand the whole intermedius, those muscles of the quadriceps femoris, and then it made sense when I saw that (cross-section)” (student 13)

An interesting point is that UoE Vet 1 did not know that some spatial MCQ questions would be based on cross-sections and so there was panic and a general feeling of being ‘unprepared.’ The feeling of being unprepared was commonly mentioned in the focus groups when discussing the spatial MCQ. Further discussions on specific spatial questions with UoE Vet 1 identified that students could, despite being unprepared, figure out the answer and it was not as bad as the initial panic.

During focus group discussions on the spatial MCQ with UoE Vet 2, it became apparent they were less surprised by the cross-sections. They said that they enjoyed the spatial MCQ, although when discussing specific questions a feeling of uneasiness and difficulty around cross-sections was identified. Since UoE Vet 2 studied from cross sections, it would be expected that scores on questions involving the cross-sections commonly encountered in textbooks would be significantly higher for this group.

This was not the case, six of the seven questions (questions 19, 20, 21, 23, 24, 25) involving cross-sections similar to ones identified in textbooks showed no significant difference in correct scores between the two cohorts. However, UoE Vet 2

commonly said the cross-section questions were ok because they had looked at and studied from cross-sections, and so perhaps memory was primarily used to answer these questions rather than spatial problem-solving.

This theory is partially supported by questions 26 (the only radiograph question), and 28 (a 'what level' question) as both of these questions involve a cross-section of a type not used in the common veterinary anatomy textbooks. For questions 26 and 28 UoE Vet 2 scored significantly higher compared to UoE Vet 1, 76.9% vs. 44.5% for question 26, and 70.2% vs. 52.1% for question 28 respectively. UoE Vet 2 would not necessarily have studied these cross-sections, and the students did voice the 'what level' questions were difficult. The finding of UoE Vet 2 scoring higher for questions 26 and 28 could also be because UoE Vet 2 had better spatial thinking of anatomy compared to UoE Vet 1, due to the 3D spatial teaching method.

### *3D Models and Spatial Teaching*

Students expressed a real enjoyment of using the 3D computer models for various reasons including visualising of muscle attachments; easier visualisation of structures such as ligaments; the images were rotatable; and the ability to study out-with the dissection room. The students expressed an interest in using more 3D models, particularly of the muscles. The spatial teaching method also included the inclusion of 3D printed models. The students also enjoyed these, but less so compared to the 3D computer models. One reason given was the sharing of the 3D printed models was difficult despite learning in small groups.

The 3D spatial teaching method incorporated tutorials and the students also expressed much enjoyment with these, one of the main reasons was the students felt the tutorials tested their knowledge; they were informal assessments which the students generally expressed a likeness for assessments. The likeness for assessments



is unsurprising since students are exam-driven. In this instance, the students were driven to understand what they did not know, not necessarily driven to learn what is required to pass the exams. The tutorials did not contribute to their final end-of-course mark.

“Personally I really like all the 3D models the most. But those were only for the bones we never really have 3D models for the muscle, so dissections were helpful” (student 5)

“I think the 3D models were good because you could visualise the muscle attachments a bit better rather than just like saying that’s where it attaches. You could see how the bones fit together quite well like in the stifle joint [...] which was good” (student 11)

“I think the 4<sup>th</sup> tutorial was particularly helpful because like we could go back to it after that tutorial session” (student 3)

“[...] with the 3D model stuff what was nice especially for hindlimb was when there’s like caudal cruciate or cranial cruciate ligament you could see it being attached, and with a 2D model you can’t see that. So that was huge [...] seeing different views of it” (student 12)

“The 3D models I found that they are really useful and they were really great to go home or to the library and sit and go through” (student 15)

“Tell you what was really useful was those classes, the ones that was almost tutorial like the anatomy 3 and 4’s” (student 15)

One student said it was difficult to orientate themselves in relation to the anatomy of the abdomen and when asked what would help the student replied:

“like you could [have] a 3D diagram [...]” (student 10)

The students were asked during one of the focus groups “would it be easier or harder if they did not have 3D computer models”, one student replied:

“I think I would have hogged a bone box from the library, I’d keep it at home yeh” (student 5)

On the contrary, the students did express the 3D computer models were difficult to follow in the lecture when the lecturer was explaining using a 3D computer model. This is discussed in the next section.

## **Sub-theme 2: Progression of Spatial Thinking**

Throughout the focus groups, students made intermittent comments that highlighted a progression of learning and understanding structures in 3D, and this identification led to the sub-theme ‘progression of spatial thinking.’ The students did identify learning from the 3D models during the lecture was difficult, but explained that as the course progressed, this became easier.

“I think I found it really complicated in the beginning of the year to follow (lecturer manipulating 3D computer models in

lecture) when we were doing the forelimb [...] then when we did the hindlimb it was either the fact that half of the year has passed, and we have got used to teaching style. But I found it much easier to follow as you were working with the 3D diagrams (models) and so I was actually getting quite a lot out of it” (student 10)

“[...] I feel like at the beginning of the year it was definitely harder (following lecturer manipulating 3D computer models in lecture) and I felt like it was almost, but not a waste of time, but it was hard to like understand or get anything from it. But I definitely feel like as the year, as the term progressed, it was easier and it was nice to be able to learn something. Then say here’s a visual representation of it because obviously we don’t have a bone in front of you so it was nice [...] just to be able to get a kind of idea. Yeh get an idea what is actually looked like rather than schematically” (student 9)

“I thought those were really difficult (‘what level’ spatial question) until you got it, and then once you got it for example the one on the other side yeh 28. I was looking at it for ages thinking what on earth is this, and then eventually it clicked that it couldn't be E because although you can only see one bone on E, there would be 5 and then all of a sudden it all became clear. Because you couldn’t, you can't really see a clear definition of your fibula that you can see. But I wasn't 100% that that was the bone. I couldn’t work out what is was. And we've never really seen anything like that before, it was a good stretch question” (student 15)

The students also made comments that CT and MRI scans were fun to look at, but difficult to follow in the lecture. They helped during revision at the end of the course or even afterward, suggesting a progression of knowledge and possibly spatial thinking.

“But if we looked at them now I might be like oh masseter muscles, jaw bone. But at the time” (student 15)

“I think during revision week I went and looked at the CT scan for the muscles of mastication and it made sense then, but in the lecture I remember being like I have no idea what I am looking at, what any of this is” (student 13)

### **Sub-theme 3: Realisation to Think Spatially About Anatomy**

It was evident the students in UoE Vet 1 had not all realised the importance of thinking spatially about anatomy, but some students had. For UoE Vet 2 students, this realisation was not evident either; many, when asked if they felt the spatial ability tests helped with anatomy, most did not consciously think so. Some felt it would be the other way around; the learning of anatomy would improve spatial ability.

Furthermore, the students agreed spatial ability was important and when asked if this was something they had thought of before learning anatomy many of them had not. All of these discussions identified the theme ‘realisation to think spatially about anatomy’; similar to UoE Vet 1, with UoE Vet 2 students not having a moment of realising that thinking spatially and having a 3D sense of anatomy was important. This may be because they received the 3D spatial teaching method with the aim of training the students to think spatially about anatomy from the start. However, it could be because the students were not asked directly if learning anatomy had improved their spatial ability (the discussions in UoE Vet 1 naturally progressed to this).

“I always thought anatomy was more just, before I started, I just thought anatomy was just memorising. But you need to like orientate yourself and a lot of spatial ability involved” (student 8)

“I didn’t really think about anatomy until I started learning it but like I can see the importance. When you are trying to orientate structures in relation to other structures” (student 6)

“I did (applying the same skills to spatial ability tests and anatomy learning) I felt equally, it wasn’t conscious. I didn’t notice until you pointed it out that there is a correlation. But I felt like it could have or would have helped me in terms of like spinning things around” (student 5)

“I didn’t put those two together (spatial ability and anatomy)” (student 12)

“I didn’t think, but I wasn’t like a conscious thing. But I think if anything it would have been the other way (anatomy learning improves spatial ability)” (student 15)

#### **Sub-theme 4: Spatial Ability Linked to Anatomy and Figuring Out Used**

The last sub-theme for theme one was ‘spatial ability linked to anatomy and figuring out used.’ When the students were describing how they tackled the spatial anatomy and the learning of anatomy it became apparent a lot of them were ‘figuring it out.’ It was not knowledge already known or factual; some of the students described their figuring out efforts to involve spatial ability. Students also spoke about finding a landmark and working out other structures based on this landmark.

“It’s [...] more of a working out thing. Whenever after dissection like the image kinda stay in your head for a really a long time. So you kinda rotate it around and try to visualise

the transverse section from what you remember from dissections” (student 8)

“This must be here in the inside, and then trochanter, and then I was thinking ok so there’s a notch. So that must be on the front and then like turn the paper round and lay it on my leg, and I think you saw me doing it, 'Ok so it's right leg” (student 15)

“I just imagined a dog and where I would be able to sort of place the section in and then sort of see what it could be potentially in there” (student 10)

Additionally, as previously mentioned, the UoE Vet 2 students found the schematic diagrams of the cross-sections easier than the photographs; this was the opposite of UoE Vet 1.

Student 8: “I think what everyone does is”

Student 1: “turn back to the drawing”

The next section of this chapter discusses the second main theme identified for UoE Vet 2 ‘2D helps for knowledge and memorisation.’

### 8.4.3 Theme 2: 2D Helps for Knowledge and Memorisation

The first part of the spatial MCQ involved non-spatial 2D questions and the students rarely commented on this part. However, when they did comment, these questions were described as ‘simpler’ and ‘taken directly from the notes.’ The students did not describe how they came to this information and often spoke of remembering or knowing. These patterns identified the second theme, ‘2D helps for knowledge and memorisation.’ When the student was asked why these questions were simpler:

“[...] it is directly given, the answers are directly given in the theory in the notes of the lectures. So it was easier to answer, cos you know it is either this or not this” (student 5).

However, two of the spatially designed MCQ questions were explained as being the same diagram as in the notes (questions 15 and 30). Thus, correctly answering could relate to the students prior knowledge to look at cross-sections before the exam and memorising cross-sectional diagrams from the lecture notes.

“It was easy because it was in her notes (for pelvis cross-section schematic)” (student 5)

## 8.5 Discussion

The thematic analysis of the focus group data from both cohorts confirms that students clearly distinguish between 3D spatial anatomy knowledge and 2D anatomy knowledge. UoE Vet 1 (2D) expressed feeling unprepared for the spatial MCQ, in particular, the cross-sections, but upon a discussion of individual questions, the students acknowledge questions were not as difficult as initially thought (i.e., the questions were ‘doable’). UoE Vet 1 commonly mentioned that 3D resources would

help with anatomy learning. UoE Vet 2 (3D) were less surprised by the spatial MCQ (possibly because UoE Vet 1 advised this cohort to look at cross-sections as part of their revision) although the spatial questions were still thought to be challenging.

UoE Vet 2 mentioned the 3D models were helpful, as were dissection and textbooks, and requested more 3D models, especially those involving muscles. UoE Vet 2 did however also acknowledge that 2D diagrams were helpful for learning anatomy. However this could reflect the anxiety of being asked to draw diagrams in the examinations and perhaps because this teaching approach was teacher-led. Students in UoE Vet 1 appeared to realise the importance of thinking about anatomy spatially, as did UoE Vet 2.

### **8.5.1 Comparison with Other Qualitative Research Findings**

Studies on the links between spatial ability and anatomy education have gathered students' views using open-ended questions and Likert scale questionnaires (Garg, Norman & Sperotable, 2001; Guillot *et al.*, 2006; Latorre *et al.*, 2007; Smith & Mathias, 2010; Brown, Hamilton & Denison, 2012; Nguyen *et al.*, 2013; Peterson & Mlynarczyk, 2016). Studies have found students report a reduced cognitive load when using computer technology such as Augmented Reality (AR) (Küçük, Kapakin & Göktaş, 2016).

Qualitative researcher efforts have also been directed towards exploring the different approaches and strategies students use to answer anatomy questions of a spatial nature. One study found the approaches taken by students to answer spatial MCQs on a trachea, oesophagus, and aorta 3D model (as identified by a self-reflective questionnaire), was varied and many different approaches were used by students (Nguyen *et al.*, 2013). Some of the approaches they described included mentally rotating the whole or parts of the structure (with students imagining themselves



looking at the object from a different perspective), using verbal and motor processes, and systematic scanning of cross-sectional images.

The focus groups in the study presented in this chapter also allude to students taking different approaches to answer spatial questions with some students using the bones for orientation purposes, while others used the muscles. Furthermore, some students found identifying left or right more challenging than identifying muscles. One study using a pictorial questionnaire found that 85% of students remembered a key view (a typical orthogonal anatomical view, e.g. dorsal, lateral, cranial) and mentally rotated this view to answer questions in a spatial MCQ on the carpal bones (Garg, Norman & Sperotable, 2001). This finding was not identified in the study presented in this chapter, possibly because the spatial MCQ on the carpal bones did not involve cross-sections but topographical projections of the carpus (i.e., lateral, medial, palmar and dorsal).

The ASSIST inventory (Approaches and Study Skills Inventory for Students) (Tait, Entwistle & McCune, 1998) has been used to investigate the approaches students use when learning anatomy across a 4-year graduate entry and 5-year undergraduate medical degree courses (Smith & Mathias, 2010). The ASSIST inventory uses questions (rated 1-5 by the learner, with 5 = high) to explore whether students use a Deep Approach, Strategic Approach, or Surface Apathetic Approach to studying. The students completed a Likert scale questionnaire around six areas: preferred learning activities; thoughts of using cadavers; difficulties of learning anatomy; how anatomical knowledge is used; students' overall views of anatomy; and some questions specific to a year of study. The authors found that students with a deep learning approach were more likely to be 'hands-on' with learning and exploring specimens, which led the authors to postulate these students may have a better spatial awareness of the anatomy.

Students with a strategic approach were described as appreciating the range of material and the depth of knowledge needed for examinations. These students were identified as feeling they had to use the knowledge quickly or it would be lost; the authors describe this as a ‘use it or lose it’ strategy. In comparison, students with a surface approach were found to be daunted by the volume of material to be learned, and felt the teaching did not work for them. These students were identified as exam-driven as they focused on what was required to pass the exam and appeared to memorise the material, reflecting that this was difficult. Smith and Mathias (2010) also found that students rated anatomy highly important and that learning using cadavers, textbooks, online resources, and mock-examinations were all important. The use of cadavers, textbooks, online resources, and mock-examinations was also identified in the focus group discussions presented in this chapter as highly important and helping to learn anatomy. This multi-modal approach has been discussed in the anatomy education literature as an approach to use to encompass all learners (Drake & Pawlina, 2014).

The study presented in this chapter found that students used diagrams and notes to help memorise anatomical structures. Studies comparing 2D to 3D teaching resources have found students prefer traditional 2D resources to help them memorise names of structures, while 3D helps achieve a better understanding of anatomy and what the structures look like in real-life (Brown, Hamilton & Denison, 2012; Peterson & Mlynarczyk, 2016). Students were also shown to reflect that working with 2D and 3D helped in their ability to visualise the anatomy topographically (Peterson & Mlynarczyk, 2016) and students preferred and enjoyed 3D resources (Latorre *et al.*, 2007; Keedy *et al.*, 2011; Tan *et al.*, 2012; Preece *et al.*, 2013; Smith *et al.*, 2018). However, students have voiced issues when working with 3D resources due to problems using the resources, such as, they are less user-friendly, and it makes it harder for some students to visualize structures. This finding was also identified in the study presented in this chapter and is consistent with the fact that there are individual differences in spatial ability in addition to preferences in relation to learning resources.

## 8.5.2 Familiarity of Structures

It has previously been proposed that students' learning anatomy perform well on tasks involving anatomical cross-sections if they are familiar and experienced with the structures (Hegarty *et al.*, 2009). Hegarty *et al.* (2009) compared the scores of first-year and fourth-year dentistry students on a Tooth Cross-Section Test. The authors' found that students in the fourth year of study performed better than students in the first year of study ( $t_{142} = 3.55$ ,  $p < 0.001$ ,  $d = 0.58$ ). In the U.S, dental students are selected based on performance on the Perceptual Ability Test (a test with different parts that examines the candidates' ability to interpret two-dimensional and three-dimensional objects) as a measure of spatial ability. When the authors included the first and fourth-year scores on the Perceptual Ability Test into the statistical analysis, the fourth-year students were still found to score significantly higher on the Tooth Cross-Section Test ( $F = 4.90$ ,  $p < 0.05$ ). These results led the authors to propose the more advanced fourth-year dental students had increased knowledge of the structures and familiarity of teeth. Thus spatial ability would no longer be as important.

The focus groups of the research in this chapter identified the students in UoE Vet 2 were less daunted by the cross-sections presented in the spatial MCQ questions and students had studied and become familiar with the cross-sections in the veterinary textbooks. Perhaps when initially studying the cross-sections the students used their spatial ability but as they became familiar with them, spatial ability was not required, and memory was used. Comparing two cohorts of students who did or did not look at cross-sections before a spatial anatomy MCQ may probe the effect of familiarity with these kinds of images. This is an area of future research for the project presented in this thesis.

### 8.5.3 Discussion of an Anatomy Threshold Concept

The idea of threshold concepts emerged from a UK wide initiative called ‘Enhancing Teaching and Learning Environment in Undergraduate Courses’ (ETL project), with the aim of identifying factors which contribute to successful learning environments across five disciplines and multiple higher education institutions.

Research on this topic, within the discipline of Economics, led Jan Meyer and Ray Land to propose the idea of threshold concepts (Meyer & Land, 2003). This work was similar to Perkins’ discussions of troublesome knowledge in his article “The Many Faces of Constructivism” (1999). In this article Perkins describes troublesome knowledge as inert (knowledge sitting in the ‘minds attic’ such as vocabulary words), ritual (knowledge with a routine to it, such as names and dates or the ‘invert and divide’ rule when dividing fractions), conceptually difficult knowledge (often most encountered in the sciences such as objects remain at a constant speed unless a force stops them), and foreign knowledge (knowledge presented as having a different perspective from the learner’s own perspective such as different views carried out in different faiths). Through interviews of students and academic colleagues within Economics and other disciplines, Meyer and Land identified that there are certain concepts that are important for a learner to grasp in order to master a subject.

