

THE IMPACT OF  
SPATIAL AWARENESS TRAINING ON  
SPATIAL ABILITY, ANATOMY LEARNING,  
MANUAL DEXTERITY AND  
SURGICAL SKILLS

Dr RENE A. GONZALES

MB Bch BAO, Ophth,

BA, MSc, HDipCT (VTCT)

A thesis submitted in partial fulfilment  
of the requirements of  
the University of Brighton  
for the degree of Doctor of Philosophy

September 2019



## ABSTRACT

Spatial ability (SA) is the capacity to understand and manipulate mental concepts of objects, remembering relationships among their parts and their surroundings. This ability is useful for orientation and navigation in anatomy and dissection, and may influence anatomy learning.

First and second-year undergraduate medical students ( $N = 194$ , 56% female) scored well in SA (70, 25 [Median, IQR], men scored more than women: 70, 20 vs 60, 30;  $p < 0.01$ ). Three months later, 104 students retook the SA test and scored higher than on the first (75, 20; Difference -5.7,  $p < 0.01$ ); 29 of them (intervention group) received two sessions of Spatial Awareness Training (SATr), their post-training scores were not significantly different than those of the control group ( $n = 75$ ) (Difference 0.24,  $p 0.95$ ). Women in the intervention group scored more in the second test ( $p 0.03$ ), than women in the control group ( $p 0.1$ ), with no post-intervention difference between them ( $p 0.3$ ). Men in both groups improved on the second test (Int.  $p < 0.001$ ; Ctrl.  $p 0.02$ ) with no post-test difference ( $p 0.1$ ). Participants with scores in the bottom quartile in SA did poorly in the final anatomy score compared to the rest ( $p < 0.001$ ), no correlation between SA scores and anatomy was apparent. In addition, the intervention group underwent testing in Manual Dexterity (MDx), the ability to use hands and fingers for specific tasks; this group improved after SATr (93.5 vs 103.3,  $p 0.02$ ). There was no correlation between MDx and SA. Post-hoc power was 0.51. Improvement in SA and MDx scores may result from a learning effect and/or learning anatomy.

Whilst SA and MDx may improve with time in the medical programme, the introduction of a compact programme of Spatial Awareness Training did not show significant benefits in spatial ability or anatomy. Future studies may aim at increasing sample size, power and duration of the intervention in a fully experimental design.



# CONTENTS

ABSTRACT .....	3
CONTENTS .....	5
LIST OF TABLES .....	11
LIST OF FIGURES .....	13
ABBREVIATIONS .....	15
ACKNOWLEDGEMENTS .....	17
AUTHOR'S DECLARATION .....	19
<b>CHAPTER 1 INTRODUCTION .....</b>	<b>21</b>
1.1 Overview .....	21
1.2 The Researcher's Perspective .....	22
1.3 Introduction .....	23
1.4 Anatomical Context .....	24
1.5 Educational Context .....	26
1.5.1 Undergraduate Education .....	26
1.5.2 Medical Education in the UK .....	26
<b>CHAPTER 2 LITERATURE REVIEW .....</b>	<b>31</b>
2.1 Introduction .....	31
2.2 Background to the Literature Review .....	31
2.2.1 Literature Review Question .....	32
2.2.2 Keyword Definition and Databases .....	33
2.2.3 Inclusion and Exclusion Criteria .....	34
2.3 Main Topics of the Literature Review .....	35
2.3.1 Spatial ability .....	35
2.3.1.1 Emergence of the Concept of Spatial Thinking .....	36
2.3.1.2 Factors in Spatial Ability .....	37
2.3.1.3 Categories or Types of Spatial Ability .....	38
2.3.1.4 Testing Spatial Ability .....	39
2.3.1.5 Gender in Spatial Ability .....	41
2.3.1.6 Medicine, Anatomy and Spatial Ability .....	44
2.3.1.7 Cerebral Function and Spatial Ability .....	47
2.3.1.8 Spatial Awareness Training .....	48
2.3.2 Manual Dexterity .....	49
2.3.3 Stereopsis .....	52

2.4	Review of Related Topics .....	55
2.4.1	Cognitive Load Theory.....	55
2.4.2	Medical Education in the UK.....	57
2.4.3	Gender Balance in Medicine.....	58
2.4.4	The Medical Student.....	58
2.4.4.1	Socioeconomic Composition of Medical Students at BSMS .....	58
2.4.4.2	Medical Student's Learning Strategies .....	59
<b>CHAPTER 3</b>	<b>METHODOLOGY AND METHODS.....</b>	<b>61</b>
3.1	Introduction.....	61
3.2	Methodology.....	61
3.2.1	Definition of Variables .....	61
3.2.2	Definition of the Research Questions.....	62
3.2.2.1	Research Question .....	63
3.2.2.2	Sub Questions .....	63
3.2.3	Definition of Hypotheses and Aims .....	64
3.2.3.1	Hypotheses of the Study.....	64
3.2.3.2	Aims of the Study.....	65
3.2.4	Research Methodology.....	66
3.2.4.1	Research Approach and Study Design .....	66
3.2.4.2	Statistical Power and Sample Size.....	69
3.2.4.3	Project Timeline .....	72
3.2.5	Ethics Committee Approval.....	72
3.2.6	Participants in the Study .....	74
3.2.7	Spatial Ability Testing .....	75
3.2.7.1	Question Challenges in the Spatial Ability Test.....	76
3.2.8	Spatial Awareness Training .....	79
3.2.9	Manual Dexterity Testing .....	81
3.2.10	Visual Performance Testing.....	82
3.2.10.1	Visual Acuity .....	83
3.2.10.2	Stereopsis.....	83
3.2.11	Academic Performance Testing.....	84
3.2.12	Data Considerations .....	85

3.2.13	Marketing and Advertising .....	85
3.3	Methods .....	86
3.3.1	Introduction.....	86
3.3.2	Statistical Analysis Plan.....	86
3.3.3	Research Technique .....	88
3.3.3.1	Power and Sample Size Calculation .....	88
3.3.3.2	Recruitment Technique and Sample Selection.....	90
3.3.3.3	Participant Eligibility .....	91
3.3.3.4	Random Allocation to the Intervention.....	92
3.3.3.5	Participant Anonymity .....	93
3.3.4	Data Management and Statistical Method .....	93
3.3.4.1	Data Collection Technique .....	94
3.3.4.2	Data Handling.....	94
3.3.4.3	Data Preparation and Clearing.....	95
3.3.4.4	Data Coding.....	96
3.3.5	Statistical Analysis Technique .....	98
3.3.5.1	Processing Data in SPSS .....	98
3.3.6	Reliability and Validity.....	100
3.3.6.1	Reliability and Validity of the Spatial Ability Test .....	102
3.3.6.2	Reliability of the Manual Dexterity Test .....	103
3.3.7	Analysis of Academic Data .....	103
3.3.8	Testing Techniques .....	105
3.3.8.1	Spatial Ability Testing Technique .....	105
3.3.8.2	Visual Function Testing Techniques .....	106
3.3.8.3	Manual Dexterity Testing Technique.....	107
3.3.9	Spatial Awareness Training Method .....	108
3.3.10	Logistics .....	109
<b>CHAPTER 4</b>	<b>RESULTS .....</b>	<b>113</b>
4.1	Introduction .....	113
4.2	Analysis of Participation .....	114
4.2.1	Participation and Group Structure.....	114
4.2.2	Participants' Age.....	118

4.3	Analysis of the First Spatial Ability Test for Group E (Full Cohort) .....	119
4.3.1	Reliability and Validity of the Spatial Ability Test .....	119
4.3.2	Descriptive Statistics of the First Spatial Ability Test .....	120
4.3.3	Comparison of Scores in the First Spatial Ability Test by Gender for Group E (Full Cohort).....	121
4.3.4	Comparison of Scores in the First Spatial Ability Test by Year of Study at BSMS	123
4.3.5	Questions Attempted in the First Spatial Ability Test.....	123
4.3.6	Measure of Difficulty of the Spatial Ability Test .....	124
4.3.6.1	Comparative Analysis of Spatial Ability Scores Based on Question Difficulty for Group E (Full Cohort).....	127
4.3.7	Comparative Analysis of the First Spatial Ability scores for Group E (Full Cohort) by participant's year of study at BSMS .....	129
4.4	Analysis of the Spatial Ability Tests for Group D (Experiment).....	130
4.4.1	Descriptive Statistics and Test Comparison for Group D (Experiment) ..	130
4.4.2	Analysis and Comparison of Spatial Ability Scores for the Intervention and Control Groups.....	133
4.4.3	Post-hoc Statistical Power Calculation.....	136
4.4.4	Comparisons of Spatial Ability Scores by Gender for Group D (Experiment)	138
4.4.5	Analysis of Spatial Ability Scores for Group D (Experiment) by Year of Study at BSMS.....	146
4.5	Analysis of Academic Performance .....	147
4.5.1	Anatomy Data.....	148
4.5.2	Analysis of Anatomy Data.....	149
4.5.2.1	Anatomy Modules and Final Score for Groups D (Experiment) and E (Full Cohort) .....	149
4.5.2.2	Scores in Anatomy Modules x03 and x04 for Groups A (Intervention) and B (Control).....	151
4.5.2.3	Comparison of Scores from Anatomy Modules x03 and x04 by Gender and by Group.....	153
4.5.2.4	Comparison of Scores in Module x04 for Group D (Experiment) by Year of Study .....	154
4.5.2.5	Comparison of Final Anatomy Scores by Spatial Ability Score Quartiles	156
4.6	Correlations between Spatial Ability and Academic Performance .....	156
4.7	Additional Tests: Manual Dexterity and Stereopsis.....	158

4.8	Limitations .....	160
4.8.1	Research Modality and Sample Size Limitation .....	160
4.8.2	Academic Score Limitations.....	161
4.8.3	Time Limitations of Medical Students .....	162
4.8.4	Time Limitations of Training.....	162
4.8.5	Time Limitations of the Spatial Ability Test.....	163
<b>CHAPTER 5</b>	<b>DISCUSSION .....</b>	<b>165</b>
5.1	Summary of results .....	165
5.2	Student Participation .....	166
5.3	Student Gender Distribution .....	168
5.4	Student Experience with Testing .....	169
5.5	Discussion on Reliability and validity findings for the Spatial Ability Test ...	170
5.6	Spatial Ability Test Related Factors.....	171
5.6.1.1	Test Duration .....	171
5.6.1.2	Test Difficulty .....	172
5.7	Spatial Ability Scores of All Participants.....	173
5.8	Spatial Ability Scores of Group D (Experiment) .....	174
5.9	Gender Characteristics in Spatial Ability.....	175
5.10	Training and Learning.....	177
5.11	Academic Performance .....	178
<b>CHAPTER 6</b>	<b>CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>181</b>
6.1	Conclusions.....	181
6.2	Recommendations .....	183
	REFERENCE LIST .....	185
	APPENDICES.....	210
	Appendix 1 – Participant Information Sheet (sample).....	210
	Appendix 2 – Participant Consent Form.....	211
	Appendix 3 – Spatial Awareness Training Workbook .....	212
	Appendix 4 – Data Collection Sheet for the Study.....	213
	Appendix 5 – Table of Coding Rules .....	214
	Appendix 6 –Spatial Ability Questions’ Index of Difficulty .....	215



## LIST OF TABLES

Table 1. Undergraduate Curriculum Map of the Degree in Medicine at BSMS .....	28
Table 2. Sample of Initial Questions on the Main Topics of the Study .....	32
Table 3. Sample of Main Keywords Considered for the Literature Review Search .....	33
Table 5. Coding of Rotation Sequence and Shape Dissimilarity, compared to the reference shape in the Spatial Ability Test.....	78
Table 6. Anatomical Relevance of Spatial Awareness Training Topics. ....	80
Table 7. Sample Size ( <i>n</i> ) as a Function of Significance, Effect Size and Power.....	89
Table 8. Timeline of Project Events and Related Academic Tests for Year 1 and 2 in Medicine.....	104
Table 9. Content of Spatial Awareness Training Sessions .....	109
Table 10. Participation Groups of the Project.....	116
Table 11. Distribution of Participants by Group, Year in Medicine and Gender .....	117
Table 12. Descriptive Statistic and Comparison of Results of the First Spatial Ability Test for Group E (Full Cohort) by Gender.....	122
Table 13. Proposed New Order of Spatial Ability Question Sequence Based on Index of Difficulty. ....	126
Table 14. Comparison of Raw and Weighted Difficulty-Based Scores from the First Spatial Ability Test for Group E (Full Cohort) by Gender. ....	128
Table 15. Descriptive Statistics and Comparison Using the Dependent Samples T-Test of Scores from the First and Second Spatial Ability Tests for Group D (Experiment). ....	131
Table 16. Scores from the First and Second Spatial Ability Tests of Group A (Intervention) and group B (Control).....	133
Table 17. Distribution of Scores of the First and Second Spatial Ability Test for Female Participants of Group A (Intervention) and B (Control).....	140
Table 18. Descriptive Statistics and Comparisons of Scores from the First and Second Spatial Ability Tests for Female Participants of the Intervention and Control Groups. ....	141
Table 19. Distribution of Scores of the First and Second Spatial Ability Test in Male Participants of the Intervention and the Control Groups. ....	143
Table 20. Descriptive Statistics and Comparisons of Scores from the First and Second Spatial Ability Tests for Male Participants of the Intervention and Control Groups. ....	144
Table 21. Spatial Ability Scores of Group D (Experiment) by Sub-group and by Year at BSMS.....	146

Table 22. Descriptive Statistics and comparisons of Anatomy Scores from Modules x03, x04 and the Final Anatomy score for Group D (Experiment).....	150
Table 23. Spatial Ability Questions Ordered by Index of Difficulty .....	215

## LIST OF FIGURES

Figure 1. Cortical Activation during Mental Rotation in fMRI Scan .....	47
Figure 2. Purdue Pegboard Test for Manual Dexterity .....	51
Figure 3. Stereopsis or Stereo Vision.....	53
Figure 4. Axes of Objects in Orthogonal Position .....	77
Figure 5. Visual Acuity and Stereopsis Tests .....	107
Figure 6. Testing Kit and Samples of Testing Room Layout for Personal Testing Sessions. ....	111
Figure 7. Distribution of Difficulty of Question in the Spatial Ability Test Based on a Sample of 159 Students by Gender.....	125
Figure 8. Histograms of the First and Second Spatial Ability Tests for Group D (Experiment). ....	132
Figure 9. Clustered Boxplot of Spatial Ability Scores Pre and Post-Intervention for the Intervention and Control Groups. ....	134
Figure 10. Grouped Scatterplot of Correlation between Scores in the Second Spatial Ability Test and Anatomy Module x04 for Group D (Experiment) by Gender. .....	157



## **ABBREVIATIONS**

<b>BSMS</b>	Brighton and Sussex Medical School
<b>CAPSULE</b>	Clinical and Professional Studies Unique Learning Environment
<b>GMC</b>	General Medical Council
<b>GP</b>	General Practitioner
<b>MRT</b>	Mental Rotation Test
<b>NCGD</b>	Non-Career Grade Doctor
<b>NHS</b>	National Health Service
<b>OSCE</b>	Objective Structured Clinical Examination
<b>PMQ</b>	Primary Medical Qualification.
<b>PSVT: R</b>	Purdue Spatial Visualization Test: Rotations (by Guay)
<b>ROT</b>	Purdue Visualisation of Rotations (by Bodner & Guay)
<b>STEMM</b>	Science, Technology, Engineering, Mathematics and Medicine



## **ACKNOWLEDGEMENTS**

I would like to thank and acknowledge the following people's advice and support in the preparation and realisation of this project and the writing process of this thesis:

Prof Claire Smith, Head of Anatomy, and Prof Gordon Ferns, Professor of Medical Education and Metabolic Medicine Head at Brighton and Sussex Medical School (BSMS), for their outstanding supervision, inspiring guidance and continuous support. I deeply appreciate their openness to innovation and their infinite patience through more than 50 meetings and innumerable revisions of documents.

The following lecturers at BSMS: Catherine Hennessy, Dr Ricardo Governo, Prabal (Charley) Chatterjee, Dr Andrew Dilley, for kindly donating part of their valuable lecture time to help with the study.

Prof Sheryl Sorby, Professor Emerita of Mechanical Engineering and Director of the Engineering Education and Innovation Research Group at Michigan Technological University, USA, for kindly providing details and advice on adapting the "Spatial awareness training programme" to this project in 2015.

Prof George M. Bodner. Professor of Chemistry Purdue University, Illinois, USA, who kindly provided the Purdue Visualization of Rotations (ROT) test for this project.

Maria Young Birch, Course Leader for the BSc Podiatry Course at the University of Brighton at Eastbourne, Dr Sergei Korneev, Senior Lecturer in Neuroscience at the University of Sussex, and Prof Michael Peters from the University of Guelph, Ontario, Canada. All of whom kindly facilitated testing for the pilot phase of the study in 2015 (report excluded from the final version of the thesis).

Students at BSMS: Year 1 and Year 2 medical students (2014-2016), anatomy demonstrators (2014-2017), and surgical trainees (2017), who gave their invaluable time for testing and training (report of surgical section excluded from the thesis).

Staff from the Anatomy Laboratory at BSMS: Tracy Young and Lucinda Evans, who kindly provided support for anatomical dissections, sample harvesting and surgical testing at the laboratory throughout the study. Staff from the school office at BSMS: Liz Kaye and Denise Nicholson, who kindly provided statistical data and logistic support for the study.

Staff and colleagues at BSMS' Medical Research Building (MRB), who gave me support and advice, and provided a welcoming and refreshing work atmosphere.

Friends and family, in particular Gertrude Cotter, for their patience and constant support throughout my studies.

## **AUTHOR'S DECLARATION**

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Signed

Dated



# CHAPTER 1 INTRODUCTION

## 1.1 Overview

The aim of this thesis is to understand the relevance of spatial ability and manual dexterity in the study of clinical anatomy. Following this goal, this project examined the impact of spatial awareness training on medical students' spatial ability, academic performance in anatomy and manual dexterity. The cohort of participants consisted of medical students in the first two years of the undergraduate Bachelor of Medicine Bachelor of Surgery (BM BS) course. The surgical section of the study, which included the participation of postgraduate surgical trainees, was excluded from the final version of the thesis to maintain focus on undergraduate results.

This thesis consists of six chapters:

The introduction provides general information about the areas relevant to this study, including the importance of anatomy in medicine, the role that medical education plays in preparing medical students for their future careers and the influence of those factors in the creation of a workforce that will ultimately provide a safe and effective health service to the public.

The literature review explores the current knowledge in medical education, spatial ability, manual dexterity and other areas that gained importance as the study developed.

The chapter on methodology looks at the research questions and proposed hypotheses, selection of research design and tests, training strategies, data considerations and recruitment. The methods section details the procedures that

facilitated collection of data, recruitment and testing, delivery of spatial awareness training and statistical procedures used to analyse data.

The results chapter reports statistical analysis findings with “section results” or summaries for each section, including participation, spatial ability tests for different groups, academic performance and additional tests (manual dexterity and stereopsis). A report on limitations found in each section follows.

The discussion chapter provides an explanation of findings in sequence, contextualising results with the existing knowledge; there is an emphasis on the impact of the training intervention on spatial ability and anatomy scores.

The conclusions chapter looks at answers to the questions of the study and considers the importance of the study in the context of medical education.

Finally, the recommendations chapter provides suggestions for implementation of future studies in this area.

## **1.2 The Researcher’s Perspective**

My interest in exploring spatial ability and manual dexterity derives from the realisation of the importance that structural understanding of human anatomy has in the practice of medicine and surgery. Coming from a medical background, I gained experience as an ophthalmologist, which involves medical and surgical care of the eye and visual function. I taught clinical optics and anatomy of the eye and orbit to first-year post-graduate trainees in my field, which showed me the challenges involved in learning, understanding and teaching complex anatomical structures like the eye. Operating under a microscope brings into play not only knowing how anatomical structures relate in space, to plan and perform surgery, but also precise manual dexterity to manipulate fragile tissues, respecting structural integrity and transparency to retain or restore

visual function. Those characteristics mirror the main areas of this project, which may offer a potential application in medical education, the practice of medicine, diagnosis and treatment.

In addition, I qualified in computing and digital design to a master level (MSc), which helps me in the delivery of content in teaching environments and allows me to envision the potential technological applications of developments in spatial ability and manual dexterity in education, communication and medical technology related to surgery.

### **1.3 Introduction**

Training doctors to be safe and effective clinicians has always been the aim of medical educational institutions as they are at the forefront in the selection and preparation of the medical workforce of the future, which ultimately will reflect in the quality of service provided to the community that they serve (GMC, 2009; BSMS, 2017a). A better understanding of the capabilities of students as they start their medical career may lead to improvements in comprehending the process of learning, and enhancing the quality of teaching by adjusting the curriculum accordingly, thus approaching the goal of maximised learning.

Optimising the acquisition of knowledge of essential areas of the medical curriculum may enhance performance of students and prevent the development of cognition gaps or misconceptions that may lead to medical error in the future (Waterson and Stewart, 2005). The study of new methods to improve teaching and learning, e.g. by promoting spatial thinking, are worthy of attention, particularly when there is evidence of potential benefits in multiple facets of academia in other scientific fields like engineering (Sorby, 2009).

In addition to understanding the characteristics of students entering medicine, this study required to investigate areas associated to medical training in anatomy, under

the working title: “The Impact of Spatial Awareness Training in Spatial Ability, Anatomy Learning and Manual Dexterity in Undergraduate Students starting Medical Training”.

The following sections provide an overview of those areas, starting with anatomical relevance and continuing with educational context in medicine.

## **1.4 Anatomical Context**

Knowledge of human anatomy is one of the essential elements of the medical profession; this field is a cornerstone of healthcare education and is central to clinical practice (Turney, 2007; Tunstall and Shah, 2012; Netter, 2014). Anatomical knowledge provides an initial structure that helps medical students understand the mechanisms of health and disease, and to interpret signs and symptoms. Later in their careers, this knowledge helps doctors taking decisions that may profoundly affect the life of a patient.

An adequate understanding of the structural aspects of the human body, learned often in the early years of medical school, and perceived as critically important to clinical medicine (Moxham and Paisant, 2007), may play a role in the prevention of medical error (Ford and Cooper, 2018). Some unfavourable outcomes during the professional life of doctors, may translate into legal cases, prompting the need of legal protection, which is now compulsory for practicing in the medical field (MPS, 2017). Enhancing spatial visualisation as a supplement to anatomy learning may provide an additional tool to help solidify anatomical knowledge and provide confidence in diagnosis and decision-making, potentially reducing erroneous actions and fear of legal consequences.

In the investigation of models of learning, reports showed that nearly half of the medical students tested used memorisation as their preferred method to learn anatomy, the rest used a combination between memorisation, understanding and visualisation (Pandey

and Zimitat, 2007). In this context, visualisation refers to recalling inter-relations of structures in the human body that students need to identify and associate to a new and complex vocabulary, forming the basis of anatomy learning, which entails a demanding intellectual effort (Pandey and Zimitat, 2007). The capacity of thinking visually, once considered at odds with verbal thinking, is now a component of the thought process (Otis, 2016). Visualisation plays an important role in understanding complex structures like the human body, which has an intricate and functional relationship between multiple organs that work in unison (Hegarty *et al.*, 2007).

Visualisation results from using one's Spatial Ability, which is the capacity to form mental models of objects, remember their relations with other objects, and being able to manipulate them abstractly (JHU, 2014). This ability is used in many scientific areas that deal with physical structures, including physics, chemistry, mathematics, engineering and medicine (Hegarty *et al.*, 2010). A number of clinical activities rely on the formation of mental models, from the examination of patients (surface anatomy) to the analysis and interpretation of medical images (CT, MRI, X-Ray, Ultrasound) and to activities with direct intervention on the human body, like surgery or diagnostic procedures (Hegarty *et al.*, 2007).

In the process of assessing a patient's state of health, physicians, use their knowledge of anatomy to keep a safe and effective clinical practice. Junior doctors use their knowledge as a visual representation (spatial representations) of the anatomy relevant to a patient's complaint, from the beginning of the physical examination (Vorstenbosch *et al.*, 2016). Knowledge of anatomy is essential for physical examination, starting from observation of posture, gait and general movement and followed by surface anatomy, which is key to understand underlying structures, interpret findings and find correlations in medical images (Tunstall and Shah, 2012; Smith *et al.*, 2018)

This study focused on the spatial aspects of learning related to visualisation of structures, which are applicable to learning anatomy. Like most abilities, visual thinking

has a personal variability that needs accounting, and support in this area maybe more beneficial for those who show an initial lower capacity for this type of thinking.

## **1.5 Educational Context**

This study took place at Brighton and Sussex Medical School (BSMS) in the United Kingdom; therefore, it is important to situate it in the wider context of undergraduate education and medical education in this country.

### **1.5.1 Undergraduate Education**

In medicine, undergraduate students face the challenge of learning vast amounts of complex material, throughout their career. One of the theories attempting to explain the learning mechanisms focuses on the acquisition of knowledge in segments. Proposed by Sweller, the “Cognitive Load Theory” indicates that learning depends on the acquisition of schemas and the transition from a deliberate cognitive effort to an automatic delivery of responses to memory and problem challenges (Sweller, 1994). These schemata require time and experience to allow the learner to “know” the answer to a given situation, rather than working through the problem systematically (Taber, 2014) (see section 2.4.1 Cognitive Load Theory).

### **1.5.2 Medical Education in the UK**

Medical education changed over time adapting to the needs of society and of the medical profession; this dynamic field closely follows scientific progress to provide up to date information to medical students and doctors so that they can deliver a better service to the public (Poynter, 1966). The General Medical Council (GMC) is the

institution that sets educational standards for doctors in the UK and for the delivery of teaching at training institutions. The GMC's goals are to protect patients and improve medical education (GMC, 2016a).

The GMC recommends medical schools develop appropriate methods of training and of quality assurance necessary to adhere to its standards and its view of "tomorrow's doctors" (GMC, 2009), which was updated in its standards of teaching, learning and assessment in 2016 with the publication of "Promoting Excellence: Standards for Medical Education and Training" (GMC, 2016b). Following the recommendations of the GMC in the curriculum of the medical school and promoting the highest possible standards in education are part of BSMS aspirations (BSMS, 2017a). The publication "Tomorrow's Doctors" in 2009 had a mixed reception in academic circles. Some felt that medical schools should retain the university element of education, teaching of a broader knowledge of medicine and assessment of understanding, rather than turning their undergraduate training programme into a tick-box model aimed at fulfilling a list of competencies directed by the GMC (Rees and Stephenson, 2010).

In compliance with the GMC directive, BSMS provides students with a three-phase programme that covers five years of study, in which instruction and early contact with patients form part of the teaching programme from the beginning. In phase one, students develop clinical skills through studies of family settings, and learn about the human body through theoretical and practical sessions of anatomy following a systemic approach. Assessment combines written tests (approximately 70%), coursework and practical exams (BSMS, 2017c). During the second phase, which consists of two years, students receive a structured rotation of clinical skills, scientific basis of medicine, pharmacology, medicine and surgery. In the third phase (last year), students complete placements in a series of specialties and prepare for a final assessment (BSMS, 2017c), (Table 1).

Phase	Year	Activity			
Phase 1	Year 1	Clinical Practice			
		Foundation of health and diseases	Heart, lungs and blood	Nutrition, metabolism and excretion	
	Year 2	Clinical Practice			
		Neuroscience and behaviour	Reproduction and endocrinology	Musculoskeletal and immune system	
Phase 2	Year 3	Clinical foundation	Scientific basis of medicine		
			Pharmacology and therapeutics		
			Medicine	Surgery	Elderly medicine and psychiatry
			Student selected component		Exams Y3 OSCE
		<i>Intercalation</i>			
	Year 4	Specialist rotations (8 rotations)		Exams Y4 OSCE	
		General practice, global and public health			
Individual research project					
Phase 3	Year 5	Regional attachments (3 attachments)	Exams Y5 OSCE	Elective and prep for practice	
		CAPSULE			
		Seminar programme			

**Table 1. Undergraduate Curriculum Map of the Degree in Medicine at BSMS**

Note: Simplified curriculum map of the five-year medical degree at BSMS. (BSMS, 2017c).

CAPSULE = Clinical and Professional Studies Unique Learning Environment  
OSCE = Objective Structured Clinical Examination

Anatomy is the area of the undergraduate medical curriculum that relates more closely to spatial ability; because of its structural content, learning anatomy may elicit the student's capacity of visualisation or visual thinking (Hegarty *et al.*, 2007). This study concentrated on the first two years of the curriculum at BSMS, where anatomy is an integral part of the system-based teaching model (BSMS, 2017c).

### **Influence of STEM in Higher Education and Medicine**

Students who obtain high scores in spatial ability tend to have an affinity with science, technology, engineering and mathematics (STEM); these academic areas evoke the

need of spatial reasoning and visualisation, which they have in common with medicine, now considered part of the acronym as “STEMM” (Hegarty *et al.*, 2010; Lubinski, 2010; Kimmel, Miller and Eccles, 2012). Project Talent, a longitudinal study that followed students up to 11 years after graduation demonstrated that highly talented students in STEM areas also scored highly in spatial ability tests, concluding that this ability might be a potentially important element in selecting students for advanced learning opportunities in STEM (Lubinski, 2010).

One of the most representative areas of STEM is engineering, where spatial ability scores were strong predictors of success in engineering graphics (among 11 other variables) and there was a clear gender difference, with men consistently obtaining higher scores (Gimmestad, 1990; Hill, Corbett and Rose, 2010; Gorska, Sheryl and Leopold, 2016). Researchers developed a training programme to assist students with poor results in spatial ability, improving their performance in graphics engineering and other subjects of their course (Sorby and Baartmans, 2000; Sorby, 2009). The training aimed at stimulating spatial thinking and therefore build upon the natural spatial ability of the students, improving their confidence and mental agility while studying complex structural objects (Sorby, 2007).

Like other STEM areas, there is evidence of medical students obtaining high scores in spatial ability, when compared with the general population (Lufler *et al.*, 2011b).

Various researchers demonstrated improvements in students’ spatial ability as an effect of learning anatomy (Hoyek *et al.*, 2009; Fernandez, Dror and Smith, 2011; Vorstenbosch *et al.*, 2013). These findings played an important part in the design and analysis of the current study.

Learning human anatomy is pivotal in medicine, although memory is essential to remember the multitude of details that medical students need to learn; understanding the relationships of those structures in space is also important, and results from a

combination of a natural ability of spatial thinking and exposure to spatial reasoning challenges (Sorby and Wysocki, 2012).

Based on the results in engineering, training in spatial ability might offer the potential to improve student performance, and although the subject of anatomy is different to graphic engineering, there is a common ground in spatial reasoning, which helps with understanding relations of body parts in space. A more solid understanding of anatomy may result from dissection, which is essential in medical education because it uses touch-mediated perception, manual dexterity, understanding of spatial relations and memory in challenging tasks (Granger, 2004).

This study will explore the effect that spatial awareness training might have on a second measurement of spatial ability and manual dexterity and search for associations with performance in anatomy. It will compare medical students' scores in spatial ability and manual dexterity with reference values from the general population, and account for known factors and the effect that learning anatomy has on spatial ability.

## **CHAPTER 2 LITERATURE REVIEW**

### **2.1 Introduction**

The literature review helped situate this project within the scientific landscape, providing a knowledge base upon which to build and contribute. Initially, this study used an organised method (inspired by formal literature reviews) to explore and acquire knowledge in the field; later, the search was more specific, but not tabulated.

The initial search topics had two categories:

**Main Topics:** Most relevant to the study were Spatial Ability, Manual Dexterity and Stereopsis. Each area covering definition, development, structure, gender variability, forms of testing and other features.

**Related Topics:** Broadly related, included Learning Theory, Medical Education, Gender Balance in Medicine and characteristics of the Medical Student. These areas containing an account of how they developed, where they stand today, and trends expected for the future.

### **2.2 Background to the Literature Review**

The search of the literature in this study started with the definition of a literature review question, selection of keywords and databases, and construction of a set of premises to select the most relevant findings (Aveyard, 2014; University of Reading, 2017). The literature selected often served as the starting point of further literature exploration; the references in the journal articles provided an extension to the review, not only as a

means of verification of the citations but also as sources of further information (Creswell, 2014).

### 2.2.1 Literature Review Question

The early stages of this project consisted of several brainstorming sessions with the purpose of looking for questions related to the study that would provide a framework to build the final “literature review research question.” Table 2 shows a sample of questions that emerged from early sessions on the main topics.

Spatial Ability	Stereopsis	Manual Dexterity
Is spatial ability related to intelligence?	What are the visual pre-requisites to test for stereopsis?	Is there a formal marking scheme for manual dexterity?
Is spatial ability a visual form of memory or is it an ability of its own?	What is stereopsis used for?	What occupations require a manual dexterity aptitude test?
Is there a biological/brain location related to spatial ability	What occupations require stereopsis?	Is manual dexterity formally required for the practice of medicine and surgery?

**Table 2. Sample of Initial Questions on the Main Topics of the Study**

During the exploration of possible questions, rather than adding terms, which increased complexity, the preferred option was to look for a more generic and inclusive phrase. In this manner, each topic would get more coverage and the results would follow filtering later using a set of inclusion and exclusion criteria. The result comprised a primary and a secondary literature review questions:

**Primary Question:**

*“Is there a relationship between Spatial Ability, Academic Performance in Anatomy, Manual Dexterity and Stereopsis in Medical Students?”*

**Secondary Question:**

*“Does spatial awareness training produce an effect in Spatial Ability, Anatomy and Manual Dexterity scores?”*

**2.2.2 Keyword Definition and Databases**

The combination of keywords forming a search phrase entered into a search engine (Google) can make a dramatic difference in results, for example, searching for “What is the impact of spatial ability on manual dexterity and surgical skills” rendered 146,000 results; adding “training” to spatial ability and “medical students” reduced the count to 12,600. Consequently, the identification of keywords became important to the study (Table 3).

---

<b>Keywords Considered for Search</b>	
• Spatial ability	• Critical looking
• Depth perception	• Proficiency-based training
• Gender differences	• 3D perception
• Human body	• Stereoscopy
• Learning curve	• Manual dexterity
• Medical student	

---

**Table 3. Sample of Main Keywords Considered for the Literature Review Search**

The use of boolean operators to narrow the results, provided an additional benefit to the combination of the main keywords above to create search phrases (Barker, 2016).

The most helpful databases included Google Scholar and Google, PubMed, Medline, ProQuest, Education Resources Information Centre (ERIC), National Centre for Biotechnology Information (NCBI) and publishers databases accessed through Academic libraries at universities.

### **2.2.3 Inclusion and Exclusion Criteria**

Although a numeric goal was not set at the start, finding 50 reports to begin with, then making a sub-selection using a set of narrowing criteria was the recommendation for early stages of a literature review (Creswell, 2014). Having a set of eligibility criteria was useful at the beginning of the study; later, inclusion became more flexible to avoid missing relevant studies.

#### **Inclusion:**

- Primary research and meta-analysis on one of the main topics
- Published literature in print or online from a credible source, e.g. a university or a health institution website would be more credible than a personal blog.
- Studies that explore the core topics on medical students.
- Publications post 1945 (post WWII): Research on spatial ability started in the late 1960s, but manual dexterity and visual perception started earlier.

#### **Exclusion:**

- Publication on a non-credible website, open to public editing or with unverifiable references.
- Publications earlier than 1945.
- Language other than English.

Due to overlap of sciences, e.g. spatial ability research emerged in many scientific fields before its application to medicine, filtering required flexibility based in common sense.

## **2.3 Main Topics of the Literature Review**

The initial search of publications, found using the technique described above, resulted in 83 articles: 42 on spatial ability, 10 on manual dexterity and 16 on stereopsis. Since the beginning of the project in 2014, there was a substantial expansion of the publications referenced in those and other related topics.

### **2.3.1 Spatial ability**

Researchers proposed a variety of definitions of spatial ability over the years; this discrepancy in opinions suggests that the subjective nature of this concept may result in different impressions based on the point of view of the descriptor. Lohman and other researchers coincide in the conclusion that spatial ability is a conceptual skill used in activities that require the mental construction of a three-dimensional artefact, generated, retained and manipulated as an abstract visual image (Macfarlane Smith, 1964; McGee, 1979a; Lohman, 1979; Clausen-May and Smith, 1998). It has the minimum requirement of an ability to encode, remember, transform, and match spatial stimuli (Lohman, 1979; Carroll, 1993). Clausen-May added that spatial ability is a collective term referring to a range of acquired skills that use memory for shape and position (Clausen-May and Smith, 1998). John Hopkins University agrees, indicating that spatial ability consists of interrelated sub-skills developed through life that help understand and remember relations among objects (JHU, 2014). In summary, most definitions accept that following observation or study of an object, the use of spatial ability involves understanding, memorisation and mental reproduction of the same

object, with the added capacity of manipulating this conceptual image as a form of mental analysis.

### **2.3.1.1 Emergence of the Concept of Spatial Thinking**

Throughout the early 1900s, the concept of spatial thinking remained embedded in intelligence testing, mostly used to categorise students and aptitudes. The earliest mention of spatial orientation appeared in an article by Alfred Binet in 1894, which relates cases of geographical disorientation but does not explore this phenomenon further (Binet, 1894). The "Stanford–Binet Intelligence Scale", standardised in 1916, included a subset of paper folding questions (Siegler, 1992); in 1938, Stern, used this test to compare the age of a child scored by the test with the biological age, to produce the "Intelligence Quotient" or IQ score (Indiana University, 2013; Duke University, 2017), which gained worldwide popularity.

In a new post-war technological era, as technical skills became more important than linguistic skills, researchers tested abilities as a method of recruitment for technical jobs and the army (Macfarlane Smith, 1964).

One of the first tests aimed specifically at measuring spatial thinking was the Minnesota Spatial Relations Test, created in late 1920s by Paterson and standardised in 1976 (Macfarlane Smith, 1964; Mohan, 2016); more tests followed as the interest in spatial ability increased. In 1935, El-Koussy found that although the majority of the tests measured general intelligence, known as factor g (Thorndike, 1921; Guilford, 1968), some of them evoked visual imagery in participants trying to solve the challenges. This property was "factor k" (after "kurtosis"), the ability to obtain and utilise visual-spatial imagery (Macfarlane Smith, 1964; McGee, 1979a). Almost simultaneously and oblivious to El-Koussy and Clarke's studies, Macfarlane Smith worked on a similar

topic, his results, published in 1937, provided further support to factor k and to sex difference because males did better than females (Macfarlane Smith, 1964).

From the 1940s to the 1960s, researchers experimented with more tests for spatial ability following the trending goal of predicting talent and selecting personnel for work or training (Mohler, 2008). Starting in the 1950s, the proposal of intelligence factors grew. Thurstone proposed nine intelligence factors, among which there were three spatial, two of memory and closure, kinaesthetic and induction (Macfarlane Smith, 1964).

### **2.3.1.2 Factors in Spatial Ability**

The acknowledgement of spatial ability as independent from general intelligence is attributable to Galton, who in 1880, collected responses to questions about seeing numbers and objects in one mind's eye (Galton, 1880). More researchers explored this area afterwards, resulting in the emergence of multiple factors involved in spatial ability and hundreds of tests; a notable example is Thurstone, who in 1951 described factors S1 (recognition of angle or rotation), S2 (recognition of change in position) and S3 (use one's body as reference to recognise object's position) (Macfarlane Smith, 1964; Mohler, 2008). In 1976, Ekstrom et al produced a revised kit that included 72 tests for 23 factors designed as guide for researchers who wished to focus on a specific spatial factor. (Ekstrom, French and Harman, 1976).

In 1957, three major types of factors emerged: (Michael *et al.*, 1957).

- Visualisation (Vz): This factor refers to the mental manipulation of observed objects in any direction and sequence of motion, which maybe simple or complex, like rotation, tilt and inversion. It involves recognition, retention and recall.

- Spatial-Relations and Orientation (SR-O): This factor represents the ability to comprehend the arrangement of elements of the observed object in reference to the point of view of the observer, remaining unconfused by changes in configuration.
- Kinesthetic Imagery (K): This factor refers to the ability to discriminate left and right in relation to the body of the observer, who uses the right or left hand as reference (Michael *et al.*, 1957; Macfarlane Smith, 1964; McGee, 1979b).

In 1979, Lohman added three further factors,

- Closure Speed: Is the speed in identifying a visual pattern without knowing in advance, what it is. In the test, the pattern is disguised or obscured.
- Flexibility of Closure: Represents the speed in identifying a visual pattern knowing in advance, what it is. The pattern is disguised or obscured.
- Perceptual speed: Is the speed in finding a known visual pattern in a test where patterns are not disguised or obscured (Lohman, 1979).

