Assessing the applied anatomy knowledge of medical students: the effect of visual resources on preparing them to become new doctors

Mandeep Gill Sagoo

UCL Institute of Education

Doctorate in Education (EdD)

2017

Declaration

I, Mandeep Gill Sagoo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Word Count (exclusive of appendices, references but including diagrams and tables but including figures and tables): 46,683 words

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr Charlie Owen and Dr Maria Kambouri for their support, guidance and patience.

I would like to acknowledge Dr Caroline Pelletier for her input in structuring my work. I am grateful to the Institute of Education, UCL and King's College London for their support throughout the journey.

I would like to thank my parents, friends and family for having faith in me, encouraging and understanding me. A special thank you to my husband Sukhdav Sagoo for the unwavering support and encouragement throughout the journey. To Sadhguru and Mahantesh, for helping me redefine my energies towards serenity. To my lovelies Navi, Priya, Rano and Manju for inspiring me and being a blessing to me, and to my most supportive mentors, Dr Richard Wingate and Professor Harold Ellis. I would also like to thank Dr Rachel Claire Aland, Dr Harvey Meller and Dr Bryan Cunningham for proof-reading and giving me their suggestions on the thesis.

To my mother Mrs Harinder Kaur Gill, father Dr Rupinder Singh Gill and granddad Dr Darshan Singh Gill, for being my ultimate pillars. Thank-you for all your blessings and prayers. It would not have been possible without all your support and trust in me.

Lastly, I would like to thank my Babaji from all of me for all of it!

Abstract

Anatomical examinations are timed examinations that assess topographical and/or applied knowledge of anatomy with or without the inclusion of visual resources i.e. cadaveric resources, cadaveric images, radiology and/or clinical findings images. Although advances in the multimedia learning theories have led to greater understanding of how we process textual and visual material during learning, the evidence with regard to the use of illustrations within written assessments is scarce. This study investigates whether the presence or absence of images (cadaveric, clinical findings and radiological images) within clinically-oriented single-best-answer questions has a significant influence on medical students' performance. A questionnaire was also included to determine the effect of students' characteristics and preferences in learning and assessments on their performance.

Second year medical students (n=175) from six UK medical schools participated voluntarily. All questions were categorised as to whether their stimulus format was purely textual or included an associated image. The type of images and deep components of images (whether the question is referring to a bone or soft-tissue on the image) was also taken into consideration. Further investigation was carried out on the question-difficulty and the regional anatomy of the questions. These examination scores were then analysed along with students' responses collected on the questionnaire. This was further illustrated with students' feedback.

The study demonstrates that inclusion of images, the deep component of an image, question difficulty and regional anatomy impact students' performance. Moreover, students' preferences play an important role in their performance.

Anatomical and radiological images are critical in the medical profession in investigating and examining a patient's anatomy, and this study set out a way to understand the effects of these images on commonly employed written assessments. This study has shown that image factors and student factors impact on the students' performance. Further research is needed to refine these examinations.

Table of Contents

List of Figures

List of Tables

Reflective report

I have been working in the field of anatomy since completing my Bachelor of Science in India in 2004. During my Master's degree in medical anatomy, I worked as a part-time anatomy demonstrator until 2006. From the early days of my undergraduate degree I have been fascinated with the subject of human anatomy and its unique cadaveric resources. This interest further developed when I had the opportunity to build my knowledge through research and teaching.

I immigrated to London around nine years ago as a newly married woman and an anatomy graduate, looking for a place in the huge world of academia. It was an enormous challenge, and I did not know where to start but I certainly knew that I was not ready to lose myself in the crowd.

For the past nine years, I have worked as an anatomy demonstrator, prosector, technician, teacher, module lead and plastinator at St. George's University of London and King's College London, and for the Royal College of Surgeons. It would not be incorrect to say that I have done every job in the field up to the position I hold now. The journey was certainly full of highs and lows, and I learnt a huge amount on the go, but my passion towards the subject never wavered.

The developments that have occurred in the areas of anatomy makes me feel like a time-traveller who had the opportunity to study anatomy in the historic style used during the 19th century in the UK (Older, 2004; Drake et al., 2009) and to apply the knowledge in the 21st century through my current position as an anatomy teaching fellow. I used the term "historic style" because the system of teaching and assessing anatomical knowledge in Northwest India at the time was inspired by English anatomists before the independence of India in 1947. This included a large number of dedicated hours for full body dissections, small group learning facilitated by a group of teachers and didactic lectures. The students' knowledge was assessed via practical and written assessment methods. The practical methods included spotter tests and oral (viva-voce) examinations with various examiners assessing the students' knowledge on various topics such as osteology, histology and topographical anatomy. The written examinations included short and long essay-type questions, which were marked by a panel of internal and external examiners. However, in the UK, the numbers of hours dedicated to teaching anatomy have been immensely reduced to accommodate other relatively new disciplines in the curriculum (Papa and Vaccarezza, 2013). The teaching has shifted from didactic lectures to small group dissecting room or laboratory based sessions and self-directed learning. Oral and long essay type written examinations are no longer popular in the UK; these have been replaced by single-best-answer questions (SBAs) and extended matched questions (EMQs) (Yaqinuddin et al. 2013; Smith and McManus, 2015).

The drive to work in the anatomy assessment area arose while I was undertaking Postgraduate Certificate of Healthcare Education (PgCert HE) in 2009. As part of this course, I chose to write an assignment on anatomy practical examination. While reviewing the literature for that assignment, I realised that in the UK anatomy is taught and assessed through a variety of methods decided by factors such as tradition, convenience and the availability of visual resources (cadaveric resources, images and/or video clips) (supported by Gunderman, 2008). A significant number of institutions no longer use cadaveric resources to teach and/or test the knowledge of anatomy (Rowland et al., 2011). Several factors have influenced this shift such as the legislative constraints of the Human Tissue Act 2004 in the UK and other respective policies in other countries; the scarcity of cadaveric material; the financial and logistical constraints of establishing and the ongoing management of a dissection room with the necessary resources (Older, 2004; Liddell and Hall, 2005). This was intriguing and highly relevant to my daily practice and inspired me to research further in the area.

In the past nine years of my working career, I have witnessed various gaps in anatomy assessment systems. The urge to seek answers, to grow in the field and to contribute meaningfully fed my interest, and around 2010, I knew the time had come to pursue further studies. I undertook the Doctorate in Education (EdD) programme in October 2011, intending to improve my understanding of the area from a wider angle. This reflective statement outlines my learning journey into and throughout the EdD programme.

Despite my original interest in the anatomy assessment systems, I chose to write my first assignment on the Human Tissue Act (HTA) 2004 for the Foundations of Professionalism (FoP) module. It was stimulating to read about the history of anatomy in England, and this gave me insight into various Anatomy Acts that had existed since 18th century. This made me realise the bond of professional continuity I always shared with anatomists of that time. In the assignment, I presented the issues and tensions produced by HTA 2004 within the anatomical profession. Reading the literature was thought-provoking and informative; however, I could not do it justice. Specifically, in presentation; and balancing the aspects on policies and my reflection - and I achieved a grade lower than I hoped for. I was disappointed at the time, and since then I have attended a number of writing workshops within and outside the Institute. In retrospect, these assisted me in refining my writing skills and equipped me with necessary academic skills for the future.

During the Methods of Enquiry (MoE) 1 module, I returned to my original interest in anatomy assessment, and wrote an assignment titled “Efforts towards Renaissance of Anatomy Assessment System: What works and what does not work and why?” It helped me to delve into the literature referring to the process and methods that require work to improve the system. Moreover, I applied the lessons learnt from my FoP assignment and paid close attention to presentation, critical analysis and balancing the information; and managed to achieve a high grade, which certainly increased my confidence in my ability to pursue this scholarly research.

For the MoE2 work, I interviewed academics to find out their views on differences in resources and methods used for testing anatomy knowledge. The process was stimulating as I never had interviewed anyone before. I passed the module with a high grade, and the feedback truly helped me in focusing the issues that needed further work.

From here my journey of Institutional Focused Study (IFS) began. For this study, I designed a summative anatomy practical examination with variations in the constructs and resources used for questions, and evaluated the students’ performance. The IFS work was demanding and time consuming, especially negotiating ethics approval, access and cooperation from the medical school, my colleagues and students, and analysing the results data. While I successfully completed the IFS, the low grade furthered my interest in investigating anatomy question design and performance. I planned to design a formative online assessment of clinically-oriented questions with and without images to assess the performance of medical students for my final EdD thesis.

I started my EdD thesis proposal in July 2014, and encountered a number of hurdles. Overcoming these hurdles refined my work, although this was difficult to realise at the time. Until early January 2015, I planned to involve a certain group of participants. After almost two months of negotiations with the Royal College of Surgeons' committee, their decision was not in my favour. This was an incredibly difficult period; both professionally and personally. During this process, I received an email from someone influential in the field with regard to my efforts towards acquiring ethics approval that definitely moved the ground under my feet. I was prohibited to contact any external boards and contacts or else I may risk damaging what I wanted to achieve with my study. After this incident, I thought it would be best to apologise for any unintentional damage I may have done in trying to get permission for the data collection. Although I had no intention to hurt anyone in the process, it may have happened because of cultural differences and/or my assertiveness.

Reflecting on the practicalities and limitations (being a little fish in the sea with big fish) and with the support of my friends and colleagues, I critically evaluated my original proposal in light of committee's rejection, and as a result, I decided to alter my study design and employ a group of undergraduates at the end of their pre-clinical years instead. This was done because anatomy is only explicitly taught in year 1 and 2, and students are expected to cover the learning objectives by the end of year 2. I approached ten UK medical schools, explained my project to the Heads of anatomy departments and/or relevant teaching and learning committees, and finally after successfully overcoming the challenges, I was granted ethics approval and formal permissions from six medical schools to proceed. To visualise the scope of the project, and the interconnections between the various elements, I drew a colour-coded mind map on the wall of my room (figure 1).

Figure 1: Mind-map of the project

My amended proposal passed the review process on 22nd July 2015. Along with acquiring necessary scholarly skills, gaining constructive feedback at various conferences and gaining a publication in Anatomical Sciences Education journal (Figure 2), this journey has undeniably given me a courage to establish my voice in the field and ability to critically reflect on feedback, and being flexible in my approach. I feel I have developed in my ability to critically evaluate the feedback that I receive, listen to different opinions and to take these into account to broaden my understanding and approach. Moreover, I feel the experience has improved me as a person and led me to acquire skills and knowledge for future scholarly endeavours. Professionally, since then I have had many opportunities to get involved in funded collaborative projects in the area, and meaningfully contribute and participate in the community of education scholars.

My journey towards investigating and evaluating anatomy assessment has enlightened me on how this area is inextricably intertwined with subject areas such as anatomy assessment models, the utility model, constructive alignment, cognitive load theory, and psychology of visuals in cognitive theory of multimedia learning. This had enhanced the strength and integrity of my present study. This experience has made me open-minded and willing to seek a cross-disciplinary approach to look for answers, and appreciate both pragmatic and constructivist viewpoints.

For my thesis work, I started with searching, understanding and describing various systems of assessing knowledge of anatomy with and without visual resources. This was followed by testing the performance of medical students from six medical schools in the UK on an online applied anatomy test particularly devised for this study. The test scores were linked and analysed with the students' views/preferences obtained on a questionnaire in order to discuss the outcome in the light of wider literature. The purpose of this study is empirical (in search of evidence), interpretative (in search of understanding) and instrumental (in search of praxis).

Figure 2: Overview of my EdD learning journey

Work

Anatomical Prosector at St. George's University of London

Anatomy demonstrator at King's College London

Anatomy Teaching Fellow at King's College London

EdD

Foundation of Professionalism (FoP)

Methods of Enquiry 1 (MoE 1)

Methods of Enquiry 2 (MoE 2)

Institutional Focused Study (IFS)

Thesis

Learning

Writing skills
NVivo skills
Time management skills
Mind-mapping for linking concepts and their categories
Mendeley and End note understanding for reference managing
Learning SPSS statistical software
keeping a daily journal called "DayOne"
Collaboration skills with people of different expertise
Designing an assessment
Experience of presenting and writing a manuscript
Experience of going through review process for publication
Learnt to listen, appreciate other people's views
Learnt to work in a team

Presentations & publications

Sagoo, M.G. (2014). Relationship between medical students' performance and format of the questions in a practical anatomy assessment. British Association of Clinical Anatomists, Summer Scientific meeting, July 3. Presented the work at Institute of Education Doctoral School Summer Conference, June 19. Presented at Learning in a Changing World Conference, November 22

Sagoo, M.G. (2015). Rethinking the assessment of applied anatomy knowledge of medical students. Anatomical Society Winter meeting, December 14-16

Sagoo, M.G., Smith, C.F., Gosden, E. (2016). Assessment of anatomical knowledge by practical examinations: The effect of question design on student performance. *Anatomical Sciences Education,* vol. 9 (5), pp. 446-452

Sagoo, M.G. (2016). Designing and assessing the applied anatomy knowledge of medical students: the effect of visual resources in single best answer format. King's College Alternative Assessment Symposium, May 12

Sagoo, M.G. (2016). Visuals in anatomy assessments. Association for Dental Education in Europe, August 24-26

Sagoo, MG. (2016). Panel moderator for Anatomy@#Roads – Applied anatomy curriculum, pros and cons of dissection based teaching versus online teaching, and Guest speak for visuals, cognition and human anatomy, International conference in anatomy-colloquium on comparative, congenital and clinical anatomy, September 23-25

Chapter 1 – Introduction and Rationale

The importance of anatomy in medicine and its impact on medical students' clinical careers is well recognized (Older, 2004; Orsbon et al., 2014). Anatomical knowledge is crucial for developing a working diagnosis, and for carrying out many clinical procedures safely and effectively (McHanwell et al., 2007). Junior doctors routinely use their anatomical knowledge during physical examination of a patient (Vorstenbosch et al. 2016) and for interpreting radiological images (Dettmer et al. 2013). With increasing expertise, this knowledge becomes "encapsulated" in clinical concepts and used more implicitly (Boshuizen and Schmidt, 1992; Schmidt and Rikers, 2007). For example, a doctor needs anatomy knowledge to understand why a fracture or a lesion leads to sensory and motor loss of specific areas, how involvement of neighbouring structures add to a patient's symptoms in a disease, why specific tests are performed to check normal working or damage/disease of certain ligaments and vessels, how different types of haemorrhages (extradural, subarachnoid, subdural etc.) appear on computed tomography (CT) scans etc. The advent of sophisticated imaging techniques and growth of medical specialties like interventional radiology provide new areas in which the knowledge of anatomy is vital (McHanwell et al, 2007; Holland et al., 2015).

On the other hand, time available for teaching and assessing anatomy within the medical curriculum has been shrinking each year (Drake et al., 2009). The evolution and expansion of other relevant disciplines fills the limited space within the curriculum (Drake et al. 2009). This has led to decades of on-going debates on its place within a crowded curriculum. Furthermore, there have been discussions on whether its subsequent reinforcement in later years of their training would be beneficial to students for integrating anatomy with clinical and other relevant sciences better (Drake et al., 2009; Gogalniceanu et al., 2009; Standring and

Larvin, 2011). There is a perception that changes in the curricula of undergraduate medical schools in UK have been implemented without rigorous research, and that these changes have affected the proficiency of future doctors (Older, 2004, Sugand et al., 2010, White and Sykes, 2012). A survey conducted by the Royal College of Surgeons (RCS) in 2009 of nearly 1,000 medical graduates from 13 universities, found that more than 50% of candidates cited “poor anatomy teaching at medical school” as a reason for not pursuing a career in surgery (Gogalniceanu et al., 2009). The Medical Protection Survey into insurance claims against surgical procedures showed that approximately half of the claims related to laparoscopic surgery were caused by inadvertent damage to adjacent structures, presumably contributed to by poor anatomical knowledge (Rainsbury et al., 2007). This is in line with Standring and Larvin (2011), who reported that inadequate knowledge of anatomy is a likely cause of increased medical errors, and consequent morbidity and mortality leading to a rise in litigation.

To compensate for loss of time dedicated to anatomy at undergraduate level, some teachers have moved from a traditional approach to an integrated one. Rather than teaching detailed topographical anatomy in pre-clinical years then focusing on applied anatomy in the clinical years, the move has been towards an integrated (basic and clinical) approach in pre-clinical years followed by application in the clinical years (Fraher and Evans, 2009; Ahmed et al., 2010). Anatomy teaching therefore has shifted towards decreased contact time, increased integration of basic and clinical sciences, and increased use of technology and electronic devices (Raftery, 2006; Sugand et al., 2010, Pawlina, 2009; DiLullo et al., 2009; Nicholson et al., 2006). Traditional didactic lectures have been replaced by teaching strategies like case-based learning, team and problem-based learning

(Chakravaraty et al., 2005; Vasan et al., 2008; 2009, 2011; Yiou and Goodenough, 2006; Philip et al., 2008).

However, assessment practices have grappled with how to assess an individual's ability to apply anatomical knowledge in a valid and reliable way while simultaneously identifying gaps in knowledge or understanding (Smith and McManus, 2015).

A variety of assessment methods have been developed and employed (Rowland et al, 2011); however, the application of these methods in anatomy has been reported by very few (Moqattash et al., 1995; Lukic et al., 2001), and those papers address more general aspects of assessment; validity, feasibility and reliability, rather than considering the multifaceted nature of anatomy, i.e. intrinsic nature of visual resources in anatomy. Broadly, assessment methods are categorised as oral examinations (viva), written examinations (paper based or online), and practical examinations (Rowland et al., 2011).

Oral examinations are now seldom employed in the UK because of perceptions of bias, low reliability per hour of testing time, lack of assessor reliability, and because these are time-consuming (Wass et al., 2001, Smith and McManus, 2015). However, they offer an emphasis on nomenclature, function and clinical and spatial relationships (Clough and Lehr, 1996), and continue to be used in the United States, Australia and New Zealand (Clough and Lehr, 1996; Fabrizio, 2013). Written assessments are common; including subtypes such as single best answer questions (SBAs), true/false multiple choice questions (MCQs), extended matching questions (EMQs), short answer questions, free response questions, essay questions, and key featured questions (Anderson, 1979; Schuwirth and Van der Vleuten, 2003, 2004; Schubert et al., 2009; Smith and McManus, 2015).

Practical assessments include Objective Structured Practical Examinations (OSPE), steeplechase, spotter tests, tag tests, think-tank, three-dimensional multiple-choice test and Integrated Anatomy Practical Paper (IAPP) (Nayar et al., 1986; Peel, 1998; Schuwirth et al., 2001; Chirculescu et al., 2007; Schubert et al., 2009; Inuwa et al., 2011; Menzes et al., 2011; Shaibah and Van der Vleuten, 2013; Yaqinuddin et al., 2013; Smith and McManus, 2015). These examinations test factual and/or applied anatomy knowledge with or without the inclusion of visual anatomical resources. In medical graduates pursuing surgical specialties, the applied knowledge of anatomy is additionally assessed through the Membership Examination of the Surgical Royal Colleges of Great Britain (MRCS). MRCS is a membership examination designed for candidates in the generality part of their speciality training in surgery. It is a crucial milestone that must be achieved if trainees are to progress to speciality surgical training as defined by the Surgical Speciality Advisory Committees. It is a multi-part examination consisting of both theory and practical assessments. Part A of the MRCS is a written examination, divided into two papers that examine applied basic science knowledge and general surgical sciences using Single Best Answer and Extended Matching Item questions. These questions are constructed with clinical case scenarios, and the scenarios consist of patients' symptoms and history followed by some information of the diagnostics such as blood test and imaging results. This is followed by a lead-in question linking to the scenario. In Paper 1 (Applied Basic Sciences) of Part A MRCS, the total number of single best questions is 135. Out of these, approximately 40-45 questions, 1/3rd of the paper relate to topographical, applied and surgical, developmental and imaging anatomy. Part B Objective Structured Clinical Examination (OSCE) integrates basic surgical scientific knowledge and its application in clinical surgery. This is done through a series of working stations reflecting elements of day-to-day clinical practice (Intercollegiate Committee for

Basic Surgical examination 2012/2013 annual report). Each timed self-contained station comprises questions relating to topographical, applied and surgical anatomy supplied with cadaveric specimens and/or radiological images. In Part B, applied knowledge of anatomy is tested through 3-4 stations. These stations assess anatomical knowledge through prosections, medical images and bones.