An example used by Meyer and Land is the principle behind heat transfer. If there were two cups of hot tea and one had milk added straight away to cool it down, and the other had a delay of a few minutes before adding milk to the other cup, which cup of tea would be the coolest? The answer is the second cup with the milk added after a few minutes delay, as it would lose heat quicker. This is because this cup had the largest temperature difference between the temperature of the tea to the outside environmental temperature. Thus, there was a steeper temperature gradient and so the second cup would cool the quickest.

Meyer and Land explain that mastery of the concept of heat transfer and temperature gradients can be utilised by chefs when selecting pots and pans that have the most appropriate properties for certain ingredients. Meyer and Land eloquently explain that threshold concepts are not the same as core concepts: “A core concept is a conceptual ‘building block’ that progresses understanding of the subject; it has to be understood but it does not necessarily lead to a qualitatively different view of subject matter” (Meyer & Land, 2003:p.4). An important feature of a threshold concept is a shift in a learner’s thought, whereas core concepts are important fundamental components of a discipline such as the learning of anatomy for medical professions.

Threshold concepts have five guiding features (*Transformative, Irreversible, Integrative, Bounded, and Troublesome*), each is explained as follows. *Transformative*, grasping a threshold concept results in a shift in the learners’ perceptions of a subject. *Irreversible*, this shift in perception once acquired is unlikely to be forgotten by a learner. Cousin (2006) identifies the *Irreversible* quality of threshold concepts that makes it difficult for experts of a subject to identify threshold concepts themselves, as they would have mentally stepped over an irreversible threshold. *Integrative*: a threshold concept reveals the interrelatedness of something, meaning a learner can identify connections that were previously hidden until mastery of the threshold concept. *Bounded*: threshold concepts ‘knock on the door’ of other threshold concepts. *Troublesome*: threshold concepts involve troublesome knowledge as discussed by Perkins (1999).

In the thematic analysis presented in this chapter, the identification of the sub-themes ‘realisation to think spatially about anatomy’ and ‘progression of spatial thinking’, have similar features to the five discussed above of a threshold concept. For instance, during the focus groups, students gave explanations to their solving of the spatial MCQ questions, the spatial ability tests, and what would help learn this *troublesome* topic. Students often discussed how they ‘figured out’ the difficult questions after ‘oh no’ moments at seeing *troublesome* cross-sections. The focus groups revealed how

*transformed*, and *integrative* students' thoughts were, "I was going to have to know it in a 3D sense", and this was *irreversible* and *bounded*, "made me want to think about everything in 3D when I revised after that more." Threshold concepts are also explained by Meyer and Land as entering students into a liminal state about a concept, meaning students go back and forth across the threshold as they navigate the troublesome knowledge.

To date, there is no methodology for the identification of threshold concepts and as discussed in "How *Not* to Identify Threshold Concepts" in "Threshold Concepts in Practice" this would "represent an important milestone in the evolution of threshold concept scholarship" (Shinners-Kennedy, 2016:p.253). However, as the "How *Not* to Identify Threshold Concepts" title entails, there are criteria on how not to identify threshold concepts and these criteria can be categorised into two main headings; asking learners and asking experts. Shinners-Kennedy (2016) explains that the central issue with asking learners to remember a critical point in their learning of a topic is the exact details of the event are often forgotten or difficult to recall. Furthermore, he explains that by asking learners in the early years of study, such as the first year, their structure of knowledge is new and could be delicate, possibly providing a 'foetal' account of a threshold concept. In comparison, a more mature understanding of the concepts may exist in advanced learners such as post-graduates, however, the issue of how they remember the change in their knowledge remains. Shinners-Kennedy goes onto explain that experts notions of threshold concepts are biased by their own knowledge. Additionally, the way experts view the material (what kind of language they use and their 'research lens') can influence conclusions and interpretations.

So far, it appears no studies have identified any threshold concepts of anatomy learning, which could be due to a lack of a simple methodology and the difficulties of identifying threshold concepts due to their nature (Barradell, 2013). However, there are discussions on the subject in one anatomy education article, (Smith,

Martinez-Álvarez & McHanwell, 2014, discussed below). Smith, Martinez-Álvarez, and McHanwell (2014) investigated anatomy learning approaches using the ASSIST inventory. As mentioned previously the ASSIST inventory has been used to assess which approach (Deep, Strategic, or Surface) students take when learning. The authors proposed that for a high volume, content-rich subject such as anatomy a surface approach is commonly utilised although, this is not necessarily ‘bad’ as a level of surface learning needs to be undertaken for deep learning and understanding to occur.

Smith, Martinez-Álvarez, and McHanwell (2014) argue that for anatomy, the amount of surface learning required is exceptionally high and this may impose a barrier to deep learning, which some students may not wish to cross . The authors relate this progression of moving from a surface (or knowledge) to a deep (or understanding) approach as a threshold concept, this may be the case (Smith, Martinez-Álvarez & McHanwell, 2014). However the threshold of going from surface to deep may not truly have all the features of a threshold concept (i.e., Transformative, Irreversible, Integrative, Bounded, and Troublesome).

The area of threshold concepts, although not new, has been little explored in the context of veterinary medical curricula. The results of this chapter potentially identify spatial 3D thinking of anatomy as a threshold concept and this would be an interesting area of future study.

# **Chapter 9 General Discussion**

## **9.1 Overview of Chapter 9**

The main research aim of the study presented in this thesis was to investigate the relationship of spatial ability to two different anatomy teaching methods specifically within veterinary medical education. The first section of this chapter will discuss the research findings from chapter 2 (psychology literature on spatial ability), chapter 5 (comparison of two teaching methods), chapter 6 (development of a spatial MCQ assessment), chapter 7 (the malleability of spatial ability after learning anatomy), and chapter 8 (student views of spatial ability and anatomy), and relate these to what has been found in the anatomy education literature.

The second section of this chapter will discuss the limitations of the research project presented in this thesis and will discuss possible solutions to these limitations. The last section of this chapter will present possible future directions for related research in both veterinary anatomy education and veterinary medical education more broadly.

## **9.2 Discussion of Findings in Relation to the Literature**

Research on the connection between a student's spatial ability and anatomy learning has primarily been researched within medical education, with a smaller number of studies within other medically related professions such as dentistry and veterinary medicine. However, the relevant psychological literature on spatial ability and the links to human intelligence are not incorporated into the majority of studies. One of



the major points emphasized in the literature review of this thesis was the understanding that the concept of spatial ability arose from research on human intelligence. A review of the literature in this area, specifically on spatial ability, further highlighted the current issue with identifying clear, and research supported, spatial ability sub-factors/categories (Hegarty & Waller, 2005; Uttal *et al.*, 2013; Carroll, 1993).

This uncertainty in defining clear terminology has so far rarely been acknowledged or discussed in the anatomy education literature and may lead to the false classification and decisions of the spatial ability skills required for anatomy learning. For instance, the Group Embedded Figures Test implemented in Berney *et al.*'s study (2015) was used to measure the sub-category of spatial visualisation in relation to anatomy learning. However, this test has been identified in the psychology literature as measuring Closure Flexibility as identified by the psychologist John Carroll's (1993) immense factor analytic study on the structure of human cognitive abilities, including over 400 datasets of intelligence test data.

Another example of confusion around the classification of spatial ability tests was the use of the Mental Rotation Test (Vandenberg & Kuse, 1978; Peters *et al.*, 1995) to assess the sub-category spatial visualisation (Nguyen, Nelson & Wilson, 2012), which may be true for some factor analytic studies, although is regarded by the Linn and Petersen's (1985) categories as mental rotation. With such uncertainty and inconsistency in the definitions of different spatial ability sub-factors/categories in the psychology literature, confusion over terminology is bound to arise. In the face of such inconsistency, work such as Carroll's (1993) extensive factor analytic study is very useful along with other studies using confirmatory factor analysis (Hegarty & Waller, 2004). Alternative non-psychometric approaches, such as Linn and Petersen's (1985) can also be used in the interim, as was done in this research (section 2.3.3).

Another important psychology finding presented in this thesis is the importance of spatial ability to human general intelligence. Spatial ability, particularly the sub-category of spatial visualisation, has been discussed in the psychology literature as strongly linked to the human general intelligence factor, known as “g” (Carroll, 1993; Johnson & Bouchard, 2005; Lohman, 1996; Wai, Lubinski & Benbow, 2009). The importance of spatial ability to intelligence means, other than the ability to rotate structures in one’s head, spatial ability is also important for creative problem-solving thinking, as discussed more fully in chapter 2 (Gottfredson, 1997).

This relationship of spatial ability to intelligence may suggest that students with high spatial ability are less likely to struggle with anatomy learning than those with lower spatial ability. Tests of the sub-category spatial visualisation are often close to tests of g, examples of tests close to g are Raven’s matrices or Figure Classification test (Carroll, 1993; Lohman, 1979; Hegarty & Waller, 2005). Interestingly, the spatial visualisation test included in the research of this thesis was significantly related to anatomy scores (the Surface Development Test). However, the two spatial ability tests of the mental rotation sub-category were not significantly related (Card Rotation Test and Mental Rotation Test). This finding supports the theory that scores on tests of spatial visualisation could help to identify students struggling with anatomy.

Another important finding from the psychology literature on human intelligence is the positive manifold effect (Spearman, 1904), i.e. all tests in a cognitive battery correlate positively with one another. The positive manifold effect means other cognitive abilities are most likely used to answer spatial ability tests and thus used for anatomy learning. Designing educational research studies involving a range of cognitive ability tests across different cognitive domains and investigating the relationship to anatomy learning could further investigate the links.

Investigations into individual differences in performance during skill acquisition have indicated that the use of cognitive abilities decreases as skills are acquired (Ackerman, 1988). However, a study investigating spatial ability and skill acquisition in laparoscopic surgery technique found that initially spatial ability and general reasoning ability were related to performance but as learning progressed, and learners became equally proficient, general reasoning ability no longer correlated but spatial ability remained significantly correlated (Keehner *et al.*, 2006). This phenomenon of skill acquisition could be occurring in students learning anatomy and has been discussed in other studies researching the links between spatial ability and anatomy learning (Berney *et al.*, 2015; Hegarty *et al.*, 2009; Peterson & Mlynarczyk, 2016). Research investigating the learning process of students could help to identify whether spatial ability is used and whether this fluctuates with learning (see below 'Future Directions of Research' for a further discussion).

Another salient finding in the psychology literature is the practice/re-test effect observed when participants take two subsequent tests of the same cognitive ability test (McCaffrey & Westervelt, 1995; Lezak *et al.*, 2012). This effect is important for investigating whether spatial ability has improved between subsequent tests rather than due to practice of a test. If the practice effect is not accounted for, it can lead to the false conclusion that spatial ability has improved (Gutierrez *et al.*, 2017). The study presented in this thesis incorporated a control group of psychology students into the study design to account for the practice effect. This study found spatial ability did improve compared to a control cohort of students that did not learn anatomy for 2D mental rotation (CRT) and spatial visualisation (SDT). So far no other studies have used the CRT or SDT in relation to anatomy education, with the MRT primarily being used (Langlois *et al.*, 2009; Lufler *et al.*, 2012; Cui *et al.*, 2016; Chatterjee, 2011; Nguyen, Nelson & Wilson, 2012). However, the effect sizes observed in this study (1.05 – 1.06 for CRT, and 1.47 – 1.87 for SDT) were small.

Medical education studies have used 3D resources such as 3D computer models, stereoscopic displays, 3D printing, and Augmented Reality to improve anatomy knowledge, particularly spatial anatomy knowledge (Garg, Norman & Sperotable, 2001; Garg *et al.*, 1999b, 1999a; Nguyen, Nelson & Wilson, 2012; Berney *et al.*, 2015; Al-Khalili & Coppoc, 2014; Cui *et al.*, 2016; de Faria *et al.*, 2016; Lim *et al.*, 2016; Smith *et al.*, 2018; Küçük, Kapakin & Göktaş, 2016). This study used a novel approach of using 3D resources (3D printing and 3D computer models) to encourage students to think spatially about anatomy in general, and found the students that received the 3D spatial method had higher exam results for short answer questions and oral examinations. The thematically analysed qualitative data in this study provides initial insight into the learning of anatomy by students and how students think about anatomy. The data showed general agreement by students that spatial ability is important for learning anatomy however not all students had connected the spatial ability tests to the learning of anatomy for both UoE Vet 1 and UoE Vet 2.

The aim of the 3D spatial teaching method was to encourage the students to think spatially about the topic of anatomy. The thematic analysis of UoE Vet 2 (3D), identified a progression of spatial thinking and knowledge as indicated by students' explanations that it became easier to manipulate and use the 3D models in the lecture. Therefore, although the quantitative results of comparing exam results to spatial ability test results did not find large effect sizes, the qualitative analysis provides evidence of a link between spatial ability and anatomy learning.

Another important aspect discussed in this thesis is whether anatomy knowledge and understanding can be divided into factual/non-spatial and spatial knowledge. The ability to divide anatomy knowledge into these two categories could help to demonstrate and distinguish the use of spatial ability for some aspects of anatomical knowledge and understanding. Several other studies have described the design of a spatial MCQ (Berney *et al.*, 2015; Nguyen *et al.*, 2013; Garg, Norman & Sperotable, 2001; Rochford, 1985; Hoyek *et al.*, 2009; Keedy *et al.*, 2011; Schubert, Schnabel &

Winkelmann, 2009). The spatial MCQ assessment developed in this study involved the incorporation of anatomy knowledge and understanding to the design of three validated spatial ability tests. This approach resulted in a successful design, as evidenced by statistical correlation and regression analysis, and was further confirmed qualitatively by students' views on the spatial MCQ assessment.

### **9.3 Limitations of Research**

A limitation of the study design on this thesis is the 3D spatial teaching method was delivered for six sections out of a total of fourteen sections, whereas the 2D traditional teaching method was delivered to all fourteen sections as a traditional 2D method. One way to account for this difference would be to compare the end-of-course examination results specifically for the six sections taught with the 3D spatial teaching method (i.e., ultrasound/diagnostic imaging, head, forelimb, pharynx, joints, and hindlimb).

A ceiling effect was exhibited in relation to students' scores on the Surface Development Test (SDT), particularly for UoE Vet 1 and UoE Vet 2. This ceiling effect means the majority of students scored highly and therefore reduced the variance of the SDT results, and the associated SDT diff score (Post minus Pre). Converting the SDT scores to a proportion and inferencing to a binomial distribution was used to overcome this. Another approach could include reducing the time to answer the SDT and thus adding a time constraint element, or perhaps an alternative spatial visualisation test such as the Paper Folding test could be used. The SDT was chosen as it was thought to potentially relate to skills a veterinary surgeon is required to perform for topographical anatomy (i.e., had elements of face validity).

One of the most difficult limitations to account for in educational research designs is the ability to control other factors influencing the results. For instance, factors such as hours of study managed per student, use of other educational sources other than those provided (such as the 3D Virtual Canine website identified to be used during focus groups), other factors that could improve spatial ability such as recreational activities for example sporting activities, societies, or hobbies, and students sharing educational resources (e.g. UoE Vet 1 sharing 2D diagrams with UoE Vet 2 receiving the 3D teaching spatial method). However, these uncontrollable factors are realistic of educational settings and thus although difficult to account for provides a realistic setting for an experimental educational intervention (Norman, 2003).

One of the limitations of this study is that regardless of the nature and format of the questions used in an examination (which may range from simple fact-based to assessing at higher levels of Blooms Taxonomy), they do not directly assess the students' learning processes during the learning of the topic. To investigate this further along with the 'progression' theme identified in the thematic analysis of this thesis, a think-aloud technique (Ericsson & Simon H, 1980), where participants articulate their inner thought processes, could be adopted to understand whether the students used spatial ability while in the learning process, and whether this was used throughout the learning process or fluctuated. Furthermore, a componential approach could be used as discussed in chapter 4.

Although a control group of psychology students was incorporated into the study design of the study presented in this thesis, it could be argued given the different degree entrance requirements that the psychology students were not an equal control group, and so is a potential limitation (as discussed in detail in section 4.4).

Given the links of spatial ability to intelligence another limitation of this study is not incorporating the intelligence of the participants into the study, this incorporation

would also have helped to answer whether the psychology control group and veterinary students were equally matched. One way to have accounted for baseline differences in intelligence would have been to give the participants an intelligence test. However this may have been time-consuming and resulted in reduced participant numbers. Alternatively, a cognitive ability test of general reasoning ability such as the Abstract Reasoning Test or Raven's Progressive Matrices (Lohman, 1979) could have been incorporated as a lighter measure of intelligence as done by Hegarty et al. (2009).

A limitation of the spatial MCQ designed for this thesis, as discussed in chapter 5, is that UoE Vet 2 were not naïve to the assessment format of cross-sectional images because the previous cohort UoE Vet 1 had communicated about the MCQ format. However, this effect was not demonstrated by all questions. Questions 26 and 28 involving a radiograph and a cross-section (not commonly encountered in anatomy textbooks) had a statistically significant higher percentage of students answering correctly for UoE Vet 2 than UoE Vet 1, suggesting the 3D spatial teaching method could have improved cohort UoE Vet 2's spatial understanding.

## **9.4 Future Directions of Research**

Chapter 5 discussed cognitive load (Ayres & Paas, 2012; van Merriënboer & Ayres, 2005) particularly extraneous cognitive load, i.e. the cognitive load involved in how tasks are presented. In the context of the studies in this thesis, this translates to the manipulation of the 3D computer models – i.e., some students who did not find this intuitive. This may relate to an increased extraneous cognitive load associated with this activity. Cognitive load is an area for further investigation as a high cognitive load could reduce the efficacy of the 3D spatial teaching method. This would be an interesting area to further develop the research of this thesis as, although not investigated in this study, discussions in the focus groups identified that some

students found the 3D computer models difficult to initially follow during lectures but that this became easier over time, perhaps indicating a reduction in cognitive load. Furthermore, the CT and MRI scans used in the lectures were also discussed in the focus groups as difficult to understand but became easier with time. Further investigations into the cognitive load associated with various models and methods would be an area of interesting future research.

Threshold concepts (Meyer & Land, 2003; Cousin, 2006) are a relatively new area of research within higher education. The qualitative findings of the study presented in this thesis proposed the ‘thinking of anatomy spatially and in 3D’ was a threshold concept. This was because the students explanations of their thinking related to the principals of a threshold concept: the focus groups revealed how *transformed*, and *integrative* students’ thoughts were, “I was going to have to know it in a 3D sense”, and this was *irreversible* and *bounded*, “made me want to think about everything in 3D when I revised after that more.” However, there are currently no methods for identifying threshold concepts (Barradell, 2013; Shinnars-Kennedy, 2016), and this is an area worthy of future research.