The factors described above have more practical importance in psychology and cognitive science research and illustrate the complexity of spatial ability as it covers many overlapping areas (Yilmaz, 2009). The factors of most relevance to this study were Spatial Orientation and Spatial Visualisation. Kinesthetic Imagery (discrimination of right and left) had some relevance but was not part of this project (Bodner and Guay, 1997).

### **2.3.1.3 Categories or Types of Spatial Ability**

The identification of spatial ability factors gave rise to the following categorisation of spatial ability based on a psychometric perspective (Linn and Petersen, 1985):

- **Spatial Perception:** Determines the orientation of objects from the observer's perspective, using kinaesthetic cues and the gravitational vertical as references. Representative test: Rod and Frame test (RFT).
- **Mental Rotation:** Involves internal cognitive processes to mentally rotate two and three-dimensional objects. Larger angles of rotation take longer to analyse. Representative test: Mental Rotation Test (MRT).
- **Spatial Visualisation:** Relates to complicated multi-step mental manipulation of objects, requiring analytical processing and flexible adaptation to find solutions. Representative test: Paper Folding test and Embedded Figure Test (EFT) (Linn and Petersen, 1985).

Spatial Visualisation and Mental Rotation appear to be more relevant to anatomy, where the study of rigid and soft tissue involves analysis of relative positions of organs and other structures. Visualising how those elements change when moved or rotated in relation to neighbouring landmarks (e.g. joint motility), and how their shape vary as part of their function (e.g. shape changes of cardiac cavities during heart-beat) may provide a more complete understanding of anatomy.

#### **2.3.1.4 Testing Spatial Ability**

The Mental Rotation Test (MRT), designed by Shepard and Metzler in 1971, was one of the first popular tests for spatial ability. The test consisted of pairs of diagrams of three-dimensional objects. Participants had to identify if two objects were representations of the sample object but depicted in different orientations or if they were diagrams of different objects (Shepard and Metzler, 1971). One of the features of the MRT is its relative simplicity; the test focuses on rotations around a single vertical axis using diagrams of objects with angular shapes, but as most of the facets are visible, comparisons maybe predictable by logic, which is likely to limit the use of

mental visualisation (Linn and Petersen, 1985). Despite its popularity, the MRT was not the test of choice for this study because of its limitation to a single axis of rotation; anatomy students commonly deal with rotations in multiple angles.

To test the spatial visualisation and mental rotation categories, Guay devised the Purdue Spatial Visualization Test: Rotations (PSVT: R) at Purdue University (USA) in 1976. This was a 20-minute, 30-item test that consisted of diagrams of composite geometrical figures that tested the participant's ability to visualise and rotate, incorporating sequential rotation and non-angular shapes (Guay, 1976). The PSVT: R test became widely accepted in the research community as a valid and reliable test for spatial ability (Ernst *et al.*, 2016). A meta-analysis conducted by Maeda & Yoon, reported that the time limit seemed to play a moderating role in the detection of gender difference, as the researchers found that minor gender differences appeared in tests with more lenient timing conditions (Maeda and Yoon, 2012). This finding agreed with Peters' report of time constraints affecting women more than men in a test similar to MRT (Peters, 2005).

Further research on time limitations of the PSVT: R led to the analysis of mental processing strategies by Bodner & Guay, according to which, solutions to visualisation problems were achievable by Analytical processing or Gestalt processing. The former breaks down the query into individual parts for analysis, while the latter results from a holistic concept of the object (Bodner and Guay, 1997). Gestalt processing, associated with spatial ability, refers to the perception of objects as complete units, which are more than the sum of their parts. For example, recognising faces relies in remembering a group of relationships between features rather than a static image (Macfarlane Smith, 1964; Bodner and Guay, 1997; Braungardt, 2018). Gestalt thinking, described as similar to "intuition," is quick, compared to analytical thinking, which takes longer to process (Braungardt, 2018). Bodner concluded that reducing the time of the PSVT: R test would maximise gestalt and minimise analytic processing, and made a short

version called the Purdue Visualisation of Rotations (ROT) test, with a strict time limit of 10 minutes (Bodner and Guay, 1997).

The Purdue Visualisation of Rotations (ROT) test, has 20 items and is different to the MRT because it includes curved facets and features that are not visible in certain positions (non-deducible), rotations are complex, based on the axes of the object rather than those of the scene (Bodner and Guay, 1997). This study, selected the ROT test because it was short, which agreed with time limitations of medical students; its sensitivity in detecting gender differences was an additional benefit as this feature became important in the analysis of results. A concern that became apparent during testing was that the time limit of 10 minutes seemed to induce psychological pressure on participants, which had the potential of affecting negatively the performance of those sensitive to this type of pressure (see section 3.2.7 Spatial Ability Test for details).

### **2.3.1.5 Gender in Spatial Ability**

Although initially this study did not aim at focusing on gender, this topic gained importance due to literature reports of a strong influence of this variable in spatial ability. The gender characteristics of the population of medical students (60% female) added to the relevance of this topic, and as the project unfolded, gender became part of the various sections of this report.

As a clarification of terms for this thesis, the selection of the term "Gender" to replace the more accurate "Sex," which refers to biologically defined males and females, follows the trend of using this term in literature that has the potential to reach non-specialist audiences. In addition, the General Medical Council (GMC) uses "gender" in official publications instead of sex, e.g. "The state of medical education and practice in the UK" (Brown *et al.*, 2015).

Numerous studies identified women as low-scorers in spatial ability when compared to their male counterparts (Anastasi, 1958; Voyer, Voyer and Bryden, 1995; Nordvik and Amponsah, 1998; Gorska, Sheryl and Leopold, 2016). Levine et al reported detecting spatial ability in children as young as four years of age (Levine *et al.*, 1999), although results tend to be equal for both genders, the first indications of difference favouring males seem to appear between ages four and five (Anastasi, 1958; Levine *et al.*, 1999). The difference in mental rotation with males outperforming females becomes significant by age 9 (Kerns and Berenbaum, 1991; Brosnan, 1998),

The gender difference in spatial ability seems to relate to the following factors:

**Intrinsic factors:**

- Reasoning and brain activation: Differences in activation of brain areas as response to different strategies to solve spatial problems (Butler *et al.*, 2006).
- Hormonal influence: The presence of abnormally elevated levels of androgenic hormones, as in congenital adrenal hyperplasia, result in a positive influence in spatial ability scores in women (Berenbaum, Korman Bryk and Beltz, 2012). Administration of androgens to transsexual females in the process of sex reassignment significantly increased spatial ability scores, while oestrogens and anti-androgens given to males deteriorated their spatial ability (Van Goozen *et al.*, 1995). Spatial ability scores were significantly better during menstruation using the Rod and Frame test, compared to poor spatial ability scores between menses (mid-luteal phase), when oestrogen levels are high (Hampson, Kimura and Thompson, 1988).
- Genetic influence: Reports of a sex-linked recessive inheritance of spatial ability (Stafford, 1961; Hartlage, 1970; Bock and Kolakowski, 1973), proved inconsistent and not supported by large family studies (McGee, 1982).

**Extrinsic factors:**

- Female's poor scoring in spatial ability might relate to insufficient exposure to hand-eye coordination activities in their youth, e.g. construction games and 3D computer games (Sorby, 2009). Gender difference in spatial ability is absent in societies where children of both sexes have more independence, e.g. Eskimo culture, as opposed to cultures where males have strong authority over females, e.g. Tenmne culture in Africa (Maccoby and Jacklin, 1974).

Poor spatial skills seem to place women at a disadvantage in some academic areas, affecting their overall performance in secondary school, which may persist at higher levels of education in areas that require spatial skills, becoming an obstacle in academic progression at university and resulting in less women completing studies up to graduation. This disadvantage may play a role, although a minor one, in the under-representation of women in the fields of engineering, mathematics, science and technology (Hill, Corbett and Rose, 2010).

Researchers estimated that the difference in spatial skills would predict a ratio of two men for every woman following a science path, but in reality, the ratio observed is of around 20 men for every woman, which shows that other factors influence women's interest in technological careers (Simon and Osborne, 2010). From 11 ability-related variables tested on first-year students of engineering, the scores in spatial ability were the best predictor of success in the engineering design graphics course (Gimmestad, 1990). Further research showed that training in spatial ability reduced gender differences in the results from the graphic engineering course, improved retention of female students and enhanced the overall performance of both genders; consequently, the institution implemented formal training in spatial ability to all their students in their curriculum (Sorby, 2009).

The training course on spatial awareness, developed by Sorby, helped students gain a skill, starting from an ability, by analysing spatial aspects of objects, and creating an awareness of a spatial perspective using manual activities (drawing or building models). It consisted of 10 sections, each with an instructional component, followed by a practice section using a workbook (Hegarty *et al.*, 2007; Sorby and Wysocki, 2012). Sorby's course was the main reference for the training intervention in this study (see section 2.3.1.8 Spatial Awareness Training).

### **2.3.1.6 Medicine, Anatomy and Spatial Ability**

In medicine, spatial ability plays a role in understanding complex structures and their distribution in space. Spatial ability may help physicians match findings from physical examination with their anatomical knowledge to determine the underlying structures they represent and if they are within normal limits. This ability may assist in the interpretation of medical imaging, e.g. scanned images, X-Ray plates or ultrasound, that show a flat transection of tissues, the understanding of which, requires visualisation to place into a three-dimensional context. Spatial ability could also be important in orientation and navigation during invasive procedures like dissection and surgery, where results depend on careful visualisation to define a plan of action.

#### **Anatomy and Spatial Ability**

There is an increasing interest in improving the quality of teaching to provide students with the best opportunity to absorb the core anatomy syllabus and ensure they acquire the knowledge to practice medicine safely (Anatomical Society, 2015; Smith *et al.*, 2015a). As structures in the human body exist in three-dimensions, it is important that students understand the relationship of anatomical features in space as they learn the anatomical language that they will use throughout their career (Smith *et al.*, 2018).

Research from the University of Southampton in 2011 focused on the relevance of spatial ability in anatomical studies. Based on 40 participants who ranged from novices to experts in anatomy knowledge and 10 controls from the general population, spatial ability was a good predictor of success in learning anatomy, based on better results obtained by more experienced participants (Fernandez, Dror and Smith, 2011). The tests were predominantly two-dimensional and proved to be useful in demonstrating the various levels of anatomical knowledge; although there was equal gender participation, there was no mention of a gender differences (perhaps not evident due to a small sample).

The relationship between innate spatial ability and the capacity of learning anatomy was also an area of interest to researchers. Scores in spatial ability were best related to learning the anatomy of the wrist in 149 anatomy students, concluding that spatial ability is an important predictor of success in anatomy (Garg, Norman and Sperotable, 2001). The opposite is also true, learning anatomy enhances spatial ability, as described by Vorstenbosch et al, who compared mental rotation test scores of medical students and educational science students before and after training in anatomy. Medical students not only scored more than science educational students but they improved the most in spatial ability after training (Vorstenbosch *et al.*, 2013). Lufler et al reported evidence of both, medical students with high scores in spatial ability were more likely to obtain higher scores in anatomy, and the average spatial ability score of all 255 students was significantly higher after their anatomy course (Lufler *et al.*, 2011b).

Using 3D graphics to teach anatomy is a growing trend but there are differences in opinions about the advantages of this method in relation to traditional teaching using flat images. Some reported no difference in performance in learning anatomy using a computerised 3D model and printed images of the shoulder joint to 86 students (Berney *et al.*, 2015), or teaching the wrist joint to 49 students (Garg *et al.*, 1999). Others indicated that teaching with computer graphic presentations of multiple views of

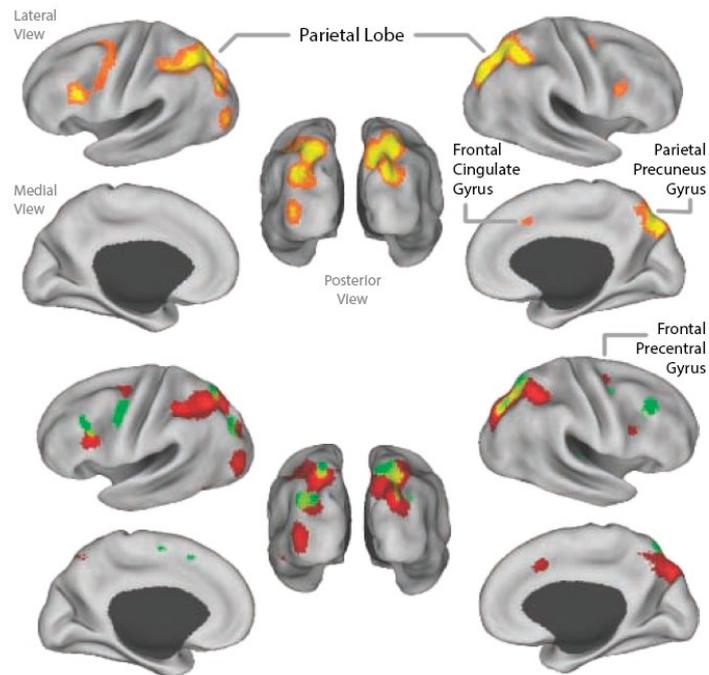
the brain had a negative learning impact on students with low scores in spatial ability, while showing key views of the brain, similar to book illustrations, resulted in better learning (Levingston *et al.*, 2007). In contrast, some researchers found a positive relation of 3D graphics and learning anatomy, even in non-medical participants with good spatial ability (MRT) (Luursema *et al.*, 2008), although there was no clear correlation of anatomy learning and Virtual Environment teaching (with head mounted displays), despite finding a correlation between spatial ability (MRT) and anatomy learning (Luursema, Vorstenbosch and Kooloos, 2017). In another study, participants who performed well in the spatial ability test, also did well in a 3D graphics test of visualisation, recommending training in spatial ability for those with poor results (Nguyen *et al.*, 2013). These reports suggest that students that start learning with good spatial ability are likely to gain better learning of anatomy with or without computer graphic representations than those with low spatial ability.

A study of 97 students concluded that students require a balanced combination of memorisation, understanding and visualisation to meet the demanding intellectual effort of learning the complexity of anatomy (Pandey and Zimitat, 2007). If this is the goal of anatomy teaching, finding the right balance of those elements is likely to become the subject of future research. Improvements in technology and computer graphics may also affect research in the future, as they may either place more demands on the student with poor spatial skills or supplement the student's spatial ability.

The relationship between anatomy and spatial ability may transfer to later stages of training in medicine and particularly in surgery, although direct links were inconclusive (Francis *et al.*, 2001; Luursema, 2010).

### 2.3.1.7 Cerebral Function and Spatial Ability

The use of Functional Magnetic Resonance Imaging (fMRI) while participants engage in mental rotation showed activation of both parietal regions, the left pre-motor region, bilateral inferior and middle frontal gyri and left middle temporal gyrus (Wanzel *et al.*, 2007). Although this is not mentioned in Wanzel's report, the involvement of the pre-motor region of the brain is of interest because this region harbours mirror motor neurons, suspected of involvement in motor intention (Purves *et al.*, 2012), which might be part of the phenomenon of mental rotation, because it includes visualisation and an intention of producing movement. In addition, a meta-analysis showed that mental rotation mainly involves activation of the intraparietal sulcus and adjacent regions, and the medial superior paracentral cortex, possibly related to motor simulation (Figure 1) (Zacks, 2008).



**Figure 1. Cortical Activation during Mental Rotation in fMRI Scan**

Note. Red: Single task mental rotation; Green: comparison of mental rotation to lose control task (fixation on crosshair); Yellow: transformation task. Image annotated based on original by Zacks (Zacks, 2008).

A refined analysis of the areas activated by gender showed that when asked to solve a mental rotation problem, men activate sensory cortices as well as regions involved in implicit learning (basal ganglia) and mental imagery (precuneus gyrus of the parietal lobe), consistent with a more automatic, “bottom–up” strategy. This refers to the activation of brain areas with minimal involvement of thought processing, which takes place at hierarchically higher areas of the brain. In contrast, women activated the dorsal-medial prefrontal and other high-order heteromodal association areas, which relate to an effortful “top-down” strategy. This strategy suggests a prominent use of thought processing at higher hierarchical areas of the brain to begin the challenge of solving the rotation problem, adding complexity and a longer route to reach a conclusion (Butler *et al.*, 2006).

### **2.3.1.8 Spatial Awareness Training**

Spatial ability refers to the state of this ability before exposure to training. Training transforms this ability into a skill, providing the learner with a tool to improve spatial reasoning and the capacity to compete with those who have a higher level of innate ability (Hegarty *et al.*, 2007).

The term that more precisely described spatial ability training in the literature was “spatial awareness training”. In such a programme, rather than practicing to answer spatial ability questions, trainees learn to analyse geometric and physical concepts that made them more aware of the intricacies of rotation, visualisation and other aspects related to the study of physical structures and motion (Sorby, 2009; Sorby and Wysocki, 2012). Those qualities make training relevant to other STEM fields (Hegarty *et al.*, 2010), (see section 3.2.8 Spatial Awareness Training in the Methodology chapter for details).

### 2.3.2 Manual Dexterity

Manual dexterity is another skill of interest to this study; it refers to the ability to coordinate vision with hands and fingers movements to grasp and manipulate objects to perform tasks (Makofske, 2016). These actions vary depending on the field of reference: everyday life demands an efficient performance (dressing, tying shoes), some occupations may require accuracy and speed (holding tools, assembly line), working in the health sector may look for precision, while art may benefit from gracility and delicacy (Stoeger and Ziegler, 2010).

Manual dexterity falls under the broad spectrum of Motor Ability, which results from the analysis of motor performance, measured with a variety of test batteries. Although the concept of a general motor aptitude (motoric “g”) exists, there is no general agreement of its validity, as specific skills seem to be independent for measuring and training (Ibrahim, Heard and Blanksby, 2011; Lorås and Sigmundsson, 2012). The taxonomy of movement skills has a hierarchical structure in which General Motor Ability (GMA) is the basis and refers to traits underlying movement, e.g. reaction time, multi-limb coordination. On the next level are Movement Skill Foundations (aspects that facilitate or limit performance, e.g. balance, flexibility that change with practice and age), followed by Movement Skill Sets (skills grouped for testing), and finally, at the top of the pyramid are Movement Skills (specific goal-directed movements, e.g. driving, writing, speaking). The latter includes the subset of manual dexterity, which can change with practice and experience (Burton and Rodgeron, 2001). A section of manual dexterity is finger dexterity, which is most relevant to medicine because procedures in this field are often delicate and precise. The Purdue Pegboard test, manufactured by Lafayette Instruments, commonly used to test finger dexterity in participants older than 15, was the test of choice for this project (Wang *et al.*, 2011).

Some researchers found that women outperformed men in visual-motor coordination and finger dexterity by 5 and 8% respectively (Maccoby and Jacklin, 1974), while others reported similar manual performance in both genders when testing Activities of Daily Living (ADL) in 9250 men and 9250 women. Men had better motor skills while women had higher processing skills but both genders were as good in completing the tasks (Merritt and Fisher, 2003). Based on the last reports, this study did not expect a gender difference in manual dexterity performance.

In higher education, some schools in the area of health require students to have a background of manual dexterity. In the United States, the American Dental Association requires a manual dexterity test as part of the Dental Admission Test (DAT) since 1957 (Cannon, 2004; Indiana University, 2007). This is also the case in the UK, where universities request proof of manual dexterity or taking a test at interview (University of Sheffield, 2017; University of Manchester, 2017; University of Dundee, 2017; UCLAN, 2017). Giuliani et al showed that training in dentistry significantly improves manual dexterity of students and therefore challenge the need of a manual dexterity aptitude test to enter dentistry, supporting direct admission to some universities (Giuliani *et al.*, 2007; University of Glasgow, 2017).

One of the most popular methods of testing manual dexterity is the Purdue Pegboard Test, developed by Tiffin and Asher, from the University of Purdue, USA, in 1948, who reported validation and reliability, and reference values for comparison (Tiffin and Asher, 1948). The test comprises of a board with two lines of holes placed along both sides of the midline up to a section marked as the beginning; it contains concavities at the top to hold small objects (pegs, washers and collars) that the participant is asked to place in the holes following instructions from the examiner. The test measures gross hand movement and fine finger dexterity (Tiffin and Asher, 1948; Lafayette Instruments, 2015), (Figure 2).



**Figure 2. Purdue Pegboard Test for Manual Dexterity**

Note. Purdue Pegboard Test designed by Tiffin and Asher in 1984 and produced by Lafayette Instruments was the test of choice for this study (Lafayette Instruments, 2015).

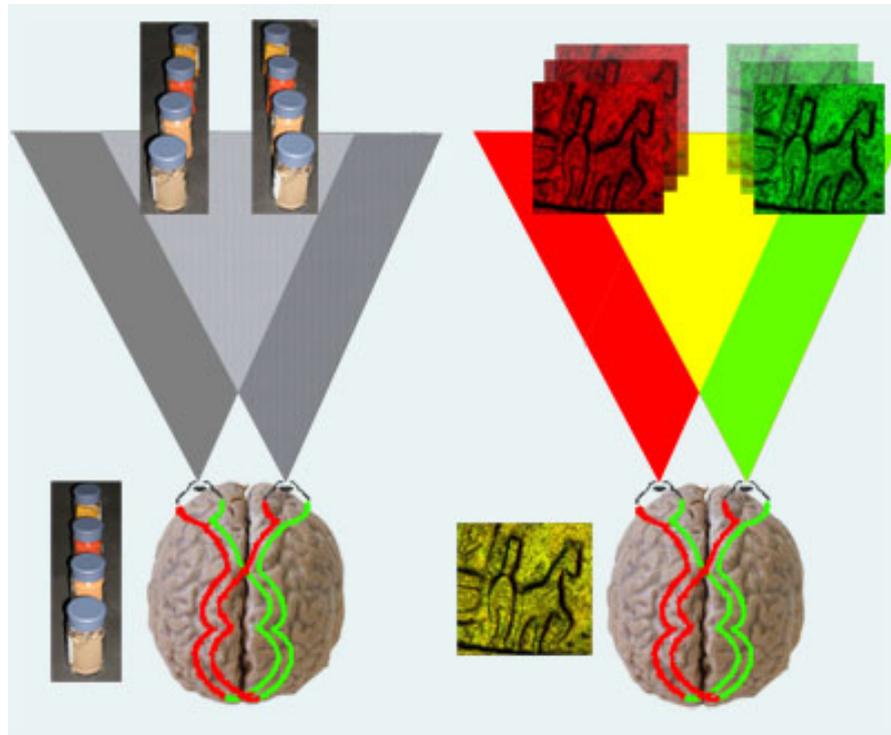
In association with manual dexterity, there are other areas considered of importance in learning anatomy and that have to do with manipulating objects or tissues. The physical contact involved in dissection, for example, provides the opportunity to reinforce the process of learning, described as touch-mediated perception; it is one of the unique experiences offered by dissecting a human cadaver (Granger, 2004). Smith and Saber-Sheikh measured manual movements during dissection, which is an indirect method of quantifying manual contact with a cadaver, closely related to touch-mediated perception (Smith and Saber-Sheikh, 2014). Touch-mediated perception and multidimensional understanding of the organisation of the human body, added to the psychological impact of dissecting a human cadaver, an experience reported as a positive and significant life event (O'Carroll *et al.*, 2002), constitute part of the gain that an anatomy laboratory can provide to students, highlighting its role in the education of future doctors (Granger, 2004).

### 2.3.3 Stereopsis

Stereopsis is the ability to appreciate depth by detecting objects closer or farther away from the viewer's eyes. The term derives from the Greek "stereo", meaning "solid" and "opsis," meaning "vision" or "power of sight," and represents solidity or completeness of vision (Ponce and Born, 2008). Stereopsis provides a unique sensation of depth within the surrounding world and is commonly taken for granted, only to become evident when exaggerated (watching 3D projections) or when it is missing, which may slow down performance and accuracy (Levi, Knill and Bavelier, 2015). Stereopsis depends on simultaneous perception with both eyes (binocular perception) but the brain also uses cues that do not rely on binocularity (monocular cues), e.g. superposition, difference in size and perspective, to understand the location of objects in the space that we see. For the medical student and physician, it is important to distinguish small levels of depth to perform delicate tasks, for example, cannulating a vessel or removing particles trapped in delicate areas e.g. near and around the eyes, or suturing a face wound. Stereopsis is also important for new forms of image display that depend on binocular perception, like virtual reality goggles and 3D displays (Luursema, Vorstenbosch and Kooloos, 2017).

The brain (mostly the occipital cortex) analyses the information coming simultaneously from both eyes, each from a slightly different point of view due to the anatomical separation of the eyes, (normal distance range between pupils: 57-65mm) (Thompson, 2002). These two images have a point or plane of interest, which is similar in both eyes (plane of fixation), while the areas beyond and closer to the observer look different from the perspective of each eye, thus called "binocular disparity." When the brain superimposes both images, the result is a plane of single images preceded and followed by areas of double vision, interpreted as depth (Purves *et al.*, 2012). The

result is a unique and dynamic visual perception of depth and a rich three-dimensional experience of the world (DeAngelis, 2000; Ponce and Born, 2008), (Figure 3).



**Figure 3. Stereopsis or Stereo Vision**

**Note** Information from images captured by the retinas partially cross to the contralateral side before arriving to the cerebral cortex. Further processing compares the input from both eyes and determines differences to define depth by disparity.

Images arriving at the brain from both eyes need to be similar in quality, with sufficient detail for the visual system to compare; if vision from one eye is blurred or obstructed, comparison is not possible. If this happens during childhood, due to misalignment of the eyes (strabismus or “squint”) or after extended obstruction to vision in one eye, the vision integration areas of the brain fail to develop. The consequence is a neuro-developmental disorder called “amblyopia”, commonly known as “lazy eye”, with a prevalence of 1-5% in the UK (Carlton and Czoski-Murray, 2010; Purves *et al.*, 2012; Rutar, 2015; Yen, 2016). In amblyopia, the vision of one eye dominates over the other,

which impedes fine perception of depth or stereopsis, because amblyopes only use one eye for visual tasks (Levi, Knill and Bavelier, 2015). Poor stereopsis may degrade visual feedback of hand position and affect visually guided hand movements, resulting in mistakes during delicate manual tasks, but most amblyopes lead a normal life as they experience depth through monocular cues, and some may not even be aware of their condition (Murdoch, McGhee and Glover, 1991; Levi, Knill and Bavelier, 2015).

### **Testing Stereopsis**

The measure of someone's stereopsis, known as "stereoacuity," refers to the minimal angle of separation in depth detected by this person, expressed in seconds of arc (arcsec); the normal range of stereopsis is 30 to 60 arcsec, or an average of 40 arcsec (Kanski, 2009; Stidwill and Fletcher, 2014). The Titmus test, one of the most popular in evaluating stereopsis, was the test of choice for this project; it uses polarised goggles to show different images to each eye and covers a range of targets from 59 minutes of arc (arcmin) to 40 arcsec (Kanski, 2009) (see section 3.2.10.2 Stereopsis).

Stereopsis is highly dependent on good visual acuity in both eyes. A difference of more than two lines between eyes in the visual acuity test (with best correction) requires further investigation, particularly if the difference persists at close distance. Based on this relationship with stereopsis and from the general concept of the importance of vision for learning, this project included visual acuity testing.

### **Testing Visual Acuity**

Visual acuity is a functional measure of the visual system and refers to the capacity of the eye to distinguish detail (ICO, 1984), and commonly receives a score that represents the size of the smallest target correctly identified by the participant (Lennie and Van Hemel, 2002). This test evaluates the integrity of the optical or transparent

parts of the eyes, the neurological portion (retina to occipital lobe of the brain) and the cognitive element that completes the visual perception. In this study, participants of Group A (Intervention) underwent visual acuity testing before stereopsis testing to rule out amblyopia (see section 3.2.10.1 Visual Acuity).

## **2.4 Review of Related Topics**

In addition to the main topics of this study (spatial ability, manual dexterity and stereopsis), additional areas provided a contextual framework within which the study took place. Those areas relate to the broad field of education and learning, medical education and a review of the medical student as the central element of this study.

### **2.4.1 Cognitive Load Theory**

Illeris proposes a continuous relation between the external or environmental interactions and the internal or individual processes, which depend on the relationship between content learnt (knowledge, skills, opinions, meaning) and incentive (mental energy to run the process, including volition and motivation). This model results from a combination of social learning theories that focus on external interactions, and classical behavioural and cognitive theories that deal with internal psychological processes (Illeris, 2018). One of the latter type is Cognitive Load Theory, which explains learning by acquisition of schemas and automation (Sweller, 1994).

In Cognitive Load Theory, schemas refer to units of knowledge acquired and organised by the learner in order of importance. Schemas are similar to the blocks of knowledge retained in short term memory, described as “chunks” by Chase, who proposed that experts differ from novices by the size of the chunks they are able to retain (Chase and Simon, 1973). Automation refers to the goal of decreasing the effort of remembering

something new by accessing permanent memory storage directly, bypassing the use of transient or working memory, which is limited and volatile. The practice of using the recalling process repeatedly creates familiarisation, which promotes automaticity, leading to remembering without conscious effort. Learning occurs by accumulating automated schemas, as single units or in groups, and storing them in long-term memory, leaving space in working memory for other functions. This slow process relies on repetition or practice to build-up automation, a sign of expertise or skill (Sweller, 1994).

One of the most simplistic forms of learning is Paired Association, where one element relates directly to another, e.g. in a new language, simply memorising words and their equivalents, or in anatomy, where it would be possible to memorise related lists, where the difficulty would arise mostly from the size of the lists. Difficulty, or cognitive demand, arises when the elements to learn are complex or with increased “interactivity” between elements, e.g. learning syntax of a language, or functional interaction of anatomical elements. In cognitive load theory, isolated elements do not constitute schemas because the latter imply interactivity of elements; these schemas may become elements of larger tasks. For a novice writer, for example, formulating a phrase with a few words may require the significant cognitive effort of remembering language rules, while a skilled author would write the same phrase effortlessly, becoming a unit in the composition of a paragraph (Sweller, 1994).

Teaching using cognitive load theory aims at reducing the cognitive load for the student, for example, by promoting goal-free problems, where instead of asking the student to solve a problem from scratch, it appears partially completed to allow the student finish it and obtain the results, or by designing self-contained graphics that do not need text to explain them. Teaching materials and methods, controlled by the teacher, impose an extrinsic cognitive load, which in addition to the learner’s intrinsic cognitive load, results in a Total Cognitive Load. From the student’s perspective, this difference is irrelevant as when the total cognitive load is high, things are difficult to

learn. Total load responds to element interactivity; when this is low, the total follows, but when it is high, it may become overwhelming for the learner, particularly if the extrinsic load is also high. In summary, reducing cognitive load aims at reducing element interactivity (Sweller, 1994), (see 5.11 Academic Performance).

#### **2.4.2 Medical Education in the UK**

Medical education in the United Kingdom is under the regulation of the General Medical Council (GMC). Since its beginning in 1858, the GMC guides the practice of medicine by setting standards for doctors, overseeing medical education (under and post graduate), managing registration for practice, keeping standards with revalidation and guarding the safety of patients by investigating and acting on concerns about doctors (GMC, 2016a).

In 2014, Kopelman published a number of “principles” to improve the undergraduate curriculum. The recommendations included definition of core areas by consensus, based on basic sciences and behaviour and avoiding overload; an emphasis in clinical skills learnt by apprenticeship, understanding teamwork and the inter-professionalism of the health service, and finally, assessment of core knowledge with constructive feedback (Kopelman, 2014). Some of those appeared in the GMC document “Promoting excellence: Standards for medical education and training,” which replaced “Tomorrow’s Doctors” in 2016 (GMC, 2016b). Kopelman’s first recommendations materialised in 2015; by consensus, a panel of experts in anatomy education, under the umbrella of the Anatomical Society, published a new undergraduate regional anatomy syllabus for medicine, identifying 156 outcomes in the field (Smith *et al.*, 2015b; Smith *et al.*, 2015a). The implementation of the new syllabus is currently under way; future reports may provide evidence of its effect on the education of medical students.

### **2.4.3 Gender Balance in Medicine**

In the UK, the first female medical doctor was Elizabeth Garrett Anderson, who entered the profession in 1865. Since then, the numbers of women in the medical register in the UK increased gradually, from 25 in 1881 to 495 in 1911; by 1965, 7% of doctors were women. Female doctors on the List of Registered Medical Practitioners (LRMP), reported by the GMC, increased from 35% in 2007 to 45.5% in 2016, although this growth seems to be slowing, in Scotland and Northern Ireland it reached parity to that of male doctors (BBC, 2017; GMC, 2017). These changes suggest a move towards equality but there is still a lot of imbalance at higher levels of the medical career that may require addressing before reaching gender equilibrium. Improvement in abilities at which female students might be at a disadvantage, e.g. spatial ability, may enhance their skills and diminish the differences in that area.

### **2.4.4 The Medical Student**

This section reviews some of the factors that influence young people in their aspiration to enter medical school, the factors that dictate the number of students that train as doctors and how students cope with the medical training programme. Because this project did not measure social factors, it will rely on findings from the literature to obtain the required information.

#### **2.4.4.1 Socio-economic Composition of Medical Students at BSMS**

The analysis of the socio-economic composition of medical students entering BSMS may provide an idea of the section of the community they represent, which may have a relation with their performance in spatial ability and manual dexterity, as both relate to

levels of education and cognitive achievement (Levine *et al.*, 2005). The widening participation programme of the medical school tries to ensure that the population of graduates reflects the full range of the society from which they are drawn (Reed, 2017; BSMS, 2017b), and therefore plays a role by promoting medicine to a variety of socio-economic levels. They account for approximately 7% of study offers by BSMS (in 2014) (MSC, 2014; UoS, 2016; Wickham, 2017; BSMS, 2017d). Like in most medical schools, the rest of the students are commonly a product of a highly selective process that offers the best education to those who were already doing well at school (McClelland, 1973). There is an association between high achievement at school, higher scores in most abilities and future achievement, which raises the expectations of performance for medical students (Guay and McDaniel, 1977; McManus *et al.*, 2003).

#### **2.4.4.2 Medical Student's Learning Strategies**

Understanding how students learn can help improve the method used for teaching; this is one of the elements guiding the change in teaching strategies, to deliver content that is intellectually challenging, like the one medical students face at university, e.g. anatomy, physiology, immunology, etc. (Pandey and Zimitat, 2007).

Students demonstrate different approaches when confronted with a learning challenge that has content and structure. Dealing with content gave rise to the concept of superficial or "surface" learning, focusing on textual content, memorising terms without context (using lists, mnemonics, etc.), and "deep" learning, which refers to the appreciation of meaning and practical use of the content. Dealing with structure takes into account either embracing the full content in a "holistic" manner or separating it in individual sections in an "atomistic" approach. Lastly, adopting a "strategic" approach, which provides the best results, students may combine surface and deep learning, depending on the material and type of test (Entwistle and Paul, 1983; Newble and Entwistle, 1986; Ramsden, 2003; Ward, 2010). Investigating the student's approach

may influence the design of assessments; the ideal test, not only has to be relevant but also provide a reason for students to select a desirable approach (Newble and Entwistle, 1986; Smith and Mathias, 2010; Smith, Finn and Border, 2017).

The deep method approach, which is most beneficial, could have an explanation through Cognitive Load Theory, where after repetition, the acquisition of schemas becomes automatic and future encounters with the subject, evoke this information, stored in long-term memory (Entwistle and Paul, 1983; Sweller, 1994; Smith and Mathias, 2010; Smith *et al.*, 2017). The value of repetition of the content at different stages of medical training reinforces learning by applying this knowledge to clinical tasks in a process called “encapsulation”, which occurs at foundation level (Boshuizen *et al.*, 1995; Smith, 2008; Smith and Mathias, 2010).

This study aims at improving understanding of anatomical structural content through enhancement of spatial skills, which might support a deep approach to learning.

# **CHAPTER 3      METHODOLOGY AND METHODS**

## **3.1 Introduction**

Methodology, the first section of this chapter, refers to the background behind the techniques used in this study, including the definition of the research questions, research design, population sampling and background on the tests and other variables measured in this project; it also covers data and marketing considerations.

The section on Methods, describes the specific techniques used in every test and procedures relevant to the study, including data management, statistical analysis and logistics required to organise the testing sessions.

## **3.2 Methodology**

### **3.2.1 Definition of Variables**

The nomenclature of variables in statistics is an issue on which not all experts agree. They recognise that there is a lack of clarity with a potential for misleading the reader to a point of even avoiding their use and proposing alternative names (Wuensch, 2004; Field, 2012; Creswell, 2014). Despite the controversy, this project identified three types of variables: Independent, dependent and confounding. The variable open to manipulation was “training in spatial awareness,” representing the Independent

variable. Its influence produces changes in the scores in spatial ability, which constitutes the Dependent variable. Other dependent variables involved in this study were the scores in anatomy, and for the intervention group, the scores in manual dexterity, and of lesser importance, scores in stereopsis and visual acuity. In addition, the variable “gender” became important, particularly in spatial ability, characterised as a Control variable, an independent modality managed during data analysis. Another variable of interest was the exposure to studying anatomy, which may indirectly influence the independent variable, therefore falling into the category of Confounding or Spurious variable, which is a form of intervening variable that is not measured or observed but exists (Creswell, 2014).

### **3.2.2 Definition of the Research Questions**

Based on the variables described above, finding a single question that encompassed all areas of this study was a challenge because there were various areas of interest. To help with the organisation of thoughts, the study took advantage of a technique that identifies its main components to build a comprehensive research question. Those components were Population, Intervention or Issue (independent variable), Comparison or Context, Outcome (dependent variable) and Time, which make the acronym PICOT and constitutes the name of the technique (Riva *et al.*, 2012).

The components identified were the following:

- (P) Population = Medical students.
- (I) Intervention = Spatial awareness training.
- (C) Comparison = Comparison between intervention and control groups.
- (O) Outcome = Outcome measured by the scores in spatial ability, manual dexterity and anatomy.
- (T) Time = Data collected between February and May 2016.

According to the technique, placing the components in the PICOT sequence provides guidance in the construction of the research question.

### **3.2.2.1 Research Question**

After adjusting the PICOT technique to this study, the resulting main research question was the following:

“How do first and second-year medical students who followed training in spatial awareness compare with those without training in their scores in spatial ability and anatomy in 2016?”

### **3.2.2.2 Sub Questions**

The following sub-questions relate to measured and analysed variables with secondary relevance.

“How does the gender of participating medical students influence results in spatial ability, manual dexterity and anatomy?”

“Is there a correlation between spatial ability and anatomy scores for the intervention and control groups, before and after training in spatial awareness?”

“How do participants of the intervention group measure in manual dexterity, before and after training in spatial awareness in 2016?”

These research questions exclude some variables that helped in the interpretation of data, e.g. age, year of study, handedness.

### **3.2.3 Definition of Hypotheses and Aims**

Building up on the research questions, this section presents a set of Null Hypotheses, stating equality between variables, and Alternative Hypotheses (also known as “research” or “experimental” hypotheses), predicting or suggesting a potential outcome (Field and Hole, 2003; Creswell, 2014; Salkind, 2017).

#### **3.2.3.1 Hypotheses of the Study**

The alternative hypotheses in this section denote a non-directional nature to match non-directional significance values used as default in the statistical analysis.

##### **Hypothesis 1**

*Null hypothesis:* There is no difference in spatial ability and anatomy scores between the intervention and control groups after training in spatial awareness.

*Alternative hypothesis:* Participants in the intervention group will obtain different scores in the spatial ability test and the anatomy exam after training in spatial awareness, than will participants in the control group who did not following training.

##### **Hypothesis 2**

*Null hypothesis:* There is no difference in spatial ability scores between male and female participants.

*Alternative hypothesis:* Male participants in all groups (full cohort, intervention and control) will obtain different scores in spatial ability than their female counterparts.

##### **Hypothesis 3**

*Null Hypothesis:* After training in spatial awareness, post-training anatomy module scores from the intervention group are no different than those from the control group.

*Alternative hypothesis:* After training, there will be a difference in anatomy scores between the intervention and control groups.

#### **Hypothesis 4**

*Null Hypothesis:* There is no correlation between spatial ability and anatomy scores for the intervention and control groups.

*Alternative hypothesis:* There will be a relationship between spatial ability and anatomy scores for students in the intervention and control groups.

#### **Hypothesis 5**

*Null hypothesis:* There is no difference between the first and second manual dexterity test scores for the intervention group after training in spatial awareness.

*Alternative hypothesis:* Participants in the intervention group will obtain different scores on the second manual dexterity test, after spatial awareness training.

### **3.2.3.2 Aims of the Study**

Based on the hypotheses, the aims or objectives of the study included:

1. Determine if the provision of training in spatial awareness influences the scores of a test – re-test experiment in spatial ability, and the scores in the anatomy test given by BSMS to participating students.