Recent work favours approaches that facilitate the application of knowledge in practice in anatomy assessments for undergraduates and graduates (McHanwell et al., 2007; Smith et al., 2016). These assessments are usually characterized either according to levels of Miller’s pyramid; knows, knows how, shows how, and does (Miller, 1990) or Bloom’s taxonomy; knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956). In principle, the use of contextual information (use of clinical case scenarios) uplifts the level of knowledge assessed from "knows" to "knows how" in Miller's pyramid (Miller, 1990), and targets up to level 3 of modified Bloom’s Taxonomy (Bloom, 1956; Palmer and Devitt, 2007). Appropriate use of contextual clinical information in measuring higher cognitive level, application of knowledge, problem solving ability and critical thinking have been supported (Page and Bordage, 1995; Case and Swanson, 2002; Schuwirth et al., 2001; Papachristodoulou, 2010; Sood and Singh, 2012; Smith and McManus, 2015). Molyneux and Robson (2012) emphasized that assessment within the clinical context enhances the expansion and application of anatomical knowledge. They assessed the students’ knowledge through an online assessment, which was developed with traditional spotter type questions, and functional and clinical image-based questions for MBBS year 1 to 4. Qualitative and quantitative data collected from students (n =96) and clinical tutors (n =23) demonstrated an overwhelming positive response, i.e. they liked clinically-oriented anatomy questions. This supported the notion that clinically oriented questions are more appropriate stimuli because these provide a closer approximation to real life (Van der Vleuten et al., 2010; Yaqinuddin et al., 2013). Smith and McManus (2015) criticised the older spotter tests for only testing identification, and MCQs for not reflecting the three dimensionality and application of clinical anatomy and thus not reflecting the case based, spiral, integrated nature of the curriculum.

They devised an integrated anatomy practical examination (IAPP) – a progression from the OSPE suggested by Yaqinuddin et al., in 2013. The IAPP tested integrated knowledge of anatomy, histology, pathology, physiology and pharmacology through a written examination. Free response format was employed in IAPP to avoid the potential downside of students guessing answers in selected response format such as SBAs and MCQs.

As surgical techniques are advancing from open (invasive) to laparoscopic and endoscopic (less invasive), there is an increasing demand for detailed knowledge of imaging anatomy, emphasising the multifaceted nature of anatomy beyond cadaveric anatomy (Phillips et al., 2013). Visual resources in anatomy provide an intrinsic, built-in meaning that makes them essentially different from, for instance, images that are used to illustrate texts (Schnotz, 2002). This refers to supplementary nature of anatomical visuals as compared to images that are only used to illustrate the textual information. To operate effectively, almost all clinical specialities rely heavily and increasingly on medical imaging. Images provide a powerful learning stimulus and help medical students understand anatomy both in health and disease (www.rcr.ac.uk). Vorstenbosch et al. (2016) emphasise that the ability to acquire adequate visual internal representations of anatomical information is an important element of learning anatomy. Considering this, testing the quality of the acquired internal visual representations and interconnections between internal and external representations ought to constitute a substantial part of anatomy assessment. This is consistent with Hegarty et al. (2009) who implied that doctors have to rely on internal representations because internal structures of the body are not directly visible. A Likert-style questionnaire study by Smith and Mathias (2010) conducted on 4th and 5th year medical students (n = 256) suggests that visualization of three-dimensional cadaveric anatomy is important in developing the skills necessary for daily clinical practice. However, in examination scenarios, visualisation is also known to promote a positive cueing effect, which is an effect that makes an examinee answer a question correctly just by recognising the correct option or an image, rather than generating the answer de novo (Schuwirth and Van der Vleuten, 2004).

Although there is extensive work reporting the role of images in learning, (Mayer, 2005b; Paivio 1986; Baddeley, 1992; Chandler and Sweller, 1991) there is little on the role of images in assessment (Vorstenbosch et al. 2013, Hegarty et al., 2009; Smith and Mathias, 2011; Hunt, 1978; Hisley et al. 2008; Inuwa et al. 2011, 2012; Khalil et al., 2005; Schubert et al., 2009; Smith and Mathias, 2010; Yaquinuddin et al., 2013; Chirculescu et al., 2007; Schoeman and Chandratilake , 2012; Holland et al., 2015). Most of this research focused on the influence of images on recognition memory or the transfer of learning content from images to text and vice versa (Standring and Smith, 1975; Brainerd et al., 1981; Ginns, 2005; Beagle, 2009; Witteman and Segers, 2010) or images working as a motivational benefit for learners (Ainsworth, 1999). Although most authors conclude that images improve learning, in the area of assessments some of these findings are ambiguous, especially in relation to the response format and/or stimulus (based on context or visual resources) involved. Most of this research is focused on text supported by images (complementary images) (Crisp and Sweiry, 2006), taking place in laboratory settings and testing mainly identification skills (Holland et al., 2015). With regard to anatomy education, generalization of these findings would become more difficult if the process of learning anatomy is considered as one of learning images supported by text, rather than vice versa (Schnotz, 2002).

Despite advances in cognitive theory of multimedia learning - that is, in the understanding of how we process verbal and visual material during learning - the research on impact of images within written assessments is still scarce. The present study aims to examine whether the inclusion of images within the stimulus format of single-best-answer questions has any significant influence on learners' performance.

In this study, I assessed the students' anatomy knowledge through their performance on clinically-oriented single-best-answer assessment with presence and absence of images. The test scores obtained from the assessment are analysed with the students' preferences obtained on a questionnaire, and illustrated with the students' feedback. The outcome is discussed within a broader literature framework. This is relevant to modern anatomy education because assessment facilitates learning (Bergman et al., 2011) by influencing students' future learning patterns, and knowledge of various multimedia props in anatomy plays a key role (Schoeman and Chandratilake, 2012). Exploring the effect of images in anatomy assessment is a valuable step, because images are among common features in increasingly multimodal assessments of anatomy. The purpose of this study is empirical (in search of evidence), interpretative (in search of understanding) and instrumental (in search of praxis).

Aims

The aim of the study is to investigate how images used in clinically-oriented anatomy single-best-answer questions affect the medical students' performance on the assessment - “What effect do anatomical and radiological resources involved in the assessment have on the students' performance?”

Assessment scores are related to participants’ views collected through a self-report instrument (closed-ended questionnaire), and these data are illustrated by participants’ feedback.

The first part of this thesis will review the assessment of anatomy knowledge. In the second part, the relationship between the use of images and participants’ scores will be investigated through an online single-best-answer assessment. As performance is influenced by factors such as cognitive level, demographics, prior knowledge, learning and assessment preferences (Mayer and Massa, 2003; Leite et al., 2010; Smith and Mathias, 2010), these factors will be investigated by the questionnaire study.

Research questions

How does students' performance vary on clinically-oriented anatomy questions in the absence and in presence of images?

Sub-questions:

To investigate whether the following factors influence students' performance:

1. Students' characteristics
2. Students' preferences and experiences (for learning and assessment of anatomy)
3. Students participation in anatomy demonstrating activities

The sub-questions were included to investigate whether those additional factors such as sex, age range, training level (school leavers or mature students), most likely prospective career (surgical, non-surgical or don't know), preferences of resources for learning anatomy (cadaveric resources, clinical findings photographs or radiology images), and participation in voluntary anatomy demonstrator programme influence the performance of students.

Researcher-practitioner role

In this study, I play an insider researcher role because I work as an anatomist. The action aspect of my work gives me the motive to identify the loopholes in anatomy assessment system and understand the reasoning in order to suggest and implement actions in future, evaluate, and be meaningfully attentive throughout the process (Crotty 1998). According to Wellington (2000), Smyth and Holian (2008), a practitioner-insider researcher has the advantages of prior knowledge, familiarity with the system and its working, easier access to the samples and huge opportunities to have a significant impact on the system; however, the possible problems of preconception, prejudice, researcher's status in the organization, too much familiarity with the system, loyalty to the workplace and dealing with sensitive information could prove to be disadvantageous. As all the data collected from the students was through an online resource, there was no face-to-face interaction; therefore, no direct impact would have interfered with the data collection. Moreover, efforts have been made to address my personal bias through clear reviewing, validating and piloting processes.

This study is focused on the assessment system to investigate the performance of students on clinically-oriented anatomy questions with and without images. To investigate and support the rationale, I reviewed the literature to facilitate my understanding and enable me to recognise gaps in my knowledge and the literature.

Thinking about the details of the data that I was planning to gather raised further questions:

- What would it mean if students perform better or worse on image-questions (and various types of images)? Does absence or presence of images make a question easy or difficult or is it context based or both?
- Would students' views, preferences and formal learning processes affect their performance on the questions?

In this study, one possibility is that images may promote positive cues and thus make a question easy to answer. Although every effort was made in the study to design the questions with applied clinical relevance and to discourage the process of only testing identification abstract knowledge, the use of images may facilitate better performance on questions with images, as suggested by Multimedia theory – this states that if an image is familiar to the students this may facilitate them to answer a question correctly because of easy retrieval of knowledge gained during learning (Mayer, 2009).

A second possibility is that interpretation of images may add to an individual's cognitive load (Sweller, 1994). Having never seen a particular kind of image and/or not having appropriate schemas/mental images built in one's memory may add to one's extraneous cognitive load to interpret an image in a question. If the previous knowledge is superficial, and the knowledge is built by using a particular type of visual resource, then the use of a different type of image in an assessment may interfere with existing knowledge and thus negatively impact on the students' performance.

A third possibility is that questions with no images may make answering a question difficult or easy. If the previous knowledge and mental images required to answer a question are not or are only partially present, then questions (without images or with images different from the ones previously encountered) may interfere and thus negatively impact on the performance of the students. It may be easier for some who have deep knowledge and have appropriate mental models (cognitive construct formed from pre-existing knowledge) to answer a question correctly.

Hence, considering the criticality of visuals in anatomy assessments, this study investigates the students' performance in clinically-oriented anatomy questions with and without visuals. It further investigates the effect of various types of images (anatomical and radiological images) on the students' performance.

Chapter 2 – Literature review

I began this journey by revisiting the literature gathered during my taught modules and institutional focused study (IFS). To look for more sources, I used Google, Google Scholar, Medline and Educational Resources Information Centre (ERIC) as search engines and databases to collate the information. The keywords used were education and anatomy assessment, formative assessment, online and practical anatomy tests, role of visuals in assessments, psychometrics of assessments, and applied anatomy tests with visuals. Later I came across the literature in the area of cognitive psychology of visuals, and cognitive theories of multimedia learning. For the questionnaire design, I looked for predictors of anatomy performance, and students' views and preferences in anatomy education. These search engines were very helpful but as soon as a few key players were identified, I adopted a snowballing process to understand the ins and outs of the subject area. This was not easy and it took me a long time to establish a workable link between the role of visuals, anatomy assessments and educational psychology of visuals.

In this chapter, I start with a brief description of the assessment system and its domains in medical education, especially in relation to human anatomy. This is followed by funnelling through approaches to the role of the visuals in educational psychology and anatomy to understand the students' performance on the test used in the study.

Assessment is an important aspect of education because it tests learners' competence, facilitates future learning patterns, and informs the quality of institutes' educational processes (Schuwirth and Van der Vleuten, 2006; Larsen et al., 2008, 2009; Vorstenbosch et al., 2014). Assessments are broadly categorised as formative and summative (Derrick et al., 2009). Formative assessment provides learners with constructive feedback; therefore, helps to develop their autonomy and sustainable learning; whereas summative assessment relates more to accountability and certification (Black and Wiliam, 1998a).

Before moving forward, I would like to express my understanding of the word “competence”. According to Epstein and Hundert (2002), the Accreditational Council for Graduate Medical Education defined six areas of competence and some means of assessing them: patient care (including clinical reasoning), medical knowledge, practice based learning and improvement (including information management), interpersonal and communication skills, professionalism, and systems-based practice (including health economics and team work). I see competence in anatomy as being a subset of competence in medical knowledge, and my understanding of anatomy competence resonates with the definition suggested by Van der Vleuten and Schuwirth (2005), i.e. the ability to handle a task by integrating the relevant cognitive (anatomy knowledge), psychomotor (skills gained while working in anatomy laboratories) and affective (skills gained while working with cadaveric resources and in a team) skills.

Assessments are often designed to assess cognitive, psychomotor and/or affective domains - the domains are categorised into “knowledge/content dimension” and “cognitive process/progress dimension”. In anatomy, the content dimension includes anatomical terminology and facts, conceptual knowledge, procedural knowledge (the knowledge of methods or procedures) and the metacognitive domains (i.e. knowledge of the principles and generalisations, theories, structures and abstraction in a certain field). The “progress dimension” demonstrates understanding of facts and ideas by organizing, comparing, translating, interpreting, and applying the knowledge gained (Brenner et al., 2015). In this study, the test is aimed to assess elements of both content and progress dimensions of applied anatomy knowledge through an online assessment.

In medical practice, there is a tendency to view assessment programmes as a whole rather than as separate assessments (Van der Vleuten and Schuwirth, 2005; Dijkstra et al., 2010). As anatomy is a multifaceted subject that ties cognitive aspects of text and visuals together, this study focuses on assessing students' performance on an assessment that tests applied knowledge of anatomy with and without visuals.

Many undergraduate medical programmes require students to acquire and display an ability to identify and interpret various types of images, i.e. histological, radiological, and anatomical images (Bloodgood and Ogilvie, 2006; O'Brien et al., 2008). Ideally, these skills should be assessed in an aligned outcomes-based curriculum (Biggs, 1996; 2003) because most qualified doctors are required to investigate and examine their patient's anatomy via physical examination or radiographic means, notwithstanding that those who specialise in areas such as surgery will go further (Benninger et al., 2014; Gunderman and Wilson, 2005; Sugand et al., 2010).

With the focus on visuals, in educational psychology, these resources have been viewed as complementary to texts, i.e. visuals illustrating the text. This means learning occurs if both verbal and visual information are simultaneously available in working memory (Paivio, 1986). Moreover, the form of visualization affects the structure of the mental model constructed during learning (Mayer, 2009), which in turn influences the patterns of performance; this eventually has an effect on the ease or difficulty of translating or applying knowledge gained from previous visual resources onto new (novel) ones (Schnotz and Bannert, 2003; Schnotz and Kurschner, 2008; Schnotz and Baadte, 2015).

In anatomical literature, the use of various anatomical visual resources has been supported as well as challenged for decades; i.e. some prefer to teach and test anatomical knowledge with cadaveric resources in a dissecting room environment, whereas others are more inclined towards computer-based tests (McWhorter and Forester, 2004; Khalil et al., 2005; Swanson et al., 2006; Fraher and Evans, 2009; Smith and Mathias, 2010, 2011; Schoeman and Chandratilake, 2012). Additionally, the literature does not provide evidence of superiority of one type of resource to another (Older, 2004; Rowland et al., 2011). In this study, I do not attempt to prove superiority of any particular resource; instead, I investigate, through the medium of an online formative test, whether or how absence or presence of various types of images in applied anatomy questions affect medical students' performance. In anatomy, it is important for students to understand visuals in depth in order to interpret the same structures and their relationships in various types of images. The reviewers support the use of visuals, although they are divided on which type of visual is most effective (some support cadaveric and/or other support online resources) (Rowland et al., 2011; Orsbon et al., 2014) but the literature on these visual resources, their effect on learners' performance, and associated complexities with them is scarce. There are a lot of opinions but not a lot of empirical research of students and their views on the system. This gave me more impetus to investigate students' performance and views to fill this gap.

This study is grounded in cognitive theories proposed by Mayer (2009) and Schnotz and Bannert (2003), and the focus is on the performance of students. In this study, I will first describe relevant models of cognitive theory of multimedia learning (Mayer, 2009), and an integrated model of text and pictorial comprehension (Schnotz and Bannert 2003). Thereafter, I will discuss the findings within a general framework of learning from multiple external representations. I will refer to related research findings and point out the need and implications for development of multimedia assessment environment.

Mayer's cognitive theory of multimedia learning and its consistency with other theories

In the 17th century, Comenius published the “Didacta Magna”, which emphasised that envisioning information is extremely important for effective learning. In the cognitive theory of multimedia learning (CTML), the central idea is that people learn better from words (spoken or written) and images (illustrations, photos, animation or videos) than from words or images alone (Multimedia principle - Mayer 2009). The key elements of CTML are: the dual-channel assumption, the limited capacity assumption, the active processing assumption, and a number of acting instructional principles that have been identified and developed in the area. This is consistent with the following:

1. Paivio's dual coding theory (Paivio1986; Clark and Paivio, 1991)
2. Baddeley's model of working memory (Baddeley, 1992)
3. Sweller's cognitive load theory (Chandler and Sweller, 1991; Sweller et al., 1990; Kalyuga and Sweller, 2004; Van Merrienboer and Sweller, 2010)

Paivio’s dual coding theory assumed that the human cognitive system includes verbal and imagery subsystems. Words and sentences are usually processed and encoded only in the verbal system, whereas pictures are processed and encoded both in the imagery system and in the verbal system. The integrative processing through referential connections is most likely to occur if verbal and visual information are simultaneously available in working memory. This enables the construction of mental models (cognitive construct formed from pre-existing knowledge) (Mayer, 2005b; Paivio 1986; Baddeley (1992; Chandler and Sweller, 1991).

Based on Baddeley's model of working memory (1992), two sensory subsystems exist in working memory: an auditory system and a visual system. Mayer's first basic assumption on multimedia learning merges these two concepts. Humans are posited to process information in working memory through two channels: an auditory-verbal channel and a visual-pictorial channel. His second basic assumption, reflecting both the work of Baddeley (1992) and of Chandler and Sweller (1991), is that these two channels have a limited capacity to convey and process information. The third basic assumption is that humans are active sense-makers; they engage in active cognitive processing to construct coherent knowledge structures, or "schemas", from both the available external information and their prior knowledge. "Schemas" are meaningful sets of connections that correspond to specific concepts and experiences, and the acquisition of expertise in an area can be characterised by development of this idiosyncratic memory (Regehr and Norman, 1996). Sweller (1994) described two critical learning mechanisms; schema acquisition and the transfer of learned procedures from controlled to automatic processing (also referred as schema construction and schema automation). The schema is a cognitive construct that organises the elements of information according to the manner they are dealt with. These schemas permit people to readily solve problems that otherwise would require immense effort. Schuwirth and Van der Vleuten (2011) described the collection of isolated facts, which combine to build schemas in medical education. These schemas are then aggregated into concise and dense illness scripts. In due course, with experience, these are further enriched into instance scripts, which enable an expert to recognise the patterns instantaneously.

In line with limited capacity assumption, Sweller's cognitive theory (Sweller, 1994) adds another element in schema acquisition, i.e. cognitive load theory (CLT). This refers to the limitation of working memory which, if overloaded, can hinder learning. Three types of cognitive load described by Sweller et al. (1994) are: Intrinsic load, extrinsic load and germane load. Intrinsic load is generated by the intrinsic complexity of the task; and it occurs during the interaction between the nature of the task and the competence of the learner. This cannot be externally manipulated. Extrinsic load depends on the impact of settings in which the information is delivered and tested; and it is not directly relevant to the task. This can be manipulated and therefore, should be minimised. Germane load is caused by learning processes that deal with intrinsic cognitive load, i.e. referring to how the information is processed (organised and integrated). This is devoted to schema acquisition and automation, and it can be manipulated (Sweller et al. 1994). According to Mayer (2009), it is important to "manage essential processing, reduce extraneous processing and foster generative processing" (p. 57). However, the distinction of intrinsic and germane cognitive load has been challenged by De Jong (2010). According to him, "intrinsic load and germane load are defined in terms of relatively similar learning processes; the difference between the two seems to be very much a matter of degree, and possibly non-existent" (p.111). However, DeLeeuw and Mayer (2008) have supported this triarchic model of cognitive load, and suggested that these can be measured by different assessment instruments.

With regard to visual resources; there is increasing evidence that although high-fidelity reproductions, or simulations maintain authenticity, they also increase cognitive load in novice learners, and that students perform better when interesting but extraneous information is excluded (Chen et al., 2015; Cook, 2006; Mayer and Monero, 2003). Although the notion of multimedia theory that multiple representations can complement each other by providing or supporting complementary information in learning has been widely accepted (Ainsworth, 1999), the parallels drawn between text processing and picture processing have been questioned (Schnotz and Bannert, 2003), as described in the next paragraph.