Another spatially demanding subject of anatomy is histology and future research exploring the connections of spatial ability to histology could be important since histology is the 2D representation of 3D structures. Histology education studies have focussed on the use of virtual histology slides (Gatumu *et al.*, 2014; Roth, Wilson & Sandig, 2015; Husmann, O’Loughlin & Braun, 2009) and investigating a spatial ability element to this use may be educationally beneficial.

In conclusion, the findings of this research could impact on future curriculum developments with the introduction of spatially orientated teaching as qualitatively the students enjoyed the 3D resources. Teaching spatially may not automatically improve anatomy learning. However it may help students understand anatomy better



and can be one way to develop anatomy teaching. The statistically significant findings of this study represented small effect sizes, and therefore the educational impact is not exactly known. The impact could represent long or short-term retention or could potentially have implications for related areas later in the curriculum e.g. surgical skills and understanding. A spatial MCQ was successfully developed identifying anatomy knowledge, and understanding can be assessed and divided into non-spatial and spatial categories. This finding provides further evidence that spatial ability is related to the learning of anatomy. Incorporation of spatial teaching across a veterinary education curriculum could help progress students learning in other areas of the veterinary education curriculum given the importance of spatial ability to human general intelligence.

## References

- Ackerman, P.L. (1988) Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117 (3), pp. 288–318. Available from: doi:10.1037/0096-3445.117.3.288.
- Al-Khalili, S.M. & Coppoc, G.L. (2014) 2D and 3D stereoscopic videos used as pre-anatomy lab tools improve students' examination performance in a veterinary gross anatomy course. *Journal of Veterinary Medical Education*, 41 (1), pp. 68–76. Available from: doi:10.3138/jvme.0613-082R.
- Altman, D.G. (2006) The cost of dichotomising continuous variables. *BMJ*, 332 (7549), pp. 1080–1080. Available from: doi:10.1136/bmj.332.7549.1080.
- Ayres, P. & Paas, F. (2012) Cognitive load theory : New directions and challenges. *Applied Cognitive Psychology*, 26 (6), pp. 827–832. Available from: doi:10.1002/acp.2882.
- Baillie, S., Warman, S. & Rhind, S. (2014) A guide to assessment in veterinary medicine. *Bristol: University of Bristol Press*. Available from: doi:10.1177/1352458509103173.
- Barradell, S. (2013) The identification of threshold concepts: a review of theoretical complexities and methodological challenges. *Higher Education*, 65 (2), pp. 265–276. Available from: doi: 10.1007/s10734-012-9542-3.
- Barrett, T.J. & Hegarty, M. (2016) Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior*, 65, pp. 220–231. Available from: doi:10.1016/j.chb.2016.06.026.
- Bergman, E.M., Prince, K.J.A.H., Drukker, J., van der Vleuten, C.P.M., et al. (2008) How much anatomy is enough? *Anatomical Sciences Education*, 1 (4), pp. 184–188. Available from: doi:10.1002/ase.35.

- Berney, S., Bétrancourt, M., Molinari, G. & Hoyek, N. (2015) How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. *Anatomical Sciences Education*, 8 (5), pp. 452–462. Available from: doi:10.1002/ase.1524.
- Bloom, B. (1984) Taxonomy of Educational Objectives. In: D. McKay (ed). *The Cognitive Domain*. 1st edition. New York: Company Inc.
- Böckers, A., Mayer, C. & Böckers, T.M. (2014) Does learning in clinical context in anatomical sciences improve examination results, learning motivation, or learning orientation? *Anatomical Sciences Education*, 7 (1), pp. 3–11. Available from: doi:10.1002/ase.1375.
- Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3 (2), pp. 77–101.
- Brewer, C., Harrower, M. & The Pennsylvania State University (n.d.) *Colorbrewer*. Available from: <http://colorbrewer2.org/> [Accessed: 30 March 2015].
- Brown, P.M., Hamilton, N.M. & Denison, A.R. (2012) A novel 3D stereoscopic anatomy tutorial. *The Clinical Teacher*, 9 (1), pp. 50–53. Available from: doi:10.1111/j.1743-498X.2011.00488.x.
- Burnett, S.A., Lane, D.M. & Dratt, L.M. (1982) Spatial ability and handedness. *Intelligence*, 6 (1), pp. 57–68. Available from: doi:10.1016/0160-2896(82)90020-4.
- Carroll, J. (1993) *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Cambridge: Cambridge University Press.
- Case, S.M. & Swanson, D.B. (2002) *Constructing Written Test Questions for The Basic and Clinical Sciences*. 3rd edition. Philadelphia: National Board of Medical Examiners.
- Chatterjee, A. (2011) *The Importance of Spatial Ability and Mental Models in Learning Anatomy*. PhD thesis. Purdue University.

- Clements, D.N., Broadhurst, H., Clarke, S.P., Farrell, M., et al. (2013) The effectiveness of 3D animations to enhance understanding of cranial cruciate ligament rupture. *Journal of Veterinary Medical Education*, 40 (1), pp. 29–34. Available from: doi:10.3138/jvme.0512.037R.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colman, A.M. (2015) *A Dictionary of Psychology*. 4th edition. Oxford: Oxford University Press.
- Cousin, G. (2006) An introduction to threshold concepts. *Planet*, 17 (1), pp. 4–5. Available from: doi:10.11120/plan.2006.00170004.
- Crandall, B.W. & Hoffman, R.R. (2013) Cognitive Task Analysis. In: J.D Lee & Alex Kirlik (eds.). *The Oxford Handbook of Cognitive Engineering*. 1st edition. Oxford: Oxford University Press, pp. 1–17.
- Crawley, M. (2012) *The R Book*. 2nd edition. Chichester, England: Wiley-Blackwell.
- Cui, D., Wilson, T.D., Rockhold, R.W., Lehman, M.N., et al. (2016) Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anatomical Sciences Education*, 10 (1), pp. 34–45. Available from: doi:10.1002/ase.1626.
- Deary, I.J. (2012) Intelligence. *Annual Review of Psychology*, 63 (1), pp. 453–482. Available from: doi:10.1146/annurev-psych-120710-100353.
- Deary, I.J. (2001) *Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.
- Deary, I.J. (2000) *Looking Down on Human Intelligence: From Psychometrics To The Brain*. Oxford: Oxford University Press.

- Deary, I.J., Cox, S.R. & Ritchie, S.J. (2016) Getting spearman off the skyhook: One more in a century (Since Thomson, 1916) of attempts to vanquish g. *Psychological Inquiry*, 27 (3), pp. 192–199. Available from: doi:10.1080/1047840X.2016.1186525.
- Deary, I.J., Penke, L. & Johnson, W. (2010) The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11 (3), pp. 201–211. Available from: doi:10.1038/nrn2793.
- Deary, I.J., Weiss, A. & Batty, G.D. (2010) Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychological Science in the Public Interest*, 11 (2), pp. 53–79. Available from: doi:10.1177/1529100610387081.
- Denoyel, A., Pinson, C. & Passet, P. (n.d.) *Sketchfab*. Available from: <https://sketchfab.com/> [Accessed: 30 March 2015].
- Drake, R.L. (2014) A retrospective and prospective look at medical education in the United States: Trends shaping anatomical sciences education. *Journal of Anatomy*, 224 (3), pp. 256–260. Available from: doi:10.1111/joa.12054.
- 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), pp. 253–259. Available from: doi:10.1002/ase.117.
- Drake, R.L. & Pawlina, W. (2014) Multimodal education in anatomy: The perfect opportunity. *Anatomical Sciences Education*, 7 (1), pp. 1–2. Available from: doi:10.1002/ase.1426.
- Ekstrom, R., French, J., Harman, H. & Derman, D. (1976) *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, New Jersey: Educational Testing Service.
- Eliot, J. & Smith, I.M. (1983) *An International Directory of Spatial Tests*. Oxford: NFER-NELSON.

- Ericsson, K.A. & Simon H (1980) Verbal reports as data. *Psychological Review*, 87 (3), pp. 215–251.
- EU (2012) *Biocidal products regulation - ECHA*. Available from: <https://echa.europa.eu/information-on-chemicals/active-substance-suppliers> [Accessed: 17 August 2017].
- Evans, C., Kandiko Howson, C. & Forsythe, A. (2018) Making sense of learning gain in higher education. *Higher Education Pedagogies*, 3 (1), pp. 1–45. Available from: doi:10.1080/23752696.2018.1508360.
- Evans, H. & DeLahunta, A. (2010) *Guide to the Dissection of the Dog*. 7th edition. St. Louis, Mo.: Saunders/Elsevier.
- de Faria, J.W.V., Teixeira, M.J., de Moura Sousa Júnior, L., Otoch, J.P., et al. (2016) Virtual and stereoscopic anatomy: when virtual reality meets medical education. *Journal of Neurosurgery*, 125 (5), pp. 1105–1111. Available from: doi:10.3171/2015.8.JNS141563.
- Fernandez, R., Dror, I.E. & Smith, C. (2011) Spatial abilities of expert clinical anatomists: comparison of abilities between novices, intermediates, and experts in anatomy. *Anatomical Sciences Education*, 4 (1), pp. 1–8. Available from: doi:10.1002/ase.196.
- Flynn, J.R. (2014) The ‘Flynn Effect’ and Flynn’s paradox. *Intelligence*, 41 (6), pp. 851–857. Available from: doi:10.1016/j.intell.2013.06.014.
- French, J. (1951) *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors*. Chicago, IL, US: University of Chicago Press.
- Galton, F. (1885) On the anthropometric laboratory at the late international health exhibition. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 14, pp. 205–221.
- Galton, F. (1880) Statistics of mental imagery. *Mind*, 5 (19), pp. 301–318.

- Gardiner, A. (2010) *Small Animal Practice in British Veterinary Medicine 1920-1956*. PhD thesis. The University of Manchester.
- Gardiner, A. & Rhind, S. (2013) Taking a history on veterinary education. *Veterinary Record*, 173 (16), pp. 388–393. Available from: doi:10.1136/vr.f6273.
- Gardner, H. (1983) *Frames of Mind: The Theory of Multiple Intelligences*. London: Paladin.
- Garg, A., Norman, G., Spero, L. & Maheshwari, P. (1999a) Do virtual computer models hinder anatomy learning? *Academic Medicine*, 74 (10), pp. S87-9. Available from: doi:10.1097/00001888-199910000-00049.
- Garg, A., Norman, G., Spero, L. & Taylor, I. (1999b) Learning anatomy: Do new computer models improve spatial understanding? *Medical Teacher*, 21 (5), pp. 519–522. Available from: doi:10.1080/01421599979239.
- Garg, A., Norman, G. & Sperotable, L. (2001) How medical students learn spatial anatomy. *The Lancet*, 357 (9253), pp. 363–364. Available from: doi:10.1016/S0140-6736(00)03649-7.
- Gatumu, M.K., MacMillan, F.M., Langton, P.D., Headley, P.M., et al. (2014) Evaluation of usage of virtual microscopy for the study of histology in the medical, dental, and veterinary undergraduate programs of a UK University. *Anatomical Sciences Education*, 7 (5), pp. 389–398. Available from: doi:10.1002/ase.1425.
- Goldberg, T.E., Harvey, P.D., Wesnes, K.A., Snyder, P.J., et al. (2015) Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1 (1), pp. 103–111. Available from: doi:10.1016/j.dadm.2014.11.003.
- Gottfredson, L.S. (1997) Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24 (1), pp. 13–23. Available from: doi:10.1016/S0160-2896(97)90011-8.

- Graham, K. & Deary, I. (1999) A role for aptitude testing in surgery? *Journal of the Royal College of Surgeons of Edinburgh*, 36 (2), pp. 70–74.
- Guilford, J.P. & Lacey, J.I. (1947) *Printed Classification Tests, Army Air Forces Aviation Psychological Progress Research Report, No 5*. Washington, DC: U.S Government Printing Office.
- Guillot, A., Champely, S., Batier, C., Thiriet, P., et al. (2006) Relationship between spatial abilities, mental rotation and functional anatomy learning. *Advances in Health Sciences Education*, 12 (4), pp. 491–507. Available from: doi:10.1007/s10459-006-9021-7.
- Gutierrez, J.C., Chigerwe, M., Ilkiw, J.E., Youngblood, P., et al. (2017) Spatial and visual reasoning: Do these abilities improve in first-year veterinary medical students exposed to an integrated curriculum? *Journal of Veterinary Medical Education*, 44 (4), pp. 669–675. Available from: doi:10.3138/jvme.0915-158R3.
- Harle, M. & Towns, M. (2011) A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education*, 88 (3), pp. 351–360. Available from: doi:10.1002/ss.
- Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T. & Moriarty Gerrard, M.O. (2007) Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *The Journal of Applied Psychology*, 92 (2), pp. 373–385. Available from: doi:10.1037/0021-9010.92.2.373.
- Hegarty, M., Keehner, M., Cohen, C., Montello, D., et al. (2007) The role of spatial cognition in medicine: applications for selecting and training professionals. In: Gary L Allen (ed.). *Applied Spatial Cognition: From Research to Cognitive Technology*. 1st edition. Mahwah, N.J.: London, Lawrence Erlbaum Associates, pp. 285–315.



- Hegarty, M., Keehner, M., Khooshabeh, P. & Montello, D.R. (2009) How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, 19 (1), pp. 61–70. Available from: doi:10.1016/j.lindif.2008.04.006.
- Hegarty, M. & Waller, D. (2004) A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32 (2), pp. 175–191. Available from: doi:10.1016/j.intell.2003.12.001.
- Hegarty, M. & Waller, D.A. (2005) Individual differences in spatial abilities. In: P Shah & A Miyake (eds.). *The Cambridge Handbook of Visuospatial Thinking*. Cambridge: Cambridge University Press, pp. 121–169.
- Horn, J.L. (1989) Models of intelligence. In: R L Linn (ed.). *Intelligence: Measurement, theory and public policy: Proceedings of a Symposium in Honor of Lloyd G. Humphreys*. 1989 Champaign, IL, US: University of Illinois Press, pp. 29–73.
- Hoyek, N., Collet, C., Rastello, O., Fargier, P., et al. (2009) Enhancement of mental rotation abilities and its effect on anatomy learning. *Teaching and Learning in Medicine*, 21 (3), pp. 201–206. Available from: doi:10.1080/10401330903014178.
- Husmann, P.R., O’Loughlin, V.D. & Braun, M.W. (2009) Quantitative and qualitative changes in teaching histology by means of virtual microscopy in an introductory course in human anatomy. *Anatomical Sciences Education*, 2 (5), pp. 218–226. Available from: doi:10.1002/ase.105.
- Johnson, W. & Bouchard, T. (2005) The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33 (4), pp. 393–416. Available from: doi:10.1016/j.intell.2004.12.002.
- Johnson, W., Bouchard, T.J., Krueger, R.F., McGue, M., et al. (2004) Just one g: Consistent results from three test batteries. *Intelligence*, 32 (1), pp. 95–107. Available from: doi:10.1016/S0160-2896(03)00062-X.

- Johnson, W., Nijenhuis, J.T. & Bouchard, T.J. (2008) Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36 (1), pp. 81–95. Available from: doi:10.1016/j.intell.2007.06.001.
- Just, M.A. & Carpenter, P.A. (1985) Cognitive coordinate systems: Accounts of mental rotation. *American Psychologist*, 92 (2), pp. 713–713. Available from: doi:10.1037/h0021465.
- Keedy, A.W., Durack, J.C., Sandhu, P., Chen, E.M., et al. (2011) Comparison of traditional methods with 3D computer models in the instruction of hepatobiliary anatomy. *Anatomical Sciences Education*, 4 (2), pp. 84–91. Available from: doi:10.1002/ase.212.
- Keehner, M., Lipka, Y., Montello, D.R., Tendick, F., et al. (2006) Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology*, 20 (4), pp. 487–503. Available from: doi:10.1002/acp.1198.
- Keehner, M.M., Tendick, F., Meng, M. V, Anwar, H.P., et al. (2004) Spatial ability, experience, and skill in laparoscopic surgery. *American Journal of Surgery*, 188 (1), pp. 71–75. Available from: doi:10.1016/j.amjsurg.2003.12.059.
- Kelley, T.L. (1928) *Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities*. Stanford University, California: Stanford University Press.
- Kilburn, K.H., Warsaw, R.H. & Shields, M.G. (1989) Neurobehavioral dysfunction in firemen exposed to polychlorinated biphenyls (pcbs): Possible improvement after detoxification. *Archives of Environmental Health*, 44 (6), pp. 345–350. Available from: doi:10.1080/00039896.1989.9935904.
- el Koussy, A.A.H. (1935) *An Investigation Into the Factors in Tests Involving the Visual Perception of Space*. PhD thesis. University of London.
- Kovacs, K. & Conway, A.R.A. (2016) Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27 (3), pp. 151–177. Available from: doi:10.1080/1047840X.2016.1153946.

- Küçük, S., Kapakin, S. & Gökteş, Y. (2016) Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. *Anatomical Sciences Education*, 9 (5), pp. 411–421. Available from: doi:10.1002/ase.1603.
- Kyllonen, P.C., Lohman, D.F. & Snow, R.E. (1984) Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology*, 76 (1), pp. 130–145. Available from: doi:10.1037/0022-0663.76.1.130.
- Langdrige, D. (2007) Reflexivity. In: *Phenomenological Psychology Theory, Research and Method*. Harlow: Pearson/Prentice Hall, pp. 58–61.
- Langlois, J., Bellemare, C., Toulouse, J. & Wells, G.A. (2017) Spatial abilities and anatomy knowledge assessment: A systematic review. *Anatomical Sciences Education*, 10 (3), pp. 235–241. Available from: doi:10.1002/ase.1655.
- Langlois, J., Wells, G.A., Lecourtois, M., Bergeron, G., et al. (2009) Spatial abilities in an elective course of applied anatomy after a problem-based learning curriculum. *Anatomical Sciences Education*, 2 (3), pp. 107–112. Available from: doi:10.1002/ase.84.
- Latorre, R.M., García-Sanz, M.P., Moreno, M., Hernández, F., et al. (2007) How useful is plastination in learning anatomy? *Journal of Veterinary Medical Education*, 34 (2), pp. 172–176.
- Lee, T., Henry, J.D., Trollor, J.N. & Sachdev, P.S. (2010) Genetic influences on cognitive functions in the elderly: A selective review of twin studies. *Brain Research Reviews*, 64 (1), pp. 1–13. Available from: doi:10.1016/j.brainresrev.2010.02.001.
- Lezak, M., Howieson, D., Bigler, E. & Tranel, D. (2012) *Neuropsychological Assessment*. 5th edition. Oxford: Oxford University Press.
- Li, C., Zhu, W. & Nuttall, R.L. (2003) Familial handedness and spatial ability: A study with Chinese students aged 14–24. *Brain and Cognition*, 51 (3), pp. 375–384. Available from: doi:10.1016/S0278-2626(03)00041-1.