To fulfil this aim, the study required the following: An initial spatial ability test given to all participants, a population sampling strategy to recruit a subgroup of participants for the experiment, a strategy for randomised allocation of the training intervention, delivery of training in spatial awareness to the intervention group and a second test in spatial ability.

2. Determine how the scores in spatial ability of male participants compare to those of female participants for every group.

To fulfil this aim, the study required to include a “gender” question in the data-collection tools (e.g. questionnaires or answer sheets) and analyse the data using the control variable of gender for every group and in comparisons of groups.

3. Determine if the scores in manual dexterity differ from the first (pre-training) to the second (post-training) test for the intervention group.

To comply with this aim, the study required measuring manual dexterity at the beginning and the end of the study, and analysing data to compare those results.

### **3.2.4 Research Methodology**

This section presents the arguments that led to the selection of a research methodology for this project.

#### **3.2.4.1 Research Approach and Study Design**

The selection of research approach, defined as “the plans and procedures needed to carry out research,” depended on the type of data required to answer the research questions and the approach most commonly found in the literature review (Creswell, 2014).

From the factors involved in spatial ability (see section 2.3.1.2), Spatial Orientation and Spatial Visualisation were the most relevant to this study because they may play a role in understanding and mentally manipulating three-dimensional structures encountered in anatomy. Researchers used a mental rotation test to explore those factors (Bodner and Guay, 1997), and for studies in medical training (Hegarty *et al.*, 2007). In addition, researchers found that the most gender-sensitive test was the Mental Rotation test, particularly when scored out of 20 points (Voyer, Voyer and Bryden, 1995). Those reports provide a strong argument supporting the use of quantitative methods for this

study, which would allow testing of theories and statistical analysis of relationships between numerical variables (Dawson, 2006; Creswell, 2014).

Within this quantitative approach there are various “strategies of inquiry” also called “research designs” (Creswell, 2014); the most relevant to this study being, the “experimental design,” selected for this project because it was ideal to determine cause an effect.

Initially, this project intended to follow a “true experimental” design, which involved the random assignment of a sub-sample from those who took the first spatial ability test, into sub-groups that received different treatments, thus creating an intervention or treatment group, and a control group, which was exempt from treatment. Both groups were to undergo measurement of the variable of interest, in this case, spatial ability, before and after the intervention and only the intervention group would receive the treatment, in this case, training in spatial awareness.

In this study, the majority of students attending a lecture (Year 1) accepted the invitation to take the first spatial ability test at the end of that class; three months later, most took the second test in a similar fashion. Within those three months, some students responded to invitations to join an experiment group. The initial expectation involved the creation of a pool of students that would undergo a random assignment of participants to the intervention and the control sub-groups, which meant that to reach the minimum sample size of 26, the experiment needed a minimum of 52 participants. By the end of the recruitment period (3 weeks), student response was extremely poor (less than a dozen). It became clear that this goal was unreachable, the population needed expansion and the strategy of sub-group assignment required a revision.

The Ethics Committee approved a request for the inclusion of Year 2 students of the medical degree, and all students who signed-up for the experiment became the intervention sub-group. The control group emerged from students who took the first and the second spatial ability tests and did not receive training. Consequently, the

inquiry design changed from “experimental,” where the researcher randomly assigns participants to the intervention and control groups, to “quasi-experimental,” where the researcher does not have complete control of the randomisation (manipulation of the independent variable). From the various types of quasi-experimental design, the “pre-test/post-test control group” or “untreated control group design with pre-test and post-test”, a form of “Between Groups” experiment design, fitted the circumstances of this project. In this design, a control group would allow a comparison of performance to determine the influence of the treatment on the intervention group (Cook and Campbell, 1979; Field and Hole, 2003; Creswell, 2014).

The loss of randomisation highlighted its importance in research, which is mainly to obtain a representative sample of a known population and ensure a comparable balanced between two groups, in a setting that supports the inference of cause and effect. The lack of randomisation induced a threat to the internal validity of this study, weakening conclusions about the intervention because participation might relate to reasons outside the control of the researcher (bias); and to the external validity, preventing results from being generalised (Cook and Campbell, 1979).

In addition to spatial ability, the central concern of this project, the original design included testing for manual dexterity and stereopsis (additional testing) as an exploration of association with spatial ability, not found in literature reports. Following the change of research method to quasi-experimental, those tests, carried out on a one-to-one basis, were no longer applicable to the control group as this group formed after taking the second spatial ability test. Only those in the intervention group underwent additional testing. Consequently, data collected from the intervention group in these areas, fell under a “one group pre-test / post-test” method, which involved a “measurement 1 > treatment > measurement 2” pattern (Field and Hole, 2003). The lack of a control group for these tests resulted in the uncertainty of not knowing what would happen without treatment, an important weakness that negatively affected its validity. Thus, reporting of additional testing in the study was minimised.

This study accepted elements of both types of quasi-experimental methods, not by design, but as a response to an inability to recruit sufficient participants, which prevented random allocation (see section 4.8 Limitations).

Regarding group naming, this study considers the combination of the Intervention (received training) and the Control (no training) groups as part of the “Experiment”. This is in contrast to some reference material, where the group receiving the treatment is the “Experimental group,” which sounds repetitive and may lead to confusion.

In summary, the methodology of this study was quantitative by approach and quasi-experimental by design.

#### **3.2.4.2 Statistical Power and Sample Size**

Statistical power supports the validity of a comparative analysis. The “null hypothesis” envisions equality between compared groups and absence of the phenomenon of interest. The researcher’s goal is to reject this null hypothesis using comparative statistical tests. The “statistical power” of a “test of null hypothesis” refers to the probability that this test will lead to the rejection of the null hypothesis. It is an indirect estimation of how good a test is at detecting a difference, which symbolises the presence of the phenomenon of interest (Cohen, 1988).

The power of a null-hypothesis statistical test depends on a) The significance value, b) The reliability of the sample, and c) The effect size.

- a) The “significance” value or “ $\alpha$ ,” is a criterion set by the researcher before gathering data. Significance is a “standard of proof” that leads to the rejection of the null hypothesis (critical region of rejection of the null hypothesis). Setting a low significance value, e.g. 0.001, provides a low risk of mistakenly rejecting the null hypothesis, therefore concluding that a difference or a phenomenon exists,

when there is no actual difference or the phenomenon is absent; this is an Alpha or Type I error. A low value restricts the probability of data meeting this standard and limits its statistical power to 0.10. The complement to this power ( $1 - \text{power}$  or  $1 - 0.10 = 0.90$ ) symbolises a high rate of failure to reject a null hypothesis, concluding that there is no phenomenon when one really exists, known as Beta or Type II error, or “*b*”. In this case, the relationship between Type I and Type II errors is  $b/a = 0.90/0.001 = 900$ , which translates into believing that mistakenly accepting that the phenomenon exists is “900 times” more serious than mistakenly missing it. This result reduces to “4 times” when the significance is set to  $a = 0.05$ , thus providing a better balance between those types of errors (Cohen, 1988; Steele, 2015). In this study, the significance value selected was  $a = 0.05$ , a conventional level commonly used in research (Deviant, 2012).

- b) Because the sample value is unknown, the reliability of a test, which is the closeness with which it approximates the population sample, is always an estimate. The larger the size of the sample, the smaller the error and the greater the reliability. This relationship extends to the power, which increases with sample size. Power also responds to the variability of observations, e.g., unreliable responses, defects on the testing method, etc., which will decrease the precision of the sample, and therefore the power. The aim of experimental design is to reduce this variability to increase the precision of observations.
- c) Starting from the premise that the null hypothesis always means that the effect size is zero, Effect Size (*ES*) is some specific value above zero, which demonstrates that the phenomenon in question is present to some degree. This degree is the excess value above that of no effect. For example, if a hypothesis states that an event is present in a reference population at an average of 50%, and the measured value in the group of interest is 52%, then the effect size is 2%, or the excess over the expected value (at which the effect would be zero).

In other words, the Effect Size is an index of the degree of departure from the null hypothesis (Cohen, 1988), or a measure of magnitude that shows “how big is big” (Salkind, 2017).

The four variables presented above (Statistical Power, Significance “ $\alpha$ ”, Sample Size “ $n$ ” and Effect Size “ $ES$ ”) depend on one another. Investigators may determine one variable by setting the other three and looking up the relevant value on reference tables (Cohen, 1988). A practical use of an extract of Cohen’s tables to calculate the statistical power in relation to the outcome for this project appears in section 3.3.3.1 Power and Sample Size Calculation.

To quantify significance, Cohen proposed calculating “index  $d$ ,” which is a value directly related to the difference between means and inversely proportional to the standard deviation, assuming that the latter is the same for each group if they are of equal numbers (Cohen, 1988). When not equal, the standard deviation is that of the control group (Salkind, 2017). Cohen’s index “ $d$ ” represents the degree of departure from the null hypothesis towards the alternate hypothesis, or “effect,” with a size expressed in units of variability, ranked as follows: 0 to 0.20 small, up to 0.50 medium and up to 0.80 or more as large (Cohen, 1988). Although the American Psychological Association (APA) recommends reporting effect size to show the magnitude of observed effects in significant and non-significant results (APA, 2013; Field, 2013; Field, 2018), this study will only use it for power calculations.

In addition to the estimation of power to calculate sample size, after the completion of an experiment, power becomes deducible from the resulting values. This post-hoc method is useful to evaluate the importance of experimental findings (Cohen, 1988).

### **3.2.4.3 Project Timeline**

This study involved first and second-year medical students. The timeline required planning to suit the students' timetable, avoiding periods commonly dedicated to preparation for exams, which include late May and early June.

The study took place between February and May of 2016. In February, the first spatial ability test provided data from students who volunteered to participate. Up to that point, the study was observational, without intervention, qualifying as a cross-sectional approach. Between March and April, active recruitment resulted in the formation of the intervention group, who underwent further testing and training. The second spatial ability test in May, produced data from students who volunteered a second time to take part and this included the intervention group. Because of the presence of a training intervention, this research approach qualified as a quasi-experimental design, rather than a longitudinal study, which is observational by nature and excludes interventions (Payne and Payne, 2004; Institute for Work & Health, 2015).

Data analysis started in June 2016 with the initial results becoming available soon after.

### **3.2.5 Ethics Committee Approval**

The Research Governance and Ethics Committee (RGEC) of Brighton and Sussex Medical School (BSMS) is the authority that governs the approval of studies carried out at the medical school.

According to the BSMS online self-assessment tool, this study required Ethical Review (BSMS, 2015a). The most important factors influencing this requirement were the following:

- Involvement of human subjects.

- Likelihood of those involved in the study to physical or psychological harm.
- Access to personal information that allows identification of the individual or corporate or company information.

Another consideration is that medical students constitute an over-researched population.

From the extensive list of risk-assessment factors reviewed by the RGEC, this study affected the following two:

1 Psychological impact:

The tests have no impact on affective or intellectual status of participants; they challenge mental abilities and manual coordination. Timed challenges may cause low levels of anxiety as participants attempt to complete the tests in the shortest possible time. Considering this, the information sheet they received beforehand contained a clear explanation of the tests and a mention of the possible anxiety factor. The sheet also explained that they could retract from the test at any time should they feel uncomfortable without any type of penalty and that if they experienced stress following the test, the researcher could provide details to contact Student Support services to help with the situation.

2 Unexpected discovery of possible clinical relevant findings:

There was a small chance of finding clinical anomalies during testing, e.g. poor vision, amblyopia or essential tremor (the latter two with similar prevalence of 1-5%), during the tests (Zesiewicz *et al.*, 2005; Carlton and Czoski-Murray, 2010; Rutar, 2015).

Considering this eventuality, the information pack sent to all participants before joining, explained that participation included vision testing and optional discussion of unexpected findings if these were to occur.

Appendices 1 and 2 of this document present a sample of the information sheet and the participant consent form respectively.

This study received RGEC approval on 3<sup>rd</sup> December 2015 (R&D Ref no 15/128/SMI). Although data collection started in February 2016, an initial poor participation, due to minimal response to the recruitment efforts, led to requesting the RGEC to allow the participation of year-2 students. The request received approval on 18th March of 2016 (R&D Ref no 15/128/SMI – Amendment 1) and data collection resumed soon after, finishing in May 2016.

### **3.2.6 Participants in the Study**

This study aimed at exploring innate abilities in participants experiencing the first stages of medical education, who were studying anatomy. Students in the Bachelor of Medicine, Bachelor of Surgery programme at BSMS matched this interest and 196 of them participated.

Students opted to take the initial Spatial Ability test at the end of a regular lecture, when the researcher offered it with the agreement of the relevant lecturer (no payment offered). Further participation in the experiment was also voluntary but a monetary incentive was on offer (£30), which, despite being a common recruitment strategy, raised the question of self-selection bias (see section 5.2 Discussion of Student Participation). A procedure similar to the first test facilitated the second spatial ability test three months later. In the time between the tests, those who joined the experiment underwent further testing and training in spatial awareness (Section 4.2, Analysis of Participation, provides more details).

### 3.2.7 Spatial Ability Testing

The Visualisation of Rotation (ROT) test, a short version of the Purdue Spatial Visualization Test: Rotations (PSVT:R) test, selected for this study was a 10-minute test designed by Bodner and Guay. The original “paper and pencil” test format was suitable to administer in typical classroom situations (Bodner and Guay, 1997), (see section 2.3.1.4 Testing Spatial Ability).

The ROT test contains object rotation questions of various levels of difficulty, presented from easy at the beginning to difficult at the end. In addition to the analysis of scores obtained in this test, this study examined question difficulty and how participants responded to each question; it explored a method of scoring based on difficulty and compared it to traditional scoring to see if it offered any benefit. Knowing which questions students answered allowed for comparisons of answering patterns, e.g. pre and post intervention answering pattern between intervention and control groups or gender subgroups.

One of the goals of the study was to determine the innate spatial ability of the participants; where “innate” refers to the state of an ability without previous training (Sorby and Wysocki, 2012). The following actions attempted to minimise the possibility of previous or recent exposure to the spatial ability test:

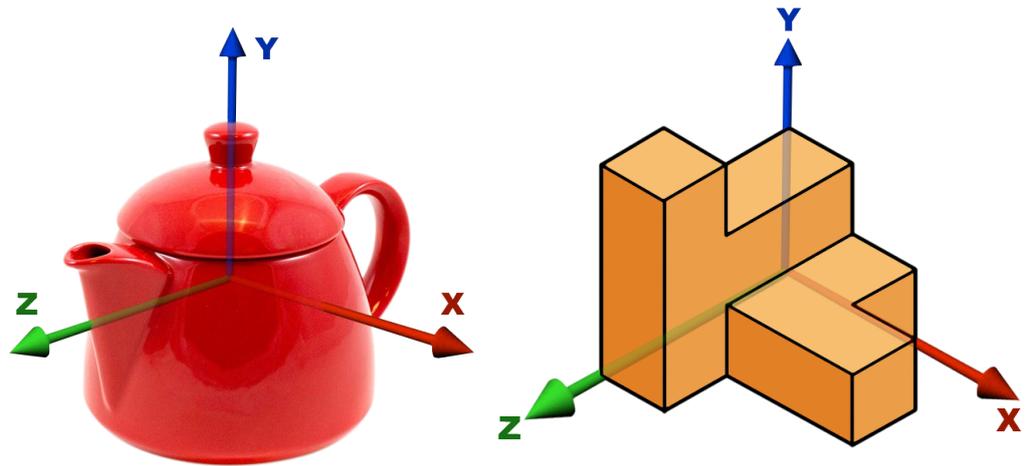
- Selection of an uncommonly used spatial ability test; the Visualisation of Rotation (ROT) test derived from the PSVT: R test.
- Participants took the spatial ability test immediately after the announcement and presentation of the project. Because there was no previous notice, they did not have the opportunity to search for the test or practice.
- The recruitment invitations to join the experiment group had an exclusion clause stating that those who experienced spatial ability testing in the last year before the first test were unable to take part.

Based on the above, it was reasonable to assume that the first test measured the participants' innate spatial ability. Section 5.4 Student Experience with Testing, discusses the possibility of exposure to spatial thinking questions in the pre-admission tests.

### **3.2.7.1 Question Challenges in the Spatial Ability Test**

In the Spatial Ability Test, the degree of difficulty escalates gradually from the first to the last questions. The increase in complexity results from adding a combination of rotations, the direction of each rotation and the dissimilarity of the shape of the object compared to the sample object.

To facilitate comparison, all objects on the test appear oriented in an "orthogonal" position, in which all edges have the same length, showing three sides at all times, front, left and top. Rotation takes place around a given axis. By convention, the horizontal axis (left to right) is X, the vertical axis is Y, and the axis towards the front is Z. All axes originate in a point at the bottom and behind the object (not visible in orthogonal view) (Figure 4)



**Figure 4. Axes of Objects in Orthogonal Position**

Note. The axes of the object originate in the centre of rotation, which is inside the object (e.g. in the kettle) or coincides with the farthest angle from the observer (e.g. in the angular object). When using the axes as references for rotation, the object “spins around” one of the axes. This spin can be “clockwise” or “+” or “anti-clockwise” or “-“. For notation purposes, objects rotate in 90-degree steps. In orthogonal position, the sides of a cube have similar lengths from the point of view of the observer.

To analyse rotation of the objects in the test, a notation system used the axis name as a symbol of a 90-degree rotation around it, e.g. Y = rotation around the vertical axis.

For the direction of rotation, the observer imagines to look down into the axis, an Anti-Clockwise rotation has a “+” sign and a Clockwise one is “-“. For example “+X” indicates a counter-clockwise rotation of 90 degrees around the X-axis (Sorby and Wysocki, 2012).

In this study, complexity depended on the number and direction of consecutive rotations, graded from one to four; and the difference in shape between the question object and the sample object, graded from one to five.

Table 4 shows examples of complexity in rotation sequence and shape difference, and the values assigned.

Rotation Sequence	Difficulty Value	Shape Difference Compared to Sample	Difference Value
			
	<b>1</b> single rotation		<b>1</b> similar cube-based
	<b>2</b> double rotation, same axis		<b>2</b> one oblique facet, otherwise cubic
	<b>3</b> double rotation, different axes/direction		<b>3</b> two oblique facets, otherwise cubic
	<b>4</b> multiple rotation, different axes/direction		<b>4</b> one curved facet, non-cubic
			<b>5</b> curved/oblique facets, non-cubic

**Table 4. Coding of Rotation Sequence and Shape Difference, compared to the sample object in the Spatial Ability Test**

Note. Difficulty takes into account the complexity of the question based on the number of rotations asked, their direction and axes involved.

The first rotation example represents a single-step rotation of 90-degrees, anti-clockwise around the X-axis, which is easy to visualise and has a difficulty value of one. The fourth example represents three sequential rotations in different directions and around different axes, a more complex visualisation that has a difficulty value of four.

The Difference value increases depending on how different the question-object appears from the sample object (presented at the top), e.g. simple angular faces make it very similar and easy to relate, thus having a value of one. In contrast, a combination of inclined and round surfaces make the question object quite different from the sample, therefore having a higher value.

Quantifying complexity with this method presented a problem when adding complex shapes to simple rotations, therefore, the study preferred to quantifying difficulty using participant responses to the questions of the test. Section 3.3.4.4.3 Coding Question Difficulty in the Spatial Ability Test, covers the details. As a comparative reference, both methods of quantifying complexity appear in Appendix 6.

### 3.2.8 Spatial Awareness Training

This study introduced an experimental training programme on spatial awareness, as a form of intervention (see section 2.3.1.8 Spatial Awareness Training).

The spatial awareness methodology, originally designed for engineering (Sorby and Baartmans, 2000) required adaptations to meet the following constraints of this study:

- **Delivery time:** The allocated duration had a restriction of two one-hour sessions. This was a big departure from the original programme, which spanned 10 weeks (40hr) (see section 5.10 Training and Learning for a discussion on this topic).
- **Relevance to anatomy:** The original content catered for engineering students. The selection of important sections for this study followed advice from the original course's author (Prof Sorby) and then analysis of possible relationship with anatomy or clinical practice. Table 5 shows the topics and their relevance, including comments on reasons for inclusion or exclusion.
- **No computer delivery:** In agreement with the original course, hand-drawing, and manipulating geometrical objects was important to enhance learning.

<b>Content</b>	<b>Anatomical Relevance</b>
Surfaces and solids of revolution	Rotation of joints, sections of viscera, sections of solid organs, analysis of transections (Included for visualisation value)
Combining solids	Visualising relations between neighbouring organs, e.g. heart shape imprint on lungs, visualisation of cavities filled with contrast on X-ray or Scans (Included for visualisation value)
Isometric drawings	Evaluation of rotation of geometric instruments, appreciation of 3D images on 2D displays (Included for manual activity value)
Orthographic Drawing	Appreciation of transparency when interpreting or drawing diagrams (Included for manual activity value)
Orthographic projection of inclined and curved surfaces	Recognition of projection of irregular objects, e.g. angular projections of bony surfaces on X-ray (Not included because it is similar to cutting planes)
Flat patterns	Reconstruction from flattened diagrams e.g. visualisation of “fit” of skin flaps (Not included due to little application in anatomy)
Rotation of objects single and multiple axes	Joint motility, locomotion, visualisation of rotation (Included for visualisation value)
Object reflection and symmetry	Analysis of projected images like X-rays or Scans (Included for visualisation value)
Cutting planes and cross sections	Medical imaging e.g. Scans, Ultrasound (Included for visualisation value)

### **Table 5. Anatomical Relevance of Spatial Awareness Training Topics.**

Note. The majority of the spatial awareness training topics had some relevance with anatomy or medical applications. Contents based on Sorby's workbook on developing spatial thinking (Sorby and Wysocki, 2012).

Aside from the above characteristics, the study required to identify a methodology to deliver this content to achieve maximum learning effect in minimum time. The teaching model selected was a “hands-on workshop,” which is a small-group teaching format that involves active participation in a comfortable atmosphere where participants can engage in useful, relevant and stimulating activities (Exley and Dennick, 2004).

According to Vygotsky, in Adey and Serret, the Strategy to teach new material in the most effective manner used three elements of teaching practice to stimulate high-level thinking: Cognitive conflict, Social construction and Metacognition.

- Cognitive conflict refers to presenting students a challenge that is beyond their current processing ability. A small amount of difficulty promotes cognitive development.
- Social construction involves the transmission of knowledge or discussion in a social context with other students.
- Metacognition relates to “thinking about your own thinking,” where students require putting thoughts into words liable to inspection. This way of thinking becomes a useful tool for the future (Adey and Serret, 2010).

These concepts were instrumental in the preparation of the teaching strategy and the materials selected for training in spatial awareness.

Efforts to avoid common problems with small group training included encouraging participation and open communication with students. Throughout the sessions, there were frequent activities after the explanation of a portion of the content, during the activity students would follow a specific short task related to the content discussed. Although there was no expectancy of preparation before the second training session, a short review of the content at the beginning, refreshed the memory of participants (Jaques, 2003). Repetition and practice were also in agreement with cognitive load theory, as they promote automation of schemas acquired by receiving information in small amounts and hands-on exercises (Sweller, 1994).

### **3.2.9 Manual Dexterity Testing**

This study tested manual dexterity in Group A (Intervention), following the change in study design, which meant that participants had the same test before and after training in spatial awareness to determine evidence of impact (see section 3.2.4.1 Research Approach)

Manual dexterity tests measure gross or fine motor skills, e.g. limb movements or finger movements respectively. From the many tests available on the market, the following are of common use in research (Makofske, 2016).

- Minnesota test: Measures gross hand movement and eye-hand coordination.
- Pegboard tests: There are different variants that can measure gross and fine dexterity. For example, the Purdue pegboard test, which combines gross and fine dexterity, the O'Connor tests for finger dexterity and tweezer dexterity and the Crawford Small Parts test that uses tweezers, screwdrivers and a pin-board.

Reports showed that the Crawford small parts dexterity test and the Purdue Pegboard test were better than other popular tests in distinguishing normal from pathological conditions of the hand (Cederlund, 1995). Both tests were reliable and practical, and commonly used for documenting the recovery from neurological injury, and progressive disease (Lafayette Instruments, 2015).

The test selected for this project was the Purdue Pegboard Test, which covers gross movements of arms and hands and fine finger movements. Participants place pegs into holes and build "assemblies" using washers and collars. A modification of this test, inspired by the administration of the Crawford test (Osborne, 1956), added common dissection tools, a straight serrated utility forceps and a curved vascular clamp, to test manual dexterity with instruments.

### **3.2.10 Visual Performance Testing**

Testing visual performance aimed at measuring stereopsis, as this is the area of interest in the study. Stereopsis is dependent on good and similar visual acuity and good alignment of the eyes; therefore, testing stereopsis required verification of both

factors. Participants in Group A (Intervention) underwent testing of visual acuity, ocular motility and stereopsis.

### **3.2.10.1 Visual Acuity**

Visual perception is one of the essential elements in the process of learning, it is also essential for the practice of medicine, from inspection of the patient, to diagnosis and in the delivery of medical and surgical treatment. The measurement of visual acuity provides a numeric value that represents the level of functionality of the visual system (Ohlsson and Villarreal, 2005).

Visual Acuity testing followed a method recommended by the International Council of Ophthalmology and the World Health Organisation to record the best “corrected” visual acuity, meaning wearing prescription glasses if needed (ICO, 1984; WHO, 2003). The method specifies the size of the largest target presented for identification, in this case the largest letter “E” on the printed Snellen chart, which should subtend an angle of 5 minutes of arc (arcmin) (Kalloniatis and Luu, 2007). Section 3.3.8.2 on Visual Function Testing Techniques, provides more detail about the test.

### **3.2.10.2 Stereopsis**

Tests for stereopsis induce a calculated visual disparity by showing a slightly different image to each eye, achieved by using polarising filters (Titmus, Wirt, Randong, Lang tests) or anaglyph glasses (TNO test) (see section 2.3.3 Stereopsis for details).

This study selected the Titmus Stereo Test (Stereo Optical Co. Inc.), most commonly used in practice. This test measures disparities from around 3500 arcsec (equivalent to 59 arcmin), to 40 arcsec. The first target of the test shows an image of a fly standing on a surface, the distance between the wings and the surface in the picture represents the

largest disparity. The ninth target shows the smallest disparity, with one of the circles separated or “elevated” by 40 arcsec from the others (chart held at 40 cm from the eyes), (see section 3.3.8.2 Visual Function Testing Techniques).

### **3.2.11 Academic Performance Testing**

Brighton and Sussex Medical School provided teaching and testing in anatomy to the participants in this study. The school supplied scores of the anatomy components from the Module Exams or “Module summative assessments” (which test the student’s knowledge using Single Best Answer and Short Answer Questions) and the overall results of Year 1 and 2 students from 2015 and 2016. This modular programme in Medicine covers the following subjects:

Year 1:

- 102 - Foundations of Health and Disease
- 103 - Heart, Lungs and Blood
- 104 - Nutrition, Metabolism and Excretion

Year 2:

- 202 - Neuroscience & Behaviour
- 203 - Reproduction & Endocrinology
- 204 - Musculoskeletal & Immune Systems

Modules 101 and 201 - Clinical Practice, carried out throughout each year, has a yearly Observed Structured Clinical Examination (OSCE). The final score results from a calculation of scores from the various modules of each year.

The school office at BSMS provided data containing results of the anatomy portion of the module exams and overall results of Year 1 and 2 during 2015 and 2016. Data only showed student numbers, without personal information, thus remaining anonymous.

### **3.2.12 Data Considerations**

This study followed recommendations by the Universities of Brighton and Sussex, keeping a data management plan to preserve integrity and availability of its information (working documents, data sets, etc.), this included regular backups to local and remote storage within the university (UoS, 2015). Datasets contained anonymous data. Any participant personal data followed destruction at the end of the study (UoS, 2014).

### **3.2.13 Marketing and Advertising**

From a marketing point of view, the testing sessions represented the advertised products while students were the target audience or consumers, who would constitute a “niche market” due to their specific affiliation. Direct marketing techniques used included posters and flyers displayed at BSMS and short presentations of the project at the end of a lecture (below 10 minutes). The study looked at creating a positive experience during testing so that personal recommendation (“word of mouth”) would play a role in increasing recruitment (Masterson and Pickton, 2014), (see section 3.3.10 Logistics for details on Recruitment).

## **3.3 Methods**

### **3.3.1 Introduction**

While Methodology described the background to the techniques used in this study, this section on Methods, presents the statistical plan and the techniques used in testing, training, handling data, statistical analysis and logistics applied in the project.

### **3.3.2 Statistical Analysis Plan**

A Statistical Analysis Plan (SAP) has the goal of aligning the analytical methods to the aims and hypotheses of the study to provide robustness to the research, and prevent it from “giving the right answer to the wrong problem”, considered an “error of the third kind” (Kimball, 1957; Schwartz and Carpenter, 1999; Olivier and Bell, 2013). Some of the benefits of a SAP include providing structure to the study, optimising statistical methods to test hypotheses, transparency of procedures and focus on main areas of interest (Thomas and Peterson, 2012).

The analysis of data will primarily focus on spatial ability scores, describing the characteristics of the collected data for the full cohort and for each group, then, exploring an alternative method of scoring based on question difficulty. Comparisons between the groups will follow, with particular attention to intervention and control. Academic scores will undergo analysis and comparisons between groups, preceding correlation analysis with spatial ability. Finally, the analysis will cover additional testing of manual dexterity and stereopsis in the intervention group.

Although there is no clear rule about the contents of a SAP, there are guidelines suggesting desirable elements as follows (Gamble *et al.*, 2017; NHS, 2018):

1. Administrative information: The name of the study and ethics approval will appear in sections 1.3 Introduction, and 3.2.5 Ethics Committee Approval.

2. Introduction: Introduction and background will appear in Chapter 2 Literature Review. Also, section 3.2.2 Definition of Research Question and 3.2.3 Definition of Hypotheses and Aims.

3. Trial methods: Research design will appear in section 3.2.4 Research Methodology, sample size in section 3.2.4.2 Statistical Power and Sample Size, and 3.3.3.1 Power and Sample Size Calculation. There was no Statistical Interim Analysis planned for this study because there was a time limit for data collection, determined by the period between anatomy exams. Recruitment stopped after the minimum goal of 26 participants in the intervention group was exceeded ( $n=31$ ), giving time to provide training and retest before the post-intervention anatomy exam (determined by BSMS). The timeline for data collection (Feb to May 2016) preceded the final analysis, which started in June 2016, details of which appear in section 3.2.4.3 Project Timeline.

4. Statistical principles: A significance level of 5% (0.05), a confidence interval of 95% and two-sided testing (t-test) were selected following assumption of normality, as detailed in section 3.3.3.1 Power and Sample Size Calculation. Multiple statistical testing, performed initially to confirm results obtained with non-parametric methods, will not appear in the current format of the thesis due to the assumption of normality. The analysis population, planned to follow organisation in groups, will undergo statistical analysis relevant to each group, as detailed in section 3.3.5 Statistical Analysis Technique and section 4.2.1 Participation and Group Structure.

5. Trial population: Eligibility criteria will appear in section 3.3.3 Research Technique. Recruitment, involving selection of participants will appear in section 3.2.4.1 Research Approach, and randomisation in section 3.3.3 Research Techniques. Baseline characteristics of the population or descriptive statistics will appear in section 3.3.5 Statistical Analysis Technique. Because skewness was high in most data, summaries will include median, interquartile range and range, as measures of central tendency.

6. Analysis of data: The outcomes will be scores in the units of the measurement, converted to percentage for comparisons. Details on handling data will appear in section 3.3.4 Data Management and include data preparation, clearing of missing data and coding for processing. The analysis method will include parametric tests, described in section 3.3.5.1 Processing Data in SPSS and the summaries will show basic descriptive statistics, confidence intervals and p-values. The rules to handle missing data appear in section 3.3.4.3 Data Preparation and Clearing; the method used for coding or replace missing values in SPSS appears in section 3.3.4.4.1 Coding Missing Data. Although there was no expectancy of harm or adverse events related to the intervention, participants received advice to contact the researcher if they occurred; as there were no reports, no actions took place. The main statistical software for analysis will be SPSS version 24 and 25, additional programmes will aid sample size calculations e.g. G\*Power, PiFace. Other methods of analysis will include references.

### **3.3.3 Research Technique**

This section covers the specific techniques used in the collection and statistical processing of data for this project.

#### **3.3.3.1 Power and Sample Size Calculation**

Following sub-section 3.2.4.2 on Statistical Power and Sample Size, Table 6 shows an extract of Cohen's tables designed to determine the Sample Size as a function of three values; it also demonstrates that effect size has a large impact on sample size.

<b>*Significance <math>\alpha_2 = 0.05</math></b>			
<b>Power</b>	<b>Effect size (<math>r</math>)</b>		
	0.10	0.40	0.50
0.70	616	37	23
0.80	783	46	28
0.90	1047	62	37

**Table 6. Sample Size ( $n$ ) as a Function of Significance, Effect Size and Power.**

Note: This table presents an example of the relationship between significance, effect size, statistical power and sample size. This sample includes selected power values to demonstrate the large impact of effect size.

\*:  $\alpha_2$  or Alpha2 significance refers to a two-tailed test, which considers a difference above or below a reference point.

The relationship most relevant to this project in Table 6 corresponds to an outcome with an effect size of  $ES = 0.50$ , a significance level of  $\alpha = 0.05$  (for a two-tailed test) and a power of 0.80, which point at a sample size of  $n = 28$ . This means 28 in the intervention group and 28 in the control group, making 56 participants in total (Cohen, 1988).

To verify the sample size suggested by Cohen's table, the project used various software tools. The first one relied on "G\*Power v3.1.9.2," a programme created by Franz Faul et al (Faul *et al.*, 2007). The output pointed at a sample size of 26, based on the parameters mentioned above. Using this software, the statistics team of the doctoral school of the University of Brighton suggested a minimum number of 24 participants, which yielded a statistical power of 80%.

Other statistical tools calculated sample size taking into account the population of interest, which in this project consisted of 268 medical students, resulting from the addition of 132 Year-1 and 136 Year-2 students. These tools relied on confidence interval rather than power for calculations. The first tool of that type was the online "Sample Size Calculator," offered by Creative Research Systems, which based on a reference population of 268, with a confidence level of 95% and a confidence interval

20, suggested a sample size of 22 participants (CRS, 2015). The confidence interval refers to the margin of error the researcher is willing to accept. The lower the number, the higher the sample size. The second tool, "PiFace", set to Confidence Interval for Proportion Analysis, using the parameters: Finite population of  $N = 268$ , Worst-case  $\pi: 0.5$ , Confidence 0.95 and Margin of error 0.09297, suggested a sample size of 79 participants (Lenth, 2009).

The average sample size provided by the various tools was 26, but considering the student population, the average climbed to 52, which this study regarded as an upper limit, aiming for the lower value as a minimum goal. Therefore, the sample size of 26 in each group, expected to yield 80% power to detect an effect size of 0.5 with a confidence level of 95% and confidence interval of 20, using a comparison test (t-test) with a 0.05 two-sided significance level.

Based on the above, the study hoped to recruit between 26 and 50 participants for the intervention group, which added to a similar number for the control group, gave 52 to 100 as the total of participants sought for the experiment.

### **3.3.3.2 Recruitment Technique and Sample Selection**

There were two main types of recruitment techniques.

The first one relied on a short presentation of the project (3 min) and a direct invitation to take part in the spatial ability test immediately after, which took place at the end of a lecture, with approval of the lecturer. As participation was voluntary, some of the students chose to leave the room but most remained. This technique was effective and therefore repeated for the second test, two months later (for Y1 and Y2).

The second recruitment technique aimed at building up the experiment group, it used posters and flyers displayed on bulletin boards and distributed by hand at BSMS.

Interested students either contacted the email posted on the advertisements or added their email addresses to a list collected by the researcher. As a response, they received an information pack with details of the project and instructions, should they decide to take part of the experiment. Participating in the experiment group meant attending four one-hour sessions, the first and last for further testing (vision and manual dexterity) and the middle two for training in spatial awareness. As an incentive, participants in the experiment group received payment of £30.

Following the research design change from “experimental” to “quasi-experimental” as described in section 3.2.4.1, Research Approach and Design, all participants in the experiment group formed the intervention group (Group A,  $n = 34$  before clearance) from a population of 268 students, the result of adding 136 students in Year-2 to the initial 132 in Year-1. In this fashion, the intervention group exceeding the minimum “participant number goal” of 26, which related to a statistical power of 80%. The control group selected after the second spatial ability test from students who also took the first test and were not in the intervention group, reached a number of 75 (Group B). The remaining students who only took the first test and were not on the intervention or the control group, formed the “single test” group (Group C,  $n = 90$ ). Section 4.2.1, Participation and Group Structure, describes the groups in detail.

### **3.3.3.3 Participant Eligibility**

The criteria used to determine participant eligibility for the study aimed at selecting a representative sample of the population of interest, medical students in the first two years of medical training (first and second year students learn anatomy and take module exams that include anatomy questions), with no recent exposure to spatial ability (although this was assumed, section 4.8. Limitations, discusses a potential source of exposure).

### **Participant Inclusion Criteria**

- Current student of the undergraduate degree in Medicine attending year 1 or 2 at BSMS in the 2015-2016 term.
- No experience of tests or training in spatial ability in the last year (assumed).
- Full mobility of upper limbs for manual dexterity test.
- Binocularity (useful vision simultaneously with both eyes) for stereopsis test.

### **Participant Exclusion Criteria**

- Those not complying with the inclusion criteria.

#### **3.3.3.4 Random Allocation to the Intervention**

The initial design of the research project was as a randomised control trial. Due to ethical reasons, participation of students had to be voluntary, which precluded a random selection from the pool of medical students. Consequently, students had to join the experiment voluntarily, mostly as a response to recruitment efforts. Section 4.8 Limitations, describes the induced bias (self-selection) that this modality may bring.

The method of randomisation planned to assign participants to receive training in spatial awareness used the “select cases” option of SPSS, which when set to 50%, can randomly assign a given number of participants into one of two groups of the same size, making a simple randomisation with 1:1 allocation of cases.

Section 3.2.4.1 Research Approach, details the reasons that led to the abandonment of randomisation and change to a quasi-experimental design. The intervention group chose to participate, while the control group relied on students volunteering to take the retest in spatial ability; both groups with a potential of self-selection bias, independent of the researcher’s intervention.

### **3.3.3.5 Participant Anonymity**

All participants received a written explanation of the project with emphasis on the requirement of anonymity to ensure confidentiality. Once they were satisfied with the information, they signed the consent form, which re-iterates the need of anonymity indicating that personal data will not enter the study. Because results were not traceable back to them, the only chance they had to ask for their results was immediately after the tests. Appendix 2 shows these details on the consent form.

After the signature of the consent form, participants' records received a random number to identify their input within the study. The online tool "Random Sequence Generator," found at [www.random.org/integers/](http://www.random.org/integers/), provided a set of random numbers (Haahr, 2015).

Participants also entered their student number, which allowed following up along the study to compare their progress and match with academic assessment scores provided by BSMS school office, who also removed personal data when making information available for research. The random number allocated at the beginning of the participation would eventually replace the student number, providing a further level of anonymity.

All personal information, including names and email addresses, used for initial communication and booking of testing sessions underwent deletion from all digital storage at the end of the project.

### **3.3.4 Data Management and Statistical Method**

Data management covers collection, handling, clearing and coding of data in preparation for statistical analysis.

### **3.3.4.1 Data Collection Technique**

The method of choice to collect data for the study was using paper forms. It was practical and quick, and due to the relatively small number of participants, appropriate for the study. The main disadvantage of this technique was readability. Calligraphic styles led to misinterpretation in a few cases, due to difficulty in deciphering handwritten text or inability to find participants in academic records.

During the testing sessions, each participant had a “data sheet” created for this project as a single document to follow his or her performance. This paper contained an area for generic details, a basic questionnaire and a table for all tests scores (see Appendix 4 for a sample of the data collection sheet).

Once transferred into digital format, this information did not contain personal details and was therefore not encrypted, only password-protected. Microsoft Excel Spreadsheets were the preferred first point of data entry; they allowed easy visualisation and the use of formulas to calculate values like the Mean, Standard Deviation providing initial values for analysis and comparison.

### **3.3.4.2 Data Handling**

Data management comprises four types of information: Personal and non-personal data, in Physical or Digital format.

Physical records (questionnaires, answer sheets, data collection sheets and consent forms) resided in a restricted access zone of the BSMS Medical Research Building. Digital information resided on a password protected PC with backups on an external drive and storage servers provided by the University of Brighton.

### 3.3.4.3 Data Preparation and Clearing

The goal of data preparation is to reduce errors and redundancy before processing or analysis. Most of the cases that qualify for removal from the database do so because of insufficient data, usually due to participant failure to complete fields in a questionnaire or non-compliance with the requirements of a study.