Alternate model of multimedia learning

Representation principle

An alternative model of multimedia learning, which gives more emphasis to representational principles, has been suggested by Schnotz and Bannert (2003). These representational principles are categorized into descriptive and depictive forms for text and pictures, which are based on different sign systems (symbols and icons) (Schnotz 1993). Textual information is represented as "descriptive representation" and pictures as "depictive representations". Words and sentences are known as examples of "symbols" and have arbitrary structure, whereas pictorial illustrations are examples of "icons" and have abstract structure. It is considered that in both text and picture comprehension, an individual combines internal and external representations. Internal representations are referred to mental models (cognitive construct formed from pre-existing knowledge). External representations contain external textual and visual displays, and these external representations are understood when a reader constructs internal mental representations of the content described in the text or shown in the pictures. Thus comprehension of text and pictures is a task-oriented construction of internal mental representations. According to Schnotz (1993), the central point in relation to text comprehension and graphics comprehension is that both types of information employ qualitatively different representation principles. Since the two formats have different proximity to propositional representations and mental models, they also contribute in different ways to the process of knowledge acquisition. A text leads to the construction of a propositional representation, which then allows constructing a corresponding mental model. A graphic, however, provides the possibility of a relatively direct construction of a mental model, in which mapping of entities from the graphic is done onto entities of a mental model.

Surface and deep structures

Furthermore, generative linguistics suggested that “sentences” have a surface structure and a deep structure. However, graphics, as explained earlier, have usually been viewed as complementary to texts and known to provide illustrative information (Paivio, 1986) and elaborate conjoint processing (Kulhavy et al., 1994). Schnotz and Baadte (2015) emphasised in their model that like textual information, graphics also have perceptual surface structure and a semantic deep structure. The surface structure of a graphic includes dots, lines, areas and their visual features; whereas, the deep structure of the graphic is a semantic construct which expresses the meaning of the image. Perceptual processing of an image includes identification and discrimination of graphic entities as well as the visual organisation of these entities according to the Gestalt laws (Winn 1994), which refers to synergy of components rather than addition. Thus comprehension of graphics is a process of schema-mediated structure mapping from external graphics on internal mental models (Schnotz and Bannert, 2003).

In order to understand a picture rather than only perceive it, semantic processing is known to play a vital role. While not directly linked to CTML and CLT, this has some connections with theories of perception organisation (seeing things and making sense of what is seen) and neuro-scientific understanding of visual pathways. According to the former, viewing of an object requires two sources of information - the sensory input from the stimuli (Gibson, 1979) and use of schemas stored in the brain (Gregory, 1973). According to Gibson's bottom-up approach, perception is an innate ability. The sensory inputs received from a stimulus are highly rich and organised; and to perceive a stimulus, optic arrays from the source fall on the receiver's retina, which are taken to the visual cortex through the visual pathway, and this is sufficient to interpret an image. It does not require contextual information or pre-existing information to make sense of a stimulus. However, according to Gregory's top-down approach, perception is an active and constructivist process i.e. it requires various cognitive processes, (pre-existing knowledge in the form of existing schemas) to interpret a stimulus. Neisser (1976) suggested a combination of the above two approaches and recommended a cyclical process i.e. we see the perceptual world through bottom-up approach; and when we pay attention, this sets in motion a search for existing schemas to find links with what is seen. The latter, neuroscience of visual pathways, explains the process of how light intensity, edges and other features of a visual stimulus form an image on the retina. The photoreceptors pick the information and convert it into electrochemical signals that are transmitted through optic nerves on either side. Some (nasal) part of the optic nerves on either side cross to form the intersection of the optic chiasma. This distinction in crossing of some part of the optic nerves and not all is the reason why in layman's language we often refers that the right side of the world is viewed by left brain (cerebral) hemisphere and vice versa. Each optic nerve is a bundle of axons from the ganglionic cells in the retina. These axons reach the lateral geniculate nucleus of the thalamus. From here the axons send signals to the visual cortex. The ventral and dorsal output pathways from the visual cortex help to process the location of a stimulus/object and also help in guiding to reach that stimulus, and process the information that leads to the identification and recognition of an object (Crossman and Neary, 2014). Hence from the explanations above, there is involvement of both bottom-up and top-down approaches in the visual perception process.

Thus, an individual has to construct a mental model of the depicted subject matter through a schema-driven mapping process, in which graphic entities are mapped onto mental entities. In other words, picture comprehension is considered as a process of analogical structure mapping between a system of visuo-spatial relations and a system of semantic relations (Schnotz, 1993). This mapping can take place in both directions i.e. it is possible to evaluate an existing mental model bottom-up from a picture and it is also possible to evaluate an existing mental model top-down with a picture.

Furthermore, pictures can be categorised into realistic and logical pictures; and one requires different cognitive schemata to evaluate these images. In understanding realistic pictures, an individual uses cognitive schemata of everyday perception. In understanding logical pictures, on the contrary, an individual requires specific cognitive schemata (so-called graphic schemata) in order to read information from the visuo-spatial configuration (Lowe, 1996).

Hence mental models are multiple layers of depictive internal representation formed from external and already existing internal representations, and are not sensory specific. For example, a mental model of a spatial configuration (say, of a room) can be constructed not only by visual perception, but also by auditory, kinaesthetic, or haptic perception. Because mental models are not bound to specific sensory modalities, they can be considered as more abstract than perceptual images. A mental model could contain more or less information than a visual image depending on what we could internally create or extract by looking at the image. So prior knowledge plays a key role.

Cognitive benefits and costs of visual resources

Schnotz and colleagues (Schnotz and Bannert, 2003; Schnotz and Kurschner, 2008; Schnotz and Baadte, 2015) raised an important point that visual resources not only come with cognitive benefits, (as explained by multimedia theory), but could also be associated with cognitive costs. A precondition to making sense of an external representation requires prior knowledge for integrating the external representation with internal representation and thus using mental models for comprehension. The literature suggests that enhancing graphics comprehension by visual design and learners' cognitive activation induced by instruction is a matter of complex interactions between numbers of factors:

- Perceptual surface structures
- Semantic deep structures
- Perspectives of different familiarity,
- Cognitive schemata associated with these perspectives,
- Interference between schemata, whereby inference depends on the cognitive load imposed by the interfering schema.

Schnotz and Kurschner (2008) suggested that these interactions co-determine the process of construction of mental models in graphics comprehension.

There are a number of reasons described in the literature for the apparent superiority of images over text for the ease of learning. According to Biedermann (1981), the general meaning of an image can usually be grasped in as little as 300 milliseconds. This may be because the elements of a visual resource can usually be processed simultaneously, whereas a text must be processed sequentially (Winn, 1987). Moreover, visual resources are likely to play a large role in the development of a student's mental model and more emphasis is likely to be placed on the ideas communicated by them than the ideas conveyed by the associated text. As Peeck (1987) states, 'too much attention may be deployed to the illustrations themselves rather than to the accompanying text' (p. 118). She describes a previous study (Peeck, 1974), in which students were presented with a story that sometimes contained a mismatch of information between text and image. During questioning, the students tended to choose the responses consistent with the visual resources more frequently than the responses that would be indicated by the text, suggesting a dominating influence of the images.

The cognitive benefits and costs of visuals with regard to learning and assessments are discussed below.

Benefits and costs of visual resources in learning

The use of appropriate visual resources in learning has been studied in a number of contexts and most authors agree that the effects are beneficial (Carney and Levin, 2002; Crisp and Sweiry, 2006; Mayer, 2009). Levie and Lentz (1982) performed a review of 55 experiments comparing learning from illustrated text with learning from text alone, and concluded that in 85% of these cases, illustrated text significantly improved retention compared to text alone. Carney and Levin (2002) also explored these concepts, reporting larger effect sizes on learning from images used for interpretational purposes, as opposed to those which were simply used for decorative purpose. The use of images is also reported to enable better visualisation and the development of spatial ability in learning (Mathai and

Ramadas, 2009). Graphics are thought to ‘simplify the complex’ and ‘make the abstract more concrete’ (Winn, 1989, p. 127). Peeck (1993) makes a similar point when she writes that images ‘might help to clarify and interpret text content that is hard to comprehend’ (p. 228). It is also argued that graphics can provide more information than can be explained in words (Stewart et al., 1979).

However, not all research has found images to be beneficial. In a review of studies on instructional texts, Levie and Lentz (1982) found that in about 15% no significant effects of including images were observed. One possible explanation is that the choice of image is important. Peeck (1987), for example, found that participants who read a text without a diagram were actually more motivated and more interested to continue reading than those who read the same text accompanied by a poor diagram. This suggests that visual resources are not always beneficial and that the quality and appropriateness of a visual resource are important. The failure of visual resources to aid instruction in some studies has often been explained as either a result of the students’ learning styles, as Ollerenshaw et al. (1997) report, or due to students not processing illustrations adequately (Weidenmann, 1989). The latter is thought to be a result of the apparent ease of processing an image giving students the false impression that they have fully understood an image when they have not (Weidenmann, 1989). In addition, Winn (1989) warns text designers of making assumptions that all students will process a particular image in a particular way. This idiosyncrasy of interpretation is also implied by Elkins (1998), an art historian, who asserts that visuals do not provide meaning via an orderly set of signs in the same way as a text.

The above research gives some insights about the positive and negative influence of visuals on learning and retention. However, as the main purpose of this study is to investigate the effect of images in an assessment rather than how it affects learning, the following literature highlights the role of visuals in assessments.

Benefits and costs of visual resources in assessments

The use of visual resources has been studied in an assessment context; including studies by Crisp and Sweiry (2006), Schnotz and Baadte (2015) and Knauff and Johnson-Laird (2002).

Crisp and Sweiry (2006) investigated the effects of visual resources in examination questions and, in particular, how and when students use images and whether subtle changes to these salient physical features steer their understanding towards the way intended by the question-setters. Sixteen-year-old students (n = 525), across four secondary schools, participated in their study. The test paper contained six questions based on past examination questions, and these involved graphical elements. For five of the six questions, two versions were designed in order to investigate the effects of changes to visual resources on processing and responses. They had two groups of students with similar ability. Twenty-seven pairs of students were interviewed afterwards. When two versions of a question were tested in parallel, the differences in the images significantly affected marks of one question and had smaller effects on marks and the nature of answers with some of the others. There were mixed views from students with regard to whether an image that is not strictly necessary should be used. Some considered it reassuring, whilst others deemed it unnecessary. It was found that if an image provides a cue to an answer, this might be used in preference to information in the text (Fisher-Hoch et al., 1997). However, there may be risks of including images in examination conditions that do not match with students' pre-existing knowledge and/or with the meaning intended by the question-setters. According to Pollitt and

Ahmed (1999), when a student reads a question, a mental model (or mental representation) is constructed as a response to the external representation being processed. This mental model is composed of images, concepts and emotions, and the relationships between concepts. It is based on ideas already known to the reader (Johnson-Laird, 1981) and hence will be the reader's own personal understanding of the text. Therefore, students' mental representations of the text and pictures may not all be the same across the board except for some salient features.

In Schnotz and Baadte's experiment (2015), 157 students (average age 23.8 years) from different faculties of a university in Germany participated. They were randomly allocated to six different treatment groups to receive different learning material and instructions. Learning content was simple and was aimed at participants with no prior knowledge. Students were asked to learn about the voting behaviour of voters with different political orientations and religions in the US presidential elections of 1956 and 1960. These two groups were further divided into three subgroups – the first group received no instructions, the second group received a party instruction (congruent with party graphs and incongruent with religious graphs) and the third group received a religious instruction (incongruent with party graphs and congruent with religious graphs). All participants received a 168-word text, combined either with party graphs or religious graphs and either with no instruction, party instruction or religious instruction. Participants were requested to memorise the information but they were not allowed to take notes. They found what they were expecting; the incongruent instruction and graph negativity affected their performance. Although they responded better in this study on what they were instructed, there was a shortage of time given to memorise the details, and these results may have differed if the experiment was repeated in an actual learning environment.

The further detailed investigation of students' performance on semantics of an image comes from graphics comprehension notion suggested by Knauff and Johnson-Laird (2002). They showed that mental models could differ from visual images and that different brain areas are involved in creating visual images and spatially organised mental models. When different graphics convey the same information in diverse ways, they look dissimilar and therefore have different surface structures. Therefore, it is not sufficient just to deliver correct information via graphics, but it is also important to choose an appropriate perceptual format for the display of information corresponding to a perspective that makes the intended schematic deep structure as transparent as possible (Schnotz and Bannert, 2003).

Although the above literature addresses some elements of cognitive benefits and associated cognitive costs of visual resources (as suggested in multimedia learning theory and by Schnotz), and the perceptual surface and semantic deep surface of graphics, the key issue of the effectiveness of these visual resources is yet to be explored in the field of medicine.

Images have a significant role in medicine and these are the basis of fields like radiology and interventional surgery, robotics and laparoscopic surgery (Dettmer et al., 2013). In anatomy, which is regarded as a multifaceted and foundation subject for medicine, surgery and radiology, visual resources play a key role (McHanwell et al., 2007; Schoeman and Chandratilake, 2012). According to Holland et al. (2014), anatomical teaching has relied upon multiple techniques to impart information, including didactic lectures, imagery and small cadaveric group tutorials. This has been supported by many studies (Fitzgerald et al., 2009; Smith and Mathias, 2011) in which it has been emphasized that junior doctors intensively use their anatomical knowledge in clinical reasoning and throughout the consultation, and that they frequently use visual representations of the anatomical information they needed (Vorstenbosch et al., 2015).

In anatomy, the particular importance of semantic processing to understand a visual resource as opposed to merely perceiving it, has also been emphasized by Schnotz (2005). Furthermore, Hegarty et al. (2009) has suggested that along with the ability to correctly make links with internal and external representations, future doctors (students studying hands-on subjects like dentistry and anatomy) are also required to have internal representations to understand the inside of a patient's body without viewing it directly. Thus building mental representation in pre-clinical years is an important aspect of medicine.

There is a lack of empirical evidence on anatomy and radiology assessments with regard to the inclusion of images within written examinations. Moreover, there is a shortage of guidance with regard to use of images in these types of questions (Case and Swanson 2002; Wood et al., 2004). The following literature highlights the differences in use of images and their effect on students' performance.

Role of visuals in anatomy assessments

Some authors propose that addition of images within written assessments has a consistent influence on performance; however, the conclusions have been conflicting. These effects depend on whether the images are considered by students to be irrelevant, helpful or essential in order to answer the question (Crisp and Sweiry, 2006). In one of the arithmetic examinations, it has been suggested that the presence of images increase item difficulty and slow down the speed at which students are able to process information, leading to increased testing time and item difficulty (Berends and Van Lieshout, 2009). However, Vorstenbosch et al. (2013) study shows that the use of images within multiple-choice format does not lead to a predictable effect but instead may have variable effects on individual items.

In the field of medicine, studies have been conducted that investigated students' responses and preferences on various types of visual resources, i.e. labelled images versus textual material; images versus textual description of images; cadaveric versus online resources; online interactive images, static line diagrams versus real objects; cadaveric and textual material; and simplistic diagrams versus histology images. Some of these studies showed consistent effects (positive, negative or no effect) whereas others showed inconsistency in students' performance and preferences (Vorstenbosch et al., 2013, 2014; Hunt, 1978; Inuwa et al., 2011, 2012; Khalil et al., 2005; Schubert et al., 2009).

In a study by Vorstenbosch et al. (2014), it was suggested that the process to answer questions with and without images requires different cognitive processes. In examining these effects with 17 first year students, by means of think-aloud protocols in an experiment, the authors proposed that textual options promote elimination of distractors and internal visualisation of answers, while visual options promote cueing and the ability to interpret visual information. In addition, they suggest that the use of some images, particularly cross-sectional anatomy, test abilities beyond anatomical knowledge or understanding, and conclude that students with high spatial ability are less influenced by the form of the response format. In a previous study, Vorstenbosch et al. (2013) analysed 39 extended-matching questions, grouped within seven themes; one version of each theme had a labelled image, while the other had an alphabetical list of textual options. On analysis, the use of images appeared to produce conflicting effects; fourteen items were more difficult when using a labelled image as opposed to textual options, while ten items were easier. Examination of item discrimination also showed contrasting effects; images reduced discrimination in five items, but increased it in two other fives. Interestingly, in these studies students expressed no clear preference for either the use of text of images, and the authors concluded that both are appropriate formats to use in examining medical students and graduates (Mayer, 2010; Vorstenbosch et al., 2013, 2014).

Hunt (1978) examined the effect of radiological images in multiple choice questions (MCQ = 70) on final year medical students. One group of students received questions containing written descriptions of the diagnostic images within their vignettes, whereas the other group received a booklet of images (anatomical, diagnostic and radiological images), containing high-fidelity reproductions of the images themselves. Overall, students who were required to interpret the original images or radiographs had a poorer performance than those provided with the written description (32.9% vs. 38.9%). Hunt explained that interpretation of radiographs is a complex extra task, which has influenced the results. However, the effect of these images was not consistent; fourty-three items were found difficult with the inclusion of an image, eighteen were easier for the students to answer correctly, and the remaining nine items showed no difference between the two groups. One example was described in detail; whereby a question with an image of a barium swallow, was answered correctly by 85% of students, as compared to a question with the written X-ray report, where only 35% chose the correct option. However, students who answered the image-based question incorrectly were all middle- and high-performers in the overall test. On further inspection, it appeared that most students had interpreted the image incorrectly, choosing the right option but for the wrong reason, and this calls for detailed understanding of cognitive processes while answering questions with and without images. Hence no consistent response on questions with and without visual resources was seen in the above studies. Moreover, students had no clear preference when images were labelled versus textual options; however, a poorer response was seen in Hunt’s (1978) study when students had to interpret an image.

However, in a study by Holland et al. (2015) first year medical students (n = 277-347 per year) over three consecutive years were tested for recognition and understanding only through questions with and without inclusion of relatively simple diagrammatic and histological images. Item analysis of three consecutive years of histology MCQ examinations were analysed (total no. of questions 195) and the mean values showed no significant difference in item discrimination or difficulty with and without inclusion of an image.

Owing to the convenience of online resources over cadaveric resources, Inuwa et al. (2011, 2012), Khalil et al. (2005) and Schubert et al. (2009) conducted the following studies and found no difference in students’ performance on different types of resources (cadaveric, online, line diagram and/or text-only resources). However, students’ preferences were inclined towards online resources in some studies (Inuwa et al., 2011, 2012; Khalil et al., 2005) and dissecting material resources in others (Schubert et al., 2009).

A study by Inuwa et al. (2011, 2012) compared first and second year students’ performance on "factual anatomy" questions with two different types of visual resources; cadaveric and online resources. The rationale for their study was lack of cadaveric resources, extreme wear and tear of the available resources and no prospects of replacing dried and damaged specimens. The experiment was conducted on two groups; one was tested with actual specimens, in dissecting room circuits through traditional “steeplechase” examination, and the other with online radiological images, prosected specimen photographs and short video clips. The same tutors taught these students, equal contact time was given to them and equal credit weighting courses were paired. The results showed no significant difference in the two groups’ mean scores; however, more than half of them preferred online over the traditional examination (Inuwa et al., 2011). Although no difference was seen in the mean scores of the two groups, it does not prove the similarity of the effect of these resources. It may be as a consequence of combination of multiple factors – adequate mental models to cope with various types of images, transition to and from two and three dimensional visual resources, and/or pros and cons of practical and online examinations: such as relative advantage of seeing specimens three dimensionally in a practical examination as compared with two dimensional images used in online examinations, relative disadvantage of inconsistency of cadaveric specimens used in different circuits in practical examinations as compared with consistency of images used in online examinations, single window to answer each question in practical examinations as compared to flexibility of moving back and forth in online examinations, and disadvantage of a set time of 1-1.5 minutes on each station in practical examinations as compared to the flexibility to answer questions in online examinations. Moreover, the administration of online examinations is relatively less demanding because unlike practical examinations, these do not require assembling and dissembling, and these can be easily changed each year without worrying about having adequately dissected specimens. As an online examination can be conducted within a single session, it is relatively more secure than practical examinations where students go through the circuits in groups.