- Lim, K.H.A., Loo, Z.Y., Goldie, S.J., Adams, J.W., et al. (2016) Use of 3D printed models in medical education: A randomized control trial comparing 3D prints versus cadaveric materials for learning external cardiac anatomy. *Anatomical Sciences Education*, 9 (3), pp. 213–221. Available from: doi:10.1002/ase.1573.
- Linn, M. & Petersen, A. (1985) Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56 (6), pp. 1479–1498.
- Lischka, M. & Gittler, G. (1997) Spatial abilities and learning modes in anatomy beginners. In: A.J.J.A. Scherpbier, C.P.M van der Vleuten, J.J Rethans, & A.F.W van der Steeg (eds.). *Advances in Medical Education*. Dordrecht: Springer, pp. 166–169.
- Lohman, D. (1988) Spatial abilities as traits, processes and knowledge. In: R J Sternberg (ed.). *Advances in the Psychology of Human Intelligence*. Hillsdale, NJ: Erlbaum, pp. 181–248.
- Lohman, D. (1979) *Spatial Ability: A Review and Reanalysis of the Correlational Literature*. Aptitude Research Project: Technical Report No. 8. School of Education: Stanford University.
- Lohman, D. (1996) Spatial ability and g. In: I. Dennis & P. Tapsfield (eds.). *Human Abilities: Their Nature and Measurement*. 1st edition. Mahwah, N.J.: Erlbaum, pp. 97–116.
- Lohman, D.F. (1986) The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception & Psychophysics*, 39 (6), pp. 427–436. Available from: doi:10.3758/BF03207071.
- Lufner, R.S., Zumwalt, A.C., Romney, C.A & Hoagland, T.M. (2012) Effect of visual-spatial ability on medical students' performance in a gross anatomy course. *Anatomical Sciences Education*, 5 (1), pp. 3–9. Available from: doi:10.1002/ase.264.

- Maeda, Y. & Yoon, S. (2013) A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualisation tests: Visualisation of rotations (PSVT:R). *Educational Psychology Review*, 25 (1), pp. 69–94. Available from: doi:10.1007/s.
- Masters, M. & Sanders, B. (1993) Is the gender difference in mental rotation disappearing? *Behavior Genetics*, 23 (4), pp. 337–341.
- Mayer, R.E. (2002) Cognitive Theory and the Design of Multimedia Instruction: An Example of the Two-Way Street Between Cognition and Instruction. *New Directions for Teaching and Learning*, 2002 (89), pp. 55–71. Available from: doi:10.1002/tl.47.
- Mayer, R.E. (2001) *Multimedia Learning*. 2nd edition. Cambridge: Cambridge University Press.
- 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), pp. 389–401. Available from: doi:10.1037/0022-0663.86.3.389.
- McCaffrey, R.J. & Westervelt, H.J. (1995) Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, 5 (3), pp. 203–221. Available from: doi:10.1007/BF02214762.
- McFarlane, M. (1925) *A Study of Practical Ability*. PhD thesis. University of London.
- McGee, M. (1979) Human spatial abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences. *Psychological Bulletin*, 86 (5), pp. 889–918.
- McGrath, H.C., Guerin, B., Harte, E., Frearson, M., et al. (2015) *Learning Gain in Higher Education*. Cambridge: RAND Corporation.

- McGrew, K.S. (2009) CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37 (1), pp. 1–10. Available from: doi:10.1016/j.intell.2008.08.004.
- 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), pp. 479–486. Available from: doi:10.1002/ase.1475.
- van Merriënboer, J.J.G. & Ayres, P. (2005) Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, 53 (3), pp. 5–13. Available from: doi:10.1007/BF02504793.
- Meyer, J. & Land, R. (2003) *Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising within the Disciplines*. Occasional Report 4, UK: ETL Project. Available from: [https://www.researchgate.net/publication/285473047\\_Threshold\\_concepts\\_and\\_troublesome\\_knowledge\\_Linkages\\_to\\_ways\\_of\\_thinking\\_and\\_practising\\_within\\_the\\_disciplines](https://www.researchgate.net/publication/285473047_Threshold_concepts_and_troublesome_knowledge_Linkages_to_ways_of_thinking_and_practising_within_the_disciplines).
- Michael, W., Guilford, J., Fruchter, B. & Zimmerman, W. (1957) The description of spatial-visualization abilities. *Educational and Psychological Measurement*, 17 (2), pp. 185–199.
- Miller, C. (1996) A historical review of applied and theoretical spatial visualisation publications in engineering graphics. *Engineering Design Graphics Journal*, 60 (3), pp. 12–31.
- Miller, G.E. (1990) The assessment of clinical skills/competence/performance. *Academic Medicine*, 65 (9), pp. S63–S67.
- Mohler, J.L. (2008) A review of spatial ability research. *Engineering Design Graphics Journal*, 72 (2), pp. 19–30.

- Mumaw, R.J. & Pellegrino, J.W. (1984) Individual differences in complex spatial processing. *Journal of Educational Psychology*, 76 (5), pp. 920–939. Available from: doi:10.1037/0022-0663.76.5.920.
- Mumaw, R.J., Pellegrino, J.W., Kail, R. V. & Carter, P. (1984) Different slopes for different folks: Process analysis of spatial aptitude. *Memory & Cognition*, 12 (5), pp. 515–521. Available from: doi:10.3758/BF03198314.
- Nguyen, N., Mulla, A., Nelson, A.J. & Wilson, T.D. (2013) Visuospatial anatomy comprehension: the role of spatial visualization ability and problem-solving strategies. *Anatomical Sciences Education*, 7 (4), pp. 280–288. Available from: doi:10.1002/ase.1415.
- Nguyen, N., Nelson, A.J. & Wilson, T.D. (2012) Computer visualizations: factors that influence spatial anatomy comprehension. *Anatomical Sciences Education*, 5 (2), pp. 98–108. Available from: doi:10.1002/ase.1258.
- Nicholson, D.T., Chalk, C., Funnell, W.R.J. & Daniel, S.J. (2006) Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education*, 40 (11), pp. 1081–1087. Available from: doi:10.1111/j.1365-2929.2006.02611.x.
- Norman, G. (2003) RCT = results confounded and trivial: The perils of grand educational experiments. *Medical Education*, 37 (7), pp. 582–584. Available from: doi:10.1046/j.1365-2923.2003.01586.x.
- Norman, G. (2009) Teaching basic science to optimize transfer. *Medical Teacher*, 31 (9), pp. 807–811. Available from: doi:10.1080/01421590903049814.
- Payton, A. (2009) The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychology Review*, 19 (4), pp. 451–477. Available from: doi:10.1007/s11065-009-9116-z.
- Pedersen, K. (2012) Supporting students with varied spatial reasoning abilities in the anatomy classroom. *Teaching Innovation Projects*, 2 (1), Article 2.

- Perkins, D. (1999) The many faces of constructivism. *Educational Leadership*, 57 (3), pp. 6–11. Available from: doi:10.1111/j.1467-8535.2009.00994\_9.x.
- Peters, M., Laeng, B., Latham, K., Jackson, M., et al. (1995) A redrawn Vandenberg & Kuse mental rotations test- Different versions and factors that affect performance. *Brain and Cognition*, 28 (1), pp. 39–58.
- Peterson, D.C. & Mlynarczyk, G.S. a (2016) Analysis of traditional versus three-dimensional augmented curriculum on anatomical learning outcome measures. *Anatomical Sciences Education*, 9 (6), pp. 529–536. Available from: doi:10.1002/ase.1612.
- Plumley, L., Armstrong, R., De Ribaupierre, S. & Eagleson, R. (2013) Spatial ability and training in virtual neuroanatomy. *Studies in Health Technology and Informatics*, 184, pp. 324–329. Available from: doi:10.3233/978-1-61499-209-7-324.
- Pooley, F. & Wapenaar, W. (2018) Barriers and motivators towards a national final year assessment in the United Kingdom and the Republic of Ireland for veterinary undergraduates. *Veterinary Record*, 183 (15), p. 472. Available from: doi:10.1136/vr.104674.
- Preece, D., Williams, S.B., Lam, R. & Weller, R. (2013) ‘Let’s get physical’: advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education*, 6 (4), pp. 216–224. Available from: doi:10.1002/ase.1345.
- Provo-Klimek, J. (2002) Gross anatomy education in veterinary medicine. *Faseb Journal*, 16 (5), p. A1110.
- Provo, J., Lamar, C. & Newby, T. (1998) *Spatial Ability, Gender, and the Ability to Visualize Anatomy in Three Dimensions*. Annual Meeting of the American Educational Research Association, p. 25. Available from: <http://eric.ed.gov/?id=ED421337>.



- Provo, J., Lamar, C. & Newby, T. (2002) Using a cross section to train veterinary students to visualize anatomical structures in three dimensions. *Journal of Research in Science Teaching*, 39 (1), pp. 10–34. Available from: doi:10.1002/tea.10007.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R foundation for statistical computing. Vienna, Austria. Available from: <https://www.r-projects.org/>.
- Reilly, D. & Neumann, D.L. (2013) Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles*, 68 (9), pp. 521–535. Available from: doi:10.1007/s11199-013-0269-0.
- Revelle, W. (2018) *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. Evanston, Illinois, USA. Available from: <https://cran.r-project.org/package=psych> Version = 1.8.4.
- Rhind, S. & Baillie, S. (2018) Assessment of veterinary students – where we’ve been and where we’re going. *Veterinary Record*, 183 (15), pp. 469–471. Available from: doi:10.1136/vr.k4381.
- Ritchie, S. (2015) *Intelligence: All That Matters*. London: John Murray Learning.
- Roach, V.A., Mistry, M.R. & Wilson, T.D. (2014) Spatial visualization ability and laparoscopic skills in novice learners: Evaluating stereoscopic versus monoscopic visualizations. *Anatomical Sciences Education*, 7 (4), pp. 295–301. Available from: doi:10.1002/ase.1412.
- Rochford, K. (1985) Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, 19 (1), pp. 13–26. Available from: doi:10.1111/j.1365-2923.1985.tb01134.x.
- Roosendaal, T. (2002) *Blender* (Version 2.79) [Computer Program]. Available from: <https://www.blender.org/> [Accessed: 9 December 2014].

- Roth, J.A., Wilson, T.D. & Sandig, M. (2015) The development of a virtual 3D model of the renal corpuscle from serial histological sections for E-learning environments. *Anatomical Sciences Education*, 8 (6), pp. 574–583. Available from: doi:10.1002/ase.1529.
- Schubert, S., Schnabel, K.P. & Winkelmann, A. (2009) Assessment of spatial anatomical knowledge with a ‘three-dimensional multiple choice test’ (3D-MC). *Medical Teacher*, 31 (1), pp. e13–e17. Available from: doi:10.1080/01421590802334325.
- Schuman, H. (1952) *Leonardo on the Human Body*. Unabridged republication from 1952, Mineola, New York: Dover publications.
- Senn, S. (2003) Disappointing dichotomies. *Pharmaceutical Statistics*, 2 (4), pp. 239–240. Available from: doi:10.1002/pst.90.
- Shepard, R.N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science*, 171 (3972), pp. 701–703.
- Shinners-Kennedy, D. (2016) How not to identify threshold concepts. In: R Land, J.H.F. Meyer, & M.T Flanagan (eds.). *Threshold Concepts in Practice*. Rotterdam: SensePublishers, pp. 253–267.
- Silverman, I., Choi, J. & Peters, M. (2007) The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior*, 36 (2), pp. 261–268. Available from: doi:10.1007/s10508-006-9168-6.
- Silverman, I. & Eals, M. (1992) Sex differences in spatial abilities: Evolutionary theory and data. In: J. H. Barkow, L. Cosmides, & J Tooby (eds.). *The adapted mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, pp. 531–549.
- Smith, C.F., Martínez-Álvarez, C. & McHanwell, S. (2014) The context of learning anatomy: Does it make a difference? *Journal of Anatomy*, 224 (3), pp. 270–278. Available from: doi:10.1111/joa.12089.

- Smith, C.F. & Mathias, H.S. (2010) Medical students' approaches to learning anatomy: Students' experiences and relations to the learning environment. *Clinical Anatomy*, 23 (1), pp. 106–114. Available from: doi:10.1002/ca.20900.
- Smith, C.F., Tollemache, N., Covill, D. & Johnston, M. (2018) Take away body parts! An investigation into the use of 3D-printed anatomical models in undergraduate anatomy education. *Anatomical Sciences Education*, 11 (1), pp. 44–53. Available from: doi:10.1002/ase.1718.
- Smith, I. (1964) *Spatial ability: Its Educational and Social Significance*. London: University of London Press.
- Somers, M., Shields, L.S., Boks, M.P., Kahn, R.S., et al. (2015) Cognitive benefits of right-handedness: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 51, pp. 48–63. Available from: doi:10.1016/j.neubiorev.2015.01.003.
- Spearman, C. (1904) 'General Intelligence', objectively determined and measured. *The American Journal of Psychology*, 15 (2), pp. 201–292.
- Speedwell (n.d.) *Speedwell software*. Available at: <https://www.speedwellsoftware.com/>
- Stephenson, W. (1931) Tetrad-differences for verbal subtests relative to non-verbal subtests. *Journal of Educational Psychology*, 22 (5), pp. 334–350. Available from: doi:10.1037/h0070857.
- Stuart, E.A. & Rubin, D.B. (2008) Best practices in quasi-experimental designs: Matching methods for causal inference. In: J Osborne (ed.). *Best Practices in Quantitative Methods*, 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc, pp. 155-176.
- Sullivan, G.M. (2011) Getting off the 'gold standard': Randomized controlled trials and education research. *Journal of Graduate Medical Education*, 3 (3), pp. 285–289. Available from: doi:10.4300/JGME-D-11-00147.1.

- Sweeney, R., Hayes, J. & Chiavaroli, N. (2014) Does spatial ability help the learning of anatomy in a biomedical science course? *Anatomical Sciences Education*, 7 (4), pp. 289–294. Available from: doi:10.1002/ase.1418.
- Tait, H., Entwistle, N.J. & McCune, V. (1998) ASSIST: a re-conceptualisation of the Approaches to Studying Inventory. In: C. Rust (eds). *Improving Students as Learners*. Oxford: Oxford Brookes University, pp. 262–271.
- Tan, S., Hu, A., Wilson, T., Ladak, H., et al. (2012) Role of a computer-generated three-dimensional laryngeal model in anatomy teaching for advanced learners. *Journal of Laryngology and Otology*, 126 (4), pp. 395–401. Available from: doi:10.1017/S0022215111002830.
- Tavakol, M. & Dennick, R. (2011) Post-examination analysis of objective tests. *Medical Teacher*, 33 (6), pp. 447–458. Available from: doi:10.3109/0142159X.2011.564682.
- The University of Newcastle Australia (n.d.). *Spatial Diagnostic*. Available at: <https://psych.newcastle.edu.au/SpatDiag/AltC/index.html> [Accessed: 26 January 2019].
- Theoret, C.L.C., Carmel, E.-N. & Bernier, S. (2007) Why dissection videos should not replace cadaver prosections in the gross veterinary anatomy curriculum: results from a comparative study. *Journal of Veterinary Medical Education*, 34 (2), pp. 151–156.
- Thurstone, L.L. (1957) *Primary Mental Abilities*. Chicago, Ill.: University of Chicago Press.
- Thurstone, L.L. (1950) Some primary abilities in visual thinking. *Proceedings of the American Philosophical Society*, 94 (6), pp. 517–521.
- Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., et al. (2013) The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139 (2), pp. 352–402. Available from: doi:10.1037/a0028446.

- Vandenberg, S. & Kuse, A. (1978) Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47 (2), pp. 599–604.
- Vernon, P. (1950) *The Structure of Human Abilities*. Methuen, NY: Wiley.
- Vorstenbosch, M.A.T.M., Bouter, S.T., van den Hurk, M.M., Kooloos, J.G.M., et al. (2014) Exploring the validity of assessment in anatomy: Do images influence cognitive processes used in answering extended matching questions? *Anatomical Sciences Education*, 7 (2), pp. 107–116. Available from: doi:10.1002/ase.1382.
- Vorstenbosch, M.A.T.M., Klaassen, T.P.F.M., Donders, A.R.T.R., Kooloos, J.G.M., et al. (2013a) Learning anatomy enhances spatial ability. *Anatomical Sciences Education*, 6 (4), pp. 257–262. Available from: doi:10.1002/ase.1346.
- Vorstenbosch, M.A.T.M., Klaassen, T.P.F.M., Kooloos, J.G.M., Bolhuis, S.M., et al. (2013b) Do images influence assessment in anatomy? Exploring the effect of images on item difficulty and item discrimination. *Anatomical Sciences Education*, 6 (1), pp. 29–41. Available from: doi:10.1002/ase.1290.
- Voyer, D., Voyer, S. & Bryden, M.P. (1995) Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117 (2), pp. 250–270. Available from: doi:10.1037/0033-2909.117.2.250.
- Wai, J., Lubinski, D. & Benbow, C.P. (2009) Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101 (4), pp. 817–835. Available from: doi:10.1037/a0016127.
- Waterhouse, L. (2006) Multiple intelligences, the Mozart Effect, and emotional intelligence: A critical review. *Educational Psychologist*, 41 (4), pp. 207–225. Available from: doi:10.1207/s15326985ep4104\_1.