Data collection involved testing large numbers of participants at the end of a lecture in the classroom, where monitoring of the test was minimal. Consequently, several issues needed verification, falling into one or more of the following categories:

1. Single participation in the spatial ability test: Most students took the test on the first round and some of them took the re-test as well (3 months apart). A number of students took the test for the first time on the second round; therefore, if participants had a single score, this data went to the column "first attempt", regardless of when they took the test.
2. Incomplete participation in the intervention group: Two of the 34 students who signed up for the experiment only attended one of the four sessions; one defaulted to the control group and the other to the single test group. One student attended three sessions, including training and was therefore included in the intervention group for spatial ability but excluded from manual dexterity calculations.
3. Failure to complete fields of the test: Nine participants failed to enter their student number on their answer sheet, two of whom had legible signatures that lead to their identification; the rest remained without number, which excluded them from academic performance calculations. Three of them also failed to enter their gender; they appeared as "missing data", leading to exclusion from the study. If only gender or age were missing, the medical school office provided this information to complete those records.

After clearing data, this information required preparation to facilitate processing in the different statistical tools selected for this project.

#### **3.3.4.4 Data Coding**

Data coding was an important process that helped with efficiency and consistency in processing data, particularly when dealing with different types of software and formats. It simplified the entry of data into spreadsheets and improved its compatibility with SPSS e.g., replacing nominal values with numeric values, as in gender, where the values Male and Female were replaced by numbers 1 and 2 respectively.

##### **3.3.4.4.1 Coding Missing Data**

Missing data may introduce bias, affecting reliability and validity, particularly in clinical trials, where it may relate to the intervention. In this study, the missing data had no relation to the intervention (e.g. missing personal information), considered as “Missing at Random”, which does not require adjustment to the data set (Thabane *et al.*, 2013).

Missing data management followed the SPSS facility to ignore cases for specific calculations. The properties section for each variable offers the option to define a code that represents missing values. When not set, SPSS assumes that all data is valid and missing values may affect the results. The recommendation is to use a code that is outside the range of expected values for that variable; for example, as adopted by the study, one of the common selections is “99” for scores that lie below this number (Field, 2012).

#### **3.3.4.4.2 Coding Test Score Variables**

A set of Coding Rules prepared for this study facilitated this process and served as a reference in the creation of new data documents, during transfer and analysis. Naming of the variables followed the criteria specified by SPSS (Hinton *et al.*, 2005; Field, 2012; IBM, 2016a). See Appendix 5 for a sample of the rules applied to the dataset.

Before transferring participants' scores from the Data Collection Forms, a second verification of calculations took place to ensure accuracy. Once in an Excel spreadsheet, raw scores (in the units of the tests) underwent further conversion into percentages, in preparation for comparisons in SPSS.

#### **3.3.4.4.3 Coding Question Difficulty in the Spatial Ability Test**

The analysis of question difficulty mentioned in section 3.2.7 Spatial Ability Testing, used scores from the first spatial ability test, which included all participants, and it required the following sequential elements:

1. Determining question difficulty: Achieved by selecting tests where the participant attempted all 20 questions ( $n = 159$ ), and determining how many students answered each question. A ratio based on this summary, provided a value of correct answers for each question or "how easy was each question". Calculating the inverse, provided a value of "how difficult each question was".
2. Coding rotation: Achieved using a nomenclature for each 90-degree rotation required to solve each question on the test. This graphic code shows the direction and axis for each step.
3. Assigning a difficulty value or index to each question: Achieved by sorting the test questions based on their difficulty index and colour coding them to denote

five ranks. Comparison by ranks also helped the analysis of performance. See Appendix 6 for a sample of Rotation Coding and Rank Colour Coding.

4. Scoring tests using difficulty values: The score resulted from the sum of values of each correct answer (maximum value 66.816, equivalent to 20 questions). These scores underwent conversion to percentage for comparisons.
5. Comparing difficulty-based scores to traditional scores: Achieved using SPSS tools.

This system used the participant responses as the basis of the calculation of a difficulty index based on the following premise: "Questions answered correctly by most participants were easy, while those answered correctly by few participants were the most difficult." Most of the calculations used Microsoft Excel features with comparisons performed in SPSS using parametric formulas and Chi-Square.

See section 4.3.6 Measure of Difficulty of the Spatial Ability Test, for results of this experimental analysis.

### **3.3.5 Statistical Analysis Technique**

The selected software for analysis was IBM's Statistical Package for the Social Sciences (SPSS) version 23 (IBM, 2016b), which required importing data from Microsoft Excel spreadsheets that were the primary software for data collection or transfer from physical data. A table of "coding rules," designed in preparation for this step facilitated this action and ensured consistency within the study (see Appendix 6).

#### **3.3.5.1 Processing Data in SPSS**

In this study, the majority of data processing followed a similar sequence of calculations, which influenced the order of reporting results. The initial steps usually

consist in the calculation of descriptive statistics, which provide key values for interpretation, including:

- Measures of Central Tendency: Mean (Arithmetic Average or *M*), Median (Value at which half of the cases fall above and below or *Mdn*) and Mode (Most frequent value in the series).
- Measures of Shape or data distribution: Skewness (Positive values show accumulation of data to the left, and negative, to the right) and Kurtosis (positive values represent steep and negative, shallow).

In this study, a number of the datasets had non-normal distribution with skewness of various degrees, therefore, the median rather than the mean, provided a better representation of the central tendency of the given set of scores. The median is less susceptible to changes by skewed distributions, extreme scores and the presence of outliers; there is a quicker convergence to normality when using medians at smaller sample sizes than when using means (Wilcox, 2010). In reports, the Inter Quartile Range (*IQR*) or “*range*” should follow the Median (*Mdn*), as much as the Standard Deviation (*SD*) should follow the Mean (*M*) (APA, 2013; Salkind, 2017; Field, 2018).

For the evaluation of normality, SPSS offers the “One Sample Kolmogorov-Smirnov” test (K-S test) (accessible via Analyse and Explore), which indicates normality of distribution when the result is  $> 0.05$ . Following the Central Limit Theorem (proposed by de-Moivre in 1738 and revived by Laplace in 1812), which indicates that if a sample size is sufficiently large (widely accepted  $n = \geq 30$ ), plotting the means of multiple samples would follow a normal curve. The assumption of normality is possible despite the appearance of the sample distribution, which is particularly true when outliers are few (Wilcox, 2010; Dunbar, 2011; Field, 2012; LaMorte, 2016).

Initially, the datasets reported as non-normal by the K-S test, underwent non-parametric testing (Mann-Whitney and Wilcoxon’s rank-sum tests). In view of the

sample size (Intervention  $n = 29$ , Control  $n = 75$ ), almost reaching the minimal value to assume normality, according to the Central Limit Theorem, the preferred method of analysis was with parametric tests, e.g. T-Test.

The comparison of scores underwent testing using the T-Test. When comparing scores of different individuals, the Independent Samples T-Test was the modality used, while comparing the performance of the same individuals on different occasions required the paired-samples T-Test. The T-test assumes normality, but when this is in doubt, it is possible to use “robust methods” to compensate for non-normal distribution and outliers. The most common is “Bootstrapping,” which samples data to mimic normality. The confidence interval selected was “Bias corrected accelerated” (BCa) set to 95% (Salkind, 2017; Field, 2018).

Processing Categorical Data like Gender and Hand Preference provided the opportunity to use Pearson’s Chi-Square, which showed to be adequate, producing a cross tabulation table with details of the comparisons.

The Academic Performance dataset, based on scores provided by BSMS school office, followed parametric calculations assuming normality in large datasets.

### **3.3.6 Reliability and Validity**

An introduction to test reliability and validity is required to understand their importance in the project. Those qualities are pertinent to the instruments (tests) used to measure the performance of participants, and therefore influence the quality of data collected and analysed, and consequently the results of the study.

The reliability of a test refers to its capacity to measure something consistently; it correlates inversely to the amount of error in the observed score and may be of various types:

1. Test-retest Reliability: Evaluates repeated tests to determine if they are reliable over time.

2. Parallel Forms Reliability: Evaluates if versions of a test are equivalent in context. This and the previous type rely on the Pearson's product-moment correlation formula.

3. Internal Consistency Reliability: Evaluates if items in the test assess a single dimension measured by Chronbach's alpha, Split-half reliability, Kuder-Richarson 20 and others.

4. Interrater Reliability: Evaluates if two observers rate an outcome consistently and in agreement, measured by the number of agreements divided by the possible agreements.

Reliability values are positive and range from low at 0, to high at 1.00 (Field, 2012; Salkind, 2017; Field, 2018).

The validity of a test refers to its capacity to measure what it claims to measure, and there are three types:

1. Content Validity: Determines if a sample of items reflect all items in a given topic based on an expert opinion.

2. Criterion Validity: Determines the correlation of scores with other measure already valid in the present (Concurrent Criterion Validity) or in the future (Predictive Validity).

3. Construct Validity: Determines the relationship of scores with a theorised outcome that reflects the topic under testing (Field and Hole, 2003; Salkind, 2017; Field, 2018).

Although a test can be reliable, it may not be valid because it could be testing something different to what it intended to measure. In contrast, a valid test requires reliability prior to obtaining validity. It is advisable to look for tests with proven reliability and validity (Salkind, 2017).

### **3.3.6.1 Reliability and Validity of the Spatial Ability Test**

#### **Reported Reliability and Validity of the Visualization of Rotations Test (ROT)**

The reliability of the ROT test, measured by its authors using the Kuder-Richardson 20 (KR-20) test and the Split/Half reliability coefficient test, was internally consistent. The former, correlates each item score with the total, while the latter, splits the scale set randomly and compares the halves. The report, based on 4,800 university students of Health Science and Engineering, produced an average of 0.79, a high reliability. Validity, compared to the Shepard-Metzler test, was 0.61 (construct validity). In addition, the Rotation test had a poor correlation with tests solvable by analytical processes, like the Minnesota Paper Form Board (MPFB) test (construct validity 0.25). This suggests that the ROT test required a similar type of ability (spatial) than the Shepard-Metzler test and that it was not solvable by simple analysis like the MPFB, which places the ROT test at similar level as the Shepard-Metzler (Bodner and Guay, 1997; Salkind, 2017; Field, 2018).

#### **Reliability of the Spatial Ability Test Using Data Obtained in this Study**

Although one of the most common measures of reliability uses the Cronbach's Alpha formula (calculated later in this section), Bodner and Guay used the KR-20 test of reliability (see above). To compare with their report, the data in this study underwent evaluation of internal consistency using the KR-20 test.

This method starts by calculating percentages of students who answered correctly and incorrectly each question; the multiplication of those values results in a product, which added to that of the other questions results in a transient total that is part of the final formula. The KR-20 of the spatial ability test was 0.75, representing a high internal reliability (Zaiontz, 2013). This is comparable to the average of 0.79 reported by

Bodner and Guay (Bodner and Guay, 1997). Instead of the KR-20, SPSS offers Cronbach's Alpha to test internal reliability, which reported 0.748, showing high reliability.

The relationship between validity and reliability has a numerical formula, by which "the maximum level of validity is equal to the square root of the reliability coefficient" (Salkind, 2014). Based on the results of the KR-20 (similar to Cronbach's Alpha) this calculation was:  $\sqrt{0.7475} = 0.8645$

This value is close to 1.0 and therefore represents a high validity, which confirms Bodner's report.

### **3.3.6.2 Reliability of the Manual Dexterity Test**

Because participants in Group A (Intervention) took the same Manual Dexterity test twice, their results formed the basis of the calculation of "Test-Retest Reliability."

The scores of the manual dexterity tests were significantly correlated for Group A (Intervention)  $r = 0.714$ , 95%BCa CI [0.521, 0.847],  $p < 0.001$ . This high correlation represents a high Test-Retest Reliability.

The calculation of validity was the following:  $\sqrt{0.714} = 0.844$

This result indicates a high validity, which together with the high reliability; make the manual dexterity test ideal for this task.

### **3.3.7 Analysis of Academic Data**

The analysis of Academic Performance focused on the Human Anatomy scores obtained by Y1 and Y2 students, who learn different modules at BSMS (see section

3.2.11 Academic Performance Testing). Table 7 shows the relationship in time between the events of the study and those of the academic programme at BSMS.

Timeframe		Event
November 2015	⇒	Knowledge Test for Module 2 (KT 102 / 202)
February and April 2016	⇒	First Spatial Ability Test (Y1 February, Y2 April)*
March	⇒	Knowledge Test for Module 3 (KT 103 / 203)
May	⇒	Training in Spatial Awareness
May and June	⇒	Second Spatial Ability Test
June	⇒	Knowledge Test for Module 4 (KT 104 / 204)
July 2016	⇒	Final score of the Year

**Table 7. Timeline of Project Events and Related Academic Tests for Year 1 and 2 in Medicine**

Note: \* The first test in Spatial Ability for Year 2 given after approval of ethics amendment to include this year in the project. This test took place shortly after their anatomy module 203 examination.

The two spatial ability test and the spatial awareness training intervention took place before the fourth module examination. Signs of any impact of the training intervention were likely to appear in that “post-training” exam and perhaps in the final score of the year. To determine the trend in knowledge scores for Group D (Experiment), the study selected three anatomy scores for analysis: Module 3, taken before training, Module 4, post-training, and the final score of the year as a reference.

To homogenise data, all scores underwent transformation to percentage. For descriptive reasons, a generic reference of modules 103 and 203, adopted the grouped name “x03,” (same treatment for x04). Although blending modules was convenient for statistic calculations, the difference in content of each module, undermined the validity of the analysis of academic scores (see section 4.8.2 Academic Score Limitations).

Following descriptive statistics and clearing of outliers, comparisons of scores used the One-way Repeated Measures ANOVA test, which controls for Type-1 error using the Bonferroni method. To analyse the differences between neighbouring scores, e.g. x03 to x04, the test of preference was the Paired Samples T-Test. The analysis of score differences focused on Group D (Experiment) and comparisons included Group E (Full Cohort).

Calculating correlations of anatomy modules with spatial ability assumed normality to use the Pearson's product-moment correlation test. Results required verification using the non-parametric Kendall's tau test because Pearson's test tends to be invalid with non-linear data (Field, 2013; Field, 2018).

### **3.3.8 Testing Techniques**

This section presents details about the techniques used for specific procedures, starting with spatial ability, followed by procedures used in the testing sessions.

#### **3.3.8.1 Spatial Ability Testing Technique**

The study used the Visualisation of Rotation (ROT) test, developed by Bodner and Guay (Bodner and Guay, 1997). Prof Bodner kindly made the test available to this study and it underwent digital enhancement to improve quality.

On the day of the test, participants received a booklet with 20 questions and a separate answer sheet with fields for student number, gender and year of study. They had to mark a cell corresponding to their selected answer. This method allowed reusing of the question booklets and marking using an answer template.

The first test took place on 1st February 2016 for Year 1 ( $n = 106$ ) and on 12th April 2016 for Year 2 ( $n = 69$ ), testing 175 participants. The re-test was on 17 May 2016 for Year 2 ( $n = 31$ ) and on 19th May 2016 for Year 1 ( $n = 91$ ), testing 122 participants. The time allocated for this test was 10 minutes, which proved to be sufficient as most participants attempted all questions.

### **3.3.8.2 Visual Function Testing Techniques**

Testing visual function started with testing visual acuity, followed by stereopsis.

For visual acuity, the participant, wearing her/his latest correction (spectacles or contacts) sat on a chair facing the Visual Acuity chart. The chart was 3 metres away, with the 6/6 vision line placed at eye-level (Figure 5). The examiner asked the participant to read aloud the letters on the chart starting with the largest while covering one eye with an occluder. The line with all the smallest characters correctly identified indicated the score (marked next to it on the chart). A reading test followed using a reading card held by the participant at 40cm from the eyes, scored in the same manner. Normal visual acuity at distance is 6/6 metres and for near 40/40 centimetres, which both values equate to 1.0 in decimal nomenclature.

For stereopsis testing, participants wore polarised goggles over their current spectacles if needed, holding the Titmus Stereopsis Test booklet or “Fly Test” at 40 cm (Figure 5). The examiner would ask if the participant noticed depth differences in the various targets on the booklet. The smallest correctly detected indicated the score (not marked on the booklet). Normal stereo acuity is 40 seconds of arc.



**Figure 5. Visual Acuity and Stereopsis Tests**

Note. For visual acuity (top), the Snellen chart is at 3 metres in front of the participant, who covers one eye with an occluder.  
For stereopsis (bottom), the participant wears polarised goggles to describe the 3D images presented on the Titmus Stereopsis Test (Stereo Optical Co. Inc.).

### **3.3.8.3 Manual Dexterity Testing Technique**

The Manual Dexterity Pegboard Test consists of a rectangular board with a row of 25 perforations running at both sides of the midline. It has four concavities or “cups” carved across the top of the board to hold the small metal parts required for the test.

During the test, the cups that lie at both extremes hold metal pegs, while the cups that lie in the middle contain either washers or collars. The collars go on the right for right-handed and on the left for left-handed participants. Following scripted instructions, the examiner would guide the participant through the sub-sections of the test. Using the dominant hand first, the participant would insert pegs into the holes along the same

side of the board, then tests the other hand, followed by both hands simultaneously on a third section. The number of pegs placed in 30 seconds provided a score for each section. Finally, the participant would be instructed to build “assemblies” using pins, washers and collars, alternating hands. The number of parts correctly placed in one minute, would count as the score.

### **3.3.9 Spatial Awareness Training Method**

After the change in research design to quasi-experimental, there was no random allocation, all participants who signed up for the experiment became Group A (Intervention) and received training in spatial awareness (see section 3.2.4.1 Research Approach). Training took place at BSMS and involved two one-hour sessions, a week apart.

Each presentation lasted 50 minutes and started with an introduction of the content and short review of the previous one (Table 8 shows the content of each session). There was a pause for questions and/or practice after the delivery of every part of the content. Practice consisted of 5-minute exercises on the same topic following a workbook. There was encouragement to use plastic blocks to reproduce the objects in the questions and practice rotation, drawing or isometric perspective.

The workbook contained selected exercises from Sorby’s manual, with adaptations to follow the presentation (Sorby and Wysocki, 2012). Participants’ comments about the presentations were positive, particularly those related to the use of building blocks to help visualise exercises. Appendix 3 shows a sample of the workbook.

Session	Topic
Session 1	<ul style="list-style-type: none"> <li>• Definition and examples of spatial ability</li> <li>• Importance of spatial ability in real world and in anatomy</li> <li>• How to stimulate your spatial ability: observe, organise, use your hands and practice to turn ability into skill</li> <li>• Study shapes: Revolution of lines and solids, Isometric drawings, Orthographic projection, Rotation part 1</li> <li>• Hands-on exercises after each section</li> </ul>
Session 2	<ul style="list-style-type: none"> <li>• Reminder from last session</li> <li>• Rotation of objects part 2</li> <li>• Reflection and symmetry</li> <li>• Cross section</li> <li>• Medical application</li> <li>• Nurturing spatial skills</li> <li>• Summary</li> </ul>

**Table 8. Content of Spatial Awareness Training Sessions**

Note: During each session, participants solved exercises in their workbooks building replicas of the objects with plastic blocks to analyse rotation sequences.

### 3.3.10 Logistics

The organisation and implementation of activities in the study involved recruitment, group testing and personal testing sessions.

#### Group Testing

The spatial ability test was suitable for group testing. Due to time limitations of medical students, it was more practical to combine presentation of the project and testing in one event (see sections 3.2.6 Participants in the Study, and 3.2.7 Spatial Ability Testing for details). This event, calculated to last 15 minutes, took place at the end of a lecture, with agreement of the lecturer. Organising this event required identification of a lecture with potential of good attendance, followed by contact with the lecturer to request permission, both tasks facilitated by the school office at BSMS. Four group events took place, two for each year to cover the test and re-test in spatial ability.

## **Personal testing**

Participants in Group A (Intervention) attended two sessions of testing, before and after training in spatial awareness. The nature of the test dictated the need of a one to one testing modality. To cater for the 31 participants who initially joined the experiment, the study required 62 one-hour sessions at BSMS. To allow participants select the testing session based on their availability, the study booked rooms for 3 hours every weekday for two months (April and May 2016). Participants received a link to an online booking facility (Doodle.com) and were able to select a single session in anonymity. The researcher confirmed the choice via email and arranged the meeting.

In the first personal testing session, participants received an explanation of the contents of the session and proceeded with testing, which covered Stereopsis (preceded by Visual Acuity testing) and Manual Dexterity. The session, which lasted 45 minutes, ended with a discussion of the participant's performance in relation to reported standards. The second session lasted 30 minutes because it only tested manual dexterity. Sections 3.2.9 Manual Dexterity Testing and 3.2.10 Visual Performance Testing present details on both tests.

## **Testing Kit and Testing Room Layout**

To facilitate testing with consistency, the researcher assembled a testing kit, which was easy to transport, flexible to adapt to different room configurations and contained all the instruments needed to test vision and manual dexterity. It mainly contained:

For visual performance: Snellen visual acuity chart (for 3 metre testing), near vision reading chart, Titmus stereopsis test (with polarising goggles), occlusors, target sticks and related tools.

For manual dexterity: Purdue pegboard test, manual dexterity testing script, a stopwatch and related tools.

Figure 6 shows the contents of the testing kit and examples of the testing environment.



**Figure 6. Testing Kit and Samples of Testing Room Layout for Personal Testing Sessions.**

Note The testing kit included accessories for general eye examination in addition to the Titmus “Fly” Stereopsis test (Stereo Optical Co. Inc) and the Purdue Pegboard Manual Dexterity test (Lafayette Instruments). There was flexibility in setting up the room as long as the visual acuity testing distance remained constant (3 metres) and there was good, diffuse illumination.

## **Advertisement and Recruitment Technique**

Considering that the audience was a niche market of university students, advertising targeted this group in two stages:

- At the time of the first spatial ability group test, the researcher presented the project and invited students to join the experiment group when advertised.
- A printed advertising campaign at BSMS followed, with posters and flyers announcing the recruitment of participants for the study. The posters included eligibility criteria and instructions to contact the researcher for information (see section 3.3.3.3. Participant Eligibility).

Recruitment stopped after signing up the expected sample size ( $n = 34$ ), with sufficient time to schedule training and re-testing before the next anatomy module exam.

## CHAPTER 4 RESULTS

### 4.1 Introduction

The project focused on a population of undergraduate medical students attending year 1 and 2 at BSMS. Testing included a repeated test of spatial ability, which took place at the start and at the end of the data collection stage of the study, which spanned from February to May 2016. Between these two tests, a sub-group of participants (Group A: Intervention), underwent an initial set of tests of Visual capacity and Manual dexterity, followed by a training intervention on spatial awareness and subsequent re-test of Manual Dexterity before the second spatial ability test. In this manner, the study obtained scores in spatial ability and manual dexterity before and after the intervention for Group A (Intervention); test and re-test scores in spatial ability for participants who took both spatial ability tests but did not receive training (Group B: Control) and scores of a single spatial ability test for those who only took one spatial ability test (Group C: Single Test).

The school office at BSMS provided scores in Anatomy extracted from module exams that took place at the beginning and after the study, and the end-of-year or final scores.

This chapter presents results of an analysis of the population of participants first, followed by analysis of results in spatial ability, anatomy, manual dexterity, stereopsis and correlations between all those scores. Each section ends with a summary of findings.

## 4.2 Analysis of Participation

### 4.2.1 Participation and Group Structure

From the population of registered medical students in Year 1 (Y1 = 132) and Year 2 (Y2 = 136), 194 participated. Section 3.2.4.1 Research Approach and Design describes limitations in the recruitment of participants and the adoption of a “quasi-experimental” design.

The following sequence of events helped define the participating groups:

#### 1. First Spatial Ability Test (Pre-intervention measurement):

Medical students in Year 1 and 2 had the opportunity to take the first spatial ability test (SA-T1) at the end of a lecture (each year separately). In that first test, 124 students from Y1 (94% of their course) and 70 from Y2 (53% of their course) accepted to take the test, making a total of 196. Because they volunteered, the selection method corresponded to “self-selection” (see section 3.3.3.1 “Population Sampling” for details).

#### 2. Intervention:

After the first test, a recruitment campaign invited students to join the experiment. From the 81 students who responded to the invitation, those from Y2 were more numerous than from Y1 (49 and 32 respectively). All students received a reply with information and instructions to book further testing should they agreed to join the experiment.

Forty-seven students (58%) did not continue communications, while 34 (42%) decided to join (18 from Y1 and 16 from Y2). This group became Group A (Intervention) (see section 3.2.4.1 Research Approach and 4.8.1 Research Modality and Sample Size Limitation for details on lack of randomisation and self-selection bias).

### **Participant Clearance in the Intervention Group**

Two of the 34 students in the intervention group did not attend the spatial ability training sessions; one of them took the pre and post-intervention spatial ability tests, and therefore moved to Group B (Control), while the other only took the first test, and moved to Group C (Single Test). Another student did not take the post-intervention test, therefore went to Group C. Two students who attended training, but only took the second spatial ability test, underwent exclusion from the spatial ability section of the study. Group A (Intervention) ended up with 29 students with complete sets of scores used in the analysis of spatial ability performance. Of the 29 participants in the intervention group, 16 were from Y1 (57%) and 13 from Y2 (43%) (Table 11).

### **3. Second Spatial Ability Test (Re-test or Post-Intervention measurement):**

Using the same method as in the pre-intervention test, the post-intervention spatial ability test took place approximately three months later. From 104 students who volunteered to take this second test (Female: 61 [59%]; Male: 43 [41%]), 73 were from Y1 (Female: 43 [59%]; Male: 30 [41%]) and 31 from Y2 (Female: 18 [58%]; Male: 13 [42%]). This number included 29 participants from Group A (Intervention), who underwent training, and 75 who did not train but took both spatial ability tests, therefore regarded as Group B (Control).

### **Definition of participating groups**

Participation in all groups was voluntary and, as clarified to participants, unrelated to academic curriculum.

**Group A (Intervention):** Participants in Group A received training in spatial ability, took the pre and post-intervention spatial ability tests, and additional tests on Manual Dexterity, Visual Acuity and Stereopsis. This group received payment for participation (£30).

**Group B (Control):** Participants in Group B took the pre and post-intervention spatial ability tests but did not receive training nor took any other test. Participation was not paid.

**Group C (Single Test):** Participants in Group C only took the spatial ability test once, either at the time of the first test (February) or the second (May-June 2016). Participation was not paid.

**Group D (Experiment):** Formed for statistical analysis by combining Group A (Intervention) and Group B (Control).

**Group E (Full Cohort):** Group E encompassed all participants in the study.

Table 9 shows the groups in the study and the activities in which they took part.

	Spatial Ability			MDx, VA, Sts		
	Test 1	Training	Test 2	Tests		
<b>Group A</b> Intervention <i>n</i> = 29	✓	✓	✓	✓	} <b>Group D</b> Experiment <i>n</i> = 104	} <b>Group E</b> Full Cohort <i>N</i> = 194
<b>Group B</b> Control <i>n</i> = 75	✓		✓			
<b>Group C</b> Single Test <i>n</i> = 90	✓					

**Table 9. Participation Groups of the Project**

Note. In addition to spatial ability testing and training, the intervention group underwent testing for Manual Dexterity (MDx), Visual Acuity (VA) and Stereopsis (Sts). Values updated after clearing.

Table 10 shows the gender composition of each group in the study.

Group	Year 1	Year 2	Total
<b>Group A Intervention</b>	<b>16</b> F: 8 = 50% M: 8 = 50%	<b>13</b> F: 9 = 69% M: 4 = 31%	<b>29</b> F: 17 = 59% M: 12 = 41%
<b>Group B Control</b>	<b>57</b> F: 35 = 61% M: 22 = 39%	<b>18</b> F: 9 = 50% M: 9 = 50%	<b>75</b> F: 44 = 59% M: 31 = 41%
<b>Group C Single Test</b>	<b>51 *</b> F: 26 = 52% M: 24 = 48%	<b>39 **</b> F: 23 = 59% M: 14 = 36%	<b>90 ***</b> F: 49 = 54% M: 38 = 42%
<b>Total ****</b>	<b>124</b> F: 69 = 56% M: 54 = 43% UG = 1%	<b>70</b> F: 41 = 59% M: 27 = 39% UG = 2%	<b>194</b> F: 110 = 56% M: 81 = 42% UG = 2%

**Table 10. Distribution of Participants by Group, Year in Medicine and Gender**

Notes. Values after Intervention Group Participant Clearance (see above)  
 The total values include students who did not disclose their gender, it is therefore larger than the sum of male and female students in the table.  
 \* Includes one student of undisclosed gender (UG)  
 \*\* Includes two students of undisclosed gender  
 \*\*\* Undisclosed gender accounts for the missing 4%  
 \*\*\*\* Undisclosed gender percentages included  
 M = Male; F = Female

The large participation of female students in the study (56%) reflected the trend of gender composition of students at BSMS. In the last few years, student intake had a larger female representation (female intake 2013: 63.7%; 2014: 58%; 2015: 54.5%), which reflected in female percentages at graduation (2015: 62%, 2016: 66%) (BSMS, 2015b) (see Background chapter for details). These values indicate that the gender composition of the participants in the study was representative of the population of students that started their medical studies at BSMS in recent years.

#### 4.2.2 Participants' Age

Ages in Group E (Full Cohort) ranged from 18 to 45 years with a median of 20 years, *IQR* 4 [19, 23]. Distribution of data was not normal (Kolmogorov-Smirnov:  $D(192) = 0.24, p < 0.01$ ) with most data accumulated towards the lower values of the scale.

When analysed by gender, the median was 20 for both groups although there was a difference in range (Female: Range 23 [18, 41], *IQR* 3 [19, 22]; Male: Range 15 [18, 33], *IQR* 5 [19, 24]).

△ **Section result 01:** The proportion of female participants in the study (56%) was close to the average gender representation of medical students registered at BSMS between 2013 and 2015 (Female 58%). This suggests that the gender distribution of participants was representative of that of the population of medical students at BSMS in recent years. The median age was 20 for both sexes with a range from 18 to 45 years.

### **4.3 Analysis of the First Spatial Ability Test for Group E (Full Cohort)**

This section reports findings on the analysis of scores from the first spatial ability test given to all participants (Group E). The spatial ability scores (maximum 20 points) underwent conversion to percentage to facilitate comparisons.

#### **4.3.1 Reliability and Validity of the Spatial Ability Test**

The usefulness of a test as research instrument depends on two important factors that define its quality: reliability and validity (Karras, 1997b; Karras, 1997a). The 20-question spatial ability test used in this study, the Visualisation of Rotation (ROT), was reliable and valid, according to its published report (Bodner and Guay, 1997), these properties underwent verification under the circumstances of this study. The set of answers to the first spatial ability test went through analysis using the Kuder and Richardson 20 test (KR-20); a formula that evaluates Internal Consistency Reliability by correlating the scores of each question to the total of all scores (Karras, 1997b; Salkind, 2014). Applying the test to the dataset in Excel, the KR20 was 0.75, which represents a high reliability.

Another common measure of scale reliability is the Cronbach's Alpha test ( $\alpha$ ), available in SPSS, that reported a value of  $\alpha = 0.748$ , which, according to Kline and Field represents good reliability (Kline, 2000; Field, 2013; Field, 2018), confirming the literature report and the KR20 result mentioned above in relation to the spatial ability test. In addition, the inter-correlation matrix produced for this test, showed that some of the last questions, e.g. Q17, 19 and 20, had a low and even negative correlation values

with the first questions e.g. Q 1 to 5, which supports the difficulty-based ranking (described later).

The validity of the ROT (and the PSVT: R) test, measured as Construct Validity was highly correlated with well-known tests in spatial ability, according to its authors (Bodner and Guay, 1997). The verification of this report used its relationship with reliability, measured using the rule “the maximum level of validity is equal to the square root of the reliability coefficient” (Salkind, 2014), and provided a result of 0.87, which is high within the possible range between 0 and 1.

△ **Section result 02:** The measurement of reliability using the Cronbach Alpha and the Kuder and Richardson tests (KR20), and of validity, using the Construct Validity test, confirmed that the “20-question (ROT) version of the PSVT:R Spatial Ability Test” is highly reliable and valid, supporting reports found in the literature.

#### **4.3.2 Descriptive Statistics of the First Spatial Ability Test**

The first spatial ability test, taken by all participants and assumed to represent a measure of their innate ability, became the reference point for future comparisons.

The test used to measure spatial ability was the Visualisation of Rotation (ROT) test, comprising 20 questions (Bodner and Guay, 1997). To facilitate comparisons and analysis, this 20-point scale underwent conversion to percentage, making the highest mark 100%.

The median of the spatial ability score of all participants was 70, *IQR* 25 [55-80], showing a range of scores between 10 and 100.

The distribution of scores was not normal (Kolmogorov-Smirnov  $D(194) = 0.12$ ,

$p < 0.01$ ) with more values on the higher end of the scale (Skewness = -0.5, Kurtosis = -0.1).

△ **Section result 03:** The median score for all participants in the first spatial ability test was 70%. The scores had a non-normal distribution with more values in the high end of the range. That value, considered the innate spatial ability of year 1 and 2 medical students, became the reference baseline for this study.

#### **4.3.3 Comparison of Scores in the First Spatial Ability Test by Gender for Group E (Full Cohort)**

Group E encompassed all undergraduate students ( $N = 194$ ); all of whom took the first spatial ability test; their gender composition was 110 (56%) female, 81 (42%) male and 3 (2%) of undisclosed gender. The following analysis for gender included 191 participants, excluding those with undisclosed gender, which in SPSS fell under the “Missing Data” category.

Female participant scored in average 11.4% (2.28 points) less than males on the first spatial ability test. The distribution of scores was not normal for both genders (Table 11).

The comparison of scores in spatial ability by gender used the independent samples T-Test, which reported:

Male participants in Group E (Full Cohort) scored higher in spatial ability ( $Mdn = 75.00$ ,  $IQR = 20$ ) than female participants ( $Mdn = 60.00$ ,  $IQR = 30$ ). This difference, 11.4, BCa 95% CI [6.4, 16.4] was significant,  $t(188.8) = 4.5$ ,  $p < 0.01$  (Table 11).

Parameter	Group E (Full Cohort) (N = 194)	Male subgroup (n = 81)	Female subgroup (n = 110)	Comparison of scores by gender
<b>Spatial Ability First Test</b>	<i>Mdn</i> = 70	75	60	<i>Independent T-Test</i> $t(188.8) = 4.5,$ $p < 0.01$
IQR	25	20	30	<i>Mann Whitney **</i>
Range	90 [10, 100]	80 [20, 100]	85 [10, 95]	$U = 2972, z = -4.16,$ $p < 0.01$
Skewness	-0.5	-0.6	-0.28	
Kurtosis	-0.1	0.7	-0.52	
*K-S Norm.	$D(194) = 0.11,$ $p < 0.01$	$D(81) = 0.15,$ $p < 0.01$	$D(110) = 0.12,$ $p = 0.01$	

**Table 11. Descriptive Statistic and Comparison of Results of the First Spatial Ability Test for Group E (Full Cohort) by Gender.**

Note. Data for Group E (Full Cohort) included for comparison only. The number of participants in this group includes three students who did not disclose their gender; it is therefore larger than the sum of male and female students in the table.

To compensate for the lack of normality, the T-Test included Bootstrapping.

\* K-S Norm = The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$ .

\*\* The Mann-Whitney test result included for comparison with a non-parametric calculation, showing an agreed in significance with the T-Test.

△ **Section result 04:** Male students from the full cohort of participants obtained higher scores than females on the first spatial ability test by an average of 2.2 out of 20 points; this difference, analysed with the T-Test was significant ( $p < 0.01$ ). These results agree with multiple reports from the literature, in which males score higher than females in spatial ability.

Section 5.9 Gender Characteristics in Spatial Ability discusses these findings further.

#### **4.3.4 Comparison of Scores in the First Spatial Ability Test by Year of Study at BSMS**

Group E (Full Cohort) included students from Year 1 (Y1) ( $n = 124$ ) and Year 2 (Y2) ( $n = 70$ ). All students sat the first spatial ability test, although testing took place at different times for each year. Those scores underwent analysis using the independent samples T-Test, with the following result:

Year 2 students obtained a lower score in spatial ability ( $Mdn = 62.5$ ) than Y1 students ( $Mdn = 70$ ), the mean difference (3.2, CI 95% [-2.6, 8.6]) was not significant  $t(136.6) = 1.05, p = 0.29$ .

Although the difference in number of participants in each year negatively affects the validity of these results, the findings suggest that participants from both years were essentially similar in their performance on the spatial ability test.

#### **4.3.5 Questions Attempted in the First Spatial Ability Test**

The original version of the test specifies a strict allocation of 10 minutes for the test (Bodner and Guay, 1997). To verify this notion, an analysis of the relationship between participants and the number of questions answered during those 10 minutes, produced the following report:

The majority of students (159 participants [81.1%]: 89 female, 68 male and 2 of undisclosed gender) attempted 20 questions; 16 students (8.2%) attempted between 17 and 19 questions; the least number of questions attempted was 10 (one female participant = 0.5%).

△ **Section result 05:** The majority of participants of the full cohort in the study attempted all the 20 questions of the first spatial ability test within the assigned

limit of 10 minutes, which supports the notion that this time was sufficient for the test.

#### **4.3.6 Measure of Difficulty of the Spatial Ability Test**

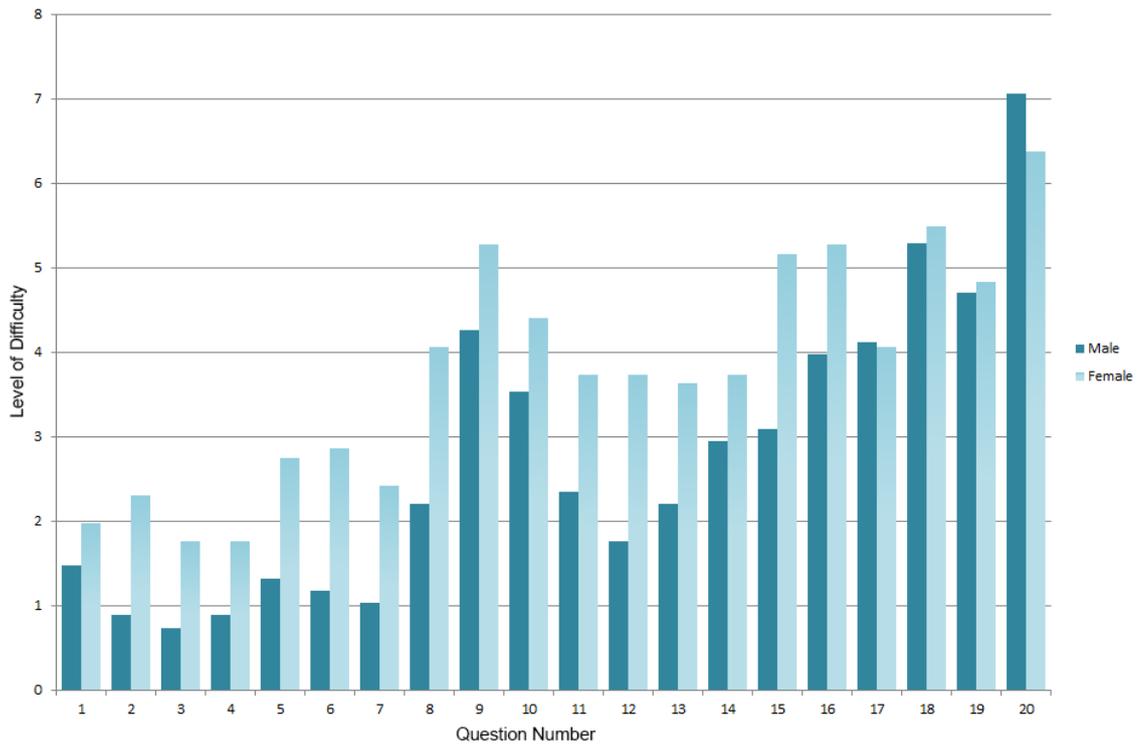
The Visualisation of Rotation (ROT) test used in this study incorporates an incremental degree of difficulty where the first questions are generally easier than the last (Bodner and Guay, 1997).

The first attempt to determine the difficulty value of each question in the test used visual assessment of challenge, a subjective organisational method that provided mixed results; consequently, the study used an analysis of participant response patterns instead. To determine if the test responses in this study agreed with the predicted levels of difficulty, this analysis looked at the proportion of correct answers for each question. The results provided a weighted scoring scheme based on question difficulty (see section 3.2.7.1 Question Challenges in the Spatial Ability Test for details).