Khalil et al. (2005) investigated undergraduates' achievement scores on the canine skull anatomy learnt by using three different types of visual resources: computer based interactive images, paper based static line drawings and paper based drawing with real objects. Sixty-four out of 67 freshman veterinary students (50 females and 14 males) volunteered to participate in the study. Group A (22 students) utilised computer based instructional material. Group B students (22) used paper-based instructional material, and Group C (20) used paper based instructional material along with real objects. Textual information was identical for all. The study used a pre-test/post-test comparison group design. Students were asked to identify structures marked by arrows and the time allowed to answer each question was 1.5 min. After the pre-test, students participated in three different types of learning for 45 minutes. Then their performance and perceptions were assessed on the two imagery strategies. The data were examined by analysis of covariance (ANCOVA) that was carried out using pre-test scores as a covariate to adjust the post-test scores and compare the effectiveness of each learning strategy. No significant differences in scores were observed in the two imagery strategies in the "immediate recall of anatomical information". The results however indicated comparable effect between computer-based interactive imagery and paper based static imagery with real objects. There was, however, a significant difference in students' opinions toward the two strategies; students' perceived computer based interactive imagery a better strategy in the assimilation of anatomical information than paper-based static imagery. Although this study resonates with Inuwa's study, it is not devoid of limitations – one major limitation is the extremely short period of learning (45 mins) between pre- and post-tests which may have been the cause of there being no significant difference in scores.

Another study that is in line with the above studies is that of Schubert et al. (2009) that compared multiple choice tag test (3D-MC) assessing "factual knowledge" with the use of prosected specimens, histological slides, models and radiographs in a dissecting room setting, with text-only multiple-choice questions (MCQs). Sixty-one medical students at the end of the first semester participated in the test. No significant differences between the mean scores of the tests were found. However, despite the fact that text-only MCQs covered exactly the same knowledge as the corresponding 3D-MC, the two tests varied significantly in “students' perception of difficulty”. Students found the 3D-MC questions easier to answer, which suggests benefits of visual resources over textual information.

Numerous empirical studies have found inconsistencies in students’ performance on assessments using factual questions with and without various types of visual resources (Khalil et al. 2005; Inuwa et al., 2011, 2012). Although students and teachers appear to prefer visual resources in anatomy (Older, 2009; Rowland et al., 2011; Orsbon et al., 2014), the evidence is inconclusive in terms of difference in students’ performance on clinically-oriented anatomy questions with and without relevant images.

Hypothesis

Based on the literature above, I hypothesise that inclusion of images in questions should have a positive effect on the students’ performance as compared to text-only questions.

According to cognitive load theory, schemas are built during learning which develop from controlled to automated mental models with repetition of knowledge and its application (Schuwirth and Van der Vleuten, 2011). In line with Mayer's cognitive theory of multimedia learning, all kinds of images would enhance performance for all kinds of tasks (Mayer, 2009). However, considering the concept of cognitive benefits and costs of images proposed by Schnotz and Bannert (2003), and the anatomical evidence based studies above on question difficulty and students' performance and views (Inuwa et al., 2011, 2012; Holland et al., 2015) I am proposing this as a hypothesis to test.

Chapter 3 – Study design

A quasi-experimental design was employed in the study - the medical schools were selected and the participants volunteered to take part in the study. The test was conducted under the same examination conditions, and the students answered questions with no visuals, and with anatomical images and radiological images. Ten UK medical schools were initially selected and approached to gain permission to access their students to conduct the study. These ten medical schools were selected on the basis of visual resources used in their anatomy teaching, and their accessibility. This information was investigated through each medical school's anatomy webpage and relevant contacts. These schools utilise either all or some combinations of available anatomical resources, i.e. dissections (dissecting cadavers), prosections (pre-dissected body parts) and radiological images. However, only six of them granted the permission in the time frame available. Fortunately, in those six medical schools, there was an acceptable distribution of anatomy resources used for teaching i.e. three used prosections and radiological images, two employed radiological images only, and one school involved dissections, prosections and radiological images for teaching anatomy.

In the study, the participants were medical students from six UK medical schools. These students were at the end of their second year, and they volunteered to participate in the study. As this test was released around two months before their final examinations, students opted to take this as a free revision tool for testing their knowledge of applied anatomy.

The rationale for selecting pre-clinical medical students is because anatomy is formally taught in the first and second years of a degree in medicine. Therefore, it was believed that the group was homogenous in regard to prior knowledge because all the students were due to take their 2nd year final examination in around one or two months' time. Moreover, it was believed that students at this stage of their medical degree have mental models to deal with the images used in the test, and therefore it was possible to assess them on the anatomical and radiological images. Furthermore, as the questions were reviewed by the anatomy leads/academics of the respective medical schools, it was confirmed that students are at a similar level with regard to the content of the test (questions-context and images) and the way it was displayed.

To ensure that the students were not coerced into participation, an introductory email and an advertisement flyer were sent to the students through their school's administrative or anatomy departments. See figure 3.

Figure 3: Advertisement flyer

On the day of the test, the students were asked to register and login through their unique medical schools' email addresses. As they logged in, the "participant information sheet (PIS)" and the "consent form" were presented to the students first. The PIS form included the information of the purpose of the study; reason for choosing them as participants; any associated risks and benefits; data protection information; participants' anonymity; right to withdraw from the study, information on how the data will be stored and how their participation will be protected. These forms were designed using British Educational Research Association (BERA) guidelines. The consent form was included to obtain their informed consent to use the data for analysis and dissemination. This was followed by the questionnaire and the applied anatomy test with and without images.

Creation and rationale for designing the online tool – My Anatomy Growth

A home coded online tool was used for the study. The coding and software building was done by my husband (who works as a software programmer) and the layout and content was designed by myself following my extensive reading and exposure in the area. This online tool was hosted on Microsoft Azura cloud for maintaining the data securely. The content was reviewed by a group of academics and it was piloted on a group of volunteers prior to releasing it to the six medical schools for the data collection.

During designing the content of the tool (participant information sheet, questionnaire and test questions) and acquiring permissions, I inspected other online assessment and survey tools available in the market. I tried and tested free versions of many online tools, for example, Articulate, Question-mark perception, Googleforms, Opinio and SurveyMonkey. Although each tool has unique proprietary components, none of them fulfilled the requirements of the study for one or the other reason. Some were merely assessment tools and there was no way to provide feedback to the students on each question to help improve their knowledge, and gain feedback from them. Some online tools provided an empty box at the end for students to provide their feedback on the test but there was no facility to provide them with feedback on each question explaining why a correct answer is correct and why a distractor is incorrect. Others had issues with uploading different formats of images. Some were good as a questionnaire tool (such as SurveyMonkey) but were incompatible to be used as a joint assessment-feedback tool. Regarding the technical aspect, some were not adequately compatible with various browsers, machines and smartphones. Others did not provide any option to customise the look to present one question per page rather than scrolling down. The ones that fit most of the criteria were too expensive. For example, the cost of a yearly access to Question-mark perception tool per student was £5, and I was envisaging around 200 or more students to take the test which would have escalated the cost. While to the best of my knowledge the information is accurate as of the time this work was done, I also recognize that softwares evolve quickly and these descriptions may have become outdated.

The pros and cons with various tools gathered through a variety of sources, including vendor websites, phone calls with technical support, and software trails are listed on the next page.

Table 1: Pros and cons of various assessment and questionnaire tools available

	Articulate	Question-mark perception	Googleforms	Opinio	Survey Monkey	My Anatomy Growth
Appropriate feedback display	✗	✓	✓	✓	✗	✓
Access to recording time taken for each question	✗	✗	✗	✗	✗	✓
Secure authentication without having their email addresses	✗	✗	✗	✗	✗	✓
Appropriate display and edit facilities for each question	✓	✗	✗	✗	✗	✓
Downloading different types of images	✓	✓	✗	✓	✗	✓
Data collection convenience	✓	✓	✗	✗	✗	✓
Integrated test-survey tool	✓	✓	✗	✓	✗	✓
Display compatibility with smartphones	✓	✓	✗	✗	✓	✓

Secure system	✓	✓	×	✓	✓	✓

Therefore, my husband kindly helped me with development of an online platform that was simple and fulfilled all the requirements for the study. We coded and designed it and named it as “My Anatomy Growth”. This facilitated inclusion of key features, i.e. the use of various images in a particular format, integrating the consent form and questionnaire within the tool, providing immediate results and feedback, and linking the data collected from the questionnaire and the test for analysis purpose. This program went through strenuous testing process which included 24*7 technical service to answer the students’ questions with regard to the functionality of the software on Macs, Windows, smartphones and various versions of different browsers. It took daily technical support of over four months to add and test all the required features of the tool and to pilot it.

During the process of piloting the tool, the focus was on using high resolution images, ensuring clarity of the text, plausibility of distractors used, layout of the tool with an emphasis on whether questions are adequately presentable electronically, and the layout of the feedback given to the students and gained from the students. On the technical front, the compatibility of the tool with multiple machines and browsers was ensured. Moreover, the capacity to allow 10s and 100s of users to take the test at the same time, and images to be loaded appropriately, and especially on the feedback page, was ensured. Hence, the designed tool had the following:

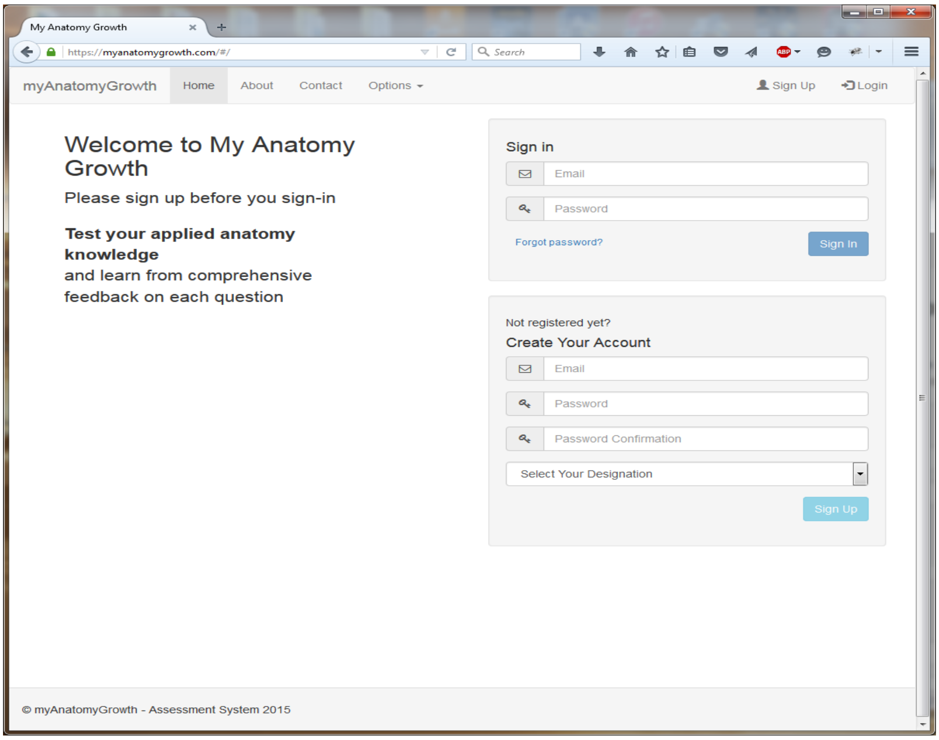
- compatibility with different browsers on windows, mac and smartphones
- clear and customised look
- 7 days 24 hours’ service to solve any technical errors

- efficiently secured system and data registration and log-in through unique medical school email addresses (not allowing Gmail, Yahoo, Hotmail, etc. access)
- authentication; i.e. password assigning through each unique email address
- incorporation of the participant information sheet, consent form and the questionnaire before the test
- asking permission to confirm the start of the test
- restricting students to take the test only once
- allocating 1 hour 30 minutes to complete the test
- providing immediate results and elaborate feedback on each question at the end of the test
- providing controlled access to go through the detailed feedback
- randomised question order for each user to avoid conditions like fatigue, boredom and lack of interest in the topic to have a significant effect on any one category of the question-design.

The layout was kept visually appealing and simple, (without accessory material to avoid cognitive overload), as suggested by Mayer (2009). Moreover, a comment box was added on the feedback page and students were encouraged to comment on their experience of taking the test questions with and without anatomical and radiology images.

Please find attached screenshots below.

Figure 4: Screenshots of "My Anatomy Growth" for authentication and compatibility with various devices



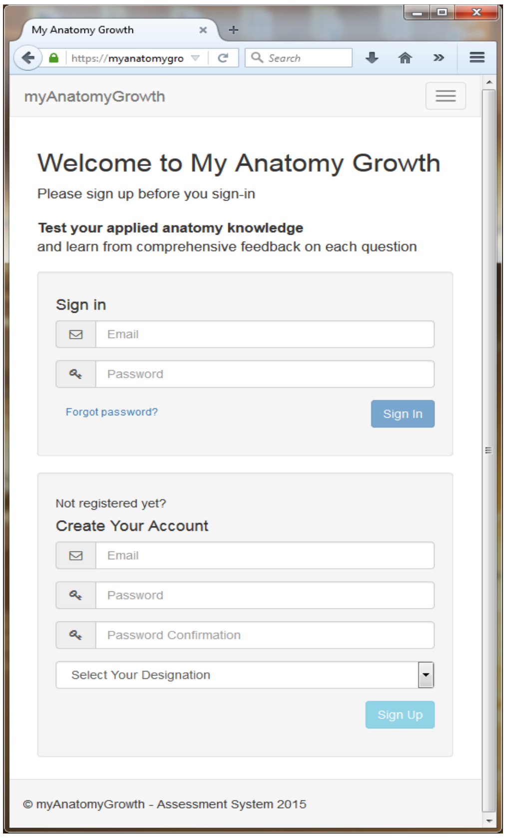
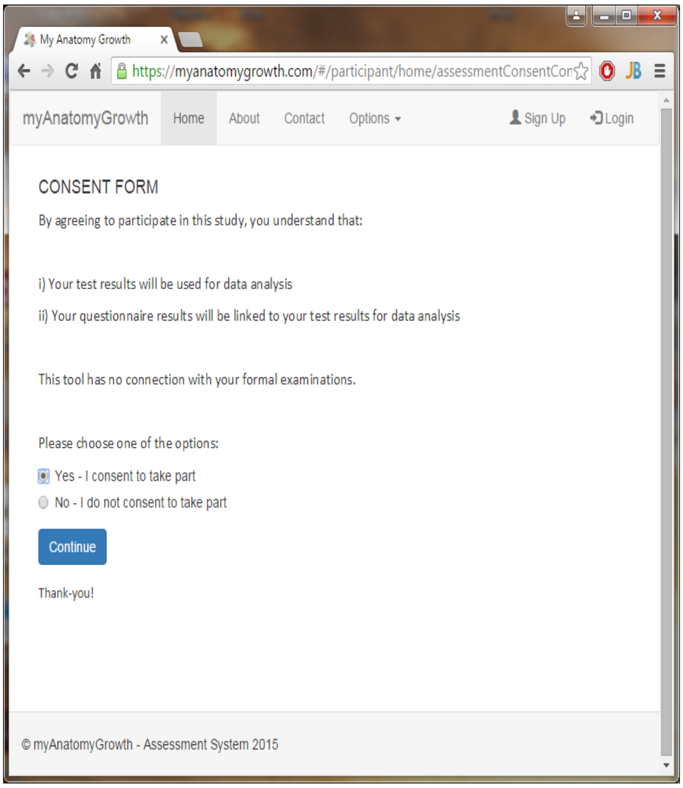


Figure 5: Screenshots of "My Anatomy Growth" for consent form and questionnaire



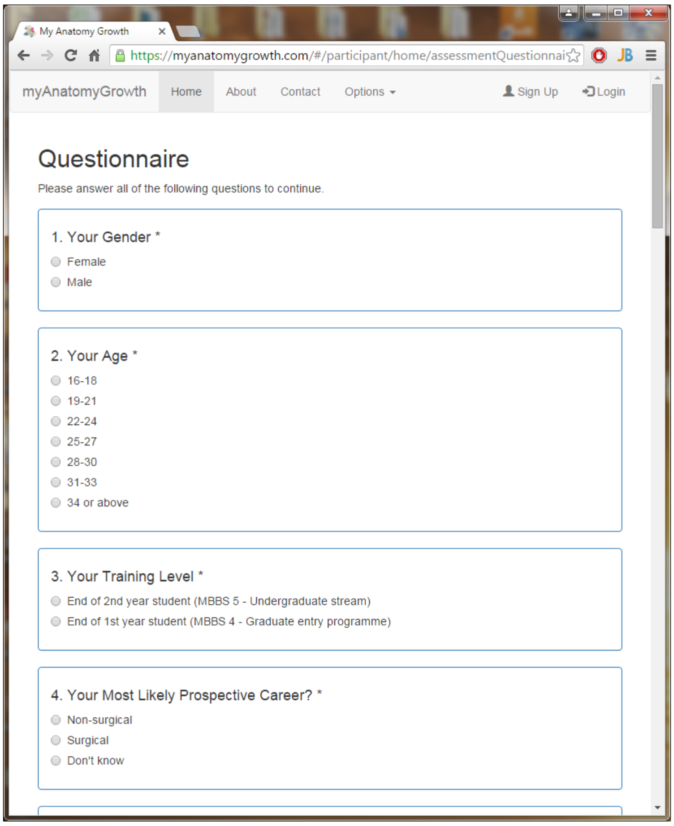
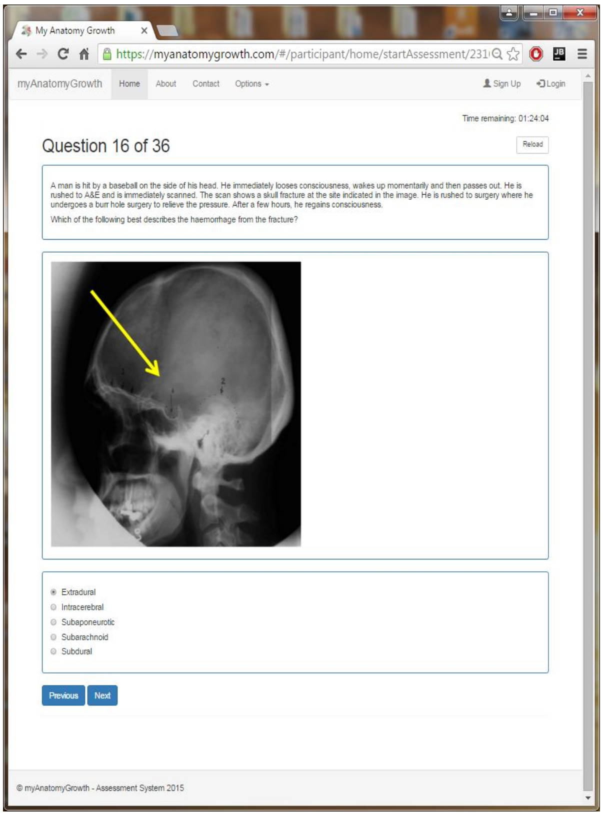