- Wechsler, D. (1958) *The Measurement and Appraisal of Adult Intelligence*. 4th edition. Baltimore, MD, US: Williams & Wilkins Co.
- Witkin, H.A., Raskin, E. & Oltman, P.K. (1971) *A Manual for the Group Embedded Figures Test*. Menlo Park, CA: Mind Garden Inc.
- Wise, L. L., McLaughlin, D. H., & Steel, L. (1979). The Project TALENT data bank. Palo Alto, CA: American Institutes for Research.
- Wright, R.J., Frier, B.M. & Deary, I.J. (2009) Effects of acute insulin-induced hypoglycemia on spatial abilities in adults with type 1 diabetes. *Diabetes Care*, 32 (8), pp. 1503–1506. Available from: doi:10.2337/dc09-0212.
- Yamine, K. & Violato, C. (2015) A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. *Anatomical Sciences Education*, 8 (6), pp. 525–538. Available from: doi:10.1002/ase.1510.
- Zell, E., Krizan, Z. & Teeter, S.R. (2015) Global gender differences can be operationalized and tested. *American Psychologist*, 70 (7), pp. 664–665. Available from: doi:10.1037/a0039590.
- Zimmerman, W.S. (1954) Hypotheses concerning the nature of the spatial factors. *Educational and Psychological Measurement*, 14 (2), pp. 396–400. Available from: doi:10.1177/001316445401400220.



# Appendices





# Appendix 1 – Participant Consent Forms

## R(D)SVS Students

Spatial Ability and Anatomy Teaching Project  
R(D)SVS  
University of Edinburgh

### CONFIDENTIAL CONSENT FORM

I agree / do not agree to take part in the following activities as part of the  
“Spatial Ability and Anatomy Teaching Research Project”

Study activities	Agree	Do not agree
Spatial ability testing		
Post course evaluations		
Focus groups or interviews		

1. I agree to participate in a research study on spatial ability undertaken by The University of Edinburgh Division of Veterinary Medical Education
2. I have been given a full explanation of the nature and purpose of the study and have been given the opportunity to ask questions about these.
3. I have been assured that my participation is entirely voluntary and I understand that I am free to withdraw my participation at any time without needing to justify my decision. I can also ask afterwards for specific comments not to be used in the research.
4. I understand that focus groups/ interviews will be audio-recorded and transcribed. These will be treated in strictest confidence and will only be accessible to the research team. They will be destroyed when no-longer required for the research.
5. I understand that anonymous data from this study may be published as research findings, including anonymised quotes, in journal articles, book chapters or a thesis / dissertation. I am aware that I can see any such material before publication upon request.

Name (please print clearly)

Signed

Date

Please complete and sign this form (whether or not you consent to participate).

**Many thanks for your help**



# UoB Veterinary Students

**Spatial Ability and Anatomy Teaching Project  
R(D)SVS  
University of Edinburgh**

## **CONFIDENTIAL CONSENT FORM**

**I agree / do not agree to take part in the following activities as part of the  
“Spatial Ability and Anatomy Teaching Research Project”**

<b>Study activities</b>	<b>Agree</b>	<b>Do not agree</b>
Spatial ability testing		

1. I agree to participate in a research study on spatial ability undertaken by The University of Edinburgh Division of Veterinary Medical Education
2. I have been given a full explanation of the nature and purpose of the study and have been given the opportunity to ask questions about these.
3. I have been assured that my participation is entirely voluntary and I understand that I am free to withdraw my participation at any time without needing to justify my decision. I can also ask afterwards for specific comments not to be used in the research.
4. I understand that focus groups/ interviews will be audio-recorded and transcribed. These will be treated in strictest confidence and will only be accessible to the research team. They will be destroyed when no-longer required for the research.
5. I understand that anonymous data from this study may be published as research findings, including anonymised quotes, in journal articles, book chapters or a thesis / dissertation. I am aware that I can see any such material before publication upon request.

Name (please print clearly)

Signed

Date

Please complete and sign this form (whether or not you consent to participate).

**Many thanks for your help**



# Psychology Students

**Spatial Ability and Anatomy Teaching Project  
R(D)SVS  
University of Edinburgh**

## **CONFIDENTIAL CONSENT FORM**

**I agree / do not agree to take part in the following activities as part of the  
“Spatial Ability and Anatomy Teaching Research Project”**

<b>Study activities</b>	<b>Agree</b>	<b>Do not agree</b>
Spatial ability testing		

1. I agree to participate in a research study on spatial ability undertaken by The University of Edinburgh School of Psychology and Royal (DICK) School of Veterinary Studies.
2. I have been given a full explanation of the nature and purpose of the study and have been given the opportunity to ask questions about these.
3. I have been assured that my participation is entirely voluntary and I understand that I am free to withdraw my participation at any time without needing to justify my decision. I can also ask afterwards for specific comments not to be used in the research.
4. I understand that anonymous data from this study may be published as research findings, including anonymised quotes, in journal articles, book chapters or a thesis / dissertation.

Name (please print clearly)

Signed

Date

Please complete and sign this form (whether or not you consent to participate).

**Many thanks for your help**



# Appendix 2 - Participant Demographic Form

## Veterinary Students and Psychology Students



### Spatial Ability and Anatomy Teaching Project

R(D)SVS

University of Edinburgh

Thank you for participating in this exciting project!

To help us analyse the data completely, please take a minute to fill in the following brief details about yourself. Please be assured that this data will then be anonymised.

Also note that if at any time you wish to remove your data from the study, all you need to do is e-mail [susan.rhind@ed.ac.uk](mailto:susan.rhind@ed.ac.uk).

1. Age \_\_\_\_\_

2. Sex \_\_\_\_\_

3. Do you have a previous degree?

Yes What was your degree title? \_\_\_\_\_

No

4. Are you right or left handed?

Right handed



Left handed

5. How often do you play videogames on your PC, tablet or phone (please select one):

Once a week

Once a month

Once every 2 or 3 months

Less Frequently than all the above

Never

We intend to re-measure your spatial ability at the end of your first year, if you agree can you please leave your name below. Again, the data will be linked via an anonymous number.

All completed surveys will be entered into a prize draw for 2 x £50 amazon vouchers.

Name \_\_\_\_\_

# Appendix 3 - Participant PhD Information Sheet

## Veterinary Students



### Spatial Ability and Anatomy Teaching Project

R(D)SVS

University of Edinburgh

#### INFORMATION SHEET

Thank you for considering taking part in this project.

#### **Background to the Study**

For qualified veterinary surgeons anatomy knowledge is pivotal, the vast number of species dealt with along with the fact that immediately after graduation, veterinary graduates are permitted to carry out major surgical procedures further emphasises the necessity for strong anatomy knowledge; perhaps even more so than in medicine. The teaching of veterinary anatomy and related clinical skills in the veterinary curriculum is also changing as in medical education. For example the introduction of imaging techniques such as radiography and ultrasonography are increasingly being used from the early stages of the curriculum, although students still lack complete understanding of interpreting these imaging techniques, a technique which takes years of experience to develop. Students' spatial ability has been shown to be related to anatomy understanding and some studies have shown that learning of anatomy can actually improve candidate's spatial abilities. This study will explore the baseline spatial ability of veterinary students and evaluate any changes in this ability following a teaching intervention using various imaging modalities and 3D printing with a spatial ability orientation.

The results of the study will be used to inform future curriculum development and teaching strategies.

**How will this study be done?**

At the start of semester, we will offer the opportunity to participate in a short series of well validated tests of your spatial ability. You will be able to access the results of these tests should you wish to do so. Following the animal body (1) course, the tests will be repeated. You will also be given the opportunity to participate in focus groups or short interviews to discuss spatial ability and the understanding of anatomy.

**Will taking part affect me in any way?**

Your decision to volunteer for this study or to withdraw at any stage will not impact on your academic performance in any way.

We will not disclose the results of the spatial ability tests to anyone other than yourself.

**How do I found out more?**

If you have any further questions about this project, please contact me by email ([julie.dickson@ed.ac.uk](mailto:julie.dickson@ed.ac.uk)).

# Psychology Students



## Spatial Ability and Anatomy Teaching Project

R(D)SVS

University of Edinburgh

### INFORMATION SHEET

Thank you for considering taking part in this project.

#### **Background to the Study**

For qualified veterinary surgeons anatomy knowledge is pivotal, the vast number of species dealt with along with the fact that immediately after graduation, veterinary graduates are permitted to carry out major surgical procedures further emphasises the necessity for strong anatomy knowledge; perhaps even more so than in medicine. The teaching of veterinary anatomy and related clinical skills in the veterinary curriculum is also changing as in medical education. For example the introduction of imaging techniques such as radiography and ultrasonography are increasingly being used from the early stages of the curriculum, although students still lack complete understanding of interpreting these imaging techniques, a technique which takes years of experience to develop. Students' spatial ability has been shown to be related to anatomy understanding and some studies have shown that learning of anatomy can actually improve candidate's spatial abilities. This study will explore the baseline spatial ability of veterinary students and evaluate any changes in this ability following a teaching intervention using various imaging modalities and 3D printing with a spatial ability orientation.

The results of the study will be used to inform future curriculum development and teaching strategies.

**How will this study be done?**

At the start of semester, we will offer the opportunity to participate in a short series of well validated tests of your spatial ability. You will be able to access the results of these tests should you wish to do so. Following 15-16 weeks later, the tests will be repeated.

**Will taking part affect me in any way?**

Your decision to volunteer for this study or to withdraw at any stage will not impact on your academic performance in any way.

We will not disclose the results of the spatial ability tests to anyone other than yourself.

**How do I found out more?**

If you have any further questions about this project, please contact me by email ([julie.dickson@ed.ac.uk](mailto:julie.dickson@ed.ac.uk)).

## Appendix 4 - Anatomy Workbook Marking Sheet

### AB1 Anatomy Workbook Feedback

form Examination no: \_\_\_\_\_ MARKING SCHEME FOR

CONSISTENCY (JUST TICKS FOR STUDENT COPIES) OUT OF 10

#### Presentation (20% of overall mark)

	Excellent	Good	OK	Poor	Very Poor
Organisation, contents page, page numbering, scales included					
Acknowledgements (others dissections etc)					
Tidy legible labelling with clear lines (handwritten or typed)					
Inclusion of instructions					

#### Content (75% of overall mark)

Please refer to specific comments in workbook regarding corrections of labelling etc.

<i>Given in order of importance</i>	Excellent	Good	OK	Poor	Very Poor
Accuracy of identification & labelling					
Inclusion of relevant structures					
Quality of dissection (adjusted according to specimen quality)					
Tissues defined (eg muscle/m., nerve/n. etc)					
Appropriate number of images (standard views) per practical					
Figure titles as requested giving species, orientation and region in consistent manner size, etc)					

Quality of images or drawings (clarity, size etc)					
--	--	--	--	--	--

**Referencing (5% of overall mark)**

	Yes	No
Reference list included		
Reference list presented as requested		

**Comments**

<b>OVERALL MARK:</b>
----------------------

- 10 Highly distinguished
- 9 Distinguished
- 8 Highly creditable
- 7 Creditable
- 6 Convincing pass
- 5 Minimum adequate
- 4 Just failed
- 3 Bad failure
- 2 Very bad failure
- 1 Almost no competent response
- 0 No answer

# Appendix 5 - 3D Teaching Method Tutorials

## Forelimb 3

There are 3 stations in this tutorial and you have 15 minutes to complete each station by working as a team and answering the questions below using the 3D printed models, any textbooks you have and the 3D computer models on the iPad provided (nominate one member of the group to log on to LEARN on the iPad).

Once the 15 minutes is finished stay seated in your groups and we will rotate the 3D printed models around for the next station.

### Station 1- Shoulder Joint

1. On the 3D printed model identify;

Glenoid cavity	Greater tubercle
Head of humerus	Infraspinous fossa
Lateral glenohumeral ligament	Medial glenohumeral ligament
Supraspinous fossa	Tendon of insertion of infraspinatus m.
Tendon of insertion of subscapularis m.	Tendon of insertion of supraspinatus m.

2. On the 3D printed model **identify** the course of the tendon of origin of the biceps brachii muscle. What holds this tendon in place and where would this be located on the model?
3. Contracture/fibrosis of the infraspinatus muscle (i.e. shortening/hardening of the muscle) can sometimes happen in dogs, most commonly working dogs. How would a dog with contracture of the infraspinatus muscle hold its forelimb?



## Station 2- Elbow Joint

1. On the 3D printed model identify;

<b>Anconeal process</b>	<b>Annular ligament</b>
<b>Humerus</b>	<b>Lateral coronoid process</b>
<b>Lateral epicondyle</b>	<b>Lateral epicondyle</b>
<b>Medial collateral ligament</b>	<b>Medial coronoid process</b>
<b>Medial epicondyle</b>	<b>Oblique ligament</b>
<b>Olecranon fossa</b>	<b>Olecranon ligament</b>
<b>Olecranon tuberosity</b>	<b>Radial fossa</b>
<b>Radius</b>	<b>Tendon of insertion of biceps brachii m.</b>
<b>Trochlear notch of ulna</b>	<b>Ulna</b>

2. What is incorrect with the collateral ligaments on the 3D printed model?

3. What is incorrect with the biceps brachii tendon of insertion on the 3D computer model?

4. Name three bony prominences which you think are important for the stability and the function of the elbow joint.

## Station 3- Forelimb Nerves

1. Use the coloured cords and the forelimb provided to show the course and distribution of each of the six main nerves of the forelimb;
  - a. Radial
  - b. Suprascapular
  - c. Axillary
  - d. Median
  - e. Ulnar
  - f. Musculocutaneous

Try to keep it in place so you have all 6 nerves on the forelimb at once.

# Forelimb 4

This is a self-directed study session with Questions A-G. At a computer working in pairs answer the questions below using the 3D models provided on LEARN under the 'Forelimb 4' lecture folder within 'Anatomy Lectures'. You will need to use your textbooks to answer some of the questions.

**Do not click on the annotations or use the annotation menu on the bottom right of the screen as the annotations reveal the answer.** You will need to rotate, zoom and move the model to find the annotations and answer the questions associated with that annotation (see below). Then click on the annotations to reveal the answers.

## Question A

Annotations 1-10: Identify the structures numbered 1-10.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

Annotation 11: find this annotation and explain the action of the structure identified.

- 11.

Annotation 12 and 13: What are these two structures formed from? What is the function of these structures?

- 12.
- 13.

## Question B

Look at the whole model and identify which one nerve is shown on the whole 3D forelimb and explain your answer (Annotations 1-5 reveal answers):

## Question C

Annotations 1-10: Identify the bony prominences 1-10 and state which can be palpated in the live dog.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

## Question D

Annotation 1: Define the term trunk:

Annotation 2: Describe how the forelimb is attached to the trunk:

Annotation 3: List which other extrinsic muscles of the forelimb are missing on this model and figure out where they would be located on the 3D model.

## Question E

Annotations 1-7: Identify the bones of the carpus.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Annotation 8: Name the articulation.

- 8.

Annotation 9: Which two articulations of the canine carpus communicate with one another?

- 9.

## Question F

Look at the whole model and identify which one nerve is shown on the 3D forelimb and explain your answer (annotations 1-7 reveal answers):

## Question G

Annotations 1-10: Identify the structures numbered 1-10.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

Annotation 11: State the excessive movement this annotation prevents.

- 11.

Annotation 12: Name the articulation.

- 12.

Annotation 13: Name the articulation.

- 13.

# A22 Hindlimb 3

There are 3 stations in this tutorial and you have 15 minutes to complete each station by working as a team and answering the questions below using the 3D printed models, any textbooks you have and the 3D computer models on the iPad provided (nominate one member of the group to log on to LEARN on the iPad).

Once the 15 minutes is finished stay seated in your groups and we will rotate the 3D printed models around for the next station.

## Station 1- Hip Joint

1. On the 3D printed model identify;

<b>Acetabulum</b>	<b>Greater ischiatic notch</b>
<b>Greater trochanter</b>	<b>Head of femur</b>
<b>Ischiatic tuberosity</b>	<b>Lesser trochanter</b>
<b>Ligament of the femoral head</b>	<b>Obturator foramen</b>
<b>Transverse acetabular ligament</b>	<b>Tuber coxae</b>
<b>Tuber sacrale</b>	

2. On the 3D printed model identify the position of the sacroiliac joint. What are the two main ligaments associated with this joint?
  
3. A 5 year old male neutered dog has a fracture of the left femur at the mid-diaphysis. The surgeon decides to place an intramedullary pin. An intramedullary pin is a rod of metal which is placed into the medulla of the bone. The surgeon decides to first insert the pin starting from the distal end of the proximal fracture fragment of the femur and then push the pin towards the proximal end of the femur (figure 1). Based on your anatomical knowledge what is the possible risk of starting the pin insertion this way compared to starting at the proximal end of the femur (figure 2)?

(You will learn more detail about this in third year)

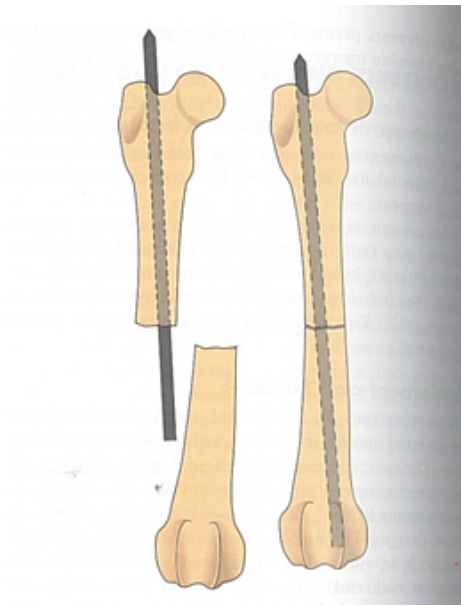


Figure 1. Fossum et al (2007). *Small Animal Surgery. 3rd Edition.*

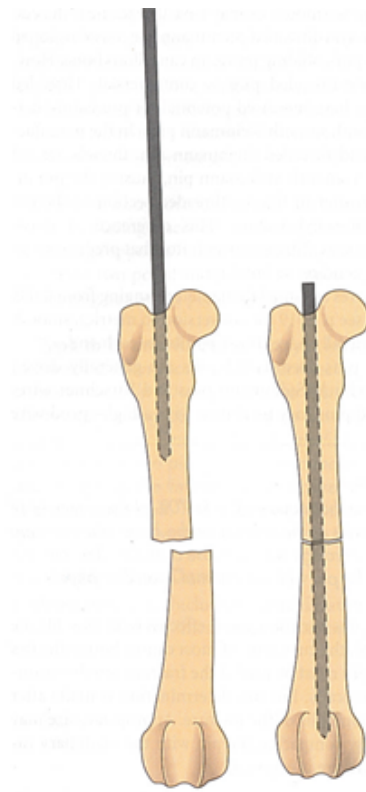


Figure 2. Fossum et al (2007). *Small Animal Surgery. 3rd Edition.*



## Station 2- Stifle Joint

1. On the 3D printed model identify;

<b>Caudal cruciate ligament</b>	<b>Cranial cruciate ligament</b>
<b>Femur</b>	<b>Fibula</b>
<b>Lateral collateral ligament</b>	<b>Lateral epicondyle of femur</b>
<b>Lateral femoropatellar ligament</b>	<b>Lateral tibial condyle</b>
<b>Medial collateral ligament</b>	<b>Medial epicondyle of femur</b>
<b>Medial femoropatellar ligament</b>	<b>Medial tibial condyle</b>
<b>Patella</b>	<b>Patellar ligament</b>
<b>Tibia</b>	

2. Identify on the 3D printed model where the following structures would be located;
  - a. Fabellae
  - b. Tendon of origin of long digital extensor muscle
  - c. Tendon of origin of the popliteus muscle
3. Which two structures are missing on the 3D printed model? Where would they be located?
4. Use the air dough provided to make the two missing structures for the 3D printed model.