Only participants who attempted the complete set of 20 questions of the test formed the sample for this assessment of difficulty (159 students). The distribution of correctly answered questions showed that most participants provided the correct response to questions three and four, and these therefore appeared to be the easiest questions of the test. In contrast, the least number of participants answered questions 20 and 18 correctly, making them the most difficult questions (Figure 7).

When separated by gender, this distribution revealed differences in general and in specific types of questions. The frequency distribution for female responses showed that they were able to provide a correct answer less frequently than males in 18 of the 20 questions, with the exception of questions 17 and 20, which females answered correctly more frequently than males.

The quantification of difficulty emerged from the assumption that difficulty constitutes the inverse of ease, which equates to the number of correctly answered questions; from this perspective, in the graphical representation of difficulty the highest bars equal to the lowest number of correctly answered questions (Figure 7).



**Figure 7. Distribution of Difficulty of Question in the Spatial Ability Test Based on a Sample of 159 Students by Gender.**

Note. The level of difficulty resulted from the inverse of the number of correctly answered questions. In this case, female participants found each question more difficult than males with exception of questions 17 and 20, in which males had more difficulty in providing the correct answer.

Re-ordering the questions by the index of difficulty calculated in this study, resulted in a new sequence of questions, from the easier to the more difficult. When analysed by groups of questions, the new sequence generally agrees with the original sequence of

the test with exception of questions 8, 9 and 10, which moved to a more difficult position in the new sequence (Table 12).

Spatial Ability Questions: New order based on difficulty index																				
Index	1.3	1.4	1.7	1.8	1.8	2.1	2.1	2.9	3.0	3.1	3.3	3.4	4.0	4.1	4.3	4.7	4.8	4.8	5.4	6.7
New	3	4	2	1	7	5	6	12	13	11	8	14	10	17	15	16	19	9	18	20
Old	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

**Table 12. Proposed New Order of Spatial Ability Question Sequence Based on Index of Difficulty.**

Note. Index values rounded up; see Method chapter for full values.  
 Arrows point at where each question moved after indexed by difficulty.  
 Colour coding represents the new five ranks of difficulty; see Method chapter for details.

This new order represents a set of questions of verified incremental difficulty for the spatial ability test (PVST: R) and constitutes a proposed change for future testing that would more precisely present a gradual increase in difficulty in the spatial ability test.

The results were the foundation to define an alternative method of marking in response to question difficulty and provide a more detailed and accurate measure of performance in spatial ability. In addition, this method may help identify areas in which the content of a spatial-awareness training programme needs to focus; it may also allow tailoring teaching to the needs of each student.

△ **Section result 06:** The analysis of question difficulty based on answers from 159 students who answered all questions of the first spatial ability test showed that questions 3 and 4, involving a single 90-degree rotation, were the easiest to

answer. In contrast, questions 18 and 20, which had three-level rotations of objects most dissimilar to the sample objects, were the most difficult. Comparing correct answers of each gender by question, male students outperformed females in the majority of questions (18 of 20 questions).

These results suggest that female students in the study found the majority of rotation challenges of the spatial ability test, more difficult than males, regardless of their complexity. As this test measures the ability to rotate objects mentally, males demonstrated a larger ability to imagine the rotation of objects from simple rotations to more complex, up a point of extremely difficulty, where both genders matched in their responses. Section 2.3.1 Spatial Ability and 5.3 Student Gender Distribution; discuss gender issues in spatial ability.

#### **4.3.6.1 Comparative Analysis of Spatial Ability Scores Based on Question Difficulty for Group E (Full Cohort)**

Comparing the difficulty-based scores with the unweighted results required conversion of the former into percentages. The goal was to quantify the difference to see if both methods provided equivalent values. The analysis assumed normality based on the number of participants despite showing non-normal values when tested. Comparisons used the paired-samples T-Test as the same population was the subject when comparing the two methods of marking the same test.

Participants in Group E (Full Cohort,  $N = 194$ ) obtained a lower weighted score when marked based on question difficulty ( $Mdn = 59$ ,  $IQR = 28.7$ ) than with traditional marking ( $Mdn = 70$ ,  $IQR = 25$ ), this difference, 7.1, was significant  $t(193) = 18.2$ ,  $p < 0.001$ .

Separating the analysis by gender, the Difficulty-Based scores were lower for both genders (male: *Mdn* = 68.6, *IQR* = 24.2, female: *Mdn* = 54.5, *IQR* = 31) than the traditional scores (male: *Mdn* = 75, *IQR* = 20, female: *Mdn* = 60, *IQR* = 30), this difference was significant for both genders (male:  $t(81) = 16, p < 0.001$ , female:  $t(109) = 11, p < 0.001$ ) (Table 13).

Parameter	Spatial Ability T1 Male ( <i>n</i> = 81)		Spatial Ability T1 Female ( <i>n</i> = 110)		Comparison of Scoring Paired Samp. T-Test	
	Raw Score	Weighted Score	Raw Score	Weighted Score	Male	Female
Median	75	68.8	60	54.6	$t(81) = 16$ $p < 0.001$	$t(110)$ $= 11$ $p < 0.001$
IQR	20	24.2	30	31		
Range	80 [20, 100]	85 [15, 100]	85 [10, 95]	86 [9, 95]		
K-S Norm.	$D(81) = 0.15,$ $p < 0.01$	$D(81) = 0.1,$ $p = 0.02$	$D(110) = 0.12,$ $p = 0.01$	$D(110) = 0.05,$ $p = 0.20$		

**Table 13. Comparison of Raw and Weighted Difficulty-Based Scores from the First Spatial Ability Test for Group E (Full Cohort) by Gender.**

Note. The paired samples T-Test assumed normality due to sample size, despite non-normal data sets (except for the female subgroup's weighted scores).  
K-S Norm = The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$ .

The comparison between genders showed that although the gender difference persisted at significant levels with the weighted marking method,  $t(191) = 3.5, p < 0.001$ , the gender gap decreased by 1.6% (Weighted difficulty-based difference male/female = 9.8%, Raw marking difference = 11.4%), with female participants showing the smallest difference.

△ **Section Result 07:** The comparison of difficulty-based marking with traditional marking of the spatial ability test showed a significant difference with lower scores when marked based on difficulty. This difference was present for the full cohort and for each gender, showing persistence of significantly higher scores in males.

These findings suggest that marking the spatial ability test based on question-difficulty might be a more stringent method than traditional marking, where every question has the same value. The analysis of difficulty showed an order of complexity, where the more complex questions were equally hard for both genders to answer, a characteristic that might help in the selection of questions for future tests, e.g. testing during spatial ability training with increments of complexity, saving the most difficult questions for higher levels of training.

When analysing spatial ability scores of all participants or comparing genders, difficulty-based marking seems to provide similar sensitivity to differences than traditional making, with the disadvantage of increasing complexity to the study without offering additional value. Consequently, the remaining sections of this thesis will only use traditionally marked scores.

#### **4.3.7 Comparative Analysis of the First Spatial Ability scores for Group E (Full Cohort) by participant's year of study at BSMS**

The analysis of the scores in spatial ability (first test) for Year-1 and Year 2 participants in Group-E (Full cohort,  $N = 194$ ), showed that Year 2 students ( $n = 70$ ) obtained lower scores than Year-1 students ( $n = 124$ ). The independent samples T-Test used to evaluate this difference, reported the following:

Year 2 participants in Group E (Full Cohort) scored less ( $Mdn = 65$ ,  $IQR = 30$ ) than Year 1 participants ( $Mdn = 70$ ,  $IQR = 28$ ) in the first spatial ability test. This difference, 3, BCa 95% CI [-2.6, 8.6], was not significant,  $t(136.6) = 1.05$ ,  $p = 0.29$ .

This result shows that the year of study at BSMS did not seem to offer performance advantages in the spatial ability test, which supports the homogeneity of the participants in the study. The analysis of participation by year in other sub-groups (Groups A, B and C) appears in section 4.4.5.

## **4.4 Analysis of the Spatial Ability Tests for Group D (Experiment)**

This section will analyse the scores of group D (Experiment,  $N = 104$ ), which includes groups A (Intervention,  $n = 29$ ) and B (Control,  $n = 75$ ). These groups took the first and second spatial ability tests, which they have in common; only the intervention group received training. See section 4.2.1 Participation and group structure for details.

### **4.4.1 Descriptive Statistics and Test Comparison for Group D (Experiment)**

To demonstrate that Group D (Experiment) is a representative sample of Group E (Full Cohort,  $N = 194$ ), a side-by-side comparison appears in Table 14, which, based on the first spatial ability test, showed similar values in their descriptive statistics, supporting their correspondence: Notably, both have a Median of 70, and a non-normal distribution.

Assuming normality due to the sample size ( $n = 104$ ), the comparison between the two spatial ability scores for Group D employed the Paired Samples T-Test, which reported the following:

Group D (Experiment) obtained a higher score in the second spatial ability test ( $Mdn = 75, IQR = 20$ ) compared to the first ( $Mdn = 70, IQR = 25$ ), this difference,  $-5.7$ , BCa 95% CI  $[-8.3, -3.0]$  was significant,  $t(103) = -4.33, p < 0.01$ .

Parameter	Group E (Full Cohort) ( $N = 194$ ) *	Group D (Experiment)		Dependent Samples T-Test
		S.A. T1 ( $n = 104$ )	S.A. T2 ( $n = 104$ )	
Median	70	70	75	$T(103) = -4.3,$ $p < 0.01$
IQR	65.4	25	20	
Range	90 [10 – 100]	80 [20 – 100]	75 [25 – 100]	
Skewness	-0.51	-0.5	-0.8	
Kurtosis	-0.1	-0.3	0.03	
***K-S Norm.	$D(194) = 0.11,$ $p < 0.01$	$D(104) = 0.13,$ $p < 0.01$	$D(104) = 0.13,$ $p < 0.01$	

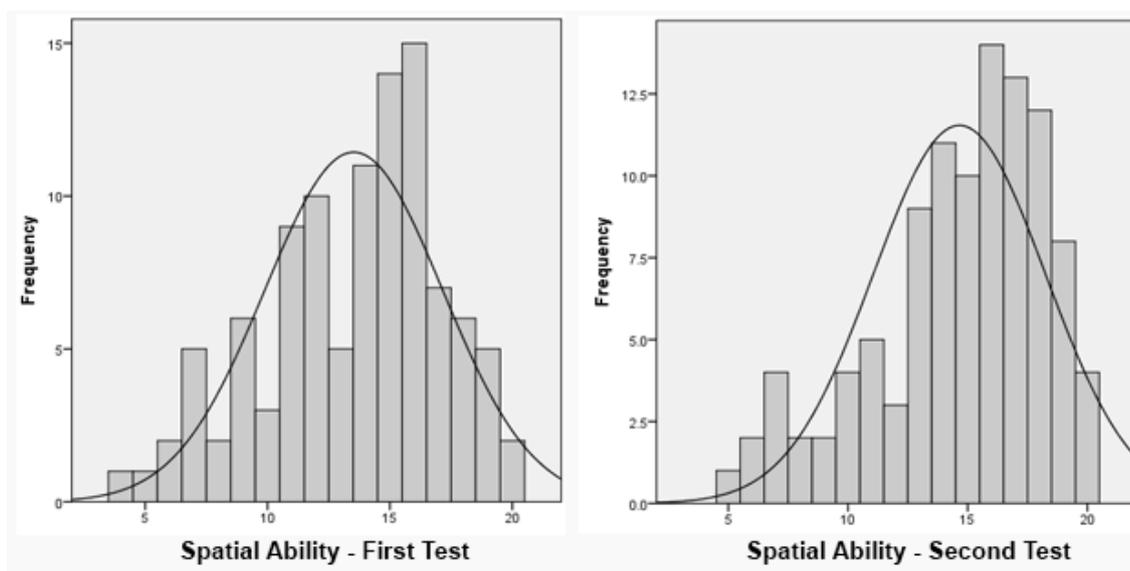
**Table 14. Descriptive Statistics and Comparison Using the Dependent Samples T-Test of Scores from the First and Second Spatial Ability Tests for Group D (Experiment).**

Note. \* For illustrative purposes, the descriptive statistics of Group E and D appear here side by side, showing a close correspondence.

\*\* The comparison of scores from the first and second spatial ability tests for Group D used the Dependent Samples T-Test, which reported a significant difference.

\*\*\* K-S Norm = The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$ ).

The histograms demonstrate the change in distribution on the second test, where there is an accumulation of scores in the higher side of the scale (Figure 8).



**Figure 8. Histograms of the First and Second Spatial Ability Tests for Group D (Experiment).**

Note. The histograms show accumulation of scores on the high end of the range, particularly on the second spatial ability test with persistence of Skewness and Kurtosis.

△ **Section result 08:** The scores of group D (Experiment) in the first spatial ability test had similar characteristics to those of group E (Full Cohort), which supports the concept of group D being a representative sample of group E. Comparing scores between tests, Group D showed a significant improvement in the second spatial ability test.

The similarity in scores between the Experiment group and the Full Cohort, helped diffuse the idea that group D (which includes the Intervention group) was a group of “good scorers”, which was a possibility derived from “self-selection,” where participants who joined the experiment group did so because they felt they could do better than the rest of their class. The analysis of group D’s subgroups: A (Intervention) and B (Control) may help define if one or both sub-groups influenced the score improvement.

#### 4.4.2 Analysis and Comparison of Spatial Ability Scores for the Intervention and Control Groups

This section will analyse the scores from the first and second spatial ability tests for groups A (Intervention) and B (Control) and will compare those scores between groups.

To prevent the influence of extreme values and improve robustness, the analysis focused on the 5% Trimmed Mean, which is a calculation that excludes 5% of the values of both extremes of a series (Field, 2012).

The statistical analysis showed that both groups had an improvement on the second test of Spatial Ability. The Intervention group showed an improvement of 9.38 % (1.87 points out of 20) in Spatial Ability scores from the first to the second test. The Control group improved by 4.76 % (0.95 points), roughly half in percentage but translates to a single point less on the test.

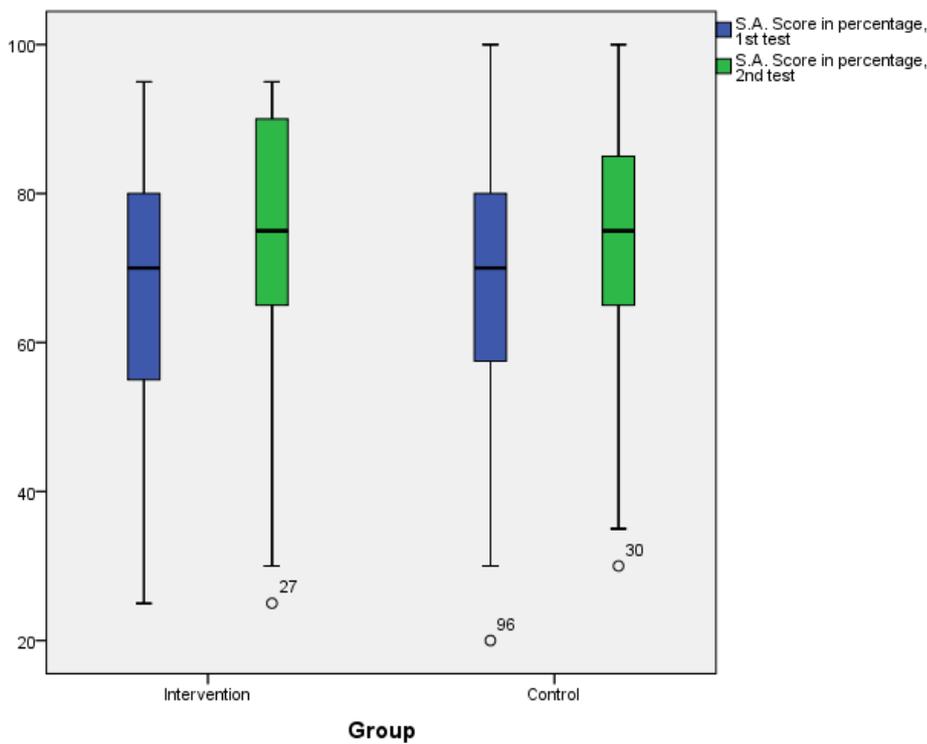
Minimum and maximum scores remained the same for the intervention group (Min 25, max 95) while the minimum for the control group increase from 20 to 30 (maximum remained at 100), (Table 15).

Parameter	Group A (Intervention) (n = 29)		Group B (Control) (n = 75)	
	S.A. Test 1	S. A. Test 2	S. A. Test 1	S. A. Test 2
Median	70	75	70	80.00
IQR	30	25	20	20
Range	70 [25, 95]	70 [25, 95]	80 [20, 100]	70 [30, 100]
K-S Norm	$D(29) = 0.18,$ $p = 0.01$	$D(29) = 0.15,$ $p = 0.07$	$D(75) = 0.13,$ $p = 0.004$	$D(75) = 0.15,$ $p < 0.001$

**Table 15. Scores from the First and Second Spatial Ability Tests of Group A (Intervention) and group B (Control).**

Note. K-S Norm = The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$

Normality tests K-S / S-W showed that distributions were not normal, which was corroborated by the Q-Q plots that showed a sinusoid distribution, characteristic of skewness. This feature appeared in both tests for the control group and the first test for the intervention group; the second test for the latter was normal (Table 15). Although the difference in distribution is evident in the Q-Q boxplots, the graphical representation using Clustered Boxplots shows these differences more clearly (Figure 9).



**Figure 9. Clustered Boxplot of Spatial Ability Scores Pre and Post-Intervention for the Intervention and Control Groups.**

Note. Although the medians (black lines on the boxes) increased by a similar amount for the intervention and control groups on the second spatial ability test, and the ranges of scores maintained their size, the distribution of data in relation to the median shows some differences. The boxplot of the second test for the intervention group show more scores above than below the median compared to the control group and to their own scores in the first test.

The graphic above suggests that the Intervention group had more scores above the median on the second spatial ability test. The comparative analysis of scores that follows will determine if those differences are significant.

### **Comparing Tests' scores for each Group**

The comparison of scores between the first and second spatial ability tests for each of the groups (A and B) required the Paired Samples T-Test, which reported the following:

Group A (Intervention) obtained a higher score on the second spatial ability test ( $Mdn = 75$ ,  $IQR = 25$ ) than on the first ( $Mdn = 70$ ,  $IQR = 30$ ), the means' difference,  $-8.9$ , BCa 95% CI  $[-13.2, -4.6]$  was significant  $t(28) = -4.2$ ,  $p < 0.001$ . Group B (Control) also obtained a higher score on the second spatial ability test ( $Mdn = 80$ ,  $IQR = 20$ ) than on the first ( $Mdn = 70$ ,  $IQR = 20$ ), the means' difference,  $-4.4$ , BCa 95% CI  $[-1.2, -2.7]$  was significant  $t(74) = -2.7$ ,  $p < 0.008$ .

Both groups scored equally higher in the second spatial ability test, which indicates that the improvement of Group D (Experiment), which includes A and B, resulted from a similar improvement in both groups.

### **Comparing Groups' scores in each Test**

The comparison between group A (Intervention) and B (Control) in their scores from the first and second spatial ability tests underwent analysis using the Independent Samples T-Test, which reported the following:

Group A and B obtained similar scores ( $Mdn = 70$  for both,  $IQR = 30$  and  $20$  respectively) in the first spatial ability test, consequently, the T-Test found no significant difference between the groups,  $t(44.8) = -1.0$ ,  $p = 0.31$ . In the second test, Group B obtained a higher score ( $Mdn = 80$ ,  $IQR = 20$ ) than Group A ( $Mdn = 75$ ,  $IQR = 25$ ), the

difference between the groups' means, 0.24, BCa 95% CI [-7.9, 8.4], was not significant,  $t(48.7) = 0.06$ ,  $p = 0.95$ .

△ **Section result 09:** The intervention and control groups obtained significantly higher scores on the second spatial ability test. When compared to one another, there was no significant difference in scores in both tests. The similarity of scores between groups in the second test (post-training), suggests that the spatial ability training provided to the intervention group did not result in a significant change in spatial ability performance when compared to the control group.

#### 4.4.3 Post-hoc Statistical Power Calculation

The section above provided information about the impact of training in spatial ability, measured by the performance of the intervention group in the spatial ability test, in comparison with the control group. The amount of such an impact relates to the statistical power of the test, which, when calculated based on the variables and results of the study, constitutes a post-hoc calculation of power (see section 3.2.4.2 Statistical Power, in the Methodology chapter). The variables for this calculation include:

- Significance value or Alpha ( $\alpha$ ) for a two-tailed study (considers the possibility of results above or below the reference group) = 0.05.
- Sample size: Although this commonly refers to the size of the intervention group ( $n = 29$ ), it assumes a control group of a similar size, in this study, the control group was larger ( $n = 75$ ). In these circumstances, Cohen recommends calculating the Harmonic Mean ( $n'$ ), using the formula:  $n' = 2n_a n_b / n_a + n_b$

(Cohen, 1988), which applied to this study would be:

$$n' = 2 (29) (75) / 29 + 75 = 4350/104 = 42$$

- Effect size (ES): Calculated using Jacob Cohen's formula:

$d = M1 - M2 / SD$  (control); using results from the second spatial ability test (post-training). Because in this study, the median (*Mdn*) provided a better measure of central tendency due to the non-normal distribution of the results, it replaced the mean (*M*) in Cohen's formula, producing the following calculation:  
 $d = 75-80/17.814 = 5/17.814 = 0.3$ , corresponding to a medium effect size.

If this analysis adhered to using the means for each group, the calculation would be:

$$d = 73.45 - 73.20 / 17.814 = 0.25 / 17.814 = 0.01$$
, which indicates a small effect size.

From these variations, the former seemed to provide a better representation of the study.

For the calculation of statistical power, using the values above (including harmonic mean), this study looked at two methods: Cohen's tables and the software G\*Power (version 3.1.9.2).

The reference table for "double-sided Alpha 0.05" for an ES 0.3 and  $n = 42$ , corresponded to a power of 0.27 (power raised to 0.30 in the hypothetical case of using a "single-sided Alpha").

Applying the same values to G\*Power, the post-hoc analysis resulted in a calculated Power of 0.51, increasing to 0.64 when set to a single-tail study (using the sample size of 29 instead of the "harmonic mean" of 42, the power output was 0.37).

Both post-hoc power calculation methods showed that the study was underpowered, most likely due to an insufficient sample size. Sections 4.8.1 Research Modality Limitations and 5.2 Student Participation, expand on this issue.

△ **Section result 10:** The post-hoc statistical power calculation using Cohen's tables and G\*Power software produced values below 50%, which indicate that the study was underpowered, possibly due to a small sample size.

#### **4.4.4 Comparisons of Spatial Ability Scores by Gender for Group D (Experiment)**

This section examines differences in spatial ability performance by gender within and between each sub-group of Group D (Experiment), which refers to Group A (Intervention) and B (Control). These results add to the analysis of gender differences in Group E (Full Cohort), covered in section 4.3.3.

The descriptive analysis of scores by gender within each group revealed that some of the sub-groups had a normal distribution; unlike the distribution of the full cohort, which was not normal. As mentioned in the methodology chapter, the project assumed normality for comparison analysis, therefore favoured the T-Test in its two varieties: paired samples, when the same population sat to both tests, and independent samples, when comparing different populations' performance in a given test (Field, 2012).

##### **Female participants:**

The distribution of scores of female participants in Group A (Intervention) was normal on the first spatial ability test. On the second test, the distribution was not normal, showing a long tail to the left and an accumulation of scores to the right, corresponding to the higher values of the range (skewness was 0.13 in the first and – 0.87 in the second test). Although the range of scores decreased slightly on the second test, the median increased, despite a lower maximum score. In contrast, the distribution of

scores of female participants in Group B (Control) was not normal on the first test, taking a normal shape on the second (skewness was - 0.50 and - 0.43 respectively), the range of scores also decreased but the median did not change, despite an increase in the maximum score (Table 16 and Table 17).

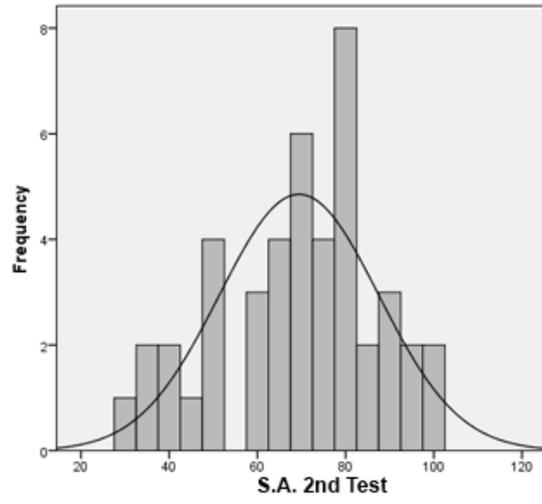
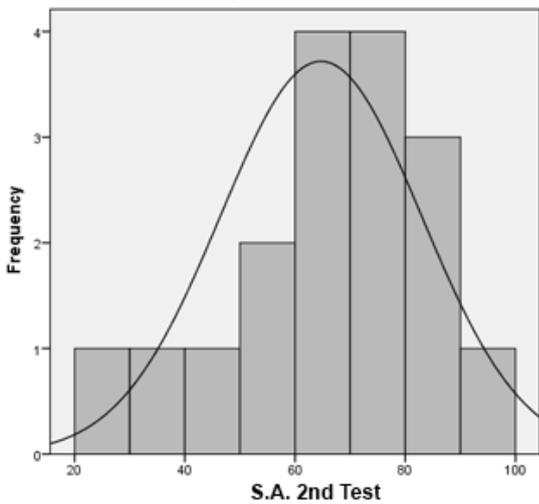
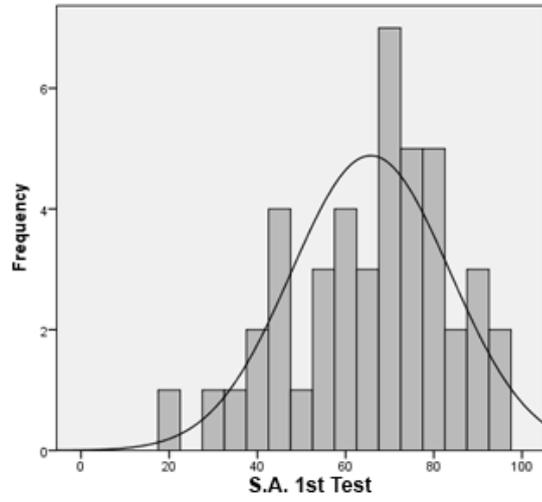
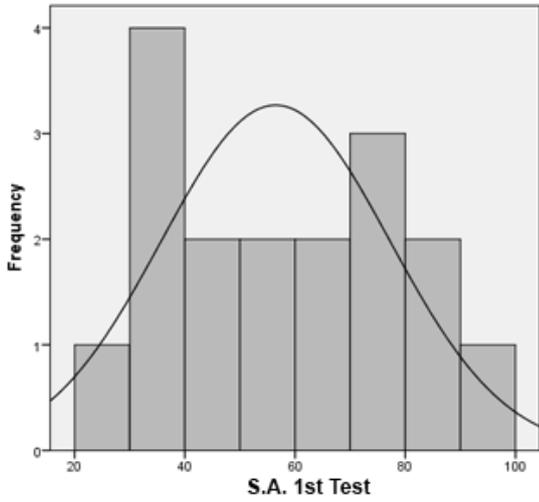
The comparison of spatial ability scores (Test 1 and 2) for female participants within in each group, showed that females in the intervention group obtained higher median scores in the second test than in the first (*Mdn* = 55 and 65 respectively), this difference was significant  $t(16) = -2.36, p = 0.03$ . In contrast, their counterparts in the control group obtained the same median score in both tests (*Mdn* = 70), therefore, there was no significant difference  $t(43) = -1.6, p = 0.1$ . These results showed that females in the intervention group improved significantly in their spatial ability scores within their group on the second test, while there was no improvement for females within the control group; it also illustrates that the improvement in the intervention group was not enough to match the median of the control group.

Comparing the scores of female participants between groups (Intervention vs Control): In the first spatial ability test, those in the Control group obtained higher scores (*Mdn* = 70) than their counterparts of the intervention group (*Mdn* = 55), this difference was not significant  $t(26) = -1.6, p = 0.12$ . In the second spatial ability test (after training of the intervention group), females in the Control group obtained similar values than on the first test (*Mdn* = 70), while those in the intervention group showed an improvement (*Mdn* = 65), this difference was also not significant  $t(29) = -0.9, p = 0.37$ . These results suggest that the training in spatial awareness did not produce a significant impact in the spatial ability performance of females in the intervention group (Table 17).

**Female**

**Intervention (n = 17)**

**Control (n = 44)**



**Table 16. Distribution of Scores of the First and Second Spatial Ability Test for Female Participants of Group A (Intervention) and B (Control).**

Note. The score distribution of females in Group A (Intervention) (Left) shifted towards higher scores on the second spatial ability test. This shift was inverse for females in Group B (Control) (Right).

Gender / Group	Parameter	1st S.A. Test Total Females <i>n</i> = 61	2nd S.A. Test Total Females <i>n</i> = 61	Comparison Paired Samples T-Test **
Female / Intervention ( <i>n</i> = 17)	Median	55.0	65	<i>t</i> (16) = -2.36, <i>p</i> = 0.03
	SD	20.7	18.2	
	Range	70 [25, 95]	65 [25, 90]	
	IQR	40 [35, 75]	23 [55, 77.5]	
	K-S Norm.	<i>D</i> (17) = 0.14, <i>p</i> = 0.20 *	<i>D</i> (17) = 0.21, <i>p</i> = 0.04	
Female / Control ( <i>n</i> = 44)	Median	70.0	70.0	<i>t</i> (43) = -1.6, <i>p</i> = 0.1
	SD	17.9	18	
	Range	75 [20-95]	70 [30-100]	
	IQR	25 [52.5, 80]	20 [60, 80]	
	K-S Norm.	<i>D</i> (44) = 0.14, <i>p</i> = 0.02	<i>D</i> (44) = 0.12, <i>p</i> = 0.07 *	
<b>Comparison Independent Samples T-Test ***</b>		<i>t</i> (26) = -1.6, <i>p</i> = 0.12	<i>t</i> (29) = -0.9, <i>p</i> = 0.37	

**Table 17. Descriptive Statistics and Comparisons of Scores from the First and Second Spatial Ability Tests for Female Participants of the Intervention and Control Groups.**

Note. \* Normal distribution: The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$ .

\*\* The Paired Samples T-Test compares variables of the same participants in parametric data sets; it is significant if  $p < 0.05$ .

\*\*\* The Independent Samples T-Test compares variables of different participants in parametric data sets; it is significant if  $p < 0.05$ .

△ **Section result 12:** Female participants in the intervention group obtained significantly higher scores in the second spatial ability test than on the first ( $p = 0.03$ ); the distribution analysis showed that they had more scores on the higher end of the scale in the second test. In contrast, the scores of females in the control group did not change significantly between tests. When compared against each other, females in the control group had higher scores in both tests than those in the intervention group, but there was no significant difference between groups in either test.

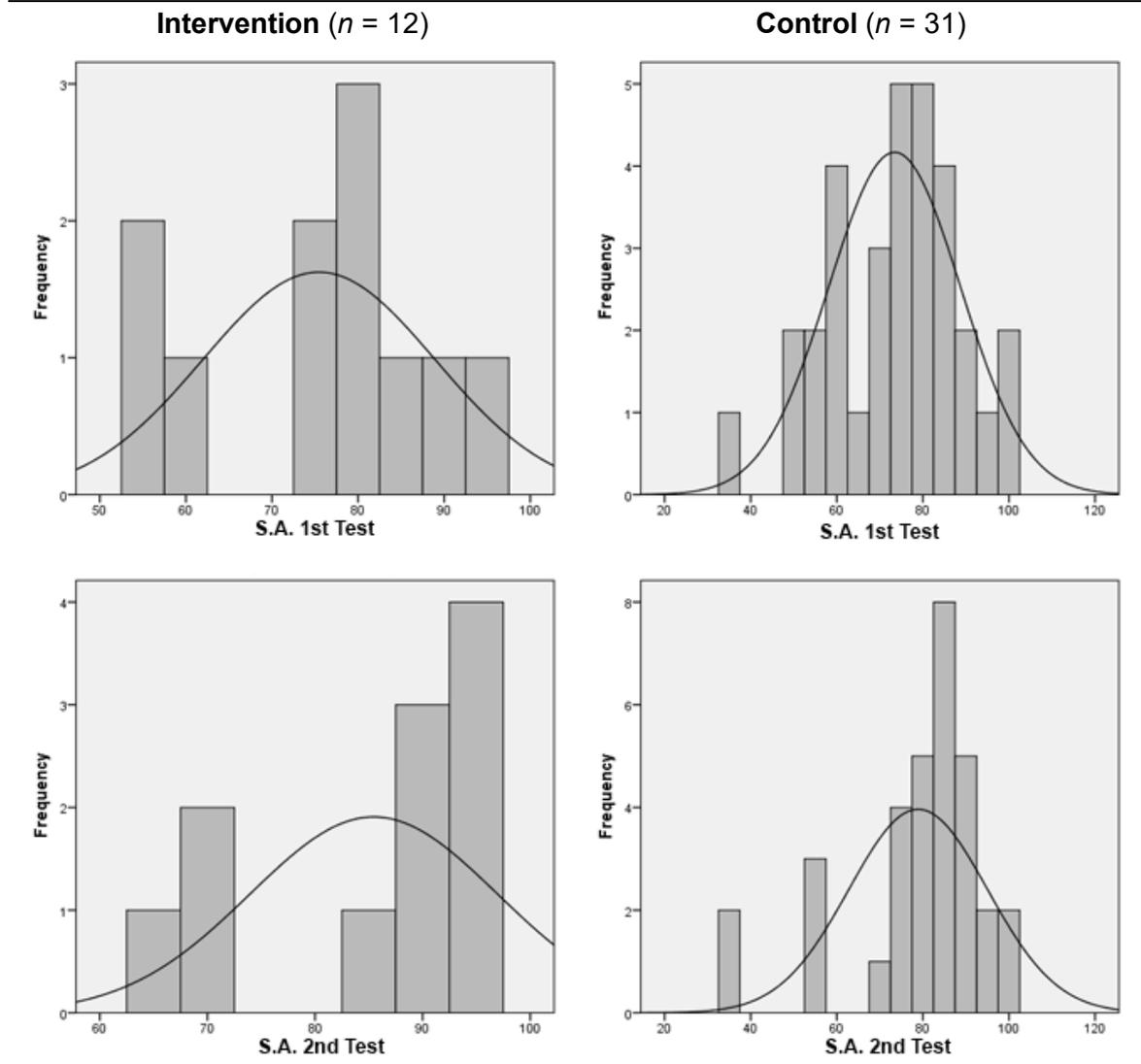
### **Male participants:**

Male participants in both groups obtained scores with a normal distribution on the first test, and not normal on the second (Table 19). In the second tests, there was an accumulation of scores in the higher end of the scale for both groups (Intervention: skewness = -0.56 on the first and -1.1 on the second test; Control: skewness = -0.35 and -1.4 respectively). In addition, the intervention group's range of scores had higher limits in the second test, while the range in the control group did not change.

Regarding the median for males within each group, there was an increase of 10 units: from 80 to 90 for the intervention group, and from 75 to 85 for the control group. Those differences were significant for both groups,  $t(11) = -6.6, p < 0.001$  and  $t(30) = -2.4, p = 0.02$  respectively (Table 18 and Table 19).

Comparing the scores of males across groups, showed that the intervention group obtained higher scores than the control group in both tests, with a difference of 5 units in both instances; this difference was not significant in both tests,  $t(24) = 0.6, p = 0.58$  and  $t(30) = 1.7, p = 0.1$ , respectively (Table 19). These results show that although male participants of both groups improved significantly on the second test, when compared with each other, there was no significant difference in either test. This suggests that the training in spatial awareness provided to the intervention group had no impact on their performance in spatial ability compared to the control group.

**Male**



**Table 18. Distribution of Scores of the First and Second Spatial Ability Test in Male Participants of the Intervention and the Control Groups.**

Note. The distribution of scores from male participants of both groups shifted towards higher scores in the range on the second spatial ability test.

Gender / Group	Parameter	1st S.A. Test Total Males <i>n</i> = 43	2nd S.A. Test Total Males <i>n</i> = 43	Comparison Paired Samples T-Test **
<b>Male / Intervention (<i>n</i> = 12)</b>	Median	80.0	90.0	<i>t</i> (11) = -6.6, <i>p</i> < 0.001
	SD	12.9	11.0	
	Range	40 [55, 95]	30 [65, 95]	
	IQR	20 [60, 80]	21 [74, 95]	
	K-S Norm.	<i>D</i> (12) = 0.22, <i>p</i> = 0.09 *	<i>D</i> (12) = 0.31, <i>p</i> = 0.02	
<b>Male / Control (<i>n</i> = 31)</b>	Median	75.0	85.0	<i>t</i> (30) = -2.4, <i>p</i> = 0.02
	SD	15.5	16.2	
	Range	65 [35, 100]	65 [35, 100]	
	IQR	25 [60, 85]	15 [75, 90]	
	K-S Norm.	<i>D</i> (31) = 0.13, <i>p</i> = 0.20 *	<i>D</i> (31) = 0.22, <i>p</i> = 0.01	
<b>Comparison Independent Samples T-Test ***</b>		<i>t</i> (24) = 0.6, <i>p</i> = 0.58	<i>t</i> (30) = 1.7, <i>p</i> = 0.1	

**Table 19. Descriptive Statistics and Comparisons of Scores from the First and Second Spatial Ability Tests for Male Participants of the Intervention and Control Groups.**

Note. \* Normal distribution: The Kolmogorov-Smirnov Normality Test (K-S Norm) indicates normality when values are above  $p = 0.05$ .

\*\* The Paired Samples T-Test compares variables of the same participants in parametric data sets; it is significant if  $p < 0.05$ .

\*\*\* The Independent Samples T-Test compares variables of different participants in parametric data sets; it is significant if  $p < 0.05$ .

△ **Section result 13:** The analysis of scores in the male population showed that there was a significant improvement in performance in the intervention and the control groups in the second spatial ability test ( $p < 0.001$ ;  $p = 0.02$ ), with the former obtaining higher scores. Comparing test results between groups, although the advantage of males in the intervention group remained on the second test; this difference was not significant in either test. These results suggest that the training in spatial awareness given to male participants of the intervention group did not produce a significant impact in their performance in

spatial ability, compared to males in the control group. The score improvement of both groups in the second test may reflect the influence of other factors, such as the learning effect resulting from exposure to the first spatial ability test.

### **Female participants compared to male participants**

The independent samples T-Test was the test of choice to compare scores in spatial ability for female and male participants in each group; the test assumed normality of distribution. The descriptive statistics appear above in tables 18 and 20.

Female participants in the Intervention group scored less than males in the first spatial ability test (Females: *Mdn* = 55; Males: *Mdn* 80), and in the second (*Mdn* = 65 and 90, respectively). The mean differences (1<sup>st</sup> test: 19.3, 2<sup>nd</sup> test: 21.1) were significant on the first test,  $t(27) = 3.1$ ,  $p = 0.005$ , and the second,  $t(27) = 3.9$ ,  $p = 0.001$ .

In the Control group, female participants scored less than males in the first spatial ability test (Females: *Mdn* = 70; Males: *Mdn* 75), and in the second (*Mdn* = 70 and 85, respectively). The mean difference (1<sup>st</sup> test: 7.5, 2<sup>nd</sup> test: 9.1), was not significant on the first test,  $t(70) = 1.8$ ,  $p = 0.056$ , but significant in the second,  $t(69) = 2.2$ ,  $p = 0.025$ .

△ **Section result 15:** The comparison of scores between genders in each group in the first and second spatial ability tests showed that male participants obtained significantly higher scores in both tests for the intervention group ( $p = 0.005$ ), while for the control group this was only true for the second spatial ability test. These results agree with a similar comparison in Group E (Full cohort) (see section 5.3.2.3), and with reports mentioned in the literature.

#### 4.4.5 Analysis of Spatial Ability Scores for Group D (Experiment) by Year of Study at BSMS.

Following an initial analysis of scores in spatial ability by year on the first spatial ability test for Group E (Full Cohort) (see section 4.3.4); this section presents the analysis of scores by year in Group D (Experiment) and its sub-groups A (Intervention) and B (Control) (Table 20).