Figure 6: Screenshots of “My Anatomy Growth” for the test questions



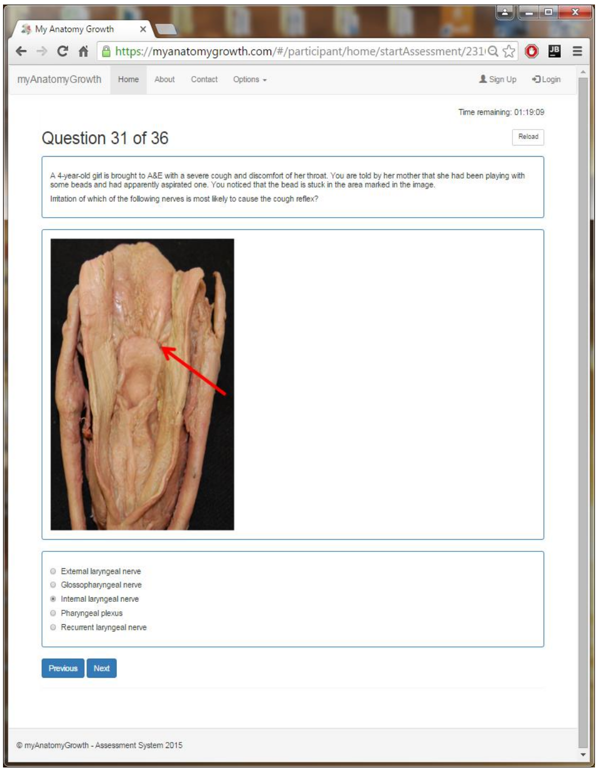
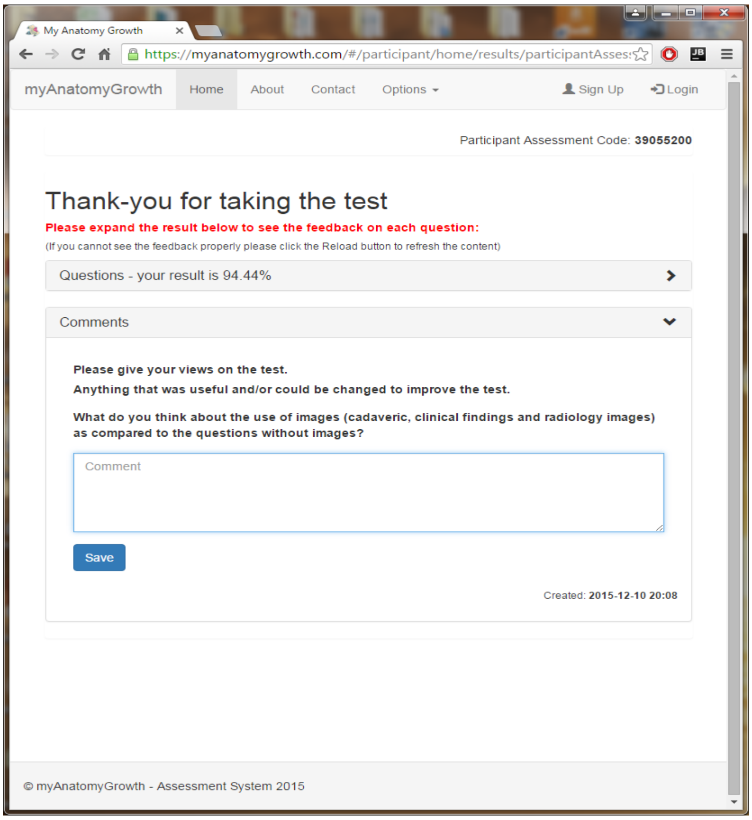
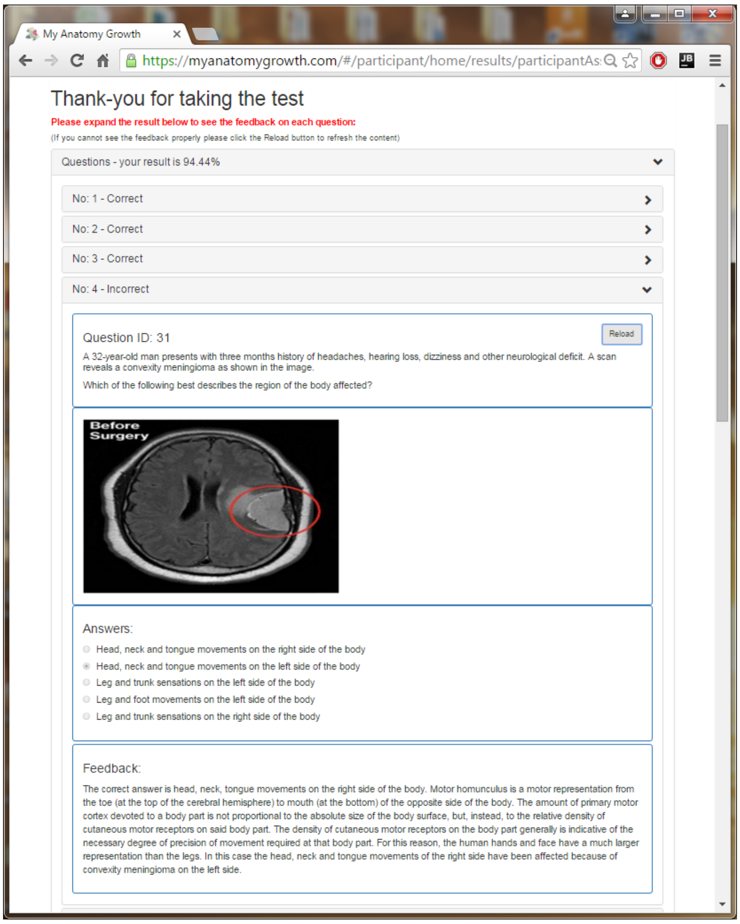


Figure 7: Screenshots of "My Anatomy Growth" for results and feedback





Internal validity of the participants

To maximise internal validity of the group, the same students acted as their own controls and tests. This design is not a regular randomised controlled trial where an intervention discriminates the control group from the test group. Here I categorised them into controls based on their performance on questions without images, test 1 based on their performance on questions with anatomical images, and test 2 based on their performance on questions with radiological images.

This was done to address any bias caused by the group (Campbell and Stanley, 1963). The components described in the literature are as follows:

- History – this refers to changes in learners' environment other than those forming a direct part of the enquiry.
- Testing – this refers to the changes occurring as a result of practice and experience.
- Instrumentation – this refers to the change in measurements between the tests.
- Regression – this depends on atypical experimentation groups.
- Mortality – this refers to participants dropping out of the study.
- Maturation – this refers to students' growth, change and development.
- Selection – this refers to initial differences between groups prior to involvement in the enquiry.
- Selection by maturation interaction – this refers to tendency of groups to grow apart.
- Ambiguity on causal direction – this refers to the actual correlation; does A cause B or B cause A.
- Diffusion of treatments - when a control group inadvertently receives aspects of a treatment intended for the test group.

- Compensatory equalization of treatments - if one group receives special treatment, it could lead to pressures from organisational sources.
- Compensatory rivalry – this refers to a change in an organisation as a competition measure. (Robson, 2011)

Although the study analysed the performances of the same group of students, the threat posed by "history" and "testing" variables cannot be eliminated. The threats posed by "instrumentation", "mortality", "diffusion of treatments", "compensatory equalization of treatments" and "compensatory rivalry" can be completely eliminated because the study was based on a single quasi-experimental design. With regard to "Regression", it was believed that participants were a mixture of students with various levels of competence. With regard to "maturation" aspect, there could have been differences in individual's development depending on their experience and practice. For the "selection" aspect, the medical schools were selected on the basis of visual resources used in their formal anatomy classes. This is investigated through each medical school anatomy webpage and relevant contacts.

Background of six medical schools

The following gives brief background of the six schools included in the study. In these schools, anatomy is only explicitly taught and tested in preclinical years (years 1 and 2).

School S –The anatomy practical sessions are organized with the use of prosected specimens/cadavers, skeletons, plastic model, medical images and surface anatomy. At the end of each semester, students are tested for their knowledge through synoptic Objective Structural Practical Examinations (OSPEs), Single Best Answers questions (SBAs) and some component in Objective Structural Clinical

Examinations (OSCEs). The synoptic element of an examination assesses students' knowledge of the topics covered in the current semester, as well as the areas covered in the previous semesters.

School K –The anatomy practical sessions are organized with the use of prosected specimens/cadavers, skeletons, plastic models, medical images, in-course dissection classes and surface anatomy. These preclinical years are assessed for the knowledge through mid-sessional assessment, final written examination and some component in Objective Structured Clinical Examinations (OSCEs) summatively. Written examinations are Single Best Answer (SBAs) based which are topographical questions (mainly identification or functional) with the use of mainly cartoon images or line diagrams. At the end of their pre-clinical year, students are assessed through three SBA papers and an OSCE.

School P – The school teaches anatomy in year 1 and 2 through radiological images, plastic models and surface anatomy, but no-cadaveric specimens are used. There are no particular examinations for testing anatomy knowledge at the end of the year. However, four progress tests are set at the standard of an FI at the end of each year, and these tests assess students' knowledge of various disciplines through progress tests.

School H – The school teaches anatomy in year 1 and 2 through prosections, plastic models, radiological images and surface anatomy, and test through single-best-answer type questions. It is a Problem Based Learning curriculum. The knowledge of anatomy is tested through anatomy spotter examinations and written examinations (combination of essay questions, multiple choice questions and extending matching questions).

School B – The school teaches anatomy in year 1 and 2 through prosections, plastic models, radiological images and surface anatomy. Anatomy knowledge is examined through knowledge tests (through multiple choice questions, extended matching questions or short answers), module tutorial tests, student selected component and anatomy viva.

School E – The school teaches anatomy in year 1 and 2 through lectures and supervised laboratory sessions in which the teaching casts and skeletal collections are used. Moreover, plastic models, radiological images and surface anatomy are used, but no-cadaveric specimens are used. There are no particular examinations for testing anatomy knowledge at the end of the year. However, four progress tests are set at the standard of an FI at the end of each year, and these tests assess students' knowledge of various disciplines through progress tests.

Ethical process

For the study, I followed the British Educational Research Association (BERA) ethical guidelines, including the criteria of informed consent, confidentiality and anonymity, their voluntary participation, their right to withdraw from the study, and any associated risks and benefits (BERA, 2011). Efforts are made to keep the schools anonymised providing it is not detrimental to the study.

I identified the following potential ethical issues and I worked determinedly for around four months to get the ethical approval:

Firstly, to obtain the access, gatekeepers and key people (Heads of anatomy departments, curriculum leads and/or educational officers) were identified and approached through email, telephone or in person depending on their availability. The purpose, aims and methods of data collection of the study were explained to them. Along with it, a brief proposal of my research work, a participant information sheet, a consent form and the test were sent to them for their consideration.

Then their permissions were gained through authorization letters from the schools after assuring them that the project will not bring any adverse effects on their students and/or school. After obtaining their formal permission, a day was planned to access their students and run the research tools. The students’ access to the online test remained open for a few days after the planned day of the test so they could go through the feedback in the time provided, and in some cases, take the test in the time convenient to them.

Secondly, to ensure students were not coerced into participation, an introductory email and an advertisement flyer were sent to the students. On the day of the test, the students were asked to register and login through their unique medical schools’ email addresses. This was followed by providing them access to "participant information sheet (PIS)" and the "consent form" before letting them answer the questionnaire and the test.

Thirdly, it was envisaged that the format of test-questions could be misleading to the students because of the timing of its release, i.e. the tool was released a month or two before their final examination, and they could perceive the standard of their final anatomy examination to be the same as this test. To address the issue, a clear indication was provided through the invitation email, flyer and participant information sheet that this test has no connection with their formal examinations, and their responses on the tool will have no effect on their future training and examinations.

Fourthly, although this is not directly related to the study, as I used participants' study time to conduct my research, I arranged to run my tools at the most convenient times for the students. Moreover, the test was offered both as a research project and as a revision tool to enable candidates to 'see how they are doing'. The correct answers and elaborate feedback was made available to the individual candidates immediately after the test to aid their revision and to pay gratitude for their participation.

Fifthly, dissemination of the findings was done carefully. Following BERA guidelines, possible attempts were made to secure the identity of individual medical schools and students' identities during analysis, reporting and dissemination. Moreover, the contact details of students and the research data were stored safely in compliance with the legal requirements of the Data Protection Act 1998. The data protection registration was done with reference no. Z6364106/2015/03/164, Section 19, medical research.

Journey of acquiring ethical approval

There were a number of hurdles in the process of acquiring ethical approval and permissions. Until early 2015, I had plans to involve medical graduates as participants, and I approached Royal College of Surgeons, the Association of Surgeons in training (AsiT), PasTest and local education and training boards (LETBs) but my initial plan did not work.

Realising the practicalities and limitations, I decided to employ a different group of participants around late January 2015 – namely, end of year 2 medical students.

Furthermore, as the test incorporated a few cadaveric images of the specimens dissected by me, the permission to use these in the test was gained from the Designated Individual at the school on March 19, 2015 in the form of an authorisation letter (attached).

Figure 8: Authorisation letter for using cadaveric images

19.3.2015

Dear Mandeep,

Re: images of anatomical specimens

I can confirm that the following anatomical specimens held at St George's, University of London all have consent for images to be taken, and that you are using these images with authorisation:

Cadaver no.	Specimen no.	Details
25	111	sagittal head and neck (right)
146	284	sagittal head and neck (right)
138	34	sagittal male pelvis (left)
59		neck
111		pharynx/larynx

With kind regards,

[redacted]

Completing and submitting the ethical approval application followed this, and the research project was ethically approved by two academics at UCL, Institute of Education on February 20, 2015 and March 24, 2015.

Obtaining permission from the gatekeepers/heads/leads was a challenging process. Some leads/heads were very supportive; whereas it was quite challenging with the others. A few medical schools never responded to my request and in some cases, it was too difficult to find the right contact. In one of the cases, I lost the opportunity to conduct the test because of lack of harmony between that school's ethics/educational committee, administrative team and the anatomy team. This delayed the whole process so much that despite having the permission acquired from their ethics committee almost a month before, unfortunately I lost the opportunity to test their 2nd year medical students. However, most staff members were very encouraging and helpful and I would not have been able to collect the data without their help.

The Joint Research and Enterprise Office of School S approved the project on behalf of the Committee on March 30, 2015. The Head of School B kindly facilitated in advertising the tool and the approval was granted on April 9 2015. School K granted the permission on April 14, 2015. The Chair of School H Ethics Committee granted the permission on April 17 2015. The Head of department of School P kindly helped to obtain agreement from the Vice-Dean of Medical Education, the ethics committee and the assessment leads, thus the permission was granted on April 20 2015. The approval from the lead of School E was granted on June 25 2015. It took almost four months to submit application, defend my case and acquire formal permissions from six medical schools.

Design of the questions

Response format and stimulus

A number of principles of an assessment have been laid out in the literature (Black and Wiliam, 1998a; Schuwirth and Van der Vleuten, 2011); however, the principles of "response format" and "stimulus" are explained below because it resonates with the design of the test.

According to Schuwirth and Van der Vleuten (2004) - Response format determines what students need to do, and indicates how their responses are captured. For assessing theoretical knowledge, assessment methods are commonly categorised into multiple-choice question types, extending matching questions, single-best-answer questions, essay questions, direct observations and free-response (open-ended) questions (Baartman et al., 2006). In this study, an online single best answer (SBAs) assessment is used to capture participants' responses. The reason for using SBAs is because multiple-choice examinations remain the primary method of assessing students' knowledge in anatomical education (Severo and Tavares, 2010), including medical education (Royal et al., 2014 and Meyer, 2016). Moreover, SBAs and multiple-choice questions (MCQs) are known to have better reliability in regard to sampling and objectivity, as compared to short or long answers questions, and these could be designed to have similar validity to free response questions in anatomy (Shaibah and Van der Vleuten, 2013). It has been known that well-constructed SBAs are better than modified essay questions (MEQs) in testing higher cognitive skills (Palmer and Devitt, 2007). Furthermore, these are feasible to conduct, easy to mark, and are capable of withstanding intellectual and statistical scrutiny. However, on the other hand these are susceptible to cueing effect. This effect is caused by poor design of the question that leads to confusion; i.e. use of implausible, heterogeneous distractors that most examinees can see are obviously wrong, increasing the odds of examinees' guessing the right answer, and thus making the questions technically flawed (Case and Swanson, 2002).

A stimulus is defined as a task that is presented to learners to trigger specific thought processes, and in this study it refers to images or no images in clinically-oriented anatomy questions. Based on my experience and literature review, I believe that the students will have seen images similar to those used in the study, but it is quite possible that they may not have been exposed to exactly the same images. Familiar images from teaching may be reassuring (Crisp and Sweiry, 2006) or they may promote positive cueing effect (Vorstenbosch et al. 2013; 2014). This cueing effect is not only limited to images but can occur in text as well (Schuwirth et al., 1996), and there is no current guidance regarding this in MCQ vignettes (Case and Swanson, 2002; Wood et al., 2004). Stimulus is a paramount in determining the type of competence being tested (Schuwirth and Van der Vleuten, 2004; Baartman et al., 2006). The question and its contents are a stimulus and indicate what the students need to know, reflecting content and validity.

Along with visuals in applied anatomy questions, 2nd year medical students' experience, views and preferences were also considered in this study investigating the effect of absence or presence of images on their performance in clinically-oriented anatomy questions.

Validity of the test-questions

Validity refers to investigating whether an assessment is measuring the competencies it is designed to examine or not (Messick, 1994). A test's validity is dependent on a number of questions; what level of students are being assessed, is the examination making grading or licensure decisions, is it for assessing low or high cognitive skills, and is it assessing narrower or broader domain (Swanson et al., 2006).

Five categories of validity are documented in the literature. These are face validity, content/direct validity, construct/indirect, concurrent validity, predictive/criterion validity and consequential validity. Face validity is the extent to which a test is compatible with its curriculum's educational philosophy (in real world situations) and makes sense to an expert in the field. As a realization of face validity, authentic and clinical images were used in the test along with contextually rich scenarios to test their performance in theoretically simulated clinical scenarios. Content/direct validity is the extent to which a test measures all the intended contents, i.e. whether a test measures all aspects of its domain. In anatomy, it refers to assessing theoretical, clinical and visual-spatial domains. For this, the test had a multifaceted design, i.e. these were integrated with valid images as well as with questions designed to test integrated knowledge of topographical and applied anatomy, basic radiology, neuroanatomy and clinical/surgical features.

Construct/indirect validity refers to an assessment supporting a sensible underpinning construct/(s) and the extent to which a test discriminates between various levels of expertise. Cronbach and Meehl (1995) suggested that construct validity is highly associated with the level of competence of people being assessed, i.e. a test that works for 1st year students may not work as efficiently for 2nd year students and so on. Secondly, it demonstrates that authenticity is not the same as validity. A resource may be authentic for making mental model but may lack validity (at the time and as the competence increases). Keeping this in mind, the test was designed to suit the level of competence of year 2 students. This was complicated because the guidelines proposed by the Anatomical Societies (McHanwell et al., 2007; Leonard et al., 1996) do not state the objectives that a 2nd year student should achieve instead, these confirm the level of anatomical knowledge that a medical graduate ought to have. Considering anatomy is only explicitly taught in year 1 and 2, it is expected that at the end of year 2, all students should have basic understanding of all documented learning objectives to build their knowledge in clinical years. Moreover, as images play an important role in medicine and anatomy, the use of various resources were used to check their understanding of both anatomical and radiology images. Besides, in the analysis section, further investigation is done to see if there are any significant differences in high and low performing students based on their performance on the whole test.

Concurrent validity is the extent to which a test correlates with existing benchmarks of that domain. However, in anatomy the gold standard measures are not clear. The Anatomical Society of Great Britain and Ireland has suggested the core syllabus in 2007 (McHanwell et al., 2007 and Leonard et al., 1996; Smith et al., 2016). It has been highlighted, however, that this needs further refinement and validation by Royal Colleges to define the level of anatomy competence of medical graduates (Standring and Larvin, 2011). Moreover, as stated above, it is not clear how students are expected to progress from start to finish of their degree. Namely, do the students need to know the overview of all the learning objectives in their early years of degree or they need to know some learning objectives in more detail than others in their preclinical and clinical years?

Predictive/criterion validity refers to the question whether the students' performance in a particular examination predicts their future performances in simulated and real situations. This requires surplus evidence of critical observations over a number of years to accumulate enough evidence to validate it. It is out of the scope of the current study. Consequential validity is considered as fundamental to the educational impact; the impact that the test has on the learners and examination-writers in preparing for an examination. This is in line with "modern assessment theory" which emphasises the importance of 360-degree feedback on all assessments, encourage reflection for deeper learning and the importance of surplus evidence for its validation (Downing, 2003; Van der Vleuten and Schuwirth, 2005; Ahmed and Pollitt, 2007). It is not possible to justify consequential validity in this study but an initial step was taken to encourage it by giving and collecting feedback to and from the participants.

Hence for validation, the test was blueprinted as recommended by the Anatomical Society of Great Britain and Ireland and the General Medical Council's "*Tomorrow's Doctors"* (McHanwell et al., 2007; GMC, 2009; Biggs and Tang, 2011; Louw et al., 2009). The domain of anatomy and the related clinical problems for designing the questions was identified. Subsequently, the basic rules for designing single best items were followed (Haladyna and Rodriguez, 2014). The anatomy test was written with the help of literature available and it was reviewed by a group of academics (Angoff, 1971). See the example on the next page.