## Station 3- Hindlimb Nerves

1. Use the coloured cords and the hindlimb provided to show the course and distribution of each of the 3 main nerves of the hindlimb;
  - a. Femoral
  - b. Obturator
  - c. Sciatic (Peroneal/Fibular and Tibial)

# A23 Hindlimb 4

This is a self-directed study session with Questions A-G. At a computer working in pairs answer the questions below using the 3D models provided on LEARN under the 'Hindlimb 4' lecture folder within 'Anatomy Lectures'. You will need to use your textbooks to answer some of the questions.

**Do not click on the annotations or use the annotation menu on the bottom right of the screen as the annotations reveal the answer.** You will need to rotate, zoom and move the model to find the annotations and answer the questions associated with that annotation (see below). Then click on the annotations to reveal the answers.

## Question A

Annotations 1-10: Identify the structures numbered 1-10.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.
- 11.

Annotation 12: find this annotation and state the classification of the joint shown and state the type of joint within this classification.

- 12.

Annotations 13 & 14: find these annotations and name the two bony structures which would be present here. State which muscle is associated with these two structures and state which two ligaments are associated with these structures.

13.

14.

## Question B

Look at the whole model and identify which one nerve is shown on the whole 3D Hindlimb and explain your answer (Annotations 1-5 reveal answers):

## Question C

Annotations 1-10: Identify the bony prominences 1-10 and state which can be palpated in the live dog.

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

## Question D

Annotations 1-3: Name the bones which make up each hipbone of the pelvis (annotations 1-3 reveal the answer). Try to trace these on the 3D model.

Annotation 4: Name and describe the articulation that occurs at annotation 4.

Annotations 5-7: Identify the structures numbered 5-7.

- 5.
- 6.
- 7.

## Question E

Annotations 1-7: Identify the bones of the tarsus.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Annotation 8: Name the articulation.

- 8.

Annotation 9: Find annotation 9. Which structures make up the common calcanean tendon?

## Question F

Look at the whole model and identify which one nerve is shown on the 3D Hindlimb (Annotation 1 reveals the answer)

Annotations 2-14: Describe the course this nerve takes (annotations 2-14 reveal the answers).

## Question G

Annotations 1-12: Identify the structures numbered 1-10.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.
- 11.
- 12.

Annotation 13: State the function of the structures annotated 4 & 8 (annotation 13 reveals the answer).



## Appendix 6 - Focus Group Interview Questions

1. How did you find the recent in-course assessment test (spatial MCQ)?
2. Did you find any of the question types particularly challenging?
3. Can you think of any particular teaching that helped you answer the questions in the assessment (spatial MCQ)?
4. You recently undertook timed spatial ability tests on two different occasions. Having done the tests once before, did you feel you were better prepared for them second time round? Was it easier the second time or did it make no difference?
5. Do you think the spatial ability tests helped with your learning of anatomy?
6. Do you feel that spatial ability is important for learning anatomy?
7. Do you have any other comments about learning of anatomy?
  - a. E.g what would help you learn it better?





## **Appendix 7 - Spatial Multiple Choice Questions**

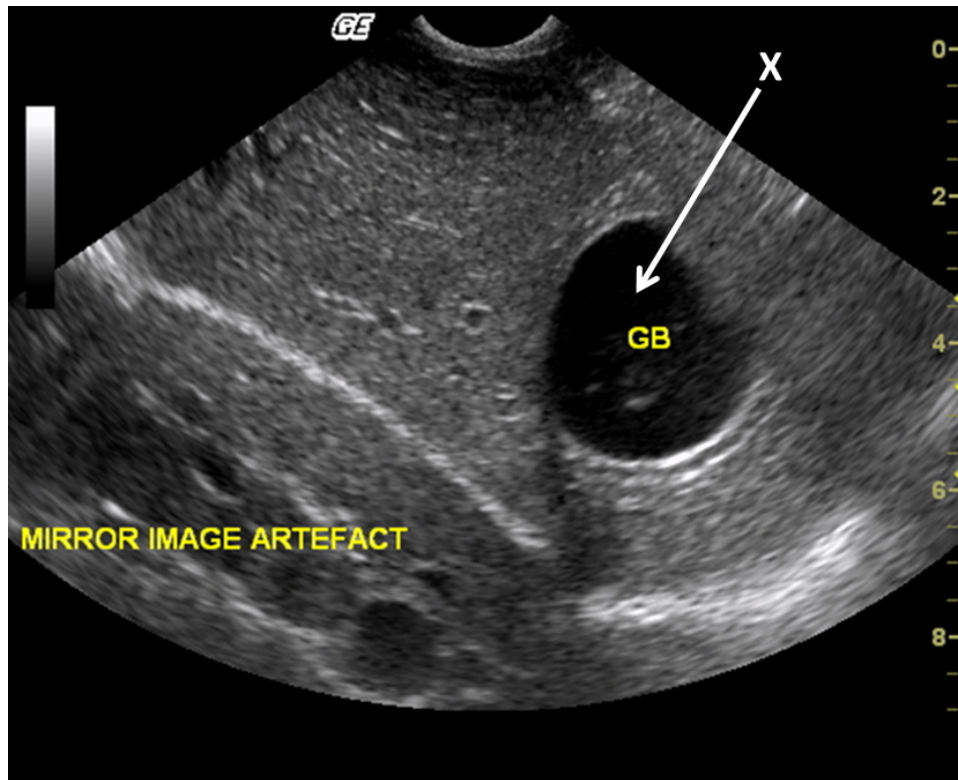
- 1. Which bones make up the pelvis?**
  - a. Ilium, sacrum and ischium**
  - b. Pubis, ilium and ischium**
  - c. Sacrum, ilium and acetabulum**
  - d. Sacrum, pelvis and acetabulum**
  
- 2. What type of joint is the hip joint?**
  - a. Ball and socket**
  - b. Condylar**
  - c. Ellipsoidal**
  - d. Hinge**
  
- 3. What are the main ligaments of the hip joint?**
  - a. Ligament of the head of femur and annular ligament**
  - b. Ligament of the head of the femur and lateral collateral ligaments**
  - c. Ligament of the head of the femur and the transverse acetabular ligament**
  - d. Medial collateral ligament and cranial cruciate ligament**
  
- 4. What type of joint is the sacroiliac joint?**
  - a. Cartilaginous and fibrous**
  - b. Cartilaginous and synovial**
  - c. Fibrocartilaginous**
  - d. Synovial and fibrocartilaginous**

- 5. What type of joint is the stifle joint?**
- a. Condylar**
  - b. Ellipsoidal**
  - c. Hinge**
  - d. Saddle**
- 6. What is the structure of the menisci?**
- a. Circular shaped, fibrocartilaginous wedges**
  - b. Circular shaped, ligamentous wedges**
  - c. Crescent shaped, fibrocartilaginous wedges**
  - d. Crescent shaped, ligamentous wedges**
- 7. What is the function of the menisci?**
- a. Attach bone to bone**
  - b. Attach muscle to bone**
  - c. Change the angle of pull of a muscle**
  - d. Reduce incongruence**
- 8. Name the ligament which prevents forward displacement of the tibia in the stifle joint.**
- a. Caudal cruciate ligament**
  - b. Cranial cruciate ligament**
  - c. Lateral collateral ligament**
  - d. Medial collateral ligament**
  - e. Patellar ligament**

- 9. Name the ligaments which prevent internal rotation of the tibia**
- a. Cranial and caudal cruciate ligaments**
  - b. Medial and lateral collateral ligaments**
  - c. Menisofemoral ligament and medial collateral ligament**
  - d. Oblique ligament and annular ligament**
- 10. What is the function of the patella?**
- a. Allows the stifle joint to flex**
  - b. Alters the angle of pull of the quadriceps femoris muscle**
  - c. Prevents over extension of the stifle joint**
  - d. Protects the stifle joint**
- 11. The fabellae are found in the muscle bellies of which muscle?**
- a. Deep digital flexor**
  - b. Gastrocnemius**
  - c. Quadriceps femoris**
  - d. Semitendinosus**
- 12. Which of the following is a part of the common calcaneal tendon?**
- a. Biceps femoris muscle**
  - b. Pectineus muscle**
  - c. Quadriceps femoris muscle**
  - d. Semimembranosus muscle**

- 13. Which statement correctly describes the perineum?**
- a. Area between the anus and scrotum**
  - b. Area between the anus and tail**
  - c. Area between the scrotum and penis**
  - d. Area between the tail and anal glands**
- 14. Which muscle comprises part of the pelvic diaphragm?**
- a. Gemelli**
  - b. Levator ani**
  - c. Middle gluteal**
  - d. Pectineus**

15. What is the term used for the area marked 'X' on the ultrasound image below?



- a. Anechoic
- b. Hyperechoic
- c. Hypoechoic
- d. Isoechoic

**16. Identify the leg and orientation shown in the photograph below.**



- a. Left leg, caudal view**
- b. Left leg, cranial view**
- c. Right leg, caudal view**
- d. Right leg, cranial view**

**17. Identify the leg and orientation shown in the photograph below.**



- a. Left leg, caudal view**
- b. Left leg, cranial view**
- c. Right leg, caudal view**
- d. Right leg, cranial view**

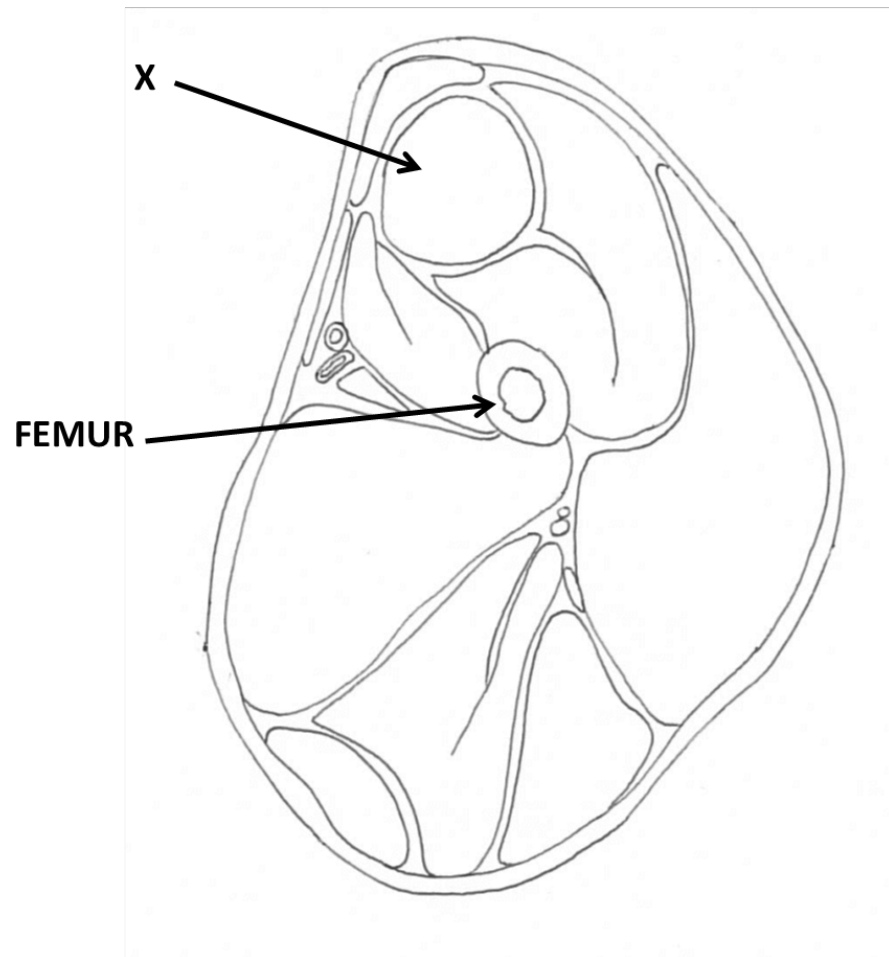


18. Identify the leg and orientation shown in the photograph below.



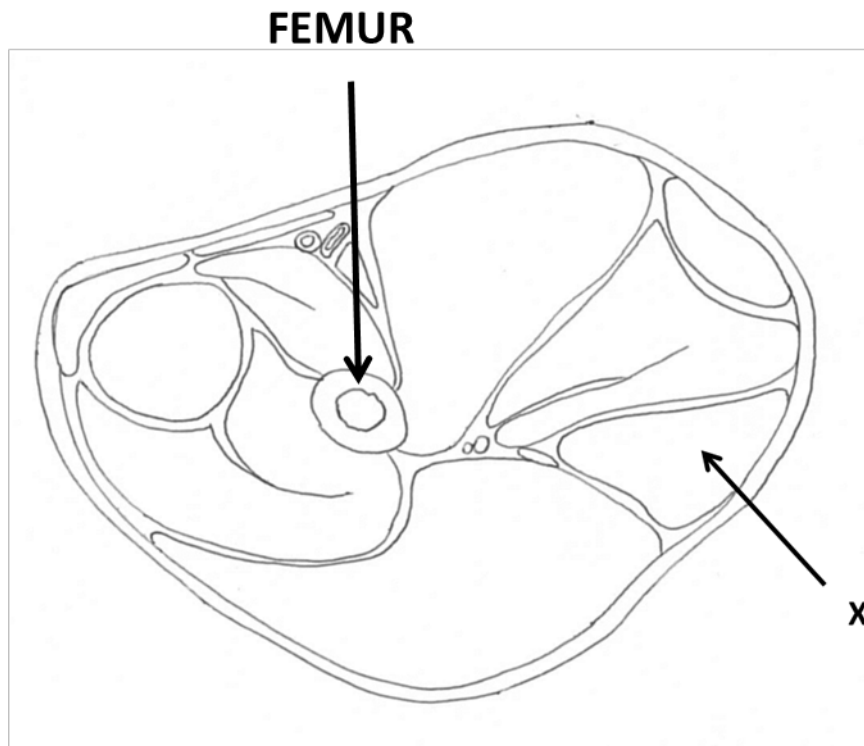
- a. Left leg, caudal view
- b. Left leg, cranial view
- c. Right leg, caudal view
- d. Right leg, cranial view

19. The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.



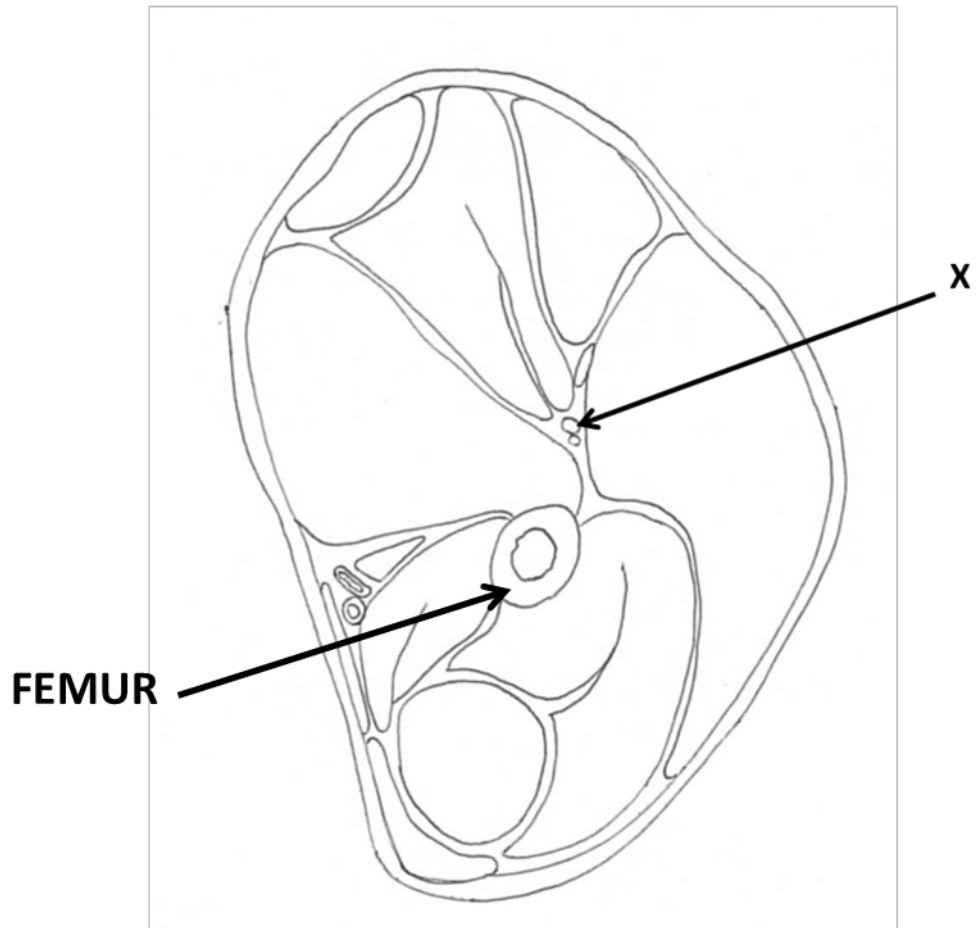
- a. Biceps femoris
- b. Cranialis tibialis
- c. Rectus femoris
- d. Semimembranosus