Group	Parameter	1st S.A. Test		2nd S.A. Test	
		Y1	Y2	Y1	Y2
<b>Group D (Experiment)</b> <i>n</i> = 104 Y1: <i>n</i> = 73 Y2: <i>n</i> = 31	Median	70	65	80	75
	SD	17.8	18.2	18.7	16.2
	Range	75 [25 - 100]	65 [20 - 85]	70 [30 - 100]	70 [25 - 95]
	IQR	28 [55 - 83]	30 [30 - 80]	23 [65 - 88]	20 [65 - 85]
	K-S Nor.	<i>D</i> (73) = 0.12, <i>p</i> = 0.007*	<i>D</i> (31) = 0.2, <i>p</i> = 0.003*	<i>D</i> (73) = 0.15, <i>p</i> < 0.001*	<i>D</i> (31) = 0.11, <i>p</i> = 0.20
<b>Group A (Intervention)</b> <i>n</i> = 29 Y1: <i>n</i> = 16 Y2: <i>n</i> = 13	Median	77.5	60	82.5	69.6
	SD	19.5	18.8	18.6	18.7
	Range	70 [25 - 95]	50 [30 - 80]	65 [30 - 95]	70 [25 - 95]
	IQR	29 [55 - 84]	38 [38 - 75]	24 [66 - 90]	28 [61 - 89]
	K-S Nor.	<i>D</i> (16) = 0.19, <i>p</i> = 0.12	<i>D</i> (13) = 0.20, <i>p</i> = 0.13	<i>D</i> (16) = 0.17, <i>p</i> = 0.20	<i>D</i> (13) = 0.17, <i>p</i> = 0.20
<b>Group B (Control)</b> <i>n</i> = 75 Y1: <i>n</i> = 57 Y2: <i>n</i> = 18	Median	70	72.5	80	77.5
	SD	17.5	17.1	18.9	14.3
	Range	70 [30 - 100]	65 [20 - 85]	70 [30 - 100]	60 [35 - 95]
	IQR	25 [58 - 83]	21 [59 - 80]	25 [60 - 85]	18 [62 - 80]
	K-S Nor.	<i>D</i> (57) = 0.12, <i>p</i> = 0.02*	<i>D</i> (18) = 0.19, <i>p</i> = 0.09	<i>D</i> (57) = 0.16, <i>p</i> = 0.001*	<i>D</i> (18) = 0.17, <i>p</i> = 0.20

**Table 20. Spatial Ability Scores of Group D (Experiment) by Sub-group and by Year at BSMS.**

Note. \* = Not normal (K-S Nor = The Kolmogorov-Smirnov Normality Test indicates normality when values are above *p* = 0.05).

The comparison of scores between participants based on their year of study in each group required the independent samples T-Test, which reported the following, regarding scores from the first and second spatial ability tests:

Year 2 participants in Group D (Experiment) scored less in both tests (1<sup>st</sup> test: *Mdn* = 65; 2<sup>nd</sup>: *Mdn* = 75) than Y1 participants (*Mdn* = 70, and 80 respectively), the mean differences (6.9 and 1.6) were not significant in both tests: 1<sup>st</sup> test:  $t(55.2) = 1.77, p = 0.08$ ; 2<sup>nd</sup>:  $t(65.3) = 0.45, p = 0.64$ .

When separated by sub-groups:

Intervention group (A): In both tests, Y2 participants scored less (1<sup>st</sup> test: *Mdn* = 60; 2<sup>nd</sup>: *Mdn* = 69.6) than Y1 participants (*Mdn* = 77.5, and 82.5 respectively), the mean differences (13 and 6.9) were not significant in both tests: 1<sup>st</sup> test:  $t(25.9) = 1.79, p = 0.08$ ; 2<sup>nd</sup> test:  $t(25.7) = 0.99, p = 0.33$ .

Control group (B): On the first test, Y2 participants scored more (*Mdn* = 72.5) than Y1 participants (*Mdn* = 70), the mean difference (2.8) was not significant,  $t(29.1) = 0.6, p = 0.55$ . On the second test, Y2 participants scored less (*Mdn* = 77.5) than Y1 participants (*Mdn* = 80), the mean difference (-0.9) was not significant,  $t(37.4) = -0.21, p = 0.83$ .

These results show that although Year 1 students in Group D (Experiment) obtained higher scores in both tests, this advantage was not significant. The difference in number of participants from each year (Y1  $n = 73$ ; Y2  $n = 31$ ) undermines the validity of these findings, preventing the selection of a better performer in spatial ability.

## 4.5 Analysis of Academic Performance

The analysis of Academic Performance focused on scores in the subject of Human Anatomy, which is part of the academic curriculum in Medicine at Brighton and Sussex Medical School (BSMS). Section 3.2.11 discussed general content of the modules and the timeline relating this project and the academic exams appeared in section 3.3.7 of the Methods chapter. Section 3.2.3.1 Hypotheses of the Study includes a hypothesis

for academic performance, which forecasts a difference between the intervention and control groups, after the intervention group receives spatial awareness training.

After the performance analysis of the experiment group (Group D), this section focuses in a comparison between intervention and control subgroups, in search of evidence to support or refute the above hypothesis.

#### **4.5.1 Anatomy Data**

The anatomy scores provided by BSMS were scores in the anatomy portion of each module or final exam. Each exam had different number of anatomy question, which provided a ceiling or maximum possible value for anatomy scores.

Year 1 Anatomy Modules 103 and 104 had a ceiling of 26 and 38 points respectively; the final score had a ceiling of 86.

Year 2 Anatomy Modules 203 and 204 had a ceiling of 35 and 54 points respectively; the final score had a ceiling of 125.

To facilitate calculations and comparisons of academic performance, scores of both years underwent conversion into percentage, which homogenised the scale.

Additionally, the nomenclature underwent simplification: modules 103 and 104 became “x03” and modules 104 and 204 became “x04” (see section 3.3.7 Analysis of Academic Data).

Because training happened before module exams 104 and 204, the analysis that follows focuses on modules immediately before and after training (x03 and x04) and the final score in anatomy.

## 4.5.2 Analysis of Anatomy Data

This section covers descriptive statistics and comparisons of Anatomy Modules x03, x04 and the Final Anatomy Score of participants in Group D (Experiment), and its sub-groups: A (Intervention) and B (Control). Normality assumed based on sample size.

### 4.5.2.1 Anatomy Modules and Final Score for Groups D (Experiment) and E (Full Cohort)

The median scores for Group D (Experiment) in Anatomy Modules x03, x04 and the Final score were 73.1, 76.3 and 74.4, respectively.

The One-way Repeated Measures ANOVA for Group D (Experiment) report for exam scores in x03, x04 and Final was: Maunchly's test indicated that the assumption of sphericity was absent  $X^2(2) = 115.1, p < 0.001$ , therefore Greenhouse-Geisser corrected tests are reported ( $\epsilon = 0.59$ ). The results show that the scores between exams x03, x04 and Final were significantly different,  $F(1.19, 122.81) = 14.9, p < 0.001$  (Table 21).

Confirming the above and analysing by pairs, the Paired Sample T-Tests of scores for Group D (Experiment), showed that the differences between exam x03 and x04, and those between x04 and the final score, were significant (Pair 1:  $t(103) = -4.06, p < 0.001$ , and Pair 2:  $t(103) = 3.9, p < 0.001$ ). Note that the differences refer to an increase in scores from modules x03 to x04 and a decrease from x04 to the final score.

In reference to Group E (Full Cohort), SPSS automatically excluded seven participants due to missing values. The median scores in Anatomy Modules x03, x04 and the Final score were 71.2, 74.1 and 72.8, respectively. The One-way Repeated Measures ANOVA for Group E (Full Cohort) report for exam scores in x03, x04 and Final was: Maunchly's test indicated that the assumption of sphericity was absent  $X^2(2) = 224.1, p$

$< 0.001$ , therefore Greenhouse-Geisser corrected tests are reported ( $\epsilon = 0.58$ ). The results show that the scores between exams x03, x04 and Final were significantly different,  $F(1.17, 218.5) = 8.37, p = 0.03$ . As confirmation, the paired samples T-Tests showed a similar pattern to those for Group D, a significant difference in the two pairs (Pair 1:  $t(186) = -3.0, p = 0.003$ ; Pair 2:  $t(186) = 2.3, p = 0.02$ ).

Parameter	Module x03	Module x04	Final Anatomy	Repeated Measures ANOVA
Median	73.1	76.3	73.5	$F(1.19, 122.81)$
SD	10.8	11.8	9.7	$= 14.9,$ $p < 0.001$
Range	50 [40, 90]	67 [30, 97]	50 [42, 92]	
IQR	13.4 [65, 79]	16.8 [69, 86]	12.5 [68, 80]	
K-S Norm.	$D(104) =$ 0.09, $p = 0.04^*$	$D(104) =$ 0.07, $p = 0.19$	$D(104) =$ 0.09, $p = 0.04^*$	

**Table 21. Descriptive Statistics and comparisons of Anatomy Scores from Modules x03, x04 and the Final Anatomy score for Group D (Experiment).**

Note. The one-way Repeated Measures ANOVA used Greenhouse-Geisser estimates because sphericity was absent, according to Maunchly's test.  
\* = Not normal (K-S Norm. = The Kolmogorov-Smirnov Normality Test indicates normality when values are above  $p = 0.05$ ).

The analysis of results from the exams immediately before and after the study's training intervention shows that there was a significant difference in the overall change, determined by repeated measures ANOVA and in the comparison of pairs using the paired samples T-Test. These differences were apparent in the full cohort and in the experiment group in similar directions, which supports the notion that the latter was a representative sample of the former.

These results do not clarify if the intervention or the control groups were different from the full cohort, therefore do not support or refute the notion of self-selection bias that

resulted from voluntary participation in the experiment groups (see section 5.2 Student Participation for a discussion on this topic).

The reasons behind this general improvement in scores between x03 and x04; may include a combination of factors, for example: More efficient methods of teaching, improved studying strategies resulting from experience gained in previous modules exams, less complexity in the content of module x04, increased motivation as module x04 constitutes the last opportunity to improve the annual score.

△ **Section result 25:** Comparing academic performance based on the Anatomy scores in Modules x03, x04 and the Final score, there was a significant increase in scores from x03 to x04 for the full cohort ( $p < 0.03$ ) and the experiment ( $p < 0.001$ ) groups. Between x04 and the Final score there was a significant decrease in scores for both groups.

It would be difficult to conclude that the difference between scores of x03 and x04 represents a genuine academic improvement because the exams were different in content for each year of study. In addition, the comparison of those exams with the final score may lack validity, as the latter was not the result of an exam but a calculation based on previous modules.

#### **4.5.2.2 Scores in Anatomy Modules x03 and x04 for Groups A (Intervention) and B (Control)**

The previous segment demonstrated a significant improvement in scores between Module x03 and x04 for Group D (Experiment), which is composed of Groups A (Intervention) and B (Control). This segment focuses on the differences in performance

between the latter two groups; as mentioned earlier, this score was not the result of a knowledge exam, and therefore of a different type than the module exams.

Group A (Intervention) scored a median of 74.3 and 75 in Modules x03 and x04 respectively. This difference of -1.8 was not significant,  $t(28) = -0.91$ ,  $p = 0.37$ .

Group B (Control) scored a median of 71.4 and 76.3 in the same modules respectively; this difference of 5.57 was significant,  $t(74) = -4.22$ ,  $p < 0.001$ .

Comparing scores of Group A (Intervention) with those of Group B (Control), the differences in scores in x03 and x04 (0.83 and -2.8) were not significant ( $t(54) = 0.36$ ,  $p = 0.71$  and  $t(40) = -0.98$ ,  $p = 0.33$ , respectively).

△ **Section result 26:** The intervention group did not score significantly more in anatomy module x04 compared to x03; in contrast, the control group did improve significantly in x04 ( $p < 0.001$ ). Comparing one group against the other in each exam showed that there was no significant difference between them in either module. The absence of difference in scores between groups in x04 (post-training in spatial awareness), suggest that the intervention had no impact in academic performance.

The final scores for Group A (Intervention) and Group B (Control) ( $Mdn = 76$  and  $74.4$  respectively) did not differ significantly when compared to each other ( $t(39.3) = -0.86$ ,  $p = 0.39$ ). When compared to their own scores in module x04, group A (x04  $Mdn = 75$ ) scored less but not significantly ( $t(27) = 1.03$ ,  $p = 0.30$ ), while group B (x04  $Mdn = 76.3$ ) had a significant difference ( $t(75) = 4.06$ ,  $p < 0.001$ ). Due to the different sample sizes, it is likely that the larger control group had more chances to include students with lower marks that influenced the final score (T-Test compares means). The same could be applicable to Group E (Full Cohort), which had a similar trend as reported earlier.

#### 4.5.2.3 Comparison of Scores from Anatomy Modules x03 and x04 by Gender and by Group

The comparison of these scores had two facets: First, the comparison of scores from each gender from Module x03 to x04 within each group, and second, the comparison of scores from female and male participants between groups (Intervention vs Control). As all distributions were normal, the calculations required the paired groups T-Test for the first comparisons and the independent sample T-Test for the second.

The T-Test report stated that female participants in the intervention group obtained higher scores in module x04 (*Mdn* = 79.6) than in module x03 (*Mdn* = 75). This difference, -1.7 *CI* [-7.6, 4.2], was not significant  $t(16) = -0.6$ ,  $p = 0.55$ . Female participants in the control group obtained higher scores in module x04 (*Mdn* = 76.3) than in module x03 (*Mdn* = 70.7). This difference, -7.57 *CI* [-10.9, -4.23], was significant  $t(43) = -4.56$ ,  $p < 0.001$ .

The difference between groups in scores of female participants in module x03, 6.24, *CI* [-.354, 12.83], was not significant  $t(59) = 1.893$ ,  $p = 0.063$ . This difference in module x04, -0.678 *CI* [-6.67, 5.29], was not significant  $t(59) = -0.22$ ,  $p = 0.82$ .

Male participants in the intervention group obtained higher scores in module x04 (*Mdn* = 70.7) than in module x03 (*Mdn* = 70.6). This difference, -1.9 *CI* [-9.5, 5.7], was not significant  $t(10) = -0.5$ ,  $p = 0.58$ . Male participants in the control group obtained higher scores in module x04 (*Mdn* = 76.9) than in module x03 (*Mdn* = 73.1). This difference, -2.82 *CI* [-6.9, 1.3], was not significant  $t(31) = -1.38$ ,  $p = 0.17$ . Comparing between groups, scores of males in x03 and x04 were not significantly different ( $t(41) = -1.2$ ,  $p = 0.23$ , and  $t(41) = -1.38$ ,  $p = 0.31$  respectively).

Comparing median scores between genders for each group: For Group A (Intervention), females scored more than males in Module x03 (Fem = 75, Male = 70.6) and x04 (Fem = 79.6, Male = 70.7). These differences were not significant ( $t(26) = -1.6$ ,

$p = 0.12$ , and  $t(26) = -1.0$ ,  $p = 0.3$  respectively). For group B (Control), females scored less than males in Module x03 (Fem = 70.7, Male = 73.1) but were close in x04 (Fem = 76.3, Male = 76.9); these differences were not significant ( $t(74) = -1.4$ ,  $p = 0.15$ , and  $t(74) = -0.4$ ,  $p = 0.66$  respectively).

As a reference, the Final scores in anatomy were not significantly different from the scores in Module x04 when separated by gender (for the intervention and control groups). Comparing female vs male in the Final score also resulted in a non-significant difference.

△ **Section result 28:** The comparison of scores by gender in Anatomy modules x03 and x04 showed a significant increase for female participants in the control group. Most other comparisons showed a modest improvement that was not significant when tested in isolation, although when tested together as Group D (Experiment), the score improvement in x04 was significant (reported earlier). These results suggest that an important part of that change might originate in a significant improvement by females in the control group. The absence of significant changes in scores for Group A after training in spatial awareness (module x04), suggests that such an intervention had no impact on academic performance.

#### **4.5.2.4 Comparison of Scores in Module x04 for Group D (Experiment) by Year of Study**

This section looks for evidence of impact of the training intervention on academic performance in Anatomy by year of study. Group D (Experiment) had 73 students from Year-1 and 31 from Year-2; each year took their respective Anatomy Module test (104

and 204 respectively), which took place after the study's intervention and second spatial ability test.

The median score in Module x04 for Year 1 students was 80.3 while Year-2 students had 72.2. Although one of the data sets was not normal, the number of participants justified an assumption of normality and therefore, the use of parametric tests.

The Independent Samples T-Test reported that in Group D (Experiment), participants from Year-1 obtained higher scores in Anatomy Module 104 than Year-2 students in Module 204. This difference, 9.1, BCa 95% CI [4.69, 13.41] was significant  $t(102) = 3.82, p < 0.01$ .

Analysing subgroups A (Intervention) and B (Control) by year of study, Year 1 students in the Intervention group scored a median of 72.1 in x03 and 80.9 in x04, and those in Year 2 scored medians of 75.7 and 74.1. In the Control group, Y1 students scored medians of 71.2 and 80.3 respectively, while Y2 students scored 72.1 and 67.1. The comparisons of scores of the Intervention vs Control groups in x03 and x04 separated by year of study yielded no significant differences.

These results suggest that there was no significant difference in performance in anatomy modules x03 and x04 when divided by year of study within the intervention and control groups, but when assembled as Group D (Experiment), Year 1 students showed a significantly higher score in anatomy than Year 2. As a reference, the comparison of scores for Group D in x03 by year of study also showed a significant difference, although in the opposite direction, as Year 2 students scored more. This variability may reflect the differences in content of each anatomy module and the variability in difficulty.

There was no significant difference in scores after the training intervention for either year of study.

#### **4.5.2.5 Comparison of Final Anatomy Scores by Spatial Ability Score**

##### **Quartiles**

Dividing participants into quartiles based on their first spatial ability score showed that those in the top quartile obtained the highest Anatomy Final scores (Q3,+∞:  $n = 55$ ,  $M = 73.7$ ), followed by the high and low quartiles (Q2,Q3:  $n = 89$ ,  $M = 71.3$  and Q1,Q2:  $n = 38$ ,  $M = 71.1$ ). The bottom quartile obtained the lowest scores (-∞,Q1:  $n = 5$ ,  $M = 50$ ).

The ANOVA test showed a significant difference in the final anatomy scores of those quartiles,  $F(3) = 8.1$ ,  $p < 0.001$ . The bottom quartile scores were significantly different to all other quartiles,  $p < 0.001$ , which were not different from each other.

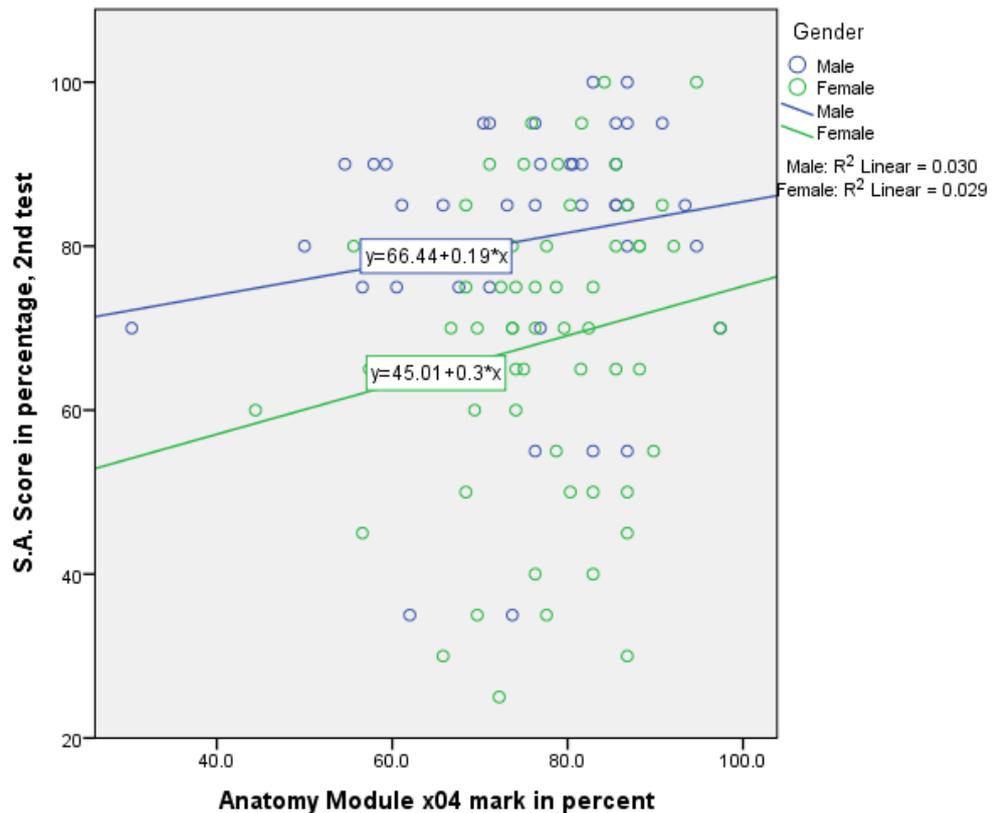
These results show that the relationship between spatial ability and final anatomy scores appears to be direct, suggesting a predictive value of spatial ability, identifying those in the low quartile (see section 5.11 Academic Performance).

#### **4.6 Correlations between Spatial Ability and Academic Performance**

The assumption of normality based on sample size for the spatial ability data coincides with the reported normality of the scores in Anatomy Modules. The recommended test to assess correlation is the "Pearson's  $r$ " or product-moment correlation. As mentioned in section 3.3.8 Analysis of Academic Data, this analysis will also include a non-parametric test (Kendall's tau) to confirm results.

Looking for relationships between Spatial Ability and Anatomy scores for the experiment group, the analysis focused on the second spatial ability test and module x04 because both took place after training in spatial awareness. Results include calculations by gender and by sub-group (Intervention and Control).

The grouped scatterplot correlating Spatial ability and Anatomy scores (post-training) showed no clear relationship; the regression lines of both genders were slightly inclined with a weak positive value for both genders (Figure 10)



**Figure 10. Grouped Scatterplot of Correlation between Scores in the Second Spatial Ability Test and Anatomy Module x04 for Group D (Experiment) by Gender.**

Note. Grouped scatterplot shows no evident correlation between spatial ability 2nd test and module x04. Regression lines of male (blue) and female (green) participants show similar inclination, which is slightly higher in males.

The Pearson's correlation reported that the relationship between scores from the second spatial ability test and anatomy module x04 for Group D (Experiment) was not significant,  $r = 0.122$ ,  $p = 0.21$ . Non-parametric Kendall's tests confirmed this result,  $\tau = 0.10$ ,  $p = 0.14$ .

The separation of Group D (Experiment) by gender also showed non-significant correlation (female:  $r = 0.17$ ,  $p = 0.19$ ; male:  $r = 0.17$ ,  $p = 0.27$ ), as did the analysis by subgroup (Intervention:  $r = -0.003$ ,  $p = 0.99$ ; Control:  $r = 0.19$ ,  $p = 0.99$ ).

As a reference, there was no relationship between scores in the first spatial ability test and module x03, both pre-training events, for Group D (Experiment),  $r = 0.31$ ,  $p = 0.75$ . This was also the case for each sub group (Intervention:  $r = 0.06$ ,  $p = 0.74$ ; Control:  $r = 0.02$ ,  $p = 0.84$ ).

△ **Section result 31:** The correlation analysis between scores in spatial ability and anatomy for Group D (Experiment) was not significant in any of the tests, which included data sets before and after spatial awareness training, even when separated by gender or by subgroups Intervention (A) and Control (B). These results suggest an absence of relationship between spatial ability and anatomy scores, which seem to be independent of one another

## 4.7 Additional Tests: Manual Dexterity and Stereopsis

In addition to testing in Spatial Ability and Training in Spatial Awareness, participants in the intervention group underwent testing in Manual Dexterity and Stereopsis.

As mentioned in section 3.2.4.1, Research Approach and Design, this section corresponds to a “one group pre-test/post-test” design that lacks a control group, which weakens its validity. Consequently, this report only covers a summary of highlights.

## Manual Dexterity Test Results

The goal with the measurement of manual dexterity was to determine if training in spatial awareness had an impact on manual dexterity performance. Although there were no reports of this type of testing in the literature, there were mentions of the association of good special ability in people with good manual skills, which in the medical field may apply to surgical skills.

Manual dexterity, tested using the Purdue Pegboard Test, produced a set of data that showed high reliability using the "Test-Retest Reliability" method, based on Pearson's Correlation Coefficient, which was significant,  $r = 0.714$ ,  $p < 0.001$ . Based on this result validity was also high at 0.844 (see section 3.3.6.2 Reliability of the Manual Dexterity Test).

Participants obtained higher scores on the second test ( $Mdn = 102$ ) than on the first ( $Mdn = 92$ ), this difference,  $-9.87$ ,  $BCa$  95%  $CI$  [13.16, -6.8], was significant,  $t(44) = -2.36$ ,  $p = 0.02$ . When separated by gender, both scored higher in the second test, with a significant difference (females:  $t(17) -4.3$ ,  $p < 0.01$ ; males:  $t(13) -4.2$ ,  $p < 0.01$ ). There was no significant difference between genders in both tests. Scores in Manual Dexterity did not correlate significantly with scores in Spatial Ability or Anatomy before or after the intervention. There was no correlation with hand dominance. The only significant correlation related the second Manual Dexterity test to the Final anatomy score for female participants (Pearson's  $r = 0.533$ ,  $p = 0.02$ ). Despite significant improvement in manual dexterity on the second test, in the absence of a control group, it is not possible to reach a tentative conclusion regarding the relationship between spatial awareness training, spatial ability and manual dexterity. It is likely that a learning effect influenced this improvement.

## **Stereopsis and Visual Performance Results**

Participants in the intervention group scored the average equivalent to full stereopsis (0.94,  $SD = 0.16$ ) and visual acuity (1.03,  $SD = 0.19$ ), with respect to the standards in the average population. The minimal variability in values resulted in no significant differences between groups.

## **4.8 Limitations**

This section presents situations that had a detrimental effect on the study or presented an obstacle to the progress of the project.

### **4.8.1 Research Modality and Sample Size Limitation**

At the beginning, the project aimed at testing spatial ability and manual dexterity in Year 1 students at BSMS to determine the effect of training in spatial awareness.

This study originally planned to use a randomised control trial design, involving a sample of participants randomly assigned to training in spatial awareness. A set of pre and post intervention tests were to produce evidence of impact of this intervention. The goal of 80% power required a sample size of 26 to 50 students per group, meaning that the study needed to recruit a minimum of 52 participants to begin randomly assigning groups (see section 3.3.2.1 Population Sampling Technique). A few weeks into the study, it became clear that this goal was unreachable as hardly a dozen students joined. Doubling the target population by including Year 2 students was barely enough to reach half of the minimum sample to randomise. The research design required changing to “quasi-experimental” with acceptance of the limitation of self-selection bias

because all those who voluntarily signed up to the experiment became the intervention group. The same bias applied to the control group, which emerged from those who volunteered to re-take the spatial ability test.

Despite reaching the theoretical sample size, a post-hoc calculation showed that the study reached a power below 50%, which showed that the inability to recruit sufficient participants was an important contributing factor in the lack of significance. This constitutes a potential Type II error, where there is failure in detecting the true impact of an intervention, mainly caused by an underpowered study.

A more severe limitation affected the intervention group. There was no control group for the manual dexterity test evaluation because everyone in the group underwent testing. The research design “one-group pre-test/post-test” allows comparisons but does not determine the presence of an effect.

#### **4.8.2 Academic Score Limitations**

Because the project did not include teaching or examining anatomy, the only measure of performance in anatomy were the scores of knowledge test provided by the medical school. As part of a modern trend of teaching in medicine, the modular approach to teaching is more inclusive, blurring the distinction between anatomy and related subjects that students learn in first and second years. Therefore, knowledge tests contain a combination of questions that include anatomy and other subjects like physiology, embryology, neuroscience, etc. Students have to prepare for all those areas as well as for anatomy. Moreover, each module teaches different regions of the body, and each year has a different programme, some modules are more complex, which makes some exams more difficult. The scores used in this study represent the anatomy portion of the knowledge exams. In an ideal scenario, anatomy would be a

separate test; or even more ambitiously, a study like this would involve teaching and testing a portion of the anatomy curriculum for a better control of the variables.

#### **4.8.3 Time Limitations of Medical Students**

Medical students at BSMS have a timetable of academic activities that includes lectures, practical labs and time allocated for sports, which covers most of the study hours from 9am to 5pm. This project aimed at offering participation within those study hours without interfering with their academic activities, using part of their non-scheduled time.

This limited time appeared to be the biggest obstacle to participate in the project; students would commonly use these gaps for personal study and revision. Despite the efforts to shorten the time required for testing and training, students that joined the intervention group had to commit to attend four hours on separate occasions for testing and training. Sessions were on a one-to-one basis because there was only a single researcher to run the tests.

#### **4.8.4 Time Limitations of Training**

One of the most important parts in the study was the Spatial Awareness Training programme, which was the factor being tested. The duration of this programme had to fit within the students' time constraints. Consequently, one of the challenges for the researcher was to filter out non-essential elements, retaining only those relevant to medicine and keeping in line with the spatial ability test, in addition to optimising the teaching method to include practice, recall and multi-sensorial stimuli (see the Methods section for details). Time restrictions allowed for two one-hour sessions, a week apart, which was preferred to a single hour session because of the importance of repetition

and practice in learning. Under these circumstances, the expectation from the beginning was that the effect of training would be smaller than the full nine-week programme that inspired the project. The question was, if this very short intervention would have any impact on the participants' performance. The results provided a clear answer. Two hours of training were not enough to produce a significant difference in spatial ability or anatomy performance, compared to those who did not receive training. The training sessions seemed to provide a memorable experience for some students. In a casual conversation with the researcher, a year later, two of the students from the intervention group, remembered the training sessions; one of them bought a set of "Goldie-Blocks", a spatial ability training game, for her daughter, after hearing from it in the training talk.

#### **4.8.5 Time Limitations of the Spatial Ability Test**

The selection of the spatial ability test was an important factor in the success of the project. One of the important goals was to find a test that would fit the time limitations of medical students (see above).

Among the many tests in spatial ability, one that stood out for its reliability and validity was the Visualization of Rotations test (ROT), which consisted of 20 questions, with a time limit of 10 minutes. It was ideal for group testing in a classroom (see section 3.2.7 Spatial Ability Testing).

One of the limitations of a paper test in a large group was that monitoring was difficult; consequently, some students ignored a number of fields in the answer sheets, which in some cases resulted in exclusion of the participant from parts of the analysis and inability to follow for re-test or correlation with anatomy scores.



## CHAPTER 5 DISCUSSION

The summary of results for discussion congregates the most salient points from the results chapter. The subsections that follow will provide a more in-depth discussion of each topic.

### 5.1 Summary of results

The proportion of female participant (56%) was close to that of medical students in the last two years (58%) and the median age for both genders was 20 years (range 18 to 45). The study confirmed the reliability and validity of the spatial ability test (20 question ROT) as reported in the literature.

The median score in spatial ability of all participants ( $N = 194$ ) was 70%, with males showing a significant 10% score advantage ( $p < 0.01$ ). Most participants (81.1%) attempted all 20 questions of the test within the allocated 10 minutes of the trial. The analysis of questions in the test showed an incremental difficulty from the initial to the last questions. Marking the test by the difficulty value of each question, scores were lower (by 5.8%) but relationships within participants and by gender remained. The experiment group ( $n = 104$ ), obtained similar scores as the full cohort on the first spatial ability test and improved significantly on the second test (difference 5.7,  $t(103) = -4.33$ ,  $p < 0.01$ ). The subgroups, intervention ( $n = 29$ ) and control ( $n = 75$ ), scored similarly in the first and second spatial ability tests. The latter being a post-training in spatial awareness test for the intervention group. The spatial ability score improvement in females was only significant for those in the intervention group ( $p = 0.03$ ), not for those in the control group ( $p = 0.1$ ). The former were able to answer more difficult questions the second time around. This difference was not evident for males, as both sub-groups,

intervention and control, improved significantly (Int  $p < 0.001$ ; Ctrl  $p = 0.02$ ). Females scored significantly less than males in both tests in the intervention group (1<sup>st</sup>  $p = 0.005$ ; 2<sup>nd</sup>  $p = 0.001$ ), while only in the second test in the control group (1<sup>st</sup>  $p = 0.056$ ; 2<sup>nd</sup>  $p = 0.025$ ). There was no difference in performance related to the year of study. Although the sample size calculation before the study aimed at 80% statistical power, a post-hoc calculation reported a power of 0.51.

Anatomy scores of the full cohort improved significantly comparing modules before to after the intervention, and worsened compared to the final anatomy score. There was no difference between intervention and control groups in any of the anatomy scores, although only females in the control group showed significant improvement between modules x03 and x04 ( $p < 0.001$ ). Participants with scores in the bottom quartile in spatial ability, did poorly in the final anatomy score compared to the rest ( $p < 0.001$ ), aside from this comparison, no correlation between scores was apparent.

The pre-test / post-test single-group additional study in the intervention group showed significant improvement in the post-intervention manual dexterity test ( $p = 0.02$ ) with no difference between genders and no correlation with anatomy scores.

## 5.2 Student Participation

The results of the main study revealed a substantial participation of medical students in the project. The majority of Year 1 students and near to half of Year 2 students at BSMS took part of the spatial ability test at least once. This level of participation might be the result of the strategy adopted to invite students to take the spatial ability tests.

Combining the presentation of the project and offering the test at the end of a class captured the attention of most students who volunteered to take the test, many of whom also agreed on a second occasion. The 10-minute duration of the test, probably made it easy to accept. In contrast, recruiting a sub-group from the above cohort for the

training experiment was quite a challenge. Participants had to commit to attend four one-hour appointments on various dates, which probably interfered with their revision or free time (see 4.8.3 Time Limitations of Students).

Response to recruitment efforts was poor, even with the addition of Year 2 students. Eventually, enrolment finalised in time to train and re-test the intervention group, reaching a sample size ( $n = 31$ ) close to the lower end of the sought range (26-50). The post-hoc calculation reported 51% power, which is far from the goal of 80%, which shows that the sample size was insufficient and the study underpowered. A review of the literature showed that underpowered studies are not uncommon, even reputable data-sources like “Cochrane Reviews”, seem to include around 70 % of underpowered sources (Turner, Bird and Higgins, 2013). Aside from increasing the sample size, some propose other methods to improve power, e.g. optimising experimental efficiency and control, and recruiting larger samples with multi-site collaborative studies (Maxwell, 2004), a consideration for future studies.

The lack of randomisation and the adoption of a “quasi-experimental” design were two frustrating factors that prevented control over important variables. Some researchers resort to accept this modality because in real life, true experimental designs are not always possible (Field and Hole, 2003). In addition, accepting self-selection bias diminishes the validity of the results, as there is a chance of misrepresenting a population by testing participants who are better than the average in that area, e.g. spatial ability (Fernandez, Dror and Smith, 2011).

Like this study, many others close without reaching a sufficient sample size due to overly optimistic recruitment expectations, time limit, dropouts or financial limitation, which may lead to seemingly important findings but with statistically insignificant effect sizes. A feasibility question that would ideally be apparent in the planning stages of the project (Olivier and Bell, 2013), an important consideration for future studies.

### 5.3 Student Gender Distribution

Since there was an important female representation in the population of this project, gender became a factor of interest in the analysis of data.

There was a larger participation of female students in the study (56%), in a similar proportion to the gender composition of students at BSMS, where the student intake from 2013 to 2015 included an average of 58.7% female participation (BSMS, 2015b). The General Medical Council (GMC) reported a similar percentage of females (55%) in the population of medical students in the UK in 2015, which was close to the rate of female doctors progressing to post-graduate foundation training (57%) (Brown *et al.*, 2015).

The training intervention used in this study, originally designed as a Spatial Awareness Training course, aimed at improving the spatial skills of those lagging behind, who were mostly women (Sorby and Baartmans, 2000; Hegarty *et al.*, 2010; Morrissey, 2017). A very short version of that course, adapted to medicine, was the basis of the training in this project.

Spatial ability has relevance in the scientific field, including health sciences and medicine. A study reported that medical students, who had borderline and failing scores in anatomy in the first and second years, were also incompetent in spatial ability and answered incorrectly most questions classified as having “spatial content”.

Although the results were statistically significant, the author did not recommend using spatial ability testing as a predictor of academic performance (Rochford, 1985).

Hegarty *et al.* indicated in 2007, that spatial ability formed part of the admission test in dentistry, which supports the recognition of the importance of this ability in a field related to medicine. In the paper the authors also mention that the advantage that good spatial ability provides to students in early years diminishes later, with the acquisition of the required medical skills (Hegarty *et al.*, 2007).

The following spatial-related competencies affects both genders, as they are part of the medical practice in the educational and professional realms, which are increasingly blending into one another:

- Spatial reasoning used to understand and learn complex structures of human anatomy e.g. training on simulators and in virtual environments (Hegarty *et al.*, 2007).
- Spatial visualisation required to practice diagnostic and therapeutic procedures, e.g. palpation, auscultation, percussion, phlebotomy.
- Spatial orientation required to comprehend and interpret medical imaging for diagnosis and treatment, e.g. CT, MRI, X-Ray, Ultrasound.
- Spatial orientation and manoeuvring needed during the performance of surgical procedures, particularly when using indirect viewing methods like microscopes, monitor screens and 3D goggles for minimal access (keyhole) surgery or robot-assisted surgery (tele surgery or remote surgery) (Abe *et al.*, 2017). These techniques are gaining popularity and may become the rule rather than the exception in the future of medicine (Morris, 2005).

In addition, the exploration of skills with potential to improve learning and academic performance in medical students was in line with the teaching aspirations at BSMS and other higher education institutions. These aspirations coincided with those of the GMC: To promote the highest possible standards of intellectual and personal development, and to improve the quality of education and training (GMC, 2016b; BSMS, 2017a; Stanford, 2017).

## **5.4 Student Experience with Testing**

One of the goals of the study was to determine the innate level of spatial ability in medical students, which ideally required no previous experience of this test.

The literature reports that short variations of the spatial ability test may form part of generic psychometric test batteries, sometimes used in job aptitude scenarios and career counselling, mostly in relation to engineering, architecture or military careers, in which this ability is important and may appear in admission assessments (Council, 2015; Cambridge, 2016).

Medical schools set their own entry requirements and the majority use an admission test (UKCAT, BMAT, GAMSAT)(MSC, 2017a). Currently, Brighton and Sussex Medical School uses the Biomedical Admissions Test (BMAT) as a pre-selection assessment for applicants (MSC, 2017b); this test does not include spatial ability questions but has sections on mathematics, physics and chemistry that may require spatial thinking (Assessment, 2017). Most students also take the University Clinical Aptitude Test (UCAT, former UKCAT), which includes spatial awareness questions under the section of Abstract Reasoning (Medical Portal, 2019; WikiJob, 2019).

Ideally, Year 1 students would have no previous experience with spatial ability testing, but considering that most medical students experienced the UCAT, it is reasonable to think that they had some exposure, which, as long as it applies to most students, only provides an advantage if compared to other populations (see section 3.2.7 Spatial Ability Test).

## **5.5 Discussion on Reliability and validity findings for the Spatial Ability Test**

The spatial ability test selected for this study had reliability and validity values reported in the original publication (Bodner and Guay, 1997). Verification of those tests provided high levels of reliability in this study. Correlations between questions and values showed that some of the last questions of the test (later proved to be the most difficult), had poor correlation with the first ones, demonstrating an absence of homogeneity,

which was a desirable feature in this case (the first part of the test was intentionally easier). Other study reported high reliability in their study of the original 30-item PSVT:R test, and agreed that no questions were worth removing (Maeda and Yoon, 2011), which supports the findings of this study.

The high reliability and validity of the PSVT: R test found in testing supports the adequacy of this test meeting time restrictions and the need for specificity to test spatial ability in this population. This test gained popularity in the research community since its design three decades ago; another testament to its adequacy as a testing tool for spatial ability. It stands out because it depicts objects with a combination of surfaces: inclined, oblique and curved, compared to other tests that only have cubic surfaces. The test also concentrates on rotation, which is more relevant to practical use than other aspects of spatial ability (folding, matching shapes, flat rotation) (Maeda and Yoon, 2011). Various researchers reported that compared to other variables investigated, the PSVT:R test was the most significant predictor of success in subjects that focused on 3-D structure manipulation, like engineering graphics, which has an affinity with anatomy because both use object visualisation (Gimmestad, 1990; Sorby, 2009; Friess *et al.*, 2016).

## **5.6 Spatial Ability Test Related Factors**

### **5.6.1.1 Test Duration**

The original rotation test (PSVT: R) underwent shortening from 30 questions in 20 minutes, into 20 questions in 10 minutes, mainly to restrict analytical thinking (Bodner and Guay, 1997). The authors make a distinction between Analytical Thinking and Gestalt Processing. Analytical thinking refers to the mapping of individual parts of the image of an object to a second image, discovering its similarities by analogy; whereas

Gestalt, refers to the formation of a concept of an object as an organised whole, recognisable from different perspectives. This test maximises Gestalt processing while minimising analytic processing (Macfarlane Smith, 1964; Bodner and Guay, 1997).