Figure 9: Example of blueprinting method

Eur J Anat, 11 (Supplement 1): 3-18 (2007)

A core syllabus in anatomy for medical students - Adding common sense to need to know

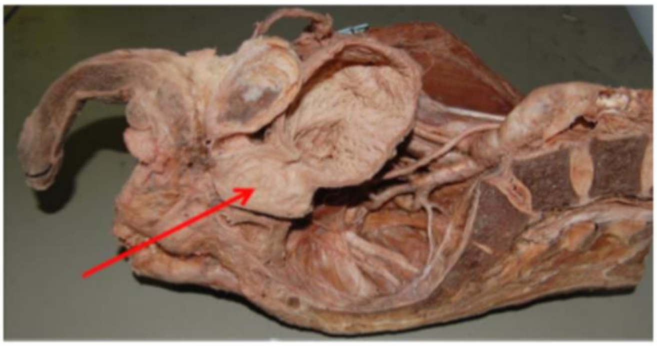
The Education Committee of the Anatomical Society of Great Britain and Ireland

Prof. S. McHanwell[1] BSc PhD
Dr. D.C. Davies[3] BSc PhD
Prof. J. Morris[5] BSc MBChB MD MA FMed Sci
Prof. I. Parkin[7] MBChB
Dr. S. Whiten[9] MA PhD
Prof. M. Atkinson[2] BSc PhD
Dr. R. Dyball[4] ScD
Prof. C. Ockleford[6] FRCPath, DSc
Prof. S. Standring[8] PhD, DSc
Dr. J. Wilton[10] BSc, MA, PhD

10. Describe the anatomy of the prostate gland, seminal vesicles and their anatomical relations. Describe the normal form of the prostate when examined per rectum and changes in relation to hypertrophy and malignancy.

32. A 69-year-old male is diagnosed with metastatic cancer. Thereafter a secondary malignant brain tumour is found on CT. However, despite best efforts the patient subsequently dies. An autopsy reveals tumour sites in the area indicated, the vertebral column and brain but no other organs.

Which of the following structures caused the cancerous cells to reach the brain?



- Anterior spinal artery
- Azygos venous system
- Thoracic duct
- Vertebral artery
- **Vertebral venous plexus**

Furthermore, for quality control and direct validation procedures (Sood and Singh, 2012), blueprinting was done as a careful analysis of the distribution of course topics (learning outcomes) within the test. Standard setting is an important criterion for establishing a standardised quality assessment. It is used to establish a threshold level of performance required to judge trainees' competence. The standard setting methods are categorised as test based (Angoff's, Ebel's method), trainee or performance based (borderline and contrasting group methods), and combined and hybrid based methods (Hofstee's method) (Case and Swanson, 2002). Angoff (1971) requires the judges to estimate the proportion of borderline candidates who were likely to respond to each question correctly. For each question, an average of the judges' estimates is calculated, and this is used to decide a cut off score for easy and difficult questions. Modified Ebel's matrix is the percentage of questions a borderline student would answer correctly. These percentages are multiplied by the relative proportion of the total questions that are assigned to each category. The results for each category are summed and to arrive at a final cut-off score (Ebel, 1983). Hosftee method (De Gruijter, 1985) involves asking judges what the maximum and minimum cut-off score and fail marks should be. In this study, the questions were put through an Angoff process. This was done by seven anatomy demonstrators (currently working as surgical and radiological registrars) and one highly experienced surgeon and anatomist. Prior to the review, the group was informed about the purpose, aims and methods of data collection of the project. This group was chosen because they were engaged in delivering anatomy knowledge to students. Moreover, as recently graduated, they were believed to be mindful of the level of anatomy knowledge required in the field, and how its delivery and assessment would benefit the students at undergraduate level. Professor Harold Ellis (Emeritus Professor of Surgery in King's College London and Royal College of Surgeons), also reviewed all the questions to help improve their quality.

Reliability of the test-questions

Reliability is a measure of whether a test is likely to yield the same results if administered to the same group of students multiple times. Another indication of reliability is that the test items should behave the same way with different population of students. This is a measure of appropriate 'sampling', 'standardisation', 'objectivity' and 'reproducibility' of an examination (Baartman et al., 2006; Sood and Singh, 2012). In the study, sampling was confirmed through blueprinting. Blueprinting is a method of writing the test-questions constructively aligned to the learning objectives suggested by national or international societies of experts. In this case, the learning objectives suggested by the Anatomical Society of Great Britain and Ireland were used (McHanwell et al., 2007). Standardization was ensured by exposing all the students to the same online environment during the test. Objectivity was achieved by electronic objective marking and by allocating equal time to all the students to complete the test. To achieve reproducibility, it requires evidence over the number of years to rule out flaws that could be raised by a number of matters, i.e. development of a subject, effect of evolution of other disciplines; development in learners' knowledge, development of educational goals, and the typical or atypical nature of the group being assessed.

Furthermore, this study followed classical test theory, which states that the observed score is a combination of the true score and an error score. The true score is the hypothetical score a student would obtain based on their competence. However, as every test induces measurement errors, the observed score may not necessarily be the same as the true score. This involves parameters such as question-difficulty, discrimination measures and taking measures to ensure the distractors are plausible (WFME, 2007; Engelhardt, 2009). In this multi-institutional study, the difficulty of the questions was decided by the Angoff method. For discrimination factor, the analysis was carried out by grouping students into high and low performing students depending on their scores on all thirty-six questions. Students who achieved 11-22 were regarded as low performers as compared to the ones who achieved between 23-34 (high performers). The plausibility of distractors was reviewed by a group of reviewers, and a group of students during the piloting phase and changes made accordingly. Hence for this study, Cronbach's alpha and item difficulty depending on high-low performing students were calculated.

For reliability, along with objectivity and standardisation, the questions were carefully sampled across the objectives.

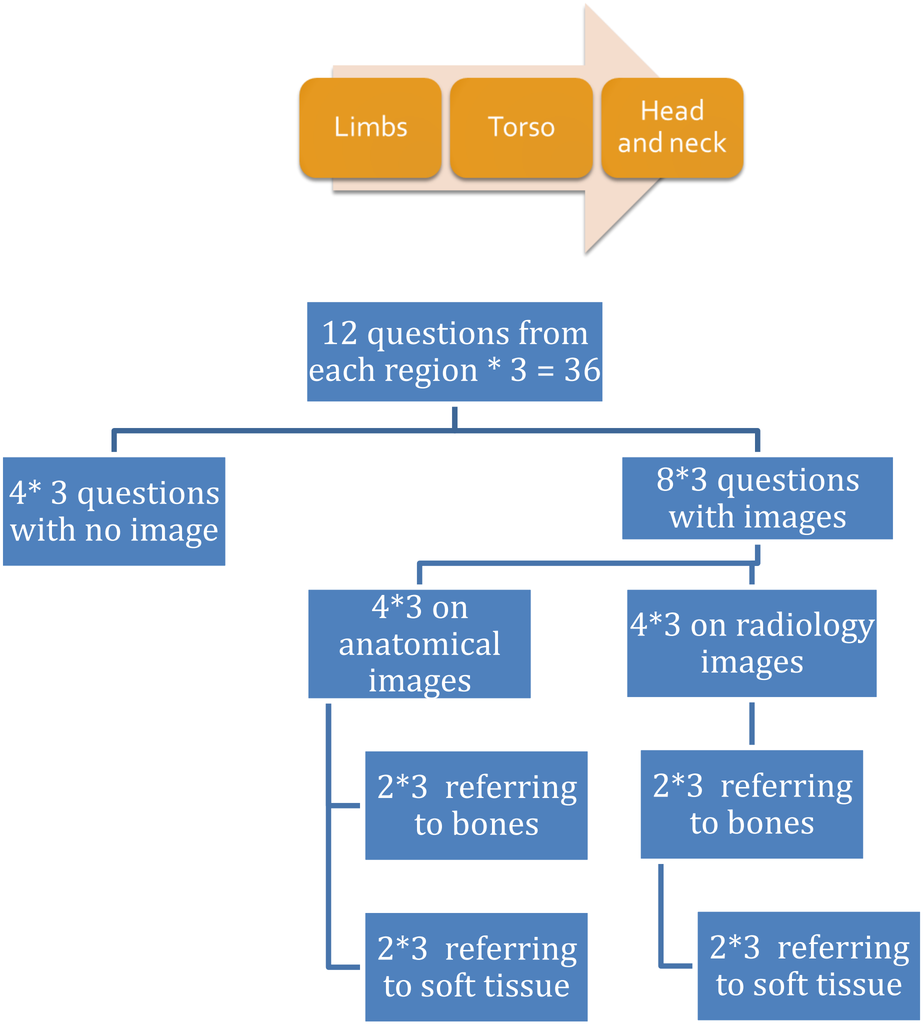
Following this procedure, a total of thirty-six questions were thematically organized with an equal distribution of questions across the following anatomical regions covered in pre-clinical years (year 1 and 2) based on the following anatomical regions:

1. Limbs (lower and upper limbs)
2. Head & neck and brain & spine
3. Torso (thorax, abdomen and pelvis)

Twelve questions were designed for each of the above anatomical regions. This was so as to have four questions in the following categories for each region:

- four questions with no resources
- four questions with images-anatomical images
    - two for identifying soft tissue
    - two for identifying bones
- four questions with images-radiology images
    - two for identifying soft tissue
    - two for identifying bones

Figure 10: Questions distribution



The reason for investigating their performance on questions with images indicating soft tissue and bones is that anatomical and radiology images are not homogenous images, i.e. bones appear different to soft tissue in these images. Especially in radiological image modalities, as X-rays or sounds waves become absorbed and/or reflect back differently from bones and soft tissue and they appear brighter or darker depending upon the density of the structure.

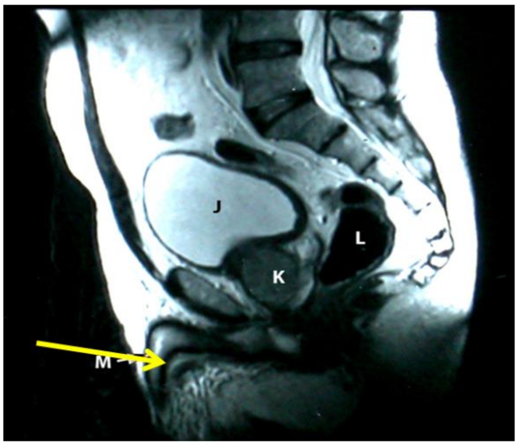
For methodological continuity, the same tool was used in all six medical schools, and its design was based on empirical research in the field and learning objectives suggested by the Anatomical Society of Great Britain and Ireland (McHanwell et al., 2007). All questions were equally weighted, and were worth one point for each correct answer.

Examples of scenario questions with and without an image are shown below. However, image and non-image versions of the same question were not used in the study. In the test, no textual material was repeated because all questions were taken by all the participants.

Example Questions

Question 1 with an image (figure 11)

Figure 11: MRI of male pelvis



Scenario: A 40-year-old man is brought to A&E with lower pelvic trauma following a road accident. An MRI is requested and the structure arrowed on the MRI is damaged.

Leading question: Which of the following best describes the site at which fluid (blood and urine) is most likely to accumulate?

Options:

- *Deep perineal pouch*
- *Ischioanal fossa*
- *Pararectal fossa*

- *Rectovesical pouch*
- *Superficial perineal pouch* *****

To answer the question above with an image (MRI), students are required to understand the scenario and identify the structure on the MRI. The structure marked on the MRI is different from the site asked in the question. For them to make sense of where pooling of blood will happen on damage of the structure, firstly they will be required to identify the structure (spongy urethra). Using the information provided, they are required to work out which pouch/fossa is the closest to the marked structure and is limited by Colle's fascia, Scarpa's fascia, dartos fascia, perineal membrane and fascia lata to allow the fluid accumulation.

Example of the same question with no image (None of the textual material was repeated in the actual test because same group of students answered the questions with and without images):

Scenario: A 40-year-old man is brought to A&E with lower pelvic trauma following a road accident. An MRI is requested and the spongy urethra is found to be damaged.

Leading question: Which of the following best describes the site at which fluid (blood and urine) is most likely to accumulate?

Options:

- *Deep perineal pouch*
- *Ischioanal fossa*
- *Pararectal fossa*

- *Rectovesical pouch*
- *Superficial perineal pouch* *****

In the above question without image, although they have been informed which structure is damaged (i.e. spongy urethra), it could be a completely new/unknown structure to some students who have never read about it. Here in this question, they do not have the advantage of seeing the structure and its neighbouring structures.

However, in question 1, for those who have never seen a sagittal pelvis MRI, the use of image would not make any positive difference. However, those who have seen these type of images may take advantage of visuals provided. Hence these questions with and without images require orchestration of internal and external representations to answer correctly.

In the study, along with images there was more layering of information in the design of these questions. These were all constructed with a clinical scenario; lead in question and five plausible, homogenous choices. Moreover, the difficulty of these questions was defined through the Angoff method.

The distribution of the questions, their anatomical regions, the use of images and the Angoff level of difficulty is shown in table 2.

Table 2: Distribution of the questions and their characteristics

Question no. (Please see the test in the appendix)	**Anatomical region**	**Use of image and type**	**Difficulty level**
1	Limbs - upper limb	No image	Easy
2	Limbs - upper limb	No image	Difficult
3	Limbs - lower limb	No image	Easy
4	Limbs - lower limb	No image	Difficult
5	Limbs - upper limb	Anatomical image	Easy
6	Limbs - lower limb	Anatomical image	Easy
7	Limbs - lower limb	Anatomical image	Easy
8	Limbs - upper limb	Anatomical image	Easy
9	Limbs - upper limb	Radiology image	Easy
10	Limbs - lower limb	Radiology image	Difficult
11	Limbs - lower limb	Radiology image	Easy
12	Limbs - upper limb	Radiology image	Difficult
13	Head and neck	No image	Difficult
14	Head and neck	No image	Easy
15	Head and neck	No image	Easy
16	Head and neck	No image	Difficult
17	Head and neck	Anatomical image	Difficult
18	Head and neck	Anatomical image	Difficult
19	Head and neck	Anatomical image	Easy
20	Head and neck	Anatomical image	Difficult
21	Head and neck	Radiology image	Easy
22	Head and neck	Radiology image	Easy
23	Head and neck	Radiology image	Difficult
24	Head and neck	Radiology image	Difficult
25	Torso - thorax	No image	Easy
26	Torso - thorax	No image	Difficult
27	Torso - abdomen and pelvis	No image	Easy
28	Torso - abdomen and pelvis	No image	Difficult
29	Torso - abdomen and pelvis	Anatomical image	Easy
30	Torso - thorax	Anatomical image	Easy
31	Torso - thorax	Anatomical image	Difficult
32	Torso - abdomen and pelvis	Anatomical image	Difficult
33	Torso - abdomen and pelvis	Radiology image	Difficult
34	Torso - thorax	Radiology image	Easy
35	Torso - thorax	Radiology image	Difficult
36	Torso - abdomen and pelvis	Radiology image	Difficult

Learning from verbal and pictorial information has also frequently been associated with individual representational preferences and cognitive styles (Mayer and Massa, 2003; Leite et al., 2010). Some learners prefer reading texts, others favour listening to an explanation, and some choose visual information from images. Although it is not in the scope of this study to discuss these relationships in much further detail, I intend to find through the questionnaire used in the study whether the students' scores have any relationship with their preferences/views and demographics.

Cognitive styles (learning styles) are not considered in the study because no matter what medical students' learning styles are, there are certain mandatory resources (especially radiological images and clinical signs/findings) they are required to understand and comprehend on account of resources' face validity.

A recent study by Schnotz and Baadte (2015) suggested that learners' recall is more accurate if the format of recall is the same as the learning format, which indicates surface structure influences. Hence schemata representing a more familiar perspective might be easier to activate, whereas less familiar perspectives might impose a higher cognitive load on working memory resulting in stronger interference effects. As anatomy is often taught in the first two years of medical curriculum to make deep structural relationships (curriculum's assumption), I have investigated whether the performance of students in this study on the questions (with and with various images) varies. I have further explored whether their scores and the resources used in their formal classes have any significant relationship.

Assessing superficial and deep knowledge of a subject is a large area of research, which will not be discussed here. However, the data collected was analysed separately for low and high performing students, based on their performance on the whole test. This helped me to delineate performances of these groups on questions with and without various images; in particular whether high performing students have deep knowledge not to show any significant difference in performance on questions with and without various types of images. This would help us understand whether high performing students have appropriate semantic deep structure to deal with and without any type of visual resources, and/or graphics' (images) surface and deep structures have any impact on their performance.

Design of the questionnaire

The questionnaire used in the study was of a simple design, collecting the students' characteristics, preferences and experiences. There was no intent to construct scores, scales or latent variables and thus no requirement for reliability or validity testing. It was adapted from tools suggested by Hisley et al. (2008) and Smith and Mathias (2010). Two clinicians (currently a surgical and a radiological trainee) reviewed it for clarity of the questions.

This was used to gather data on the features of the participants that may influence the test analysis; age, sex, training level, preferences, participation in voluntary anatomy programmes, and anticipated career choice. Demographics were thought to be an important factor for the study. The training level information was obtained because students for medicine degrees are selected by two methods; UKCAT (UK Clinical Aptitude Test) and GAMSAT (Graduate Medical School Admissions Test). UKCAT is an aptitude test taken by school leavers and GAMSAT by graduates (in science or non-science fields). Regarding preferences, the questions were set to investigate participants' interests in cadaveric specimens; body parts photographs and radiographs for their learning and assessment. The common voluntary anatomy programmes organised by various medical schools are summer dissection programme, anatomy demonstrator programme and revision classes. The voluntary demonstrating activity question was included in the questionnaire because in my previous study (Sagoo et al. 2016), interest or participation in demonstrating activity was found to have a significant positive effect on students' performance. Moreover, in anatomy, teaching junior students/demonstrating methods are often used in a number of medical schools, and are known to provide an effective learning environment (Evans and Cuffe, 2009).

Piloting the study

Once reviewed, the test was piloted by a group of seven 2nd year volunteer medical students at King's College. This pilot was conducted approximately two months before the final release of the online tool, and their informal views were incorporated to improve the test. The pilot group unanimously found the clinical format of the test and the use of images very useful. They all asked for more questions because they found it a useful learning resource. Moreover, they found the elaborate feedback, explaining why the given answer is correct or incorrect, provided at the end of the test as most useful since it helped them to revise related topics and improve their knowledge. Some of them found orientation of some images (both radiological and anatomical images) difficult, so, where appropriate, these were changed with better images followed by reviewers' consensus. Some students also commented on the appropriateness of the difficulty level of the questions.

Some of the comments are shown below:

> *"Thought the test was really good. Only a couple of minor grammatical improvements to be made and one regarding the median nerve question - it asks which structure is being compressed but then the answers are all nerves".*

"I did find some of the images hard to interpret particularly the MRI & CT scans. The images of cadaveric material were sometimes difficult to interpret as well, particularly 27 and 31. The image for question 40 isn't necessary once you see the available answers and again it's difficult to interpret actually what it's pointing at. So my only real criticism is just clearer images or a description of what's going on particular if it's an MRI or CT. Otherwise I thought the skeletal ones were great and the feedback was excellent".

"I thought it was great- a few technical issues with changing question and pictures taking time to load despite questions changing".

"I did find that when I went back to change answers, in the feedback they showed the previous answer I had selected so changing that would be useful".

"My only suggestion would be to try to find a way to include the images from the questions in the feedback as it would be useful to see when going back through the questions. Although, this may have been an error with mine as it did display an error message at the top of the feedback page".

A number of hours were invested to make the suggested changes before the release of the tool for the study.

Furthermore, all students who took the test were encouraged to give feedback at the end of the test on the tool, and many of them expressed their views and

preferences in detail. This feedback is used to illustrate the quantitative data in the discussion chapter. A thematic analysis of the feedback was not carried out as part of this thesis but a similar work was carried out in my Methods of Enquiry 2.
This shows the careful attention was paid to the design of the study and the online tool containing the applied anatomy test, consent form, participant information sheet and the questionnaire.

Chapter 4 – Results and Analysis

Analysis was carried out by repeated measure analysis of variance (ANOVA) because the same group of students took the test consisting of questions with and without images (anatomical images and radiological images). This was done by using SPSS statistical package, Mac Version 22.0 (IBM Corp., Armonk, NY) and Robson (2011) to investigate the impact of independent variables (visuals, and questionnaire variables) on the dependent variable (students' scores).

As the classical theory was adopted, the reliability of the online test was investigated through Cronbach's alpha. It is a value that expresses the amount of variance between students that is genuinely due to true differences between them. It shows how much of the variability in the scores is due to other sources of variance such as inconsistencies between questions-difficulty and level of the students' competence etc.

The high performing students were separated from low performing students on the basis of their performance on the whole test. This was followed by investigating their performance on questions with and without images in the above two groups (low and high performers).