20. The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.



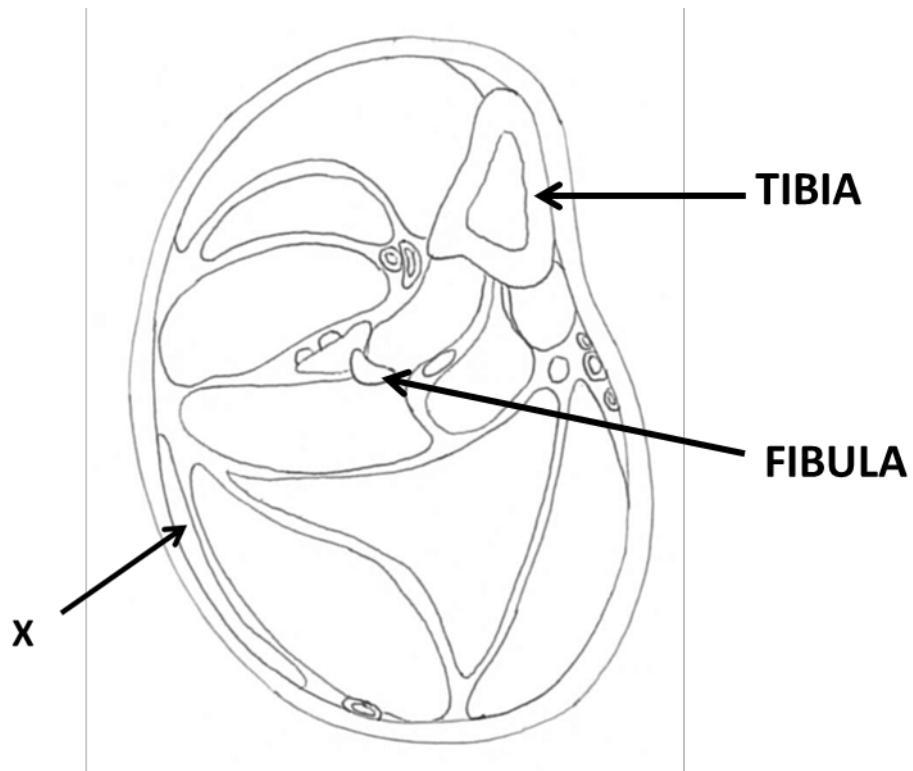
- a. Long digital extensor
- b. Rectus femoris
- c. Semitendinosus
- d. Superficial digital flexor

21. The diagram below shows a cross section through a canine hindlimb. Identify structure 'X'.



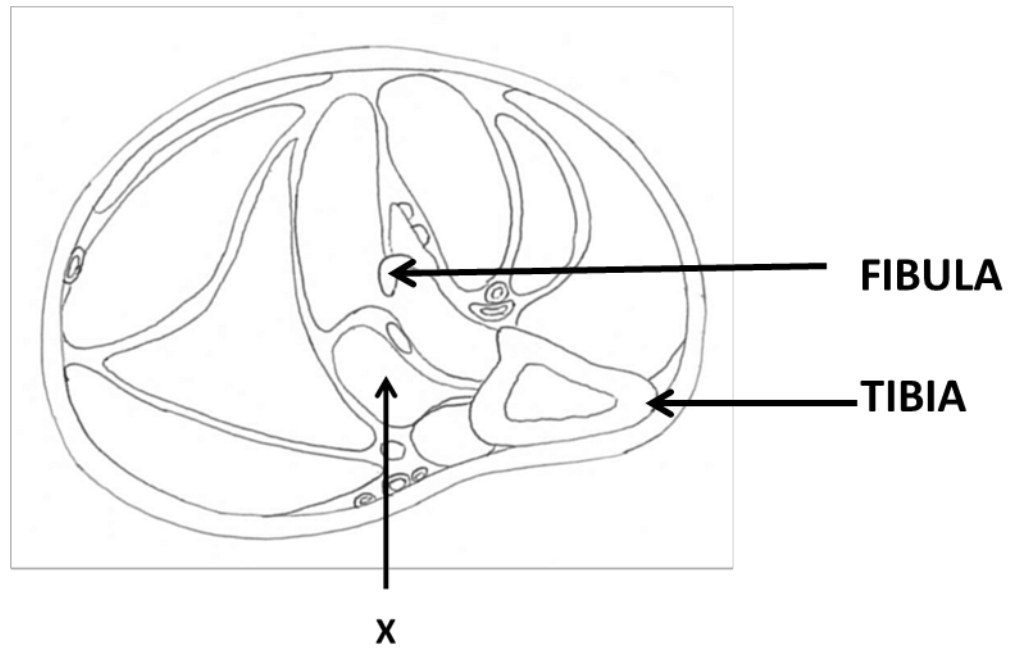
- a. Femoral artery
- b. Femoral nerve
- c. Gluteal artery
- d. Sciatic nerve

22. The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.



- a. Biceps femoris
- b. Cranialis tibialis
- c. Semitendinosus
- d. Tensor fasciae latae

23. The diagram below shows a cross section through a canine hindlimb. Identify muscle 'X'.



- a. Cranialis tibialis
- b. Deep digital flexor, lateral head
- c. Deep digital flexor, medial head
- d. Vastus intermedius

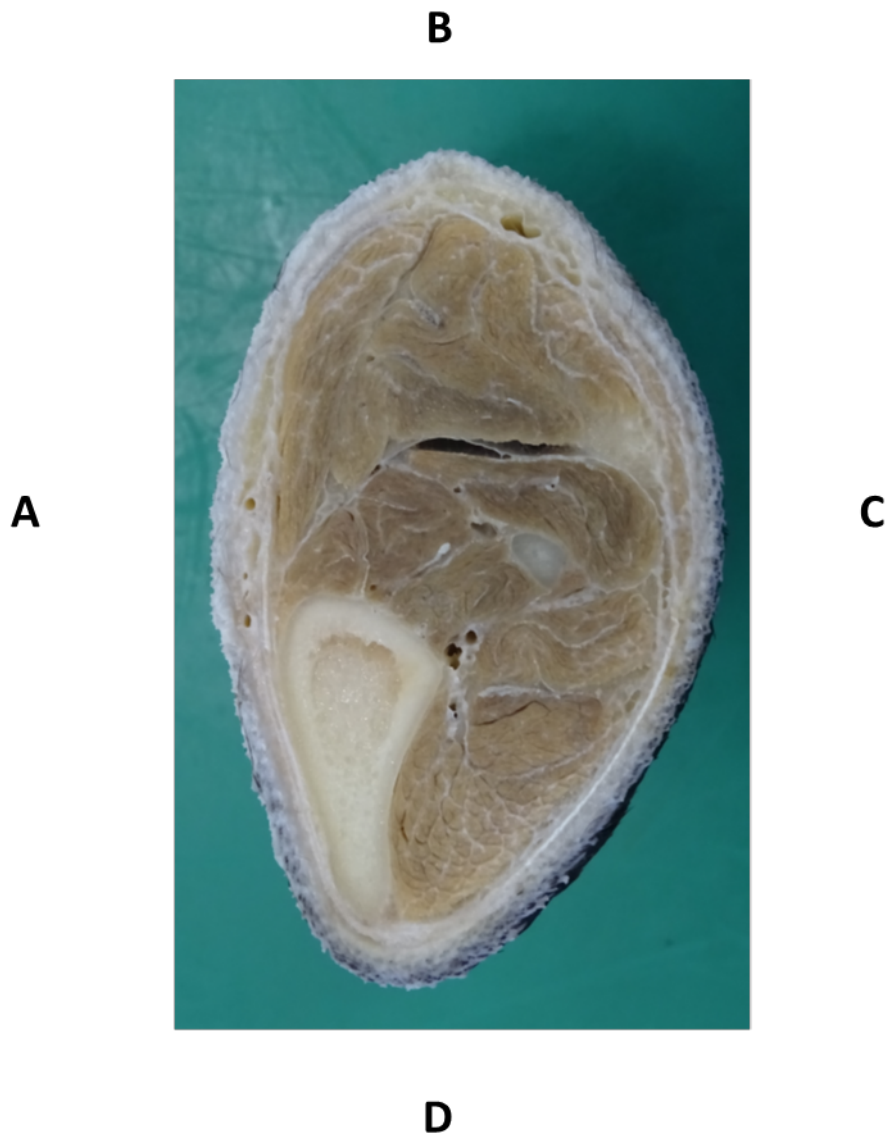
24. The photo below is a transverse section through a canine hindlimb. Which side is medial?

A

B

C

D



25. The photo below is a transverse section through a canine hindlimb. Which side is cranial?

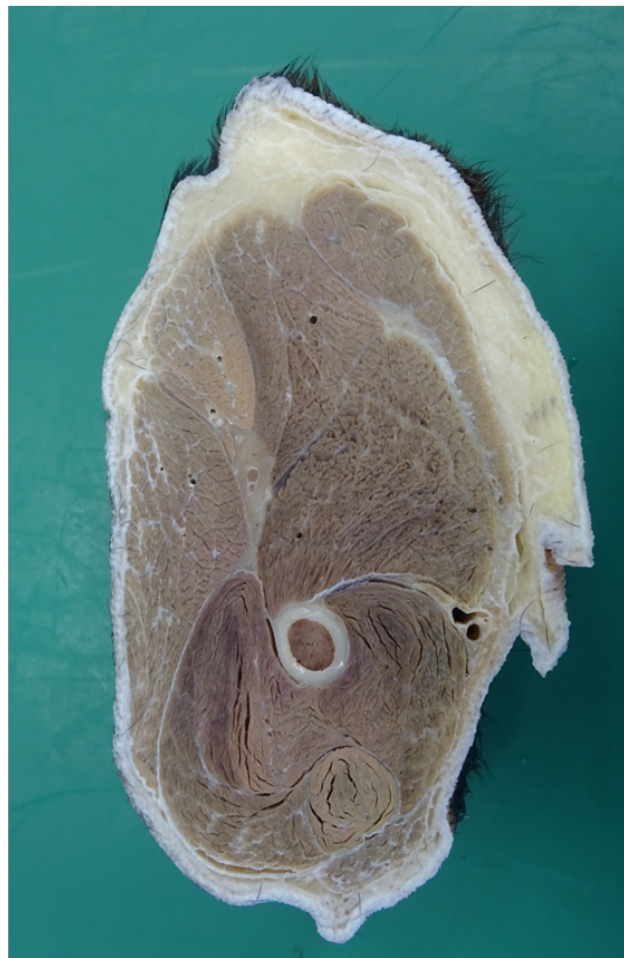
A

B

C

D

B



A

C

D



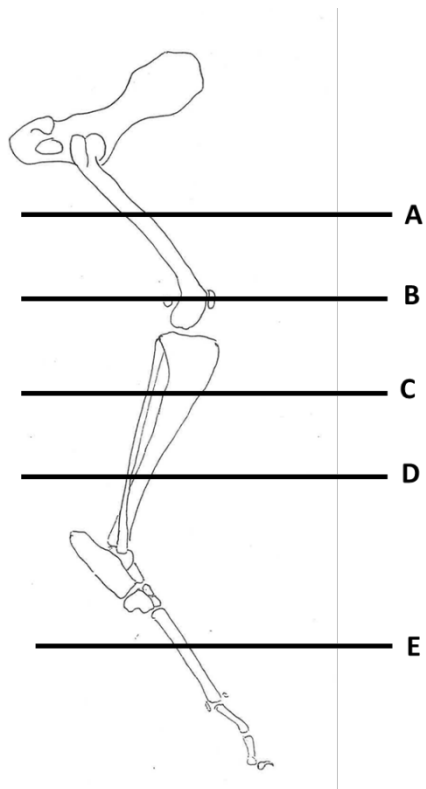
**26. In relation to this X-ray of a normal stifle, which of the following statements could be correct?**

- a. Right leg, cranial view**
- b. Right leg, caudal view**
- c. Left leg, cranial view**
- d. Left leg, caudal view**

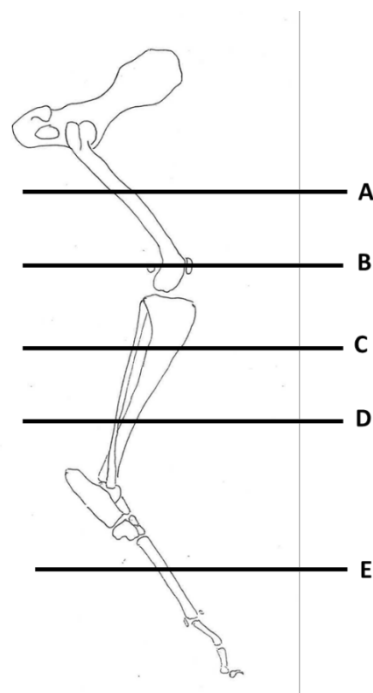
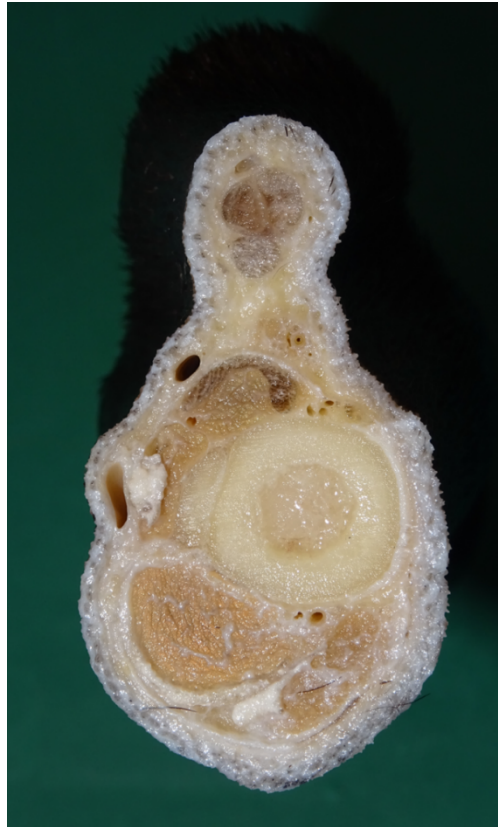


- a. 1 and 2 could be correct**
- b. 2 and 3 could be correct**
- c. 1 and 3 could be correct**
- d. 2 and 4 could be correct**
- e. 1 and 4 could be correct**

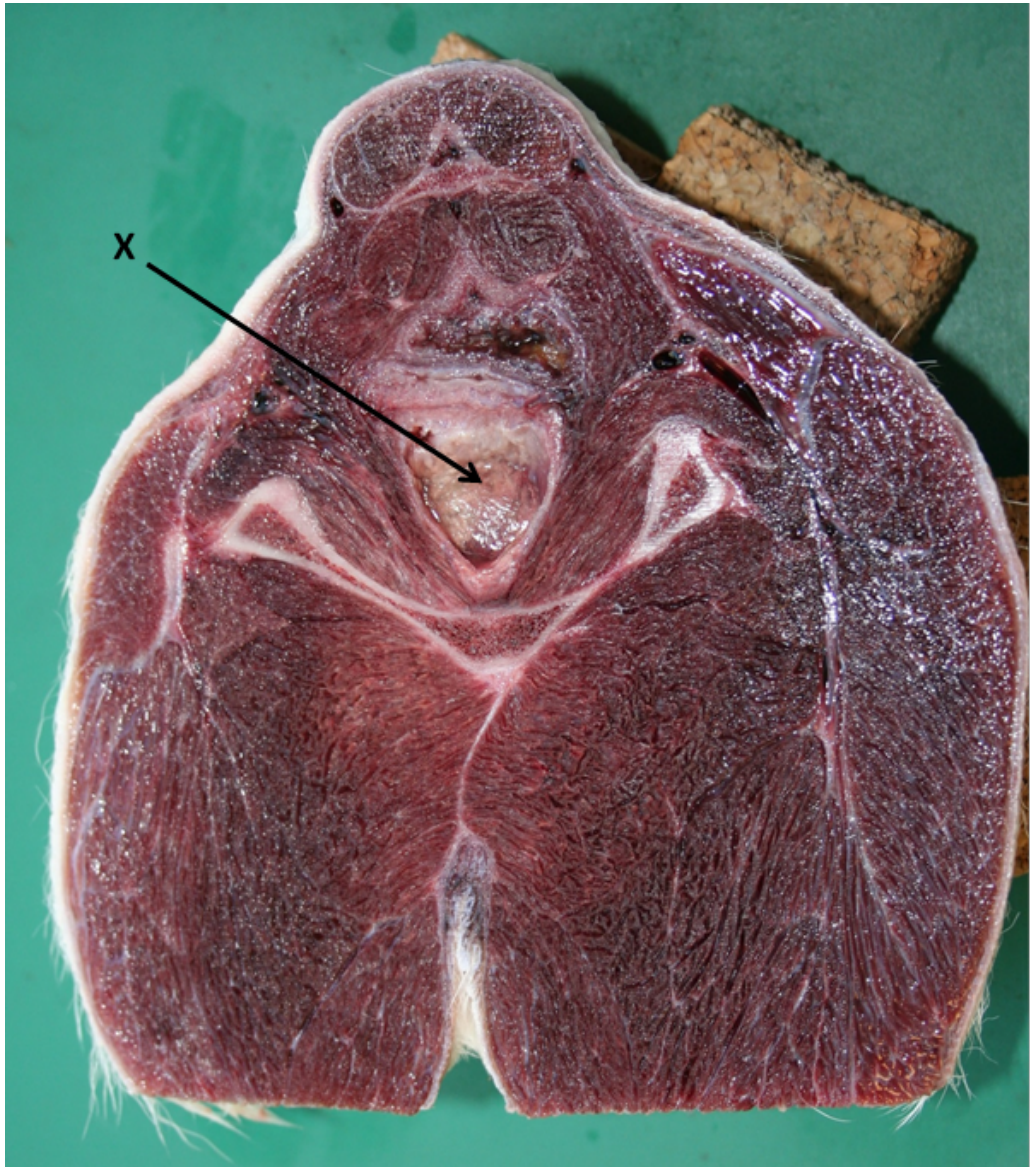
27. At which level of the limb is this cross section taken at?



28. At which level of the limb is this cross section taken at?



29. Identify structure 'X'.



- a. Colon
- b. Gall bladder
- c. Urinary bladder
- d. Uterus

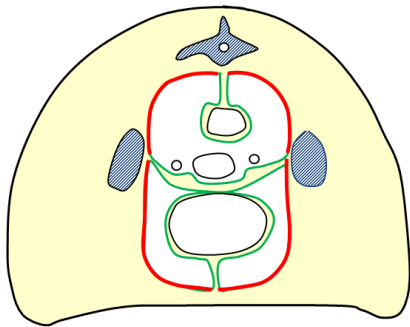
30. Which picture best represents the visceral peritoneum of the FEMALE pelvis?