Results from the analysis of attempted questions in the spatial ability questionnaire showed that a majority of students (89%) attempted between 17 and 20 questions. The least number of questions attempted was 10. These results suggest that the 10 minutes assigned to complete the 20 questions of the test were sufficient for students to attempt most questions of the questionnaire.

On the day of the test, introduction, instructions and distribution of the question booklet and answer sheets took around seven minutes, followed by the 10-minute test, making a 17-minute event (see section 4.8.5 Time Limitations of the Spatial Ability Test).

### **5.6.1.2 Test Difficulty**

This study analysed the difficulty of questions in the ROT test, based on which questions were the most commonly answered ( $n = 159$ ). The resulting indexing method allowed for a new sequencing of questions and a new way of marking the tests; neither of which provided a clear advantage over the original format (see 4.3.6 Measure of Difficulty of the Spatial Ability Test). Maeda & Yoo tested the 30-item PSVT:R test (precursor of the ROT) and found that it was generally organised from easy to difficult. They assigned an index of item difficulty to each of the questions (from -2.87 to 0.64) and rated the overall test as easy, based on the average difficulty (-1.45) (Maeda and Yoon, 2011). This position, based on statistics, may not coincide with the opinion of the participants in this study, informal commentaries by participants tended to focus on the surprising difficulty of the last questions, which resulted in a frustrating experience for some (Students, 2016).

It is possible that the contrast of easy questions at the start and difficult ones at the end, together with the time restraint, intensified the levels of anxiety induced by the test, enhancing the perception of complexity. Test anxiety in medical students when facing multiple-choice questions, can be debilitating and affects those with vulnerable personalities (Powell, 2004). Female medical students tend to be more vulnerable to stress than males, despite being more resilient, abstract and adaptive in their personalities than the general population (Meit, Borges and Early, 2007). These personality traits may affect their performance during tests.

## **5.7 Spatial Ability Scores of All Participants**

The average spatial ability score obtained by all participants (65.5%;  $N = 196$ ) is above the considered pass rate of 60% (Czapka, Moeinzadeh and Leake, 2002). Its significance depends on reports from studies of similar populations.

Studies from 1990s using the ROT test, found that students of chemistry/health-science scored an average of 60.9%, science/engineering 69.5% and pre-medicine/biology 70.8% (Bodner and Guay, 1997). A comparison of PSVT:R test results from 10 selected reports between 1985 and 2006 found that for engineering, the average score was 67.21% (Ault and John, 2010). Those references place the students in this study, close to the average in engineering and below those in pre-medical courses in USA, two decades ago.

A more recent study by Vorstenbosch et al reported that 242 first year medical students scored 72% (14.4 of a maximum of 20 points) in spatial ability using the Mental Rotation Test (MRT) (Vorstenbosch *et al.*, 2013). This reference score is higher than the one obtained by students at BSMS. Reasons for this disparity are difficult to determine, the most evident difference being the test type. The MRT tests rotations around a single axis and the drawings do not have hidden parts, while the ROT

presents multiple axes of rotation and parts of the images are not visible (Bodner and Guay, 1997), features that may increase the difficulty of the latter.

## 5.8 Spatial Ability Scores of Group D (Experiment)

When compared to Group E (Full Cohort), Group D (Experiment) obtained similar results on the first spatial ability test, which supports the assumption of group D being representative of the full cohort.

Group D scored significantly more on the second spatial ability test ( $p < 0.01$ ), which was promising as this group was composed by the intervention and control groups. There was no difference between these groups in either test, which ruled out the effect of training. An explanation of the improvement of Group D pointed at practice effect, which was possible, although similar studies gathered practice effect as minimal, explaining 0.2% of the improvement after three months (Sorby and Baartmans, 2000).

It was important to take into account that all participants in group D were studying anatomy, a known factor in the improvement of spatial ability in medical students (Hoyek *et al.*, 2009; Lufner *et al.*, 2011b; Vorstenbosch *et al.*, 2013), which would partially explain the improved average score of all students tested.

The lack of significance in the difference between the intervention and control group after training, indicated a lack of impact of the training intervention. In addition, the post-hoc power calculation suggested that the study was underpowered (51%), both results likely linked to a small sample size (section 4.4.2 Analysis of Comparisons). Other factors that negatively affected the validity of the study were the lack of randomisation and insufficient training time. Some of these elements affected a similar study, which reported improvement in mental rotation in medical students after four months, favouring males; they also suspected a self-selection bias due to voluntary participation and thought that learning anatomy played a role in the improvement. In

addition, they tested all participants who volunteered to the intervention group; a control group was not possible because all medical students were taking the anatomy course. (Lufler *et al.*, 2011b). Aside from the similarities, this study found no correlation between mental rotation scores and high scores in anatomy (see section 5.11 Academic Performance).

Reported training programmes in spatial ability had positive results. The programme that inspired this study had a ten-fold contact time (40 hours) in 10 weeks, resulting in 26.4% improvement in post-intervention scores. The shortest training time reported was four weeks, also with positive results (Sorby, 2009). A meta-analysis of 2017 studies categorised training in three types: course, video game and spatial task; all of which produced improvement in spatial ability scores for both genders (maintaining a gender difference favouring males) (Uttal *et al.*, 2013).

In view of the literature evidence of the impact of training and learning from the insufficient effect found in this project, future studies might benefit from extended training time and a different research approach to achieve a true experimental design.

## **5.9 Gender Characteristics in Spatial Ability**

In Group E (Full Cohort), males outperformed females significantly ( $p < 0.01$ ) by an average of 2 points. Incorporating the second test, the gender difference was also significant in the intervention group (pre, and post intervention:  $p = 0.005$ ,  $p = 0.02$ ) and the control group (pre, and post intervention:  $p = 0.056$ ,  $p = 0.025$ ). These findings agreed with the literature (see below) and answered the first research sub-question of the study, and reject the second null hypothesis (see 3.2.3.1 Hypotheses of the Study, and 4.3.3 and 4.4.4 Comparisons of Scores by Gender for the Full cohort and the experiment group respectively).

The number of participants in the intervention group ( $n = 29$ ) was close to the lower value sought to obtain 0.8 power. Although it was sufficient to reveal major significant differences, like those related to gender, it was insufficient to show smaller differences, possibly including an effect of the training intervention, therefore undermining gender and overall results (see 5.2 Student Participation).

Spatial ability is an area known for persistently showing that males obtain better scores than females. Reports appeared since 1958, with support from many researchers (Anastasi, 1958; Macfarlane Smith, 1964; Maccoby and Jacklin, 1974; McGee, 1979b; Nyborg, 1983; Linn and Petersen, 1985; Voyer, Voyer and Bryden, 1995; Butler *et al.*, 2006; Vuoksimaa, Kaprio and Kremen, 2010; Heil *et al.*, 2011). Only a few reported no difference (Fairweather, 1976; Sherman and Fennema, 1978; Lord and Garrison, 1998; Puts *et al.*, 2010). Consequently, most studies in spatial ability have to engage in discussing gender characteristics (see section 2.3.1.5 Gender in Spatial Ability).

The designers of the ROT test also reported higher scores in males by academic field; the difference in health-sciences was 2.85 points, in engineering 2.47 points and in pre-med 1.44 points (Bodner and Guay, 1997). A meta-analysis by Maeda and Yoon from 2012, based on 40 selected studies that used the PVST:R test, also found a mean difference in scores by gender favouring males of 2.32 points (11.6%), which is in line with the study above (Maeda and Yoon, 2016).

This study did not account for hormonal factors that may affect spatial ability performance. Studies reported significant improvements in mental rotation test scores during menstruation, when oestrogen levels are low (Hampson, Kimura and Thompson, 1988; Silverman and Phillips, 1993). Others described monthly performance cycles in men and women with more variation in scores in women (20%) than in men (10%); men performed better, when testosterone was lower, while women did when testosterone was higher. High levels of oestradiol matched the lowest scores in spatial ability cycles in women, which supports Silverman's report (above) (Lord and

Garrison, 1998; Courvoisier *et al.*, 2013). Testosterone has two types of influence: a. Organisational or prenatal, affecting in utero growth with permanent effects, and b. Activational, which depends on transient blood levels of the hormone. Prenatal testosterone influences male traits, including the length of fingers, characteristically making the 4th finger longer than the 2nd (a low 2D:4D ratio of less than 1), which relates to higher mental rotation scores. As reference, females have no exposure to prenatal testosterone and commonly have a high 2D:4D ratio (shorter 4th digit) (Manning and Taylor, 2001). Females exposed prenatally to testosterone, e.g. cases with a male twin, score better in mental rotation than females who had a female twin (Vuoksima, Kaprio and Kremen, 2010; Heil *et al.*, 2011; Bütikofer *et al.*, 2019). Transient increased of blood testosterone following a single dose, showed an increase in mental rotation scores and orientation tasks success in women (Pintzka *et al.*, 2016). In contrast, a report on a large cohort of 160 women and 177 men, challenges the above by claiming that circulating testosterone does not predict mental rotation in either sex (Puts *et al.*, 2010). Although the influence of hormones in spatial ability is likely to be minimal, this area may gain more interest in the future.

## **5.10 Training and Learning**

The results showed that training in spatial ability did not improve participants' performance in spatial ability or anatomy.

The training intervention in this study consisted of a very short adaptation (2 hours) of a training programme in spatial awareness designed by Sorby and Baartmans in 1993.

That version consisted of a 40-hour course delivered in 10 weeks, aimed at engineering students. Their results after 6 cycles of the course (over 500 students in 6 years) were clearly positive, with an improvement of 26.4% in post-training scores.

They selected the PSVT:R spatial ability test (precursor of ROT), because it was more

sensitive to changes compared to other tests. They re-tested after 3 months and thought that the practice effect was negligible (0.2) in relation to the net gain (10 times larger, from 1.7 to 2.6 points) (Sorby and Baartmans, 2000; Sorby, 2007). Other researchers reported significant improvements in rotation scores after training Physical Education students in spatial ability (they were learning anatomy); their course was 12 hr over 3 weeks on a small sample ( $n = 32$ ) (Hoyek *et al.*, 2009). Both studies had considerably longer training times and larger population samples, which clearly providing enough statistical power.

Preparing the training lectures required care to ensure the best use of the short teaching time. With that goal, the training sessions included theory and practice using manual activities. Most of the diagrams in the presentation were self-explanatory and in demonstrating sequences, they showed a systematic approach, minimising the cognitive load of the learner by dividing a process in small sections (e.g. notation for rotation based on 90-degree steps). For practice, learners made physical models with plastic blocks to recreate each rotation step, which reduced the demand of mental work and enforced the correlation between viewing a 2D drawing and handling a 3D model. This method aimed at reducing the extrinsic load presented by the teaching material and minimising the element interaction through segmentation of the rotation process, in hope to make it easier to understand and remember (Sweller, 1994) (see Cognitive Load Theory in section 1.5.1 Undergraduate Educational Context).

Training and testing in anatomy, provided by BSMS staff, were outside the control of the study.

## **5.11 Academic Performance**

There was a significant improvement in scores in module x04 compared to x03 for both groups (intervention and control). Although the comparison with the final score also

showed a significant difference, the comparison was invalid because the final score was the result of a calculation based on module scores, not an exam.

The comparison between groups showed no significant differences, particularly for the intervention group after training, which supports the null hypothesis (no effect of the training intervention) (see 3.2.3.1 Hypotheses of the Study).

This study found no correlations between spatial ability and anatomy scores. Some researchers reported a significant correlation between spatial ability and functional anatomy scores, and although they used a different test (MRT and others), they tested larger samples, obtaining sufficient power (Garg, Norman and Sperotable, 2001; Guillot *et al.*, 2006; Fernandez, Dror and Smith, 2011; Lufler *et al.*, 2011a; Vorstenbosch *et al.*, 2013; Berney *et al.*, 2015). Other researchers considered low scores in spatial ability tests predictive of anatomical spatial weakness because of their association with low scores in spatial-anatomy questions (Rochford, 1985). Meanwhile, others described an association with practical anatomy exams but not with non-spatial anatomy questions (MCQs) (Langlois *et al.*, 2017). Although no correlations between spatial ability and anatomy scores were apparent in this study, the comparison of the final anatomy scores after dividing participants into quartiles based on their scores in spatial ability, showed an almost direct correspondence. Those in the top quartile score the most and the few in the bottom quartile scored significantly less (see section 4.5.2.5 Comparison of Final Anatomy Scores). Those findings support reports from the literature that give spatial ability a predictive value (Lufler *et al.*, 2011a). Identifying students with weak performance early in their studies might provide an opportunity to provide extra tuition.

The difference in content of the anatomy components negatively affected the validity of this analysis because it goes against the ideal of giving all participants the same test, compounded by including two different courses.



# CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

## 6.1 Conclusions

The spatial ability score of the full cohort of participants (Y1 and Y2 medical students) was above the accepted pass level in the general population, placing those tested, between engineering and pre-medicine undergraduates (see section 5.7 Spatial Ability Scores of All Participants).

In relation to the main research question: There was no significant evidence to indicate that training in spatial awareness had an impact on spatial ability or anatomy performance (see section 3.2.2.1 Research Question). This answer requires some clarifications. First, training was a very short adaptation of a programme known to be effective. Second, due to the small number of participants in the experiment, the study was underpowered. The combination of those two points suggested a Type II error situation. A point that future studies might be able to address with a larger sample size to obtain sufficient power (see section 4.8 Limitations).

Overall, there was a significant improvement in spatial ability scores on re-testing, perhaps beyond the influence of a practice effect, and clearly not a result of training. This suggested the possible effect of studying anatomy, a factor reportedly linked to improvement in spatial ability (see section 5.8 Spatial Ability Scores of Group D (Experiment)).

Regarding the first sub question, in the comparison of scores in spatial ability, anatomy and manual dexterity by gender. Males significantly outperformed females in spatial ability, a difference absent in anatomy performance and manual dexterity (see sections

3.2.2.2 Sub Questions of the Study, 5.9 Gender Characteristics in Spatial Ability, 5.11 Academic Performance and 4.7 Additional Tests).

Answering the second sub question involved correlation analysis, mainly between spatial ability and academic scores, where no correlation appeared, possibly a reflection of lack of power. However, the comparison of spatial ability scores by quartile showed that the few participants in the bottom quartile had the lowest final scores in anatomy, a significant difference that suggest a predictive quality of the spatial ability test (see section 5.11 Academic Performance).

For the third sub question, the comparison of manual dexterity scores before and after the intervention, produced no significant difference, which had limited validity, due to the absence of a control group (see sections 3.2.2.2 Sub Questions of the Study).

In the broader context of research, despite the questionable validity, this study showed that in spatial awareness training, a two-hour programme might be below the effective training threshold, perhaps providing an excessive learning load within insufficient time to acquire learning schemas and achieve automation in accessing them from permanent memory stores (see section 2.4.1 Cognitive Load Theory). This reasoning may point at extending the training time in future projects.

The successes and failures of this study resulted in an enriching experience for the researcher, highlighting the importance of selecting a research design, a realistic expectation of participation and following a statistical plan. The value of simplicity in testing, avoiding multiple tests and complex analysis, while avoiding unnecessary expansions that could lead to confusion and loss of focus.

Learning these lessons may improve the quality and efficacy of future studies.

## 6.2 Recommendations

The following sections may benefit from recommendation emerging from the experience gained in this study:

**Participation:** Although offering the test immediately after the presentation of the study at the end of a lecture worked well to facilitate voluntary participation, recruitment of a group for the experiment was difficult. This perhaps could improve if incentives emerged from student suggestions e.g. ask students what is of value and their realistic availability. Alternatively, recruit a “student ambassador” to bridge communication with student, or perhaps include other medical schools to increase sample size. Ensure random allocation to the intervention and control groups in similar numbers. Because most medical programmes offer anatomy in year 1, a control group not exposed to anatomy, may involve inviting students from other science schools, where teaching anatomy may happen later.

**Data Collection:** At the time of the test, clearly instruct participants to complete all fields of information using boxes (student number, age, sex, etc.) to minimise missing data. Include closed questions about recent exposure to the field (spatial ability).

**Training:** The option of offering longer training may not be viable due to time limitations of students. Online tutorials may not provide sufficient motivation to ensure effectivity. Alternatively, the introduction of spatial ability elements to parts of the curriculum that may benefit from spatial reasoning, e.g. revolving solids or ration as part of the introduction to functional anatomy, joint motility and gait; or the analysis of cross-sections in medical imaging, might be possible. Monitoring spatial ability could involve implementing a 10-minute test for all students at the beginning and end of year 1 and 2, supported by reports of positive effects resulting from anatomy learning, which possibly affected this study. A concern about formalising this test would be that students might “prepare” for it or practice, which would distort the validity of the test as a monitoring tool but might benefit students by helping them practice their spatial skills.

Anatomy: An alternative to measure the impact of spatial awareness training in anatomy learning may involve teaching the same content to the experiment group, e.g. major vessels at the base of the neck, which requires appreciation of relations in space and orientation (taught and tested before they see it in the curriculum). Scores may show differences between participants who received training in spatial awareness and those who did not.

## REFERENCE LIST

Abe, T., Raison, N., Shinohara, N., Khan, S., Ahmed, K. and Dasgupta, P. (2017) 'The Effect of Visual-Spatial Ability on the Learning of Robot-Assisted Surgical Skills', *Journal of Surgical Education*.

Adey, P. and Serret, N. (2010) 'Science Teaching and Cognitive Acceleration', in Osborne, J. and Dillon, J. (eds.) *Good Practice in Science Teaching: What Research Has to Say*. 2nd ed. Maidenhead: McGraw-Hill, pp. 82-107.

Anastasi, A. (1958) *Differential Psychology - Individual and Group Differences in Behaviour*. New York: Macmillan.

Anatomical Society (2015) *Undergraduate Medical Anatomy Syllabus: New Anatomy Syllabus*: Anatomical Society. Available at: <http://www.anatsoc.org.uk/education/core-curriculum/medical-anatomy-syllabus> (Accessed: 10 January 2018).

APA (2013) *Publication Manual of the American Psychological Association*. 6th edn. London: American Psychological Association (APA).

Assessment, C. (2017) *About Biomedical Admissions Test (BMAT)*. Cambridge: Cambridge Assessment Admissions Testing. Available at: <http://www.admissionstesting.org/for-test-takers/bmat/about-bmat/>.

Ault, H. K. and John, S. (2010) 'Assessing and Enhancing Visualization Skills of Engineering Students in Africa: A Comparative Study', *Engineering Design Graphics Journal*, 74(2), pp. 12 - 20.

Aveyard, H. (2014) *Doing a Literature Review in Health and Social Care: A practical guide*. 3rd edition edn. Berkshire: Open University Press.

Barker, J. (2016) *Boolean Operators*: University of California, Berkeley. Available at: <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Boolean.pdf> (Accessed: 1 March 2016).

BBC 2017. History. British Broadcasting Corporation (BBC).

Berenbaum, S., Korman Bryk, K. and Beltz, A. (2012) 'Early Androgen Effects on Spatial and Mechanical Abilities: Evidence from Congenital Adrenal Hyperplasia', *Behav Neurosci*, 126(1), pp. 86-96.

Berney, S., Betrancourt, M., Molinari, G. and Hoyek, N. (2015) 'How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models', *Anat Sci Educ*, 8(5), pp. 452-62.

Binet, M. A. (1894) 'Reverse illusions of orientation', *Psychological Review*, 1(4), pp. 337-350.

Bock, R. D. and Kolakowski, D. (1973) 'Further Evidence of Sex-linked Major-Gene Influence on Human Spatial Visualising Ability', *Am J Hum Genet*, 25(1), pp. 1-14.

Bodner, G. M. and Guay, R. B. (1997) 'The Purdue Visualization of Rotations Test', *The Chemical Educator*, 2(4), pp. 1-17.

Boshuizen, H. P. A., Schmidt, H. G., Custers, E. J. F. and Van de Wiel, M. W. (1995) 'Knowledge Development and Restructuring in the Domain of Medicine: The Role of Theory and Practice', *Learning and Instruction*, 5(4), pp. 269-89.

Braungardt, J. (2018) 'Gestalt Psychology', *Philosophical Explorations*, Available: Jürgen Braungardt (Accessed 10 January 2018).

Brosnan, M. (1998) 'Spatial ability in children's play with Lego blocks', *Perceptual and Motor Skills*, 87, pp. 19-28.

Brown, S., Bruce, L., Clarke, D., Cruise, H., Darton, D., Kazmierczak, A., Koulama, D., Harmati, G., Maraki, I., Phillips, K., Slawinski, J., Stocker, T., Summerwill, K. and Thornton, J. (2015) *The state of medical education and practice in the UK: 2015*: General Medical Council. Available at: <http://www.gmc-uk.org/publications/somep2015.asp> (Accessed: 16 September 2017).

BSMS (2015a) *Does your research require ethical review? A self-assessment checklist*: Brighton and Sussex Medical School (BSMS). Available at: <http://www.sussex.ac.uk/staff/research/governance/checklist> (Accessed: 28 August 2015).

BSMS 2015b. *RE: Entry and graduation student numbers at BSMS*.

BSMS (2017a) *About BSMS*: Brighton and Sussex Medical School. Available at: <https://www.bsms.ac.uk/about/index.aspx> (Accessed: 20 August 2017).

BSMS (2017b) *Social Mobility Tsar Comes to BSMS*: Brighton and Sussex Medical School (BSMS). Available at: <https://www.bsms.ac.uk/about/news/2017/11-07-social-mobility-tsar-comes-to-bsms.aspx> (Accessed: 10 November 2017).

BSMS (2017c) *Undergraduate Course Structure - BSMS*: Brighton and Sussex Medical School (BSMS). Available at: <https://www.bsms.ac.uk/undergraduate/our-course/course-structure.aspx> (Accessed: 10 November 2017).

BSMS (2017d) *Widening participation to medicine*: Brighton and Sussex Medical School (BSMS). Available at: <https://www.bsms.ac.uk/about/info-for-schools-teachers-parents/widening-participation-to-medicine.aspx> (Accessed: 10 November 2017).

Burton, A. W. and Rodgerson, R. W. (2001) 'New Perspectives on the Assessment of Movement Skills and Motor Abilities', *Adapted Physical Activity Quarterly*, 18(4), pp. 347-365.

Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y.-S., Stern, E. and Silbersweig, D. (2006) 'Sex differences in mental rotation: Top-down versus bottom-up processing', *NeuroImage*, 32(1), pp. 445-456.

Bütikofer, A., Figlio, D. N., Karbownik, K., Kuzawa, C. W. and Salvanes, K. G. (2019) 'Evidence that prenatal testosterone transfer from male twins reduces the fertility and socioeconomic success of their female co-twins', *Proceedings of the National Academy of Sciences*.

Cambridge, U. o. (2016) 'Architecture Admissions Assessment', Available: University of Cambridge.

Cannon, B. (2004) 'A study of manual dexterity in relation to achievement in various facets of a dental hygiene program - Thesis', Available: University of Southern California.

Carlton, J. and Czoski-Murray, C. (2010) 'The value of screening for amblyopia revisited', in Lorenz, B. and Brodsky, M. (eds.) *Pediatric Ophthalmology, Neuro-Ophthalmology, Genetics: Strabismus - New Concepts in Pathophysiology, Diagnosis, and Treatment*. Berlin: Springer, pp. 95111.

Carroll, J. B. (1993) *Human Cognitive Abilities: A Survey of Factor Analytic Studies* Cambridge: Cambridge University Press. Available at: /ebook.jsf?bid=CBO9780511571312 (Accessed: 15 September 2015).

Cederlund, R. (1995) 'The Use of Dexterity Tests in Hand Rehabilitation', *Scandinavian Journal of Occupational Therapy*, 2(3-4), pp. 99-104.

Chase, W. G. and Simon, H. A. (1973) 'Perception in chess', *Cognitive Psychology*, 4(1), pp. 55–81.

Clausen-May, T. and Smith, P. (1998) *Spatial Ability: A Handbook for Educators*. The National Foundation for Educational Research.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*. J Grad Med Educ New York: Lawrence Erlbaum Associates.

Cook, T. D. and Campbell, D. T. (1979) *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.

Council, N. R. (2015) 'Spatial Abilities | |', *Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession*. Washington DC: The National Academies Press.

Courvoisier, D. S., Renaud, O., Geiser, C., Paschke, K., Gaudy, K. and Kirsten, J. (2013) 'Sex hormones and mental rotation: An intensive longitudinal investigation', *Hormones and Behaviour*, 63(2), pp. 345-351.

Creswell, J. W. (2014) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th edn. London: SAGE Publications.

CRS (2015) *Sample Size Calculator*. Creative Research Systems. Available at: <http://www.surveysystem.com/sscalc.htm> (Accessed: 1 March 2015).

Czapka, J. T., Moeinzadeh, M. H. and Leake, J. M. 'Application of Rapid Prototyping Technology to Improve Spatial Visualization'. *The 2002 American Society for Engineering Education Annual Conference & Exposition*, Montreal, Canada.

Dawson, C. (2006) *A Practical Guide to Research Methods: A user friendly manual for mastering research techniques and projects*. 2nd edn. Oxford: How To Books Ltd.

DeAngelis, G. (2000) 'Seeing in three dimensions: the neurophysiology of stereopsis', *Trends in Cognitive Sciences*, 4(3), pp. 80-90.

Deviant, S. (2012) 'Alpha Level (Significance Level): What is it?', *Statistics How To*. Available at: <https://www.statisticshowto.datasciencecentral.com/what-is-an-alpha-level/> (Accessed 10 October 2018).

Duke University (2017) 'Guide to the William Stern Papers, 1937-1938, undated', *Collection Guides*, Available: Duke University Libraries.

Dunbar, S. R. (2011) 'The de Moivre-Laplace Central Limit Theorem.', *Topics in Probability Theory and Stochastic Processes.*, Available: University of Nebraska-Lincoln (Accessed 10 August 2018).

Ekstrom, R. B., French, J. W. and Harman, H. H. (1976) *Manual for Kit of Factor Referenced Cognitive Tests* Princeton, NJ, USA: Educational Testing Service. Available at: [http://www.ets.org/Media/Research/pdf/Manual\\_for\\_Kit\\_of\\_Factor-Referenced\\_Cognitive\\_Tests.pdf](http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf).

Entwistle, N. J. and Paul, R. (1983) *Understanding Student Learning*. New York: Nichols Publishing Company.

Ernst, J. V., Williams, T. O., Clark, A. C. and Kelly, D. P. 'Psychometric Properties of the PSVT:R Outcome Measure: A Preliminary Study of Introductory Engineering Design Graphics', *70th Engineering, Design, Graphics Division (EDGD) Midyear Conference*, Daytona Beach Shores, Florida. USA, 24-26 January 2016: American Society for Engineering Education (ASEE).

Exley, K. and Dennick, R. (2004) *EBL Reader - Small Group Teaching : Tutorials, Seminars and Beyond*. London: Taylor and Francis, p. 137 - 149.

Fairweather, H. (1976) 'Sex differences in cognition', *Cognition*, 4(3), pp. 231-280.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007) 'G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', *Behavior Research Methods*, 39(2), pp. 175-191.

Fernandez, R., Dror, I. E. and Smith, C. (2011) 'Spatial abilities of expert clinical anatomists: comparison of abilities between novices, intermediates, and experts in anatomy', *Anat Sci Educ*, 4(1), pp. 1-8.

Field, A. (2012) *Discovering Statistics Using IBM SPSS Statistics and sex and drugs and rock 'n' roll*. 4th edn. London: Sage Publications Ltd.

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics and Sex and Drugs and Rock 'n' Roll*. 4th edn. London: Sage Publications Ltd.

Field, A. (2018) *Discovering Statistics Using IBM SPSS Statistics*. 5th edn. London: SAGE Publications Ltd.

Field, A. and Hole, G. (2003) *How to Design and Report Experiments*. London: Sage. Reprint, 2006.

Ford, K. E. and Cooper, L. R. L. (2018) 'Learning from lawsuits: Ten-years of NHS litigation authority claims against 11 surgical specialities in England.', *The Surgeon: Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*, 16(1), pp. 27-35.

Francis, N., Hanna, G., Cresswell, A., Carter, F. and Cuschieri, A. (2001) 'The performance of master surgeons on standard aptitude testing', 182(1), pp. 30–33.

Friess, A., Martin, E. L., Esparragoza, I. E. and Lawanto, O. 'Improvements in Student Spatial Visualization in an Introductory Engineering Graphics Course using Open-ended Design Projects Supported by 3-D Printed Manipulatives'. *2016 American Society for Engineering Education (ASEE) Annual Conference & Exposition*, New Orleans, Louisiana: American Society for Engineering Education.

Galton, F. (1880) 'Visualised Numerals', *Nature*, 21(533), pp. 252-256.

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszcak, E., Doré, C., Williamson, P. R., Altman, D. G., Montgomery, A., Lim, P., Berlin, J., Senn, S., Day, S., Barbachano, Y. and Loder, E. (2017) 'Guidelines for the Content of Statistical Analysis Plans in Clinical Trials', *JAMA*, 19(318(23)), pp. 2337-2343.

Garg, A. X., Norman, G., Spero, L. and Taylor, I. (1999) 'Learning Anatomy: Do New Computer Models Improve Spatial Understanding?', *Medical Teacher*, 21(5), pp. 519-522.

Garg, A. X., Norman, G. and Sperotable, L. (2001) 'How medical students learn spatial anatomy', *The Lancet*, 357(9253), pp. 363-364.

Gimmestad, B. J. 'Gender differences in spatial visualization and predictors of success in an engineering design course'. *National Conference on Women in Mathematics and the Sciences*, St. Cloud State University, St. Cloud, MN, 133-136.

Giuliani, M., Lajolo, C., Clemente, L., Querqui, A., Viotti, R., Boari, A. and Miani, C. (2007) 'Is manual dexterity essential in the selection of dental students?', *British Dental Journal*, 203(3), pp. 149-155.

GMC (2009) *Tomorrow's Doctors 2009*: General Medical Council UK. Available at: [http://www.gmc-uk.org/Tomorrow\\_s\\_Doctors\\_1214.pdf\\_48905759.pdf](http://www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf).

GMC (2016a) *Our role - General Medical Council UK*: General Medical Council. Available at: <http://www.gmc-uk.org/about/role.asp> (Accessed: 10 March 2016).

GMC (2016b) 'Promoting excellence: standards for medical education and training', Available: General Medical Council. DOI: <http://www.gmc-uk.org> (Accessed 2016-03-04T12:30:59).

GMC (2017) *List of Registered Medical Practitioners - statistics*: General Medical Council (GMC). Available at: [https://www.gmc-uk.org/doctors/register/search\\_stats.asp](https://www.gmc-uk.org/doctors/register/search_stats.asp) (Accessed: 10 November 2017).

Gorska, R., Sheryl, S. and Leopold, C. (2016) 'Gender Differences in Visualization Skills - An International Perspective', *Engineering Design Graphics Journal*, 62(3), pp. 9 - 18.

Granger, N. A. (2004) 'Dissection laboratory is vital to medical gross anatomy education', *The Anatomical Record*, 281B(1), pp. 6-8.

Guay, R. B. (1976) *Purdue Spatial Visualization Test: Rotations*: Purdue Research Foundation. Available at: <https://lphscadd.wikispaces.com/file/view/Purdue+Spatial+Test.pdf> (Accessed: 10 October 2016).

Guay, R. B. and McDaniel, E. D. (1977) 'The Relationship between Mathematics Achievement and Spatial Abilities among Elementary School Children', *Journal for Research in Mathematics Education*, 8(3), pp. 211-215.

Guilford, J. P. (1968) 'Intelligence Has Three Facets', *Science*, 160, pp. 615-620.

Guillot, A., Champely, S., Batier, C., Thiriet, P. and Collet, C. (2006) 'Relationship Between Spatial Abilities, Mental Rotation and Functional Anatomy Learning', *Advances in Health Sciences Education*, 12(4), pp. 491-507.

Haahr, M. (2015) *True Random Number Service*: Random.org. Available at: <https://www.random.org/> (Accessed: 10 February 2015).

Hampson, E., Kimura, D. and Thompson, R. F. (1988) 'Reciprocal Effects of Hormonal Fluctuations on Human Motor and Perceptual-Spatial Skills', *Behavioral Neuroscience*, 102(3), pp. 456-459.

Hartlage, L. C. (1970) 'Sex-Linked Inheritance of Spatial Ability', *Perceptual and Motor Skills*, (31), pp. 610.

Hegarty, M., Crookes, R., Dara-Abrams, D. and Shipley, T. (2010) 'Do All Science Disciplines Rely on Spatial Abilities? Preliminary Evidence from Self-Report Questionnaires', *Spatial Intelligence and Learning Center (SILC)*, [www.spatiallearning.org](http://www.spatiallearning.org) [To appear in C. Hölscher, T. F. Shipley, M. Olivetti, J. Bateman, & N. Newcombe (Eds.), Proceedings of the 7th international conference on Spatial cognition. Lecture Notes in Computer Science 6222 (pp. 85-94). Springer.], Available: Spatial Intelligence and Learning Center (SILC) (Accessed 4 December 2014).

Hegarty, M., Keehner, M., Cohen, C., Montello, D. and Lippa, Y. (2007) 'The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals', *Spatial Cognition and Medicine*, Available: Michigan Technology University (Accessed 18 March 2015).

Heil, M., Michael, K., Rolke, B., Beste, C. and Jansen, P. (2011) 'Mental rotation in female fraternal twins: Evidence for intra-uterine hormone transfer?', *Biological Psychology*, 86(1), pp. 90–93.

Hill, C., Corbett, C. and Rose, A. (2010) *Why So Few? Women in Science, Technology, Engineering, and Mathematics* Washington: AAUW. Available at: <http://www.aauw.org/files/2013/02/Why-So-Few-Women-in-Science-Technology-Engineering-and-Mathematics.pdf> (Accessed: 26 June 2015).

Hinton, P., Brownlow, C., McMurray, I. and Cozens, B. (2005) *SPSS Explained*. Hove: Routledge.

Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P. and Guillot, A. (2009) 'Enhancement of Mental Rotation Abilities and Its Effect on Anatomy Learning', *Teaching and Learning in Medicine*, 21(3), pp. 201-206.

IBM (2016a) *IBM Knowledge Center - Variable names and labels*: IBM. Available at: [http://www.ibm.com/support/knowledgecenter/SSLVQG\\_7.0.0/com.spss.ddl/savdsc\\_varnames.htm](http://www.ibm.com/support/knowledgecenter/SSLVQG_7.0.0/com.spss.ddl/savdsc_varnames.htm) (Accessed: 10 August 2016).

Author (2016b) *Statistical Package for the Social Sciences (SPSS) (Version 23)*. Available at: <https://www.ibm.com/marketplace/cloud/statistical-analysis-and-reporting/us/en-us>.

Ibrahim, H., Heard, N. P. and Blanksby, B. (2011) 'Exploring the General Motor Ability Construct', *Perceptual and Motor Skills*, 113(2), pp. 491-508.

ICO (1984) *Visual Acuity Measurement Standard*: International Council of Ophthalmology. Available at: <http://www.icoph.org/dynamic/attachments/resources/icovisualacuity1984.pdf> (Accessed: 8 August 2015).

Illeris, K. (2018) 'A comprehensive Understanding of Human Learning', in Illeris, K. (ed.) *Contemporary Theories of Learning: Learning Theorists ... In Their Own Words*. 2nd Edition ed. Abingdon: Routledge, pp. 1-14.

Indiana University (2007) *Manual Dexterity - Health Professions and Prelaw Center*: Indiana University Bloomington. Available at: <http://hpplc.indiana.edu/ohp/d-md.shtml> (Accessed: 18 August 2015).

Indiana University (2013) *William Stern*. Biographical Profiles: Indiana University (USA). Available at: <http://www.indiana.edu/~intell/stern.shtml> (Accessed: 10 December 2017).

Institute for Work & Health (2015) 'What researchers mean by cross-sectional vs. longitudinal studies', *At Work*, 81(Summer 2015), pp. 2, Available: Institute for Work & Health (Accessed 10 June 2018).

Jaques, D. (2003) 'Abc Of Learning And Teaching In Medicine: Teaching Small Groups', *British Medical Journal (BMJ)*, 326(7387), pp. 492-494.

JHU (2014) 'What is spatial ability? - John Hopkins University', 2014 [PDF], Available: John Hopkins University (Accessed 13 October 2014).

Kalloniatis, M. and Luu, C. (2007) 'Visual Acuity', *Webvision*, Available: University of Utah (Accessed 20 February 2015).

Kanski, J. J. (2009) *Clinical Ophthalmology: A Synopsis*. Windsor: Elsevier Limited.

Karras, D. J. (1997a) 'Statistical methodology: II. Reliability and validity assessment in study design, Part B', *Acad Emerg Med*, 4(2), pp. 144-7.

Karras, D. J. (1997b) 'Statistical methodology: II. Reliability and validity assessment in study design, Part A', *Acad Emerg Med*, 4(1), pp. 64-71.

Kerns, K. and Berenbaum, S. (1991) 'Sex differences in spatial ability in children', *Behavior Genetics*, 21(4), pp. 383-396.

Kimball, A. W. (1957) 'Errors of the third kind in statistical consulting', *Journal of the American Statistical Association*, 52(277), pp. 133-42.

Kimmel, L. G., Miller, J. D. and Eccles, J. S. (2012) 'Do the Paths to STEMM Professions Differ by Gender?', *Peabody Journal of Education*, 87(1).

Kline, P. (2000) *The handbook of psychological testing*. 2nd edn. London: Routledge.

Kopelman, P. (2014) 'The future of UK medical education curriculum – what type of medical graduates do we need?', *Future Healthcare Journal*, 1(1), pp. 41-46.

Lafayette Instruments (2015) *Purdue Pegboard Test: User Instructions* Lafayette, USA: Lafayette Instrument Company. Available at: <http://www.limef.com/downloads/MAN-32020A-forpdf-rev0.pdf> (Accessed: 10 June 2017).

LaMorte, W. W. (2016) 'Central Limit Theorem', *The Role of Probability.*, Available: Boston University. (Accessed 10 August 2018).

Langlois, J., Bellemare, C., Toulouse, J. and Wells, G. (2017) 'Spatial abilities and anatomy knowledge assessment: A systematic review.', *Anatomical Sciences Education*, 10(3), pp. 235-241.

Lennie, P. and Van Hemel, S. B. (eds.) (2002) *Visual Impairments: Determining Eligibility for Social Security Benefits.*: National Centre for Biotechnology Information (NCBI).

Author (2009) *Java Applets for Power and Sample Size*. Available at: <http://homepage.stat.uiowa.edu/~rlenth/Power/> (Accessed: 10 February 2016).

Levi, D. M., Knill, D. C. and Bavelier, D. (2015) 'Stereopsis and Ambliopia: A Mini-Review', *Vision Research*, 114, pp. 17-30.

Levine, S., Huttenlocher, J., Taylor, A. and Langrock, A. (1999) 'Early Sex Differences in Spatial Skill', *Developmental Psychology*, 35(4), pp. 940 - 949.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S. and Huttenlocher, J. (2005) 'Socioeconomic status modifies the sex difference in spatial skill', *Psychological Science*, 16(11), pp. 841-5.

Levingston, A. J., Weaver, B., Garside, S., McGinn, H. and Norman, B. R. (2007) 'Virtual Reality and Brain Anatomy: A Randomised Trial of E-Learning Instructional Designs', *Medical Education*, 41, pp. 495-501.

Linn, M. and Petersen, A. (1985) 'Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis', *Child Development*, 56(6), pp. 1479-1498.

Lohman, D. F. (1979) 'Spatial Ability: A Review and Reanalysis of the Correlational Literature.', (TR-8).

Lord, T. R. and Garrison, J. (1998) 'Comparing Spatial Abilities of Collegiate Athletes in Different Sports', *Perceptual and Motor Skills*, (86), pp. 1016-1018.

Lorås, H. and Sigmundsson, H. (2012) 'Interrelations between Three Fine Motor Skills in Young Adults:' *Perceptual & Motor Skills: Physical Development & Measurement*.

Lubinski, D. (2010) 'Spatial ability and STEM: A sleeping giant for talent identification and development', *Personality and Individual Differences*, 49(4), pp. 344–351.

Lufler, R. S., Zumwalt, A. C., Romney, C. A. and Hoagland, T. M. (2011a) 'Effect of Visual–Spatial Ability on Medical Students' Performance in a Gross Anatomy Course', *Anatomical Science Education*, 5, pp. 3-9.