The data were analysed:

1) To assess the effect of the question-design (with and without images) on the total number of correct responses in the assessment.

- in the group as a whole
- subgroups of high and low performing students
- within and in-between individuals acting as controls and tests

Scores were compared between questions with:

- images and without images
- anatomical images and radiological images
- soft tissue represented in anatomical and radiological images
- bones represented in anatomical and radiological images

2) To assess the influence of questionnaire, the following variables were used:

- sex (coded as female and male)
- range of age (coded as 16-18, 19-21, 22-24, 25-27, 28-30, 31-33, and 34 or above)
- training level (coded as end of 2nd year student – MBBS 5 undergraduate programme, and end of 1st year – MBBS 4 graduate entry programme)
- most likely prospective career choice (coded as non-surgical, surgical and don't know)
- medical school study in (six medical schools)
- resources used to teach anatomy (coded as dissection of human cadavers only, prosections only, radiology images only, all of the above, dissection of human cadavers and radiology images, prosections and radiology images)
- preferences and participation in the voluntary programmes were investigated through questions 7-21 in the questionnaire. (See appendix)

Furthermore, the following was investigated:

- whether variability in the students' scores was dependent on questions with or without images and/or other variables generated from the questionnaire study
- which elements of test items were significantly affecting the performance of the students
- what was the effect size and direction

In parametric tests based on the normal distribution, the assumption is that data points are independent. However, in this case a repeated measure design was chosen because data of the students' performance on different types of questions has come from the same group of students. This meant that data of performance on different question-types would be related. Hence I assumed that the relationship between pairs of performance on different question types may be similar (i.e. the level of dependence between pairs of groups is roughly equal). This assumption is known as the assumption of sphericity. Sphericity is met when these variances (differences between pairs of scores in all combination – variance) are roughly equal. In three treatments; if two have similar variances then the data have local circularity (or local sphericity), because two of the variances of differences are similar.

The effect of violating sphericity is a loss of power (i.e. an increased probability of a Type 2 error) and a test statistic (F-ratio) that simply cannot be compared to tabulated values of the F-distribution. Departures from sphericity can be measured in three ways:

- Greenhouse and Geisser
- Huynh and Feldt
- The lower bound estimate

The Greenhouse-Geisser and Huynh-Feldt estimates can both range from the lower bound (the most severe departure from sphericity possible given the data) and 1 (no departure from sphericity at all).

- If Mauchly's test statistic is nonsignificant (i.e. $p > .05$) then it is reasonable to conclude that the variances of differences are not significantly different (i.e. they are roughly equal).

- If Mauchly's test statistic is significant (i.e. has a probability value less than .05) it is concluded that there are significant differences between the variance of differences: the condition of sphericity has not been met. In this scenario we cannot trust the F-ratios produced by SPSS.

Fortunately, if data violate the sphericity assumption, the degrees of freedom are adjusted for the effect by multiplying it by one of the aforementioned sphericity estimates. This makes the degrees of freedom smaller; by reducing the degrees of freedom we make the F-ratio more conservative (i.e. it has to be bigger to be deemed significant). SPSS statistical package applies these adjustments automatically, as follows.

- use the Huynh-Feldt correction when $\varepsilon > .75$
- use the Greenhouse-Geisser correction when $\varepsilon <. 75$

With any significance test, the power of Mauchly's test depends on the sample size.

- In small samples, large deviations from sphericity might be deemed non- significant.
- In large samples, small deviations from sphericity might be deemed significant.

To represent Mauchly's test:

X2 (df) = approximately Chi-square, $p > .05$.

In case of less than three conditions, Mauchly box only shows a dot because at least three conditions are required for sphericity to be an issue.

These tests of within-subjects tell us if the difference is significant; however, it does not tell us the direction of the effect. To understand the direction and effect size, pairwise comparisons were carried out in this study (SPSS statistical package, Version 22.0 and Robson, 2011).

Measures of effect size in ANOVA are measures of the degree of association between the effect (e.g., a main effect, an interaction, and a linear contrast) and the dependent variable. They can be thought of as the correlation between an effect and the dependent variable. If the value of the measure of association is squared, it can be interpreted as the proportion of variance in the dependent variable that is attributable to each effect. The partial Eta squared used in this study can be defined as the ratio of variance accounted for by an effect, and that effect plus its associated error variance with an ANOVA study; i.e. SS effect / (SS effect + SS error). In the literature, 0.01 <= partial eta squared < 0.06 is considered as small effect, 0.06 = partial eta squared < 0.14 is considered as medium effect, and partial eta squared >= 0.14 is large effect.

Test results

The test had 36 questions in total. On reliability statistics, it has an acceptable reliability with Cronbach's alpha being .728. The impact on the overall table for dropping each individual question from the calculation is shown in table 3.

Table 3: Cronbach's alpha if item deleted

Item-Total Statistics	
Test Questions	Cronbach's Alpha if Item Deleted
TR1	.721
TR2	.715
TR3	.713
TR4	.728
TR5	.723
TR6	.722
TR7	.723
TR8	.721
TR9	.723
TR10	.722
TR11	.722
TR12	.728
TR13	.721
TR14	.714
TR15	.718
TR16	.724
TR17	.720
TR18	.717
TR19	.727
TR20	.725
TR21	.723
TR22	.713
TR23	.722
TR24	.720
TR25	.727
TR26	.733
TR27	.716
TR28	.726
TR29	.738
TR30	.725
TR31	.720
TR32	.715
TR33	.715
TR34	.724
TR35	.726
TR36	.729

174 students completed the tool (test and questionnaire). Out of 36 questions, 12 were without images, 12 had anatomical images and 12 had radiological images.

The means and standard deviations of each group of questions is shown in table 4.

Table 4: Means and standard deviations of each group of questions

Descriptive Statistics			
	Mean	Std. Deviation	N
answered correctly on 12 no image questions	7.34	2.202	174
answered correctly on 12 anatomical image questions	8.05	1.787	174
answered correctly on 12 radiology image questions	7.95	2.059	174

On the correlation scale (Pearson correlation), the above three categories had significant correlation (p<. 001) as shown in table 5. This means those students who performed better in one category performed better in other categories.

Table 5: Pearson correlation of each group of questions

Correlations				
		answered correctly on 12 no image questions	answered correctly on 12 anatomical image questions	answered correctly on 12 radiology image questions
answered correctly on 12 no image questions	Pearson Correlation	1	.498**	.446**
	Sig. (2-tailed)		.000	.000
	N	174	174	174
answered correctly on 12 anatomical image questions	Pearson Correlation	.498**	1	.470**
	Sig. (2-tailed)	.000		.000
	N	174	174	174
answered correctly on 12 radiology image questions	Pearson Correlation	.446**	.470**	1
	Sig. (2-tailed)	.000	.000	
	N	174	174	174
**. Correlation is significant at the 0.01 level (2-tailed).				

Eight academics reviewed the test (seven anatomy demonstrators who are currently working as surgical and radiological trainees and registrars, and one surgical and anatomical expert). To define the difficulty of each question, they were asked to Angoff the questions, i.e. what percentage of borderline students would answer each question correctly. This was followed by averaging the percentage of each question.

Considering the data, 56% was considered as a cut-off score. The questions with score 56 or below were considered "difficult" and the questions with score 56 above were considered "easy". Considering the students' performance, the group was divided into high and low performing students. The students who achieved 11-22 marks were considered as “low performing group” and the students with 23-34 were considered “high performing group”.

The first repeated measure ANOVA was run using “within-subjects’ variable” as a level of difficulty of questions as defined by the academics (“easy” and “hard”) and “between-subject factor” as “high performing group” and “low performing group”.

For 1st general linear model (analysis of question-difficulty and low-high performing students):

- Within-subjects’ factors (IVs): the level of questions-difficulty as put through an Angoff process
    - 1 = 56 and below (difficult)
    - 2 = 56 above (easy)
- Between-subjects’ factors: low and high performing students’ scores
    - 1 = students with scores between 11 and 22
    - 2 = students with scores between 23 and 34

Descriptive data below shows the means and standard deviations of low and high performing students on “difficult” and “easy” questions. See table 6.

Table 6: Means and standard deviations of low and high performing students on “difficult” and “easy” questions

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
56 and below (difficult)	Low performing (11-22)	7.78	1.718	78
	High performing (23-34)	11.71	2.303	96
	Total	9.95	2.839	174
56 above (easy)	Low performing (11-22)	11.22	2.16	78
	High performing (23-34)	15.17	1.434	96
	Total	13.4	2.662	174

Mauchly's test of sphericity does not apply here because the dependent variable (level of difficulty of questions) has only two variables. Tests of within-subjects' effects and contrasts showed that there was a significant difference in questions-difficulty (F (1, 172) = 308.88, $p < .001$), partial Eta Squared is = .642 (indicates very large effect size). There was a significant difference between high and low performing students (F (1, 172) = 320.44, $p<.001$), partial eta squared = .651 (indicates very large effect size). However, the interaction between the difficulty of the questions and low-high performing student groups was not statistically significant.

In Figure 12, 1 = 56 and below (difficult questions), 2 = 56 above (easy questions)

Figure 12: High and low performers' scores on easy and difficult questions

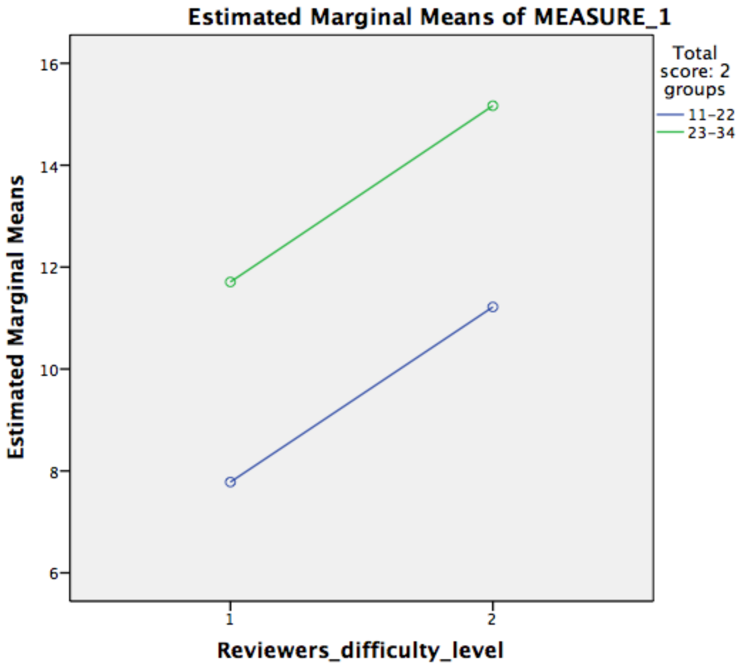


Figure 12 shows a significant difference in easy and difficult questions, and high and low performing students i.e. students performed significantly better on easy questions as compared to difficult questions; and the performance of high and low performing students was significantly different. However, there was no significant interaction between the level of difficulty of the questions and high-low performing students.

For 2nd general linear model (analysis of with and without image questions and low-high performing students):

- Within-subjects' factors (IVs):
    - question types –
        - 1 = questions with no images
        - 2 = questions with images
- Between-subjects' factors:
    - low and high performing students' scores
        - 1 = students with scores between 11 and 22
        - 2 = students with scores between 23 and 34

Within-subject factors were computed. TR refers to test question.

- Has_no_image=TR1 + TR2 + TR3 + TR4 + TR13 + TR14 + TR15 + TR16 + TR25 + TR26 + TR27 + TR28 (total 12 questions)
- Has_ image=TR5 + TR6 + TR7 + TR8 + TR9 + TR10 + TR11 + TR12 + TR17 + TR18 + TR19 + TR20 + TR21 + TR22 + TR23 + TR24 + TR29 + TR30 + TR31 + TR32 + TR33 + TR34 + TR35 + TR36 (total 24 questions). These scores were divided by two.

Descriptive data below shows the means and standard deviations of low and high performing students on two question types. See table 7.

Table 7: Means and standard deviations of low and high performing students on two question types

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
answered correctly on 12 no image questions	11-22 scores	5.81	1.588	78
	23-34 scores	8.58	1.816	96
	Total	7.34	2.202	174
answered correctly on 24 image questions (divided by two)	11-22 scores	6.59	1.093	78
	23-34 scores	9.14	1.023	96
	Total	8.00	1.650	174

Mauchly's test of sphericity does not apply here because the dependent variable (level of difficulty of questions) has only two variables. Tests of within-subjects' effects and contrasts showed that there was a significant difference in questions-types ($F (1, 172) = 21.811$, $p < .001$), partial Eta Squared is = .113 (indicates medium effect size). There was no significant interaction between the above question-types and low-high performing student groups.

In Figure 13, 1 = questions without images, 2 = questions with images

Figure 13: High and low performers' scores on questions with and without images

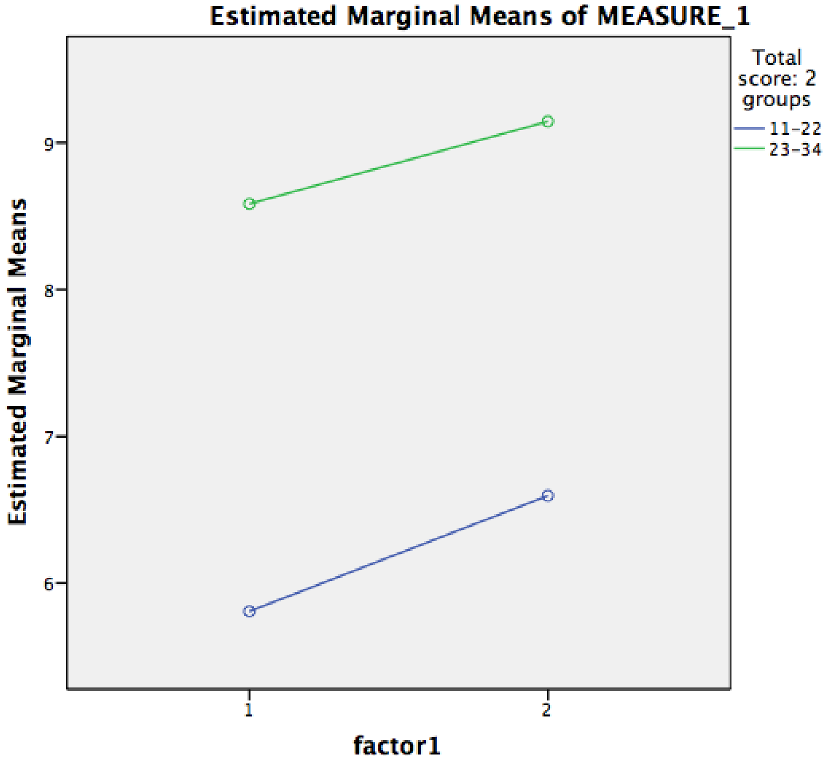


Figure 13 shows a significant difference in questions with and without images, and high and low performing students, i.e. the students' performance on questions with images was significantly better than questions without images, and the performance of high and low performing students was significantly different. However, there was no significant interaction between the above question-types and high-low performing students.

For 3rd general linear model (analysis of three question-types and low-high performing students):

- Within-subjects' factors (IVs):
    - question types –
        - 1 = questions with no images
        - 2 = questions with anatomical images
        - 3 = questions with radiological images
- Between-subjects' factors:
    - low and high performing students' scores
        - 1 = students with scores between 11 and 22
        - 2 = students with scores between 23 and 34

Within-subject factors were computed. TR refers to test question.

- Has_no_image=TR1 + TR2 + TR3 + TR4 + TR13 + TR14 + TR15 + TR16 + TR25 + TR26 + TR27 + TR28 (total 12 questions)
- Has_anatomical_image=TR5 + TR6 + TR7 + TR8 + TR17 + TR18 + TR19 + TR20 + TR29 + TR30 + TR31 + TR32 (total 12 questions)
- Has_radiology_image= TR9 + TR10 + TR11 + TR12 + TR21 + TR22 + TR23 + TR24 + TR33 + TR34 + TR35 + TR36 (total 12 questions)

Descriptive data below shows the means and standard deviations of low and high performing students on three question types. See table 8.

Table 8: Means and standard deviations of low and high performing students on three question types

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
answered correctly on 12 no image questions	11-22	5.81	1.588	78
	23-34	8.58	1.816	96
	Total	7.34	2.202	174
answered correctly on 12 anatomical image questions	11-22	6.76	1.487	78
	23-34	9.10	1.235	96
	Total	8.05	1.787	174
answered correctly on 12 radiology image questions	11-22	6.44	1.640	78
	23-34	9.19	1.453	96
	Total	7.95	2.059	174

Mauchly's test of sphericity is not significant so the assumptions are met. Tests of within-subjects' effects and contrasts showed that there was a significant difference in question-types ($F (2, 344) = 12.24$, $p < .001$), partial eta squared = .066 (indicates medium effect size). There was a significant difference between high and low performing students ($F (1, 172) = 320.44$, $p<.001$), partial eta squared = .651 (indicates very large effect size). However, the interaction between the question-types and low-high performing student groups was not statistically significant. See table 9.

Table 9: Pairwise comparisons of three groups of questions

Pairwise Comparisons						
Question types	Question types	Mean Difference	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
No image	Anatomical image	-.735*	.154	.000	-1.108	-.362
	Radiology image	-.616*	.172	.001	-1.031	-.201
Anatomical image	No image	.735*	.154	.000	.362	1.108
	Radiology image	.119	.152	1.000	-.248	.485
Radiology image	No image	.616*	.172	.001	.201	1.031
	Anatomical image	-.119	.152	1.000	-.485	.248

Table 9 shows a significant difference in performance of students on question with:

- anatomical images than no images, partial eta squared = .116 (large effect size)
- radiology images than no images, partial et squared = .070 (medium effect size)

In figure 14, 1 = questions without image, 2 = questions with anatomical images, 3 = questions with radiology images

Figure 14: High and low performers' scores on questions with and without anatomical and radiological images

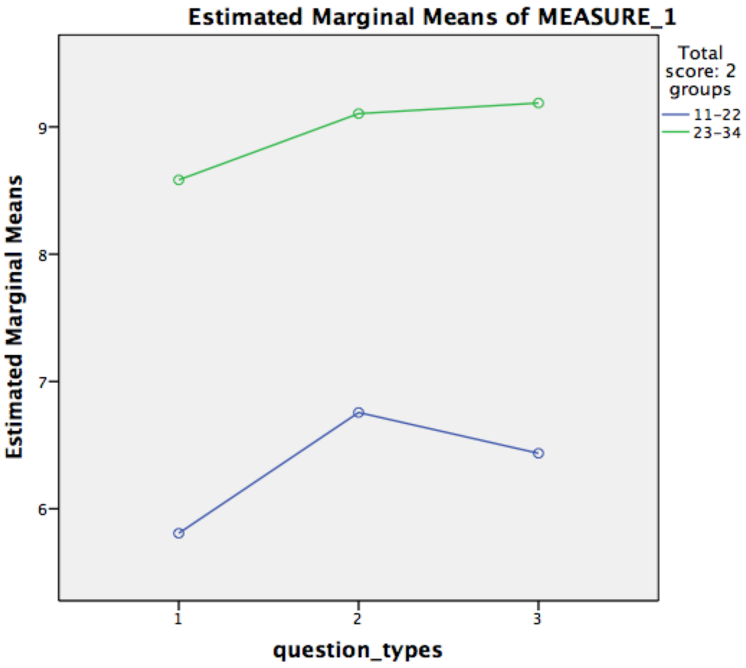


Figure 14 shows a significant difference in questions with and without anatomical and radiological images, and high and low performing students, i.e. the students' performance on questions with anatomical and radiological images was significantly better than questions without images, and the performance of high and low performing students was significantly different. However, there was no significant interaction between the above question-types and high-low performing students.

For 4th general linear model (analysis of three question-types and level of questions-difficulty):

- Within-subjects' factors (IVs):
  - question types
    - 1 = questions with no images
    - 2 = questions with anatomical images
    - 3 = questions with radiological images
- questions-difficulty
  - 1 = 56 and below (difficult)
  - 2 = 56 above (easy)

The variables were computed as shown in table 10.