A

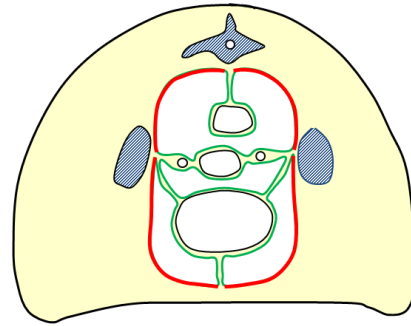
B

C

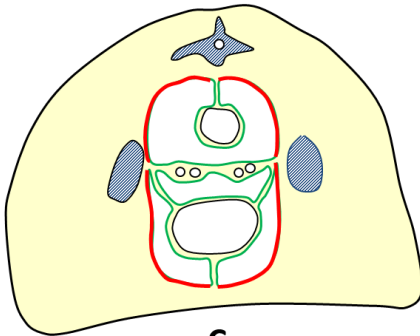
D



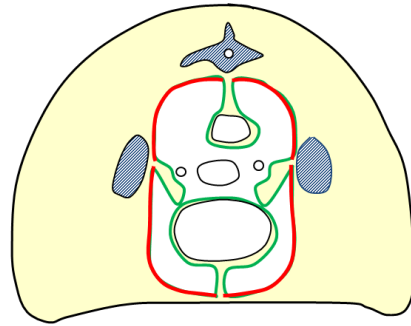
A



B



C



D

# Appendix 8 – Abstracts and Awards

## Abstracts

### **R(D)SVS Postgraduate Research Day 2015**

Is spatial ability teachable? New dimensions in veterinary anatomy education

Dickson, J. , Rhind, S. , Gardiner, A. R(D)SVS, University of Edinburgh

Spatial abilities can be defined as ‘the ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment’ (Wright et al 2009). This ability to mentally manipulate objects is fundamental to many medical disciplines such as surgery, radiography, cardiology and emergency medicine. All of these specialised skills in the medical and veterinary professions require a sound understanding of anatomy. Studies in medical education have shown that student’s spatial ability can be a predictor of success in anatomy examinations and help students learn anatomy however the phenomenon is poorly understood and has been little researched in veterinary students. This research aims to determine the spatial ability of two cohorts of first year veterinary students using a series of well validated spatial ability tests and explore links between this spatial ability and performance in anatomy examinations of different formats. In addition, the study will explore whether anatomy teaching incorporating diagnostic imaging/ 3D images and 3D printing improve students’ spatial ability and anatomy understanding to a greater extent than a more traditional teaching approach.

### **VetEd 2015**

Exploring Spatial Ability in Veterinary Students

Dickson, J. , Rhind, S. , Gardiner, A. R(D)SVS, University of Edinburgh

In recent years, anatomy teaching has been the focus of much research within medical education as curriculum changes have tended to result in a greater emphasis on clinical and professional skills and less on traditional disciplines such as anatomy. Relatively little research has been conducted in this area of veterinary education, yet it could be argued that sound anatomy knowledge is more important compared to medical education due to the multispecies nature of veterinary curricula and the level of surgical ability expected of new graduates.

Many medical studies have explored students' visual-spatial ability and its relationship to anatomy learning. One study found that students with a high spatial ability were 2.2 times more likely to score greater than 90% on practical examinations than students with a lower spatial ability (Lufler et al, 2012). In our study, we measured the visual-spatial ability of 92 first year veterinary students before and after an anatomy course. The students were tested with three different well validated spatial ability tests before receiving any anatomy teaching and were then re-tested with the same three spatial ability tests 16 weeks later, near the end of the anatomy course. Students with any previous anatomy knowledge, e.g. re-sitting students or students with a previous degree, were excluded from the study. The performance of the students on the three spatial ability tests was correlated with their performance on an in-course MCQ and the end of course examinations. The results of the study and links to future work will be presented.

### **R(D)SVS Postgraduate Research Day 2016**

Is spatial ability teachable? New dimensions in veterinary anatomy education

Dickson, J. , Rhind, S. , Gardiner, A. R(D)SVS, University of Edinburgh

Spatial abilities can be defined as 'the ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment' (Wright et al. Diabetes Care 2009; 32 1503-1506). This ability to

mentally manipulate objects is fundamental to many medical disciplines such as surgery, radiography, cardiology and emergency medicine. All of these specialised skills in the medical and veterinary professions require a sound understanding of anatomy. Studies in medical education have shown that student's spatial ability can be a predictor of success in anatomy examinations and help students learn anatomy however the phenomenon is poorly understood and has been little researched in veterinary students. This research aims to determine the spatial ability of two cohorts of first year veterinary students using a series of well validated spatial ability tests and explore links between this spatial ability and performance in anatomy examinations of different formats. In addition, the study will explore whether anatomy teaching incorporating diagnostic imaging/ 3D images and 3D printing improve students' spatial ability and anatomy understanding to a greater extent than a more traditional teaching approach. Results of the demographic details of cohort 2015/16 will be presented and compared to cohort 2014/15.

### **European Association of Veterinary Anatomists 2016**

Is spatial ability teachable? New dimensions in veterinary anatomy education

Julie Dickson, Susan Rhind and Andrew Gardiner, Veterinary Medical Education Division, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, UK

Introduction: Spatial ability can be defined as 'the ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment' (Wright et al. *Diabetes Care* 2009; 32 1503-1506). Studies in medical education have shown that good spatial ability can be a predictor of success in anatomy examinations. However the phenomenon has not been explored in veterinary students. This research aims to determine the spatial ability of two cohorts of first year veterinary students using a series of well validated spatial ability tests and will explore links between this spatial ability and performance in anatomy



examinations of different formats. Results from the first cohort of students will be discussed.

**Materials and Methods:** 114 first year veterinary students' spatial ability was tested prior to any regional anatomy teaching. Three well validated paper and pencil psychometric spatial ability tests were used; Card Rotation Test (Ekstrom et al. Manual for kit of factor-referenced cognitive tests 1976), Surface Development Test (Ekstrom et al. Manual for kit of factor-referenced cognitive tests 1976) and the 3D Mental Rotations Test (Vandenberg et al. Percept Mot Skills 1978; 47 599-604). Demographic details of the students were also obtained.

**Results:** The results of the pre-teaching spatial ability test results show:

1. All three spatial ability tests were positively correlated with one another ( $p < 0.001$ ).
2. Students found the Surface Development Test easier.
3. Male students performed significantly better on the 3D Mental Rotations Test ( $p = 0.008$ ) compared to female students.
4. Students found the 3D Mental Rotations Test harder.

**Conclusion:** The 3D Mental Rotations Test may help to distinguish between stronger and poorer students although it has a bias towards male students. In contrast, the Surface Development Test has a higher baseline and therefore may not be as useful.

All voluntary student participations were approved by the institutional ethics committee.

## **R(D)SVS Postgraduate Research Day 2017**

Is spatial ability teachable? New dimensions in veterinary anatomy education

Dickson, J. , Rhind, S. , Gardiner, A. R(D)SVS, University of Edinburgh

Ritchie, S. School of Psychology, University of Edinburgh

Human spatial abilities can be defined as ‘the ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment’ (Wright et al. Diabetes Care 2009; 32 1503-1506). This ability to mentally manipulate objects is fundamental to many medical disciplines such as anatomy, surgery and radiography. Yet this important cognitive ability is poorly understood in veterinary students. This research aims to address this by analysing the results of three cohorts of first year veterinary students, from two universities, on three well-validated tests of spatial ability (test 1= Card rotation test, test 2= 3D mental rotation test and test 3= surface development test). These tests specifically measure the sub-categories of mental rotation and spatial visualisation. Cohorts 1 and 2 are first year veterinary students from the University of Edinburgh academic years 2014/15 and 2015/16 respectively. Cohort 3 are first year veterinary students from the University of Bristol academic year 2016/17. Demographic details of the students were obtained including age, gender, left/right handed and any previous degree studied. An analysis and comparison of the performance of the 3 cohorts on the tests will be presented.

### **VetEd 2017**

Testing anatomy: dissecting spatial and non-spatial knowledge in MCQ assessment

Dickson, J. , Rhind, S. , Gardiner, A. R(D)SVS, University of Edinburgh

Ritchie, S. School of Psychology, University of Edinburgh

In recent years much research has been conducted in the medical field on the relationship between the cognitive ability of spatial ability and anatomy education. Although many studies have shown that spatial ability is linked to anatomy learning, other studies have shown no link (Lischka & Gittler 1997 and Sweeney 2014). Recently Langlois et al (2016) completed a systematic review of the literature on spatial abilities and the assessment of anatomy knowledge. They concluded that anatomy knowledge could be assessed both spatially and non-spatially, but that ‘the

relationship between spatial ability tests and anatomy knowledge assessment using spatial MCQs was unclear.’ In the study reported here an anatomy MCQ test was designed to test the anatomy knowledge of 1st year veterinary students on the canine hindlimb, pelvis and the theory of ultrasonography both spatially and non-spatially. The MCQ test consisted of 30 MCQs with an equal 50:50 split of non-spatial and spatial anatomy questions. Three cohorts of 1st year vet students completed the MCQ test (cohort 1= academic year 2014-15, cohort 2= academic year 2015-16 & cohort 3= academic year 2016-17). Additionally the spatial ability of students in cohorts 1 and 2 was tested along with collection of anatomy examination results and compared to the MCQ test. Initial preliminary findings suggest the MCQ test could assess anatomy knowledge both spatially and non-spatially, the full results of this study and links to future work will be presented.

## **Young Generation of Veterinary Anatomists 2017**

### DOES SPATIAL ABILITY VARY BETWEEN VET STUDENTS?

Dickson J<sup>1\*</sup>, Rhind S<sup>1</sup>, Gardiner A<sup>1</sup>, Ritchie S<sup>2</sup>, Baillie S<sup>3</sup>, Richens I<sup>3</sup>

<sup>1</sup>Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, UK

<sup>2</sup> School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, UK

<sup>3</sup>School of Veterinary Sciences, The University of Bristol, Bristol, UK

### INTRODUCTION

Human spatial abilities can be defined as ‘the ability to generate, retain, retrieve and transform or manipulate structural images to orientate and interpret the surrounding environment’ (Wright et al. 2009). This ability to mentally manipulate objects is fundamental to many medical disciplines such as anatomy, surgery and radiography, yet this important cognitive ability is poorly understood in veterinary students. Several studies have tested the spatial ability of medical students and compared this

to exam results on different types of anatomy assessment, as a means to evaluate if a relationship exists but there appears to be no research on how stable a trait this is between veterinary students across different universities. This research analyses the results of three cohorts of first year veterinary students, from two universities, on three well-validated tests of spatial ability. The tests used measure the spatial ability sub-categories of mental rotation and spatial visualisation. Participation was voluntary and human ethics approval was granted by the two institutions.

## MATERIALS AND METHODS

Cohorts 1 and 2 were first year veterinary students from the University of Edinburgh academic years 2014/15 (n=108) and 2015/16 (n=98) respectively. Cohort 3 were first year veterinary students from the University of Bristol academic year 2016/17 (n=74). Demographic details of the students were obtained including age, gender, left/right handed and any previous degree studied by means of a questionnaire given prior to testing. Participants completed three timed paper and pencil tests on spatial ability: 1. Card rotation test (CRT), 2. 3D mental rotation test (MRT) and 3. Surface development test (SDT). The tests were given twice in the same order, once at the start (Pre) of academic teaching and then 15/16 weeks later (Post). Test instructions were read out to the students. No negative marking was used. Participants repeating studies, those who had a previous degree relating to anatomy and/or those who had incomplete exam results were excluded from the study.

## RESULTS

Pre spatial ability test results were used for the statistical analysis to assess the baseline spatial ability of newly enrolled undergraduates on a veterinary degree programme. Kolmogorov-Smirnov statistical tests were used to compare whether there was a difference in the distributions between the three cohorts; all possible pairwise combinations for each spatial ability test were analysed with no statistically significant differences found between the cohorts. This analysis was further divided by gender, with no statistically significant differences noted. Although the distribution of the SDT was left skewed, this was found for all three cohorts. To compare whether there were any differences in the student scores obtained for the

three spatial ability tests between the cohorts a Kruskal-Wallis analysis was performed for each spatial ability test. Again, no statistically significant differences were found between the three veterinary student cohorts for each spatial ability test.

## DISCUSSION

The results of the analysis show there are no statistically significant differences in the Pre results of the spatial ability tests between the three cohorts. This suggests that spatial ability, as measured by the 3 instruments used, is a stable trait among undergraduate veterinary students in this study. Although sample sizes of each cohort are not large, this is difficult to control as participation is voluntary and institutions vary in the number of students accepted on to the veterinary degree programme. Further analysis involving the same institutions and other new institutions may help to continue to explore spatial ability among veterinary students. Additionally this research provides evidence of the utility of these well validated tests of spatial ability but within the context of veterinary student cohorts.

## REFERENCES

Wright, R., Frier, B. & Deary, I., 2009. Effects of Acute Insulin-Induced Hypoglycemia on Spatial Abilities in Adults With Type 1 Diabetes. *Diabetes Care*, 32(8), 3–6.

## TESTING ANATOMY: DISSECTING SPATIAL AND NON-SPATIAL KNOWLEDGE IN AN MCQ ASSESSMENT

Dickson J<sup>1\*</sup>, Rhind S<sup>1</sup>, Gardiner A<sup>1</sup>, Ritchie S<sup>2</sup>

<sup>1</sup>Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, UK

<sup>2</sup> School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, UK

## INTRODUCTION

In recent years much research has been conducted in the medical field on the relationship between spatial ability and anatomy learning. Although many studies have shown that spatial ability is linked to anatomy learning, others have shown no link (Lischka & Gittler 1997; Sweeney et al. 2014). Recently Langlois et al (2017) completed a systematic review of the literature on spatial abilities and the assessment of anatomy knowledge. They concluded that anatomy knowledge could be assessed both spatially and non-spatially, but that ‘the relationship between spatial ability tests and anatomy knowledge assessment using spatial MCQs were unclear.’ In the study reported here an anatomy MCQ test was designed to test the anatomy knowledge of first year veterinary students on the canine hindlimb, pelvis and the theory of ultrasonography. The test included both spatial ability (SA) and non-spatial ability (nSA) questions. Two cohorts of first year veterinary undergraduate students completed the MCQ as part of their in-course assessment. Cohort 1 (n=108) comprised students in academic year 2014/15 and cohort 2 (n=98) were in academic year 2015/16. The spatial ability of each cohort was assessed at the start of the academic year. Subsequently, each cohort received a different teaching method: cohort 1 were taught predominantly using 2D materials (T-2D), cohort 2 with 3D materials (T-3D).

## MATERIALS AND METHODS

The MCQ test consisted of 30 MCQs with an equal 50:50 split of nSA and SA anatomy questions. nSA questions consisted of text only questions and the SA questions consisted of questions involving a combination of diagrams and/or photographs of anatomical cross-sections, bones and radiographs. The questions were piloted with 6 academic staff and refined. The assessment was administered as a 60-minute written examination. Demographic details including age, gender, left or right handedness and any previous degree were collected, along with results on three paper and pencil tests of spatial ability: 1. Card rotation test (CRT), 2. 3D mental rotation test (MRT) and 3. Surface development test (SDT). Participants repeating studies, those who had a previous degree relating to anatomy and/or those who had incomplete exam results were excluded from the study.

## RESULTS

Correlations between the nSA questions and each of the three spatial ability tests showed no statistically significant correlation for each cohort (cohort 1: CRT  $p=0.8768$ , MRT  $p=0.2494$ , SDT  $p=0.6467$ ; cohort 2: CRT  $p=0.9445$ , MRT  $p=0.9280$ , SDT  $p=0.0971$ ). Correlations were conducted for the SA questions and the spatial ability tests and in contrast, only the MRT for cohort 2015 was not statistically significantly correlated to the SA questions (cohort 1: CRT, MRT, SDT  $p<0.01$ ; cohort 2: CRT  $p<0.05$ , MRT  $p=0.1842$ , SDT  $p<0.05$ ). In addition, all three spatial ability tests significantly correlated with one another ( $p<0.001$ ). The nSA questions and the SA questions for both cohorts were statistically significantly correlated to one another (cohort 1  $p<0.05$ ; cohort 2  $p<0.001$ ). To determine whether the correlation between each spatial ability test and the SA questions, and the spatial ability tests and the nSA questions, are significantly different from one another a dependent correlation comparison was conducted for each cohort. The nSA and SA questions were found to be statistically significantly different to one another for each spatial ability test (cohort 1: CRT  $p<0.01$ , MRT  $p<0.01$ , SDT  $p<0.001$ ; cohort 2: CRT  $p=0.01$  and SDT  $p<0.001$ ) apart from the MRT for cohort 2015 ( $p<0.16$ ).

## DISCUSSION

Initial findings from this MCQ study show that it is possible to design questions which appear to test anatomy spatially, as evidenced by correlations between spatial ability and test performance. This in turn may help to demonstrate whether students have gained a deeper learning on the topic of anatomy, as spatially demanding questions theoretically require acquired knowledge to answer plus problem solving spatial abilities. The SDT distribution is skewed to the left and this may not reflect an accurate comparison as a ceiling effect is being exhibited with all students performing well on this test making comparisons difficult. Further analysis needs to be conducted on this MCQ with other veterinary student cohorts and potentially other tests of spatial ability to further analyse and confirm the nSA and SA assessment of this MCQ.

## REFERENCES

Langlois, J. et al., 2017. Spatial abilities and anatomy knowledge assessment: A systematic review. *Anatomical sciences education*, 10(3), 235–241.

Lischka, M. & Gittler, G., 1997. Spatial abilities and learning modes in anatomy beginners. In A. J. J. A. Scherpbier et al., eds. *Advances in Medical Education*. 166–169.

Sweeney, R., Hayes, J. & Chiavaroli, N., 2014. Does Spatial Ability Help the Learning of Anatomy in a Biomedical Science Course? *Anatomical sciences education*, 7(4), 289–94.

## **Awards and Nominations**

2018 – Nominated ‘Best overall Teacher’ by Edinburgh University Students' Association

2017 – Nominated ‘Best Course Award’ by Edinburgh University Students' Association

2017 – Nominated ‘Best overall Teacher’ by Edinburgh University Students' Association

2016 – Awarded ‘Best overall Teacher’ by Edinburgh University Students' Association

2015 – Nominated ‘Best overall Teacher’ by Edinburgh University Students' Association

2015 – Awarded ‘PG Research Day 1st year poster prize winner’ by R(D)SVS, University of Edinburgh

2014 – Nominated ‘Best overall Teacher’ by Edinburgh University Students' Association