Lufler, R. S., Zumwalt, A. C., Romney, C. A. and Hoagland, T. M. (2011b) 'Effect of Visual–Spatial Ability on Medical Students' Performance in a Gross Anatomy Course', *Anatomical Science Education*, 5, pp. 3-9.

Luursema, J.-M. (2010) *See me, Touch me, Heal me: The Role of Visuo-spatial Ability in Virtual Anatomical Learning and Endoscopic Simulator Training* Tilburg: Medialaan BV, the Netherlands. Available at: [http://doc.utwente.nl/71593/1/thesis\\_J\\_Luursema.pdf](http://doc.utwente.nl/71593/1/thesis_J_Luursema.pdf) (Accessed: 1 March 2016).

Luursema, J.-M., Verwey, W. B., Kommers, P. A. M. and Annema, J.-H. (2008) 'The role of stereopsis in virtual anatomical learning'.

Luursema, J.-M., Vorstenbosch, M. and Kooloos, J. (2017) 'Stereopsis, Visuospatial Ability, and Virtual Reality in Anatomy Learning', *Anatomy Research International*, 2017.

Maccoby, E. E. and Jacklin, C. L. N. (1974) *The Psychology of Sex Differences*. Stanford, California: Stanford University Press.

Macfarlane Smith, I. (1964) *Spatial Ability, its Educational and Social Significance*. London: University of London.

Maeda, Y. and Yoon, Y. (2011) 'Measuring Spatial Ability of First-Year Engineering Students With the Revised PSVT:R', *American Society for Engineering Education*, 2011.

Maeda, Y. and Yoon, Y. (2012) 'A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R)', *Educational Psychology Review*, 25(1), pp. 69-94.

Maeda, Y. and Yoon, Y. (2016) 'A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R) | SpringerLink', *Educational Psychology Review*, 25(1), pp. 69-94.

Makofske, B. (2016) 'Manual Dexterity', *Encyclopedia of Clinical Neuropsychology*, pp. 1522-1523.

Manning, J. T. and Taylor, R. P. (2001) 'Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans', *Evolution and Human Behavior*, 22(1), pp. 61-69.

Masterson, R. and Pickton, D. (2014) *Marketing: An Introduction*. 3rd edn. London: Sage.

Maxwell, S. E. (2004) 'The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies', *Psychological Methods*, 9(2), pp. 147-163.

McClelland, D. C. (1973) 'Testing for competence rather than for "intelligence."', *American Psychologist*, 28(1), pp. 1-14.

McGee, M. G. (1979a) 'Human Spatial Abilities: Psychometric Studies and Environmental, Genetic, Hormonal, and Neurological Influences', *Psychological Bulletin*, 86(5), pp. 889-918.

McGee, M. G. (1979b) *Human Spatial Abilities: Sources of Sex Difference*. New York: Praeger Publishers.

McGee, M. G. (1982) 'Spatial Abilities: The Influence of Genetic Factors', in Potegal, M. (ed.) *Spatial Abilities: Development and Physiological Foundations*. New York: Academic Press, pp. 199-222.

McManus, I. C., Smithers, E., Partridge, P., Keeling, A. and Fleming, P. R. (2003) 'A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study', *British Medical Journal (BMJ)*, 327, pp. 139.

Medical Portal (2019) *UCAT (UKCAT) - a Free Guide to the UCAT Test*. The Medical Portal. Available at: <https://www.themedicportal.com/application-guide/ukcat/> (Accessed: 10 January 2019).

Meit, S., Borges, N. and Early, L. (2007) 'Personality Profiles of Incoming Male and Female Medical Students: Results of a Multi-Site 9-Year Study', 12.

Merritt, B. K. and Fisher, A. G. (2003) 'Gender differences in the performance of activities of daily living', *Archives of Physical Medicine and Rehabilitation*, 84(12), pp. 1872-7.

Michael, W., Guilford, J., Fruchter, B. and Zimmerman, W. (1957) 'The Description of Spatial Visualisation Abilities', *Educational and Psychological Measurement*, 17(2), pp. 185-199.

Mohan, R. (2016) *Measurement, Evaluation and Assessment in Education*. Delhi, India: PHI Learning Pvt. Ltd.

Mohler, J. L. (2008) 'A Review of Spatial Ability Research', *Engineering Design Graphics Journal*, 72(3), pp. 19-30.

Morris, B. (2005) 'Robotic Surgery: Applications, Limitations, and Impact on Surgical Education', *MedGenMed*, 7(3), pp. 72.

Morrissey, S. R. (2017) 'Educate To Innovate', *Chemical & Engineering News*, 88(4), pp. 25-26.

Moxham, B. J. and Paisant, O. (2007) 'Perception of medical students towards the clinical relevance of anatomy', *Clinical Anatomy*, 20(5), pp. 560–564.

MPS (2017) *About Medical Protection: Introducing MPS*: Medical Protection Society (MPS). Available at: <https://www.medicalprotection.org/uk/about-mps> (Accessed: 10 December 2017).

MSC (2014) *Selecting for Excellence: Final Report*: Medical Schools Council (MSC). Available at: <https://www.medschools.ac.uk/media/1911/selecting-for-excellence-final-report.pdf>.

MSC (2017a) *Applications to Medical Schools*: Medical Schools Council. Available at: <https://www.medschools.ac.uk/studying-medicine/applications> (Accessed: 10 October 2017).

MSC (2017b) *Entry requirements for UK medical schools - 2018 entry*: Medical Schools Council (MSC). Available at: <https://www.medschools.ac.uk/media/2357/msc-entry-requirements-for-uk-medical-schools.pdf> (Accessed: 10 October 2017).

Murdoch, J. R., McGhee, C. N. and Glover, V. (1991) 'The relationship between stereopsis and fine manual dexterity: Pilot study of a new instrument', *Eye*, 5(5), pp. 642-643.

Netter, F. H. (2014) *Atlas of Human Anatomy*. 6th edn. Philadelphia, PA, USA.

Newble, D. I. and Entwistle, N. J. (1986) 'Learning styles and approaches: implications for medical education', *Medical Education*, 20(3), pp. 162–175.

Nguyen, N., Mulla, A., Nelson, A. J. and 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.

NHS (2018) *Statistical Analysis Plan Template*: Cambridge Clinical Trials Unit (CCTU). Available at: <https://www.cuh.nhs.uk/node/343> (Accessed: 10 February 2019).

Nordvik, H. and Amponsah, B. (1998) 'Gender Differences in Spatial Abilities and Spatial Activity Among University Students in an Egalitarian Educational System', *Sex Roles*, 38(11 - 12), pp. 1009 - 1023.

Nyborg, H. (1983) 'Spatial ability in men and women: Review and new theory - ScienceDirect', *Advances in Behaviour Research and Therapy*, 5, pp. 89–140.

O'Carroll, R. E., Whiten, S., Jackson, D. and Sinclair, D. W. (2002) 'Assessing the emotional impact of cadaver dissection on medical students', *Medical Education*, 36(6), pp. 550–554.

Ohlsson, J. and Villarreal, G. (2005) 'Normal visual acuity in 17-18 year olds', *Acta Ophthalmologica Scandinavica*, 83(4).

Olivier, J. and Bell, M. (2013) 'The Importance of Statistics in Cancer Research', *World Journal of Cancer Research*, 1, pp. 78-81.

Osborne, T., Sanders, Wilma (1956) 'The Crawford Small Parts Dexterity Test as a Time-Limit Test', *Personnel Psychology*, 9(2), pp. 177-180.

Otis, L. (2016) 'A New Look at Visual Thinking', Available: Psychology Today (Accessed 10 Jan 2018).

Pandey, P. and Zimitat, C. (2007) 'Medical students' learning of anatomy: memorisation, understanding and visualisation', *Medical Education*, 41(1), pp. 7-14.

Payne, G. and Payne, J. (2004) *Longitudinal and Cross-sectional Studies*. SAGE.

Peters, M. (2005) 'Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems', *Brain and Cognition*, 57, pp. 176–184.

Pintzka, C. W. S., Evensmoena, H. R., Lehn, H. and Håberg, A. K. (2016) 'Changes in spatial cognition and brain activity after a single dose of testosterone in healthy women', *Behavioural Brain Research*, 298, pp. 78-90.

- Ponce, C. and Born, R. (2008) 'Stereopsis', *Current Biology*, 18(18), pp. R845–R850.
- Powell, D. H. (2004) 'Behavioral treatment of debilitating test anxiety among medical students.', *Clinical Psychology*, 60(8), pp. 853–865.
- Poynter, F. N. L. (1966) *The evolution of Medical Education in Britain*. 5th edn. London: British Congress on the History of Medicine.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S. and White, L. E. (2012) *Neuroscience*. 5th edn. Sunderland: Sinauer Associates, Inc.
- Puts, D. A., Cardenas, R. A., Bailey, D. H., Burriss, R. P., Jordan, C. L. and Breedlove, S. M. (2010) 'Salivary testosterone does not predict mental rotation performance in men or women', *Hormones and Behavior*, 58(2), pp. 282-289.
- Ramsden, P. (2003) *Learning to teach in higher education*. 2nd edn. London: RoutledgeFalmer.
- Reed, M. (2017) 'Why widening participation matters', *InPulse*(88), Available: Brighton and Sussex Medical School (BSMS) (Accessed 10 November 2017).
- Rees, J. and Stephenson, A. (2010) 'The future of medical education in the UK', *Br J Gen Pract*, 60(580), pp. 795-6.
- Riva, J., Malik, K., Burnie, S., Endicott, A. and Busse, J. (2012) 'What is your research question? An introduction to the PICOT format for clinicians', *J Can Chiropr Assoc*, 56(3), pp. 167-71.
- Rochford, K. (1985) 'Spatial learning disabilities and underachievement among university anatomy students', *Medical Education*, 19(1), pp. 13–26.
- Rutar, T. (2015) 'Amblyopia', *BMJ Best Practice*. DOI: monograph/1162 (Accessed 10 September 2015).

Salkind, N. J. (2014) *Statistics for People Who (Think They) Hate Statistics*. 5th edn. London: Sage.

Salkind, N. J. (2017) *Statistics for People Who (Think They) Hate Statistics*. 6th edn. London: SAGE Publications Limited.

Schwartz, S. and Carpenter, K. M. (1999) 'The right answer for the wrong question: Consequences of type III error for public health research', *Am. J. Public Health*, 89(8), pp. 1175-80.

Shepard, R. and Metzler, J. (1971) 'Mental Rotation of Three-Dimensional Objects', *Science*, 171(3972), pp. 701 - 703.

Sherman, J. A. and Fennema, E. (1978) 'Distribution of Spatial Visualization and Mathematical Problem Solving Scores: A Test of Stafford's X-linked Hypotheses', *Psychology of Women Quarterly*, 3(2), pp. 157-167.

Siegler, R. S. (1992) 'The Other Alfred Binet', *Developmental Psychology*, 28(2), pp. 179-190.

Silverman, I. and Phillips, K. (1993) 'Effects of estrogen changes during the menstrual cycle on spatial performance', *Ethology and Sociobiology*, 14(4), pp. 257-269.

Simon, S. and Osborne, J. (2010) 'Students' attitudes to science', in Osborne, J. and Dillon, J. (eds.) *Good Practice in Science Teaching: What Research has to Say*. 2nd ed. Maidenhead: Open University Press, pp. 238-258.

Smith, C. F. (2008) *Learning Anatomy at University: effectiveness, issues and implications for the future education of doctors*. PhD thesis, University of Southampton, Southampton 10 December 2017).

Smith, C. F., Dilley, A., Mitchell, B. S. and Drake, R. L. (2018) *Gray's Surface Anatomy and Ultrasound: A Foundation for Clinical Practice*. London: Elsevier.

Smith, C. F., Finn, G. M. and Border, S. (2017) 'Learning Clinical Anatomy', *European Journal of Anatomy*, 21(4), pp. 269-278.

Smith, C. F., Finn, G. M., Stewart, J., Atkinson, M. A., Davies, D. C., Dyball, R., Morris, J., Ockleford, C., Parkin, I., Standing, S., Whiten, S., Wilton, J. and McHanwell, S. (2015a) 'The Anatomical Society core 156 regional anatomy syllabus for undergraduate medicine', *Journal of Anatomy*, 228(1), pp. 15-23.

Smith, C. F., Finn, G. M., Stewart, J. and McHanwell, S. (2015b) 'Anatomical Society core regional anatomy syllabus - Delphi process', *Journal of Anatomy*, 228(1), pp. 2-14.

Smith, C. F. and Mathias, H. S. (2010) 'What impact does anatomy education have on clinical practice?', *Clinical Anatomy*, 24(1), pp. 113-119.

Smith, C. F. and Saber-Sheikh, K. 'Measuring human movement during an expert exploration of the thorax, does this have implication for understanding how touch mediated perception works in learning anatomy?'. *Anatomical Society Summer Meeting 2014*, Bradford: *Journal of Anatomy*, 189-213.

Smith, C. F., Stokholm, C., Sinha, R., Poinkwer, F., Carter, M. and Birch, M. (2017) 'Interplays of psychometric abilities on learning gross anatomy', *MedEdPublish*, 6(2), pp. 42.

Sorby, S. (2009) 'Educational Research in Developing 3-D Spatial Skills for Engineering Students', *International Journal of Science Education*, 31(3), pp. 459 - 480.

Sorby, S. and Baartmans, B. (2000) 'The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students', *Journal of Engineering Education*, (July 2000), pp. 301-307.

Sorby, S. and Wysocki, A. (2012) *Developing Spatial Thinking: Workbook*. Workbook edn. New York: Delmar.

Sorby, S. A. (2007) 'Developing 3D Spatial Skills for Engineering Students', *Australasian Journal of Engineering Education*, 13(1), pp. 1.

Sorby, S. A. (2012) *Developing spatial thinking*. 1st edn. New York, NY, USA: Delmar Cengage Learning.

Stafford, R. E. (1961) 'Sex Differences in Spatial Visualization as Evidence of Sex-Linked Inheritance', *Perceptual and Motor Skills*, (13), pp. 428.

Stanford (2017) *Defining Your Own Teaching Goals*. Stanford, California, USA: Stanford University. Available at: <https://teachingcommons.stanford.edu/resources/teaching/planning-your-approach/defining-your-own-teaching-goals> (Accessed: 20 August 2017).

Steele, C. (2015) 'Understanding Hypothesis Tests: Significance Levels (Alpha) and P values in Statistics.', *The Minitab Blog*. Available at: <http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests-significance-levels-alpha-and-p-values-in-statistics> (Accessed 10 October 2018).

Stidwill, D. and Fletcher, R. (2014) *Normal Binocular Vision*. Wiley.

Stoeger, H. and Ziegler, A. (2010) 'How Fine Motor Skills Influence the Assessment of High Abilities and Underachievement in Math', *Journal for the Education of the Gifted*, 34(2), pp. 195-219.

Students, B. 2016. *RE: Conversation about the Spatial Ability test given to medical students at Brighton and Sussex Medical School*. Type to Gonzales, R.

Sweller, J. (1994) 'Cognitive load theory, learning difficulty, and instructional design', *Learning and Instruction*, 4(4), pp. 295-312.

Taber, K. S. (2014) *Student Thinking and Learning in Science: Perspectives on the Nature and Development of Learner's Ideas*. Arbingdon: Routledge.

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangratorio, L., Dennis, B., Kosa, D., Debono, V. B., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G. and Goldsmith, C. H. (2013) 'A tutorial on sensitivity analyses in clinical trials: the what, why, when and how', *BMC Medical Research Methodology*, (13), pp. 92.

Thomas, L. and Peterson, E. D. (2012) 'The Value of Statistical Analysis Plans in Observational Research: Defining High-Quality Research From the Start', *JAMA*, 308(8), pp. 773-774.

Thompson, P. (2002) 'Eyes wide apart: Overestimating interpupillary distance', *Perception*, 31, pp. 651-656.

Thorndike, E. L. (1921) 'On the organization of the intellect', *Psychological Review*, 28(2), pp. 141-151.

Tiffin, J. and Asher, E. J. (1948) 'The Purdue Pegboard: Norms and studies of reliability and validity', *Journal of Applied Psychology*, 32(3), pp. 234-247.

Tunstall, R. and Shah, N. (2012) *Surface Anatomy*. JP Medical Publishers.

Turner, R. M., Bird, S. M. and Higgins, J. P. T. (2013) 'The Impact of Study Size on Meta-analyses: Examination of Underpowered Studies in Cochrane Reviews', *PLOS-ONE*, 8(3), pp. e59202.

Turney, B. (2007) 'Anatomy in a Modern Medical Curriculum', *Ann R Coll Surg Engl*, 89(2), pp. 104-7.

UCLAN (2017) *Dentistry (Graduate Entry) BDS*: University of Central Lancashire (UCLAN). Available at: [https://www.uclan.ac.uk/courses/bds\\_dentistry\\_graduate\\_entry.php](https://www.uclan.ac.uk/courses/bds_dentistry_graduate_entry.php) (Accessed: 10 December 2017).

University of Dundee (2017) *Dentistry: Overview*: University of Dundee. Available at: <https://www.dundee.ac.uk/study/ug/dentistry/> (Accessed: 10 December 2017).

University of Glasgow (2017) *Dental School: New Applicant - Frequently Asked Questions*: University of Glasgow. Available at: <https://www.gla.ac.uk/schools/dental/undergraduate/newapplicant/#portfolio> (Accessed: 10 December 2017).

University of Manchester (2017) *BDS Dentistry: Application and selection*: The University of Manchester. Available at: <http://www.manchester.ac.uk/study/undergraduate/courses/2017/00398/bds-dentistry-first-year-entry/application-and-selection/> (Accessed: 10 December 2017).

University of Reading (2017) 'Literature Reviews', *Study Advice*, Available: University of Reading.

University of Sheffield (2017) *The School of Clinical Dentistry: Entry Requirements and Selection Procedure*: The University of Sheffield. Available at: <https://www.sheffield.ac.uk/dentalschool/undergraduate/bds/selection#PersonalQualities> (Accessed: 10 December 2017).

UoS (2014) 'Code of Practice for Research - University of Sussex', Available: The University of Sussex (Accessed 24 August 2015).

UoS (2015) *Long-term storage and preservation - University of Sussex*: The University of Sussex. Available at: <http://www.sussex.ac.uk/library/researchdatamanagement/store/longtermstorageandpreservation> (Accessed: 24 August 2015).

UoS (2016) 'Children graduate from "inspirational" University of Sussex programme', *Bulletin*, Available: University of Sussex (UoS) (Accessed 10 November 2017).

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, Christopher and Newcombe, N. S. (2013) 'The malleability of spatial skills: A meta-analysis of training studies', *Psychological Bulletin*, 139(2), pp. 352-402.

Van Goozen, S. H. M., Cohen-Kettenis, P. T., Gooren, L. J. G., Frijda, N. H. and Van de Poll, N. E. (1995) 'Gender differences in behaviour: Activating effects of cross-sex hormones', *Psychoneuroendocrinology*, 20(4), pp. 343-363.

Vorstenbosch, M., Klaassen, T., Donders, R., Kooloos, J., Bolhuis, S. and Laan, R. (2013) 'Learning anatomy enhances spatial ability', *Anat Sci Educ*, 6(4), pp. 257-62.

Vorstenbosch, M. A. T. M., Kooloos, J. G. M., Bolhuis, S. M. and Laan, R. F. J. M. (2016) 'An Investigation of Anatomical Competence in Junior Medical Doctors', *Anatomical Sciences Education*, 9(1), pp. 8-17.

Voyer, D., Voyer, S. and Bryden, P. (1995) 'Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables', *Psychological Bulletin*, 117(2), pp. 250-270.

Vuoksima, E., Kaprio, J. and Kremen, W. S. (2010) 'Having a male co-twin masculinizes mental rotation performance in females.', *Psychological Science*, 21(8), pp. 1069–1071.

Wang, Y.-C., Magasi, S. R., Bohannon, R. W., Reuben, D. B., McCreath, H. E., Bubela, D. J., Gershon, R. C. and Rymer, W. Z. (2011) 'Assessing dexterity function: a comparison of two alternatives for the NIH Toolbox', *Journal of Hand Therapy*, 24(4), pp. 313-20.

Wanzel, K. R., Anastakis, D. J., McAndrews, M. P., Grober, E. D., Sidhu, R. S., Taylor, K., Mikulis, D. J. and Hamstra, S. J. (2007) 'Visual-spatial ability and fMRI cortical activation in surgery residents', *Am J Surg*, 193(4), pp. 507-10.

Ward, P. J. (2010) 'First year medical students' approaches to study and their outcomes in a gross anatomy course', *Medical Education*, 24(1), pp. 120–127.

Waterson, S. W. and Stewart, I. J. (2005) 'Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students - Waterston - 2005 - Clinical Anatomy - Wiley Online Library', *Clinical Anatomy*, 18(5), pp. 380-384.

WHO (2003) *Consultation on Development of Standards for Characterization of Vision Loss and Visual Functioning. Prevention of Blindness and Deafness* Geneva: World Health Organisation [PDF]. Available at: [http://apps.who.int/iris/bitstream/10665/68601/1/WHO\\_PBL\\_03.91.pdf](http://apps.who.int/iris/bitstream/10665/68601/1/WHO_PBL_03.91.pdf) (Accessed: 8 August 2015).

Wickham, S. 2017. *RE: The role of Widening Participation, Outreach and Recruitment at BSMS*. Type to Gonzales, R.

WikiJob (2019) *What is the UKCAT Test*: Wikijob. Available at: <https://www.wikijob.co.uk/content/aptitude-tests/test-providers/what-ukcat> (Accessed: 10 January 2019).

Wilcox, R. R. (2010) *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer.

Wuensch, K. L. 2004. Independent Variables and Dependent Variables. Greenville NC, USA: Department of Psychology, East Carolina University.

Yen, K. G. (2016) 'Amblyopia', Available: Medscape (Accessed 10 January 2018).

Yilmaz, H. B. (2009) 'On the development and measurement of spatial ability', [PDF] (Accessed 13 October 2014).

Zacks, J. M. (2008) 'Neuroimaging Studies of Mental Rotation: A Meta-analysis and Review', *Journal of Cognitive Neuroscience*.

Zaiontz, C. (2013) 'Kuder and Richardson 20 Formula', *Real Statistics Using Excel*, Available: [www.real-statistics.com](http://www.real-statistics.com) (Accessed 10 September 2017).

Zesiewicz, T. A., Elble, R., Louis, E. D., Hauser, R. A., Sullivan, K. L., Dewey, R. B., Ondo, W. G., Gronseth, G. S. and Weiner, W. J. (2005) 'Practice Parameter: Therapies for essential tremor: Report of the Quality Standards Subcommittee of the American Academy of Neurology', *Neurology*, 64(12).

# APPENDICES

## Appendix 1 – Participant Information Sheet (sample)



Research Study:

### Exploring the Impact of Spatial Awareness Training on Learning Anatomy and Basic Surgical Skills

Dear Student

We would like to invite you to take part in our research study. Before you decide, we would like you to understand why the research is being done and what it would involve for you if you took part.

One of our team members will follow up your initial contact and provide you with this information and answer any questions you may have.

Talk to others about the study if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

**1. What is the purpose of the study?**

This study will help us understand the relationship between three types of personal abilities and the effect that a training intervention might have on performance.

A pilot study, carried out earlier this year with the assistance of other students at university, provided us with some initial information. This time we will concentrate on medical students, to compare results and to explore avenues that might influence the teaching of medical skills in the future.

The tests in this study consist of standard methods of measuring abilities and capacities, some especially adapted for medical participants.

The training component is a special section created to enhance awareness of the ability to understand and manipulate 3-D structures in your mind (spatial ability). All participants will receive training, although some will train before the second testing round, while others will train afterwards.

**2. When and where will the study take place?**

The timeframe of the study spans the first half of 2016.

The first and second rounds of testing and training will happen during the first trimester of 2016 (during modules 103 and 104 of your programme).

All testing and training sessions will take place at Brighton and Sussex Medical School.

**8. What are the possible benefits of taking part?**

The tests will measure your abilities in four areas. If you wish, we can give you the scores you obtain during the testing session. For some of the tests, we can compare your results with those from the average population. Please note that we will not be able to find your scores after the testing session because your participation is anonymous (your student number will be replaced by a random number for data storage).

When we complete the study, you may have the satisfaction of knowing that your contribution was crucial in the design of future training programmes.

**9. Are there any possible disadvantages or risks of taking part?**

Some portions of the tests require a time-limit. Some participants may experience tension and nervousness, which we expect will be comparable to that of playing a board game. Considerably less than what you experience at university exams.

Should you feel too anxious or stressed, please feel free to let us know so we can stop the test or even cancel your participation.

In case you feel stressed afterwards because of the test, please contact the University of Sussex's Student Life Centre, where they can provide support and help.

Tel: 01273 87 6767

Email: studentlifecentre@sussex.ac.uk

Web: www.sussex.ac.uk/studentlifecentre/

**10. What about confidentiality?**

Your contribution will be made anonymous. We do not need your personal data for the study other than to track your progress.

We will use the contact details that you provide only to arrange bookings of time-slots during the study. All your results will enter our data sets under a random number instead of your name.

Your contact details will be in password-protected storage and will not be shared with any other party. Your personal data will be deleted with a secure method at the end of the experiment.

Due to the anonymity of your data, we will not be able to trace your results after they enter our system. That is why your will only have the opportunity to access your results during the testing session.

**11. What will happen if I don't want to carry on with the study?**

You are free to withdraw at any time and without giving a reason. Withdrawing or not joining the study will not have any effect on your academic course.

**3. Who is organising and funding the research?**

Brighton and Sussex Medical School is funding this research project.

**4. Why have I been invited?**

You received an invitation because you recently started training in medicine or in your medical speciality and are building up experience in surgical skills. You have an association with the department of Anatomy at BSMS where new methods of training are under exploration and development. We think that you might have an interest in learning about your own abilities and in helping us endeavour into new areas in medical learning.

**5. Do I have to take part?**

It is up to you to decide whether or not you wish to join the study. We will provide you with information and answer your questions and if you agree to take part, we will ask you to sign a consent form. Your participation has no relation with your academic course.

**6. What will happen to me if I take part?**

We do not expect that the test will have any effect on you, other than a little nervousness because we will ask you to perform tasks within certain limits of time.

**7. What will I have to do?**

Your participation will be required in three stages:

- One initial testing session (baseline) = approx. 45min
- One final testing session (evaluative) = approx. 30min
- Two training sessions (intervention) = approx. 1hr each

On the testing session:

First, we will test your vision (please bring your spectacles or contacts if these were prescribed by your optician), followed by your ability to perceive depth. We will ask you to read text characters off a chart and look at images on a booklet.

Secondly, we will ask you to perform simple tasks with your hands, using small items on a testing board.

Lastly, we will ask you to perform manual tasks using instruments to manipulate small objects on the testing board.

Once you confirm your withdrawal, we will delete your personal data securely from our records. We will not save the data you provide.

**12. What if there is a problem?**

If you have any concerns about any aspect of this study or complaints about the way you have been treated during the study or possible harm you might suffer, you should ask to speak with the researchers who will do their best to answer your questions. The researchers' contact details are at the end of this sheet. However, as we do not intend to ask you to do anything above or beyond your abilities or your comfort, we do not expect any problems.

**13. Harm**

The Universities of Brighton and Sussex have insurance in place to cover their legal liabilities in respect of this study.

**14. What will happen to the results of the research study?**

The results of the study will be published in a scientific journal. We will also present the results at conferences, locally and internationally.

**15. Who has approved this study?**

This study has received ethical approval from the Brighton and Sussex Medical School Research Governance and Ethics Committee (BSMS RGEC).

**15. Contact Details:**

Main Researcher	Research Supervisor
<b>Dr Rene Gonzales</b> PhD candidate at BSMS Email: bsms6349@bsms.ac.uk Mobile Phone: (01273) 877886	<b>Dr Claire Smith</b> Head of Anatomy at BSMS Email: medicineanatomyDA@bsms.ac.uk Phone: (01273) 877890

Thank you for taking the time to read this information sheet.

## Appendix 2 – Participant Consent Form

Sample of participant consent form.



### CONSENT FORM

Title of Project: **Exploring the Impact of Spatial Awareness Training on Learning Surgical Skills**

Researcher: **Dr Rene Gonzales (Supervisor; Dr Claire Smith)**

Please check box

I confirm that I have read and understood the information sheet dated ..... for the study: **Exploring the Impact of Spatial Awareness Training on Learning Surgical Skills**. I have had the chance to ask questions about the study and am satisfied with the answers I have been given.

I understand that my participation in this study is voluntary and that I am free to stop at any time, and I do not have to give a reason for doing so. I also understand that if I ask to stop the study my legal rights will not be affected in any way.

I understand that the personal data I provided (name, email address, phone number) will only be used for booking the test and will not be part of the study. This personal data will not be made public or shared with other parties.

I understand that occasionally an external regulator or funding body may ask to look at the data for this study to check that it is being run correctly.

I understand that if there are any unexpected findings that need further investigation you will inform me so that I can decide if I wish to contact my health care providers.

I understand that my interview might be recorded. I give permission to use the recorded digital material (images, video and audio) for educational purposes and for dissemination of the study.

I agree to take part in the study as described above.

\_\_\_\_\_  
Name of Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

- 
- I have explained the information in this document and encouraged the participant to ask questions and provided adequate time to answer them.

\_\_\_\_\_  
Dr Rene Gonzales  
PhD student BSMS  
Email: bsms6349@bsms.ac.uk

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

# Appendix 3 – Spatial Awareness Training Workbook

Sample of Spatial Awareness Training Workbook, based on Sorby's workbook (Sorby, 2012).

Spatial Ability Study – BSMS 2016 Researcher: Dr Rene Gonzales Student Workbook – Part 1

## Spatial Awareness - Student Workbook – May 2016

Based on Dr Sheryl Sorby's 'Developing Spatial Thinking,' 2012.

### 1. Revolving Solids to Create Objects

Please circle the letter corresponding to the object or objects that were formed by revolving the shape shown on the left about an axis. There may be more than one answer per problem.

1

Spatial Ability Study – BSMS 2016 Researcher: Dr Rene Gonzales Student Workbook – Part 1

### 3. Isometric Drawing & Coded Plans

How to draw an isometric view from a coded plan

When sketching a corner view of an object (Isometric View), sketch one surface at a time until the view is complete. The "X" marks the viewpoint of the observer.

Notes & sketches:

3

Spatial Ability Study – BSMS 2016 Researcher: Dr Rene Gonzales Student Workbook – Part 1

### 2. Combining Solid Objects

The objects on the left are to be combined, with the result shown on the right. Please circle the appropriate word (cut, join or intersect) indicating the operation that was performed.

2

Notes & sketches:

2

Spatial Ability Study – BSMS 2016 Researcher: Dr Rene Gonzales Student Workbook – Part 1

### 3. Isometric Drawing & Coded Plans (Continued)

Please circle the letter beneath the isometric sketch of the object that corresponds to the coded plan shown on the left.

Please complete the coded plan for the object shown in an isometric sketch on the right.

4

## Appendix 4 – Data Collection Sheet for the Study

### Sample of Data Collection Sheet for the Study.

#### Participant Information & Data

STUDENT No (UniCard): ... ASSIGNED No: ... Year at BSMS: ( )1 ( )2

AGE: ... GENDER: ( ) Male ( ) Female GLASSES/CONTACTS: ( ) Yes, ( ) No

HAND DOMINANCE: ( ) Right-handed ( ) Left-handed ( ) Ambidextrous

#### PERCIEVED VISUAL CAPACITY

How do you rate your visual capacity (wearing spectacles if prescribed)?

( ) Poor ( ) Moderate ( ) Good

#### PERCIEVED CAPACITY FOR MENTAL MANIPULATION OF OBJECTS

How do you rate your capacity to visualise objects and turn them around in your mind?

( ) Poor ( ) Moderate ( ) Good

#### PERCIEVED MANUAL DEXTERITY

How do you rate your capacity to manipulate small objects with your hands and fingers?

( ) Poor ( ) Moderate ( ) Good

#### INTEREST IN ANATOMY

What is your level of interest in learning Human Anatomy for your profession?

( ) Little ( ) Moderate ( ) High

#### INTEREST IN CAREER PROGRESSION

What type of medical field would you like to pursue after you graduate from medicine?

( ) Medical ( ) Surgical ( ) Mixed medical/surgical ( ) No progression

#### NOTES:

#### TEST SCORES (to complete by examiner)

TEST	Right Hand Eye	Left Hand Eye	Both Hands Eyes	Sum	Assembly	Final Score	Code Value
Visual Ac Dist.				--	--		
Visual Ac Near	--	--		--	--		
Stereopsis	--	--	--	--	--		
Spatial Ability	--	--	--	--	--		
Manual Dexterity							
Instrument Dext			--		/		

## Appendix 5 – Table of Coding Rules

Table of “Coding Rules” for Variables in the Study.

Item	Variable Label	Variable	Measurement and Code Value
1	Group	<b>Group</b>	Nominal (Intervention=1, Control=2, Base=3)
2	Age	<b>Age</b>	Scale
3	Gender	<b>Gender</b>	Nominal (male=1, female=2)
4	Hand Dominance	<b>HandDom</b>	Nominal (right=1, left=2, ambidextrous=3)
5	Perceived visual acuity	<b>Per_VA</b>	Ordinal (poor=1, moderate=2, good=3)
6	Perceived spatial ability	<b>Per_SpAb</b>	Ordinal (poor=1, moderate=2, good=3)
7	Perceived manual dexterity	<b>Per_MDx</b>	Ordinal (poor=1, moderate=2, good=3)
8	Interest in Anatomy	<b>IntAnat</b>	Ordinal (little=1, moderate=2, high=3)
9	Interest in Career Progrs	<b>IntProg</b>	Nominal (surgical=1, medical=2, mixed =3, no-prog=4)
10	Visual Corr. specs/contacts	<b>VisCorrect</b>	Nominal (yes=1, no=2)
11	Visual Acuity at DISTANCE (corrected, both eyes)	<b>VA_Dist</b>	Scale: Low=0.1; 0.2, 0.32, 0.4; 0.5; 0.6; 0.8; Normal = 1.0, to Above average = 1.25
12	Visual Acuity at NEAR	<b>VA_Near</b>	Scale: Low = 4m = 0.1; 2m = 0.2; 1.4m = 0.28; 1.2m = 0.33; 1.0m = 0.4; 80cm = 0.5; 60cm = 0.66; 50cm = 0.8 Normal = 40cm = 1.0
13	Stereopsis	<b>Stereo</b>	Scale: (max 40seconds or arc = 40/chart value) Low = 800sec = 0.05; 400sec = 0.1; 200sec = 0.2; 140sec = 0.28; 100sec = 0.4; 80sec = 0.5; 60sec = 0.66; 50sec = 0.8; Normal = 40sec of arc = 1.0
14	S.A. Score 1 <sup>st</sup> or 2 <sup>nd</sup>	<b>SpAb_1st (2nd)</b>	Scale (score from 0 to 20)
15	S.A. Questions Attempted	<b>Atmp_1st (2nd)</b>	Scale (number of questions attempted)
16	Spatial Ability in Percent	<b>SA1_Perc</b>	Scale (score * 100 / 20 questions)
17	S.A. Difference Ctrl or A.P.	<b>SA_Dif_Ctr (Int)</b>	Scale (2 <sup>nd</sup> <b>Minus</b> 1 <sup>st</sup> score. Neg. values in <b>Red</b> )
18	Right Hand	<b>R_hand_1 (L)</b>	Scale (test score)
19	Both Hands	<b>BothH_1</b>	Scale (test score)
20	Sum of Hands	<b>Sum_1</b>	Scale (Sum of R + L + Both)
21	Assembly	<b>Assmb_1</b>	Scale (test score).
22	Measured Manual Dext.	<b>MDx_1</b>	Scale ( <b>Sum</b> of Hands + Assembly)
23	Right Hand Instrument	<b>Rh_In_1 (Lh)</b>	Scale (test score)
24	Sum of Hands Instrument	<b>Sum_In_1</b>	Scale (test score)
25	Assembly Instrument	<b>Asm_In_1</b>	Scale ( <b>Average</b> of both trials)
26	Measured ManDex Instr.	<b>MDx_In</b>	Scale ( <b>Sum</b> : Hands Instr.+ <b>Avg</b> Assembly Inst.)
27	Manual Dexterity Differ.	<b>MDx_Dif</b>	Scale (2 <sup>nd</sup> score <b>Minus</b> 1 <sup>st</sup> score MDx)
28	Man. Dext. Instrum. Diff.	<b>MDxIn_Dif</b>	Scale (2 <sup>nd</sup> score <b>Minus</b> 1 <sup>st</sup> score MDx_In)
29	ManDex Global	<b>MDx_Glb</b>	Scale ( <b>Sum</b> of MDx + MDx+In)
30	ManDex Global Difference	<b>MDx_Glb_Dif</b>	Scale (Diference of global scores <b>in %</b> )
31	Anatomy Module Mark	<b>AM_102_of22</b>	Scale (Module N <sup>o</sup> + Mark out of N <sup>o</sup> of questions)
32	Anatomy Final Mark	<b>An_Final_Y1</b>	Scale ( <b>Sum</b> of Modules + Mark out of 86/125 quest/yr)
33	Anatomy Module in Percent.	<b>AM_x02_Perc</b>	Scale (mark * 100 / max N <sup>o</sup> questions per module)
34	Anatomy Final in Percent	<b>An_Final_Perc</b>	Scale (mark * 100 / max N <sup>o</sup> questions / year)
35	Position in own Year	<b>Position_Year</b>	Scale (position in student ranking of own year)
33	Decile in own Year	<b>Decile_Year</b>	Ordinal (position in decile rank of year: 1-10)

## Appendix 6 –Spatial Ability Questions’ Index of Difficulty

Question Number & Object	Rotation sequence	Rotation & Shape Disparity Complexity	Participant Response Based Difficulty Index
3 	$\curvearrowright +X$	$1 \times 4 = 4$	<b>1.32</b>
4 	$\curvearrowright +Y$	$1 \times 2 = 2$	<b>1.39</b>
2 	$\curvearrowright +Z$	$1 \times 1 = 1$	<b>1.70</b>
1 	$\curvearrowleft -Z$	$1 \times 3 = 3$	<b>1.76</b>
7 	$\curvearrowleft -Y$ $\curvearrowleft -Y$	$2 \times 2 = 4$	<b>1.82</b>
5 	$\curvearrowleft -Z$	$1 \times 1 = 1$	<b>2.14</b>
6 	$\curvearrowleft -Y$ $\curvearrowleft -Y$	$2 \times 5 = 10$	<b>2.14</b>
12 	$\curvearrowleft -Y$ $\curvearrowleft -Z$	$2 \times 5 = 10$	<b>2.89</b>
13 	$\curvearrowright +X$ $\curvearrowright +Y$	$2 \times 4 = 8$	<b>3.02</b>
11 	$\curvearrowleft -Z$ $\curvearrowleft -X$	$2 \times 2 = 4$	<b>3.146</b>
8 	$\curvearrowleft -X$ $\curvearrowleft -X$	$2 \times 2 = 4$	<b>3.27</b>
14 	$\curvearrowleft -Y$ $\curvearrowright +X$	$3 \times 1 = 3$	<b>3.40</b>
10 	$\curvearrowright +Y$ $\curvearrowright +Y$	$2 \times 4 = 8$	<b>4.03</b>
17 	$\curvearrowright +Y$ $\curvearrowright +Z$ $\curvearrowright +Z$	$4 \times 3 = 12$	<b>4.09</b>
15 	$\curvearrowright +Y$ $\curvearrowright +Z$	$3 \times 3 = 9$	<b>4.28</b>
16 	$\curvearrowright +Y$ $\curvearrowright +Y$ $\curvearrowright +Z$	$4 \times 2 = 8$	<b>4.72</b>
19 	$\curvearrowleft -X$ $\curvearrowright +Y$ $\curvearrowright +Z$	$4 \times 2 = 8$	<b>4.78</b>
9 	$\curvearrowright +Z$ $\curvearrowright +Z$	$2 \times 2 = 4$	<b>4.84</b>
18 	$\curvearrowleft -Y$ $\curvearrowright +Z$ $\curvearrowright +X$	$4 \times 3 = 12$	<b>5.41</b>
20 	$\curvearrowright +Z$ $\curvearrowright +Z$ $\curvearrowright +X$	$4 \times 3 = 12$	<b>6.67</b>

**Table 22. Spatial Ability Questions Ordered by Index of Difficulty**

Note: The numbers on the left column are the question number on the Spatial Ability Test (ROT), re-ordered based on the difficulty index in column 5 (far left). The rotation sequence expressed by symbols corresponding to 90-degree rotation steps. Rotation and Shape complexity added in column 4. The participant-response-based index of difficulty resulted from the inverse of questions most commonly answered correctly.