Between-subjects' factors: low and high performing students' scores

- 1 = students with scores between 11 and 22
- 2 = students with scores between 23 and 34

Table 10: Within-subjects factors for question types and question difficulty

Within-Subjects Factors		
Question types	Question difficulty	Dependent Variable
1	1	No image hard
	2	No image easy
2	1	Anatomical image hard
	2	Anatomical image easy
3	1	Radiology image hard
	2	Radiology image easy

Within-subject factors were computed as follows:

- Has_no_image_hard= TR2 + TR4 + TR13 + TR16 + TR26 + TR28 (total 6 questions)
- Has_no_image_easy= TR1 + TR3 + TR14 + TR15 + TR25 + TR27 (total 6 questions)
- Has_anatomical_image_hard= TR17 + TR18 + TR20 + TR31 + TR32 (total 5 questions)
- Has_anatomical_image_easy= TR5 + TR6 + TR7 + TR8 + TR19 + TR29 + TR30 (total 7 questions)
- Has_radiology_image_hard= TR10 + TR12 + TR23 + TR24 + TR33 + TR35 + TR36 (total 7 questions)
- Has-radiology_image-easy= TR9 + TR11 + TR21 + TR22 + TR34 (total 5 questions)

Descriptive data below shows the means and standard deviations of low and high performing students on three question-types and two question difficulty. See table 11.

Table 11: Means and standard deviations of low and high performing students on three question types and two question difficulty

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
No image hard	11-22	2.62	1.154	78
	23-34	3.70	1.291	96
	Total	3.21	1.341	174
No image easy	11-22	3.19	1.239	78
	23-34	4.89	.993	96
	Total	4.13	1.392	174
Anatomical image hard	11-22	1.73	.976	78
	23-34	3.05	1.118	96
	Total	2.46	1.243	174
Anatomical image easy	11-22	5.03	1.105	78
	23-34	6.05	.639	96
	Total	5.59	1.014	174
Radiology image hard	11-22	3.44	1.076	78
	23-34	4.96	1.123	96
	Total	4.28	1.336	174
Radiology image easy	11-22	3.00	1.269	78
	23-34	4.23	.827	96
	Total	3.68	1.212	174

Mauchly's test of sphericity is not significant so the assumptions are met with two independent variables (question types and question difficulty) and their interaction. Tests of within-subjects' effects and contrasts showed that there was a significant difference in the above question-types ($F(2, 344) = 12.24$, $p < .001$), partial eta squared = .066 (indicates medium effect size), and question difficulty ($F(1, 172) = 308.88$, $p < .001$), partial eta squared = .642 (indicates a very large effect size). The interaction between the above question-types and question-difficulty was significant ($F(2, 344) = 267.99$, $p < .001$), partial eta squared = .609 (indicates very large effect size). The interactions between the above question-types, question-difficulty and high low performing student was statistically significant ($F(2, 344) = 5.18$, $p < .05$), partial eta squared = .029 (indicates small effect size).

The analysis showed a significant difference in performance of students and:

- on question types
  - with anatomical images better than no images, partial eta squared = .116 (large effect size)
  - with radiology images better than no images, partial et squared = .070 (medium effect size)
- on question difficulty
  - with easy questions better than difficult questions, partial eta squared = .642 (indicates very large effect size)
- on the interactions between question-types and question-difficulty
  - with no image easy and hard questions versus anatomical image easy and hard questions
  - with no image easy and hard questions versus radiology image easy and hard questions
- on the interactions between question types, question difficulty and high low performing student as shown in table 12:
  - in high-low performing students - with no image easy and hard questions versus anatomical image easy and hard questions
  - in high-low performing students - with no image easy and hard questions versus radiology image easy and hard questions

Table 12: Interactions between question types, question difficulty and high low performing students

Total score: 2 groups – question types * question difficulty					
Total score: 2 groups	Question-types and question-difficulty	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
11-22	No image hard	2.615	0.139	2.34	2.891
	No image easy	3.192	0.126	2.944	3.44
	Anatomical image hard	1.731	0.12	1.495	1.967
	Anatomical image easy	5.026	0.099	4.829	5.222
	Radiology image hard	3.436	0.125	3.19	3.682
	Radiology image easy	3	0.119	2.766	3.234
23-34	No image hard	3.698	0.126	3.45	3.946
	No image easy	4.885	0.113	4.662	5.109
	Anatomical image hard	3.052	0.108	2.839	3.265
	Anatomical image easy	6.052	0.09	5.875	6.229
	Radiology image hard	4.958	0.113	4.736	5.18
	Radiology image easy	4.229	0.107	4.018	4.44

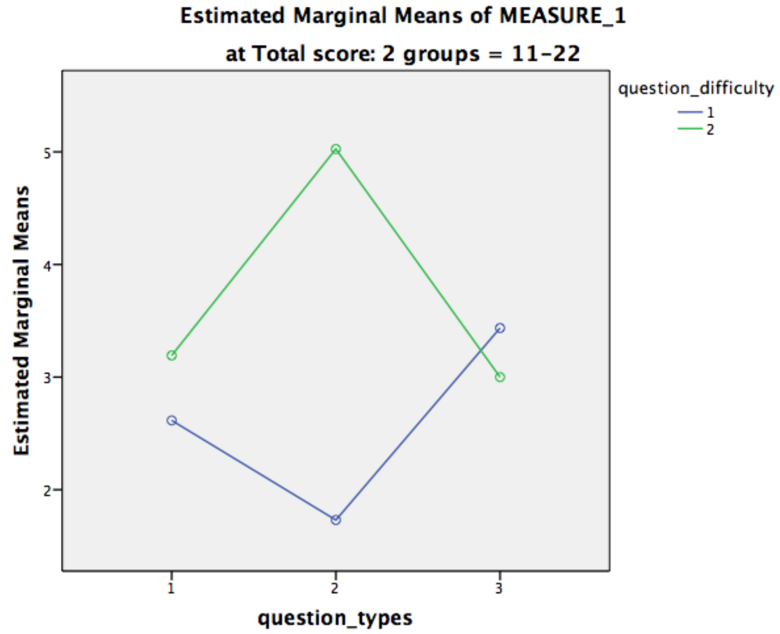
In both low and high performing students, the order of mean scores is as follows: Anatomical image easy > radiology image hard > no image easy > radiology image easy > no image hard > anatomical image hard (not significant).

In figure 15, for question-types on x-axis: 1 = questions without images, 2 = questions with anatomical images, 3 = questions with radiological images

For question-difficulty: 1 (blue line) = difficult questions, 2 (green line) = easy questions

1st graph is of low performing students and 2nd graph is of high performing students.

Figure 15: High and low performers' scores (as two separate figures) on easy and difficult questions with and without anatomical and radiological images



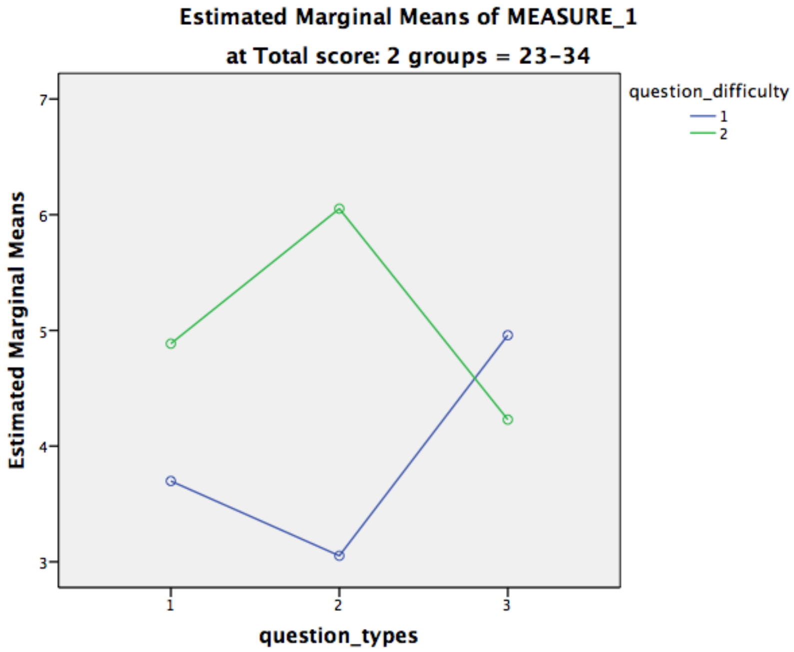


Figure 15 shows highly significant difference in questions with and without anatomical and radiological images, easy and difficult questions, and high and low performing students, i.e. the students' performance on questions with images was significantly better than in questions without images, the students' performance on easy questions was better than difficult questions, and the performance of high and low performing students was significantly different. Moreover, the interaction between the above question-types and the level of question-difficulty was highly significant. This meant that the students' performance on question-types was significantly dependent on the level of questions' difficulty. The students performed significantly better on anatomical-image easy-and-difficult questions as compared to no image easy-and-difficult questions. The students performed significantly better on radiology-image easy-and-difficult questions as compared to no image easy-and-difficult questions. Moreover, there was a significant interaction between the above question types, the level of question difficulty and high-low performing students but it was a relatively low significance.

<u>For 5th general linear model (analysis of anatomical and radiology image-questions and questions-subtypes):</u>

Within-subjects' factors (IVs):

- Image-questions (two types with images)
    - 1 = questions with anatomical image
    - 2 = questions with radiology image
- Questions-subtypes
    - 1 = questions referring to bones
    - 2 = questions referring to soft-tissue

The variables were computed as shown in table 13.

Between-subjects' factors: low and high performing student groups

- 1 = students with scores between 11 and 22 o 2 = students with scores between 23 and 34

Table 13: Within-subjects factors for image types and image subtypes

Within-Subjects Factors		
Image types	Image subtypes	Dependent Variable
1	1	Has anatomical image for bone
	2	Has anatomical image for soft tissue
2	1	Has radiology image for bone
	2	Has radiology image for soft tissue

Within-subject factors were computed as follows:

- Has_anatomical_image_for_bone=TR5 + TR6 + TR18 + TR19 + TR29 + TR30 (total 6 questions)
- Has_anatomical_image_for_softtissue=TR7 + TR8 + TR17 + TR20 + TR31 + TR32 (total 6 questions)
- Has_radiology_image_for_bone=TR9 + TR10 + TR21 + TR22 + TR33 + TR34 (total 6 questions)
- Has_radiology_image_for_softtissue=TR11 + TR12 + TR23 + TR24 + TR35 + TR36 (total 6 questions)

Descriptive data below shows the means and standard deviation of low and high performing students on image questions and questions subtypes. See table 14.

Table 14: Means and standard deviations of low and high performing students on image questions and questions subtypes

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
answered correctly on 6 anatomical image questions for bones	11-22	4.29	1.175	78
	23-34	5.16	.621	96
	Total	4.77	1.005	174
answered correctly on 6 anatomical image for soft tissue	11-22	2.46	1.113	78
	23-34	3.95	1.109	96
	Total	3.28	1.333	174
answered correctly on 6 radiology image for bones	11-22	3.78	1.213	78
	23-34	5.27	.900	96
	Total	4.60	1.285	174
answered correctly on 6 radiology image for soft tissue	11-22	2.65	1.079	78
	23-34	3.92	1.121	96
	Total	3.35	1.267	174

Mauchly's test of sphericity is met. There are only two levels of image questions and questions subtypes. Tests of within-subjects' effects and contrasts showed that there was a significant difference in the above question subtypes ($F(1, 172) = 277.31$, $p < .001$), partial eta squared = .617 (indicates very large effect size). The interaction between the image questions, the above question subtypes and low-high performing students ($F(1, 172) = 7.09$), $p < .05$), partial eta squared = .040 (indicates small effect size).

This analysis showed a significant difference in performance of students on:

- Questions referring to bone better than questions referring to soft tissue, partial eta squared = .617 (very large effect size)
- On the interaction between image-questions, questions-subtypes and low-high performing students:
    - in high-low performing students - with anatomical image questions on bones and soft tissues versus radiology image questions on bones and soft tissues (relatively low significance)

Table 15: Interactions between image types, image subtypes and high low performing student

Total score: 2 groups - image types * image subtypes					
Total score: 2 groups	Image type and image subtypes	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
11-22	Has anatomical image for bone	4.295	0.103	4.091	4.499
	Has anatomical image for soft tissue	2.462	0.126	2.213	2.71
	Has radiology image for bone	3.782	0.119	3.547	4.017
	Has radiology image for soft tissue	2.654	0.125	2.407	2.9
23-34	Has anatomical image for bone	5.156	0.093	4.973	5.34
	Has anatomical image for soft tissue	3.948	0.113	3.724	4.172
	Has radiology image for bone	5.271	0.107	5.059	5.483
	Has radiology image for soft tissue	3.917	0.113	3.695	4.139

In low performing students, the order of mean scores is as follows:

Anatomical image bone > radiology image bone > radiology image soft tissue > anatomical image soft tissue.

In high performing students, the order of mean scores is as follows:

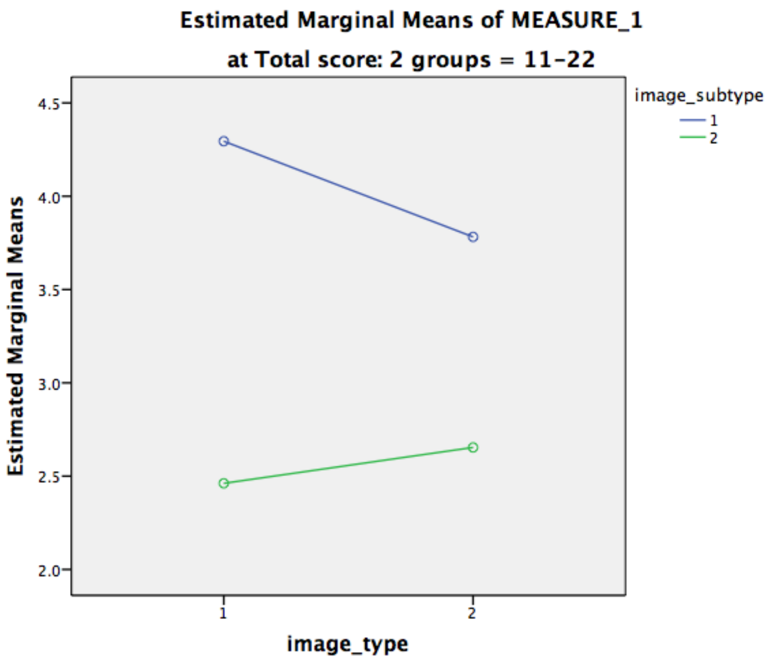
Radiology image bone > anatomical image bone > anatomical image soft tissue > radiology image soft tissue.

In figure 16, for image-types on x-axis: 1 = questions with anatomical images, 2 = questions with radiology images

For image-subtypes: 1 = questions referring to bones, 2 = questions referring to soft tissues

1st graph is of low performing students and 2nd graph is of high performing students.

Figure 16: High and low performers' scores (as two separate figures) on questions with anatomical and radiological images and their subtypes



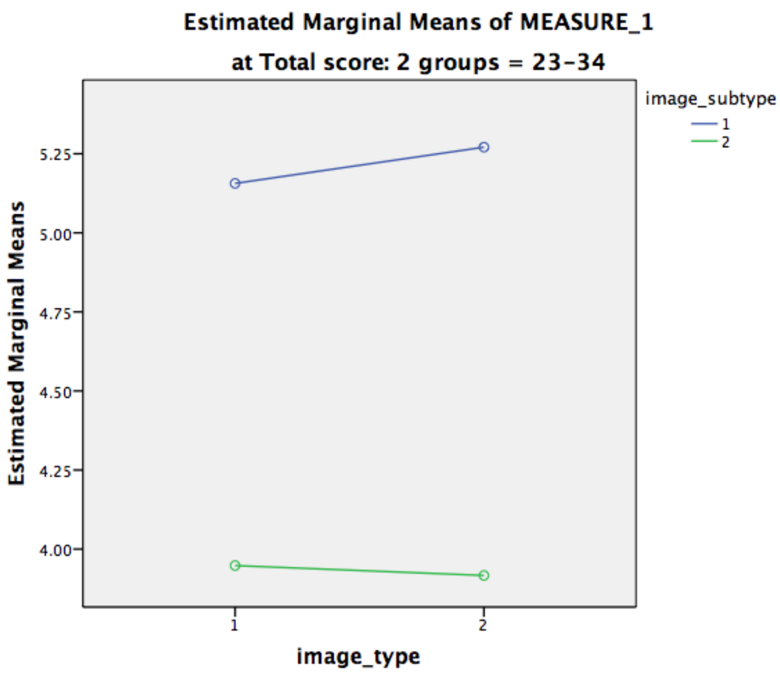


Figure 16 shows significant difference in the above question subtypes, and high and low performing students; i.e. the students' performance on image-questions referring to bones was significantly better than image questions referring to soft tissues, and the performance of high and low performing students was significantly different. The interaction between the image questions, the above question subtypes and low-high performing students has relatively low significance.

<u>For 6th general linear model (analysis of three question-types and regional anatomy):</u>

- Within-subjects' factors (IVs):
    - Question types
        - 1 = questions with no images
        - 2 = questions with anatomical images
        - 3 = questions with radiological images
    - Regional anatomy

- 1 = limb questions
- 2 = torso questions
- 3 = head neck brain and neuroanatomy questions (HN)

The variables were computed as shown in table 16.

Between-subjects' factors: low and high performing students' scores

- 1 = students with scores between 11 and 22
- 2 = students with scores between 23 and 34

Table 16: Within-subjects factors for question types and regional anatomy

Within-Subjects Factors		
Question types	Regional anatomy	Dependent Variable
1	1	No image limbs
	2	No image torso
	3	No image HN
2	1	Anatomical image limbs
	2	Anatomical image torso
	3	Anatomical image HN
3	1	Radiology image limbs
	2	Radiology image torso
	3	Radiology image HN

Within-subject factors were computed as follows:

- No_image_limbs=TR1 + TR2 + TR3 + TR4 + TR5 + TR6 + TR7 + TR8 + TR9 + TR10 + TR11 + TR12 (total 12 questions)
- No_image_torso=TR25 + TR26 + TR27 + TR28 + TR29 + TR30 + TR31 + TR32 + TR33 + TR34 + TR35 + TR36 (total 12 questions)
- No_image_HN=TR13 + TR14 + TR15 + TR16 + TR17 + TR18 + TR19 + TR20 + TR21 + TR22 + TR23 + TR24 (total 12 questions)
- Anatomical_image_limbs=TR1 + TR2 + TR3 + TR4 + TR5 + TR6 + TR7 + TR8 + TR9 + TR10 + TR11 + TR12 (total 12 questions)

- Anatomical_image_torso=TR25 + TR26 + TR27 + TR28 + TR29 + TR30 + TR31 + TR32 + TR33 + TR34 + TR35 + TR36 (total 12 questions)
- Anatomical_image_HN=TR13 + TR14 + TR15 + TR16 + TR17 + TR18 + TR19 + TR20 + TR21 + TR22 + TR23 + TR24 (total 12 questions)
- Radiology_image_limbs=TR1 + TR2 + TR3 + TR4 + TR5 + TR6 + TR7 + TR8 + TR9 + TR10 + TR11 + TR12 (total 12 questions)
- Radiology_image_torso=TR25 + TR26 + TR27 + TR28 + TR29 + TR30 + TR31 + TR32 + TR33 + TR34 + TR35 + TR36 (total 12 questions)
- Radiology_image_HN=TR13 + TR14 + TR15 + TR16 + TR17 + TR18 + TR19 + TR20 + TR21 + TR22 + TR23 + TR24 (total 12 questions)

Descriptive data shows the means and standard deviations of low and high performing students on three question types and regional anatomy. See table 17

Table 17: Means and standard deviations of low and high performing students on three question types and regional anatomy

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
No image limbs	11-22	2.18	1.003	78
	23-34	3.22	0.784	96
	Total	2.75	1.027	174
No image torso	11-22	2.23	0.882	78
	23-34	2.83	0.914	96
	Total	2.56	0.946	174
No image HN	11-22	1.4	0.888	78
	23-34	2.53	1.036	96
	Total	2.02	1.122	174
Anatomical image limbs	11-22	2.92	0.849	78
	23-34	3.75	0.435	96
	Total	3.38	0.771	174
Anatomical image torso	11-22	1.87	0.727	78
	23-34	2.61	0.826	96
	Total	2.28	0.864	174
Anatomical image HN	11-22	1.96	0.959	78
	23-34	2.74	0.771	96
	Total	2.39	0.942	174

Radiology image limbs	11-22	2.29	0.955	78
	23-34	3.15	0.821	96
	Total	2.76	0.978	174
Radiology image torso	11-22	2.06	0.944	78
	23-34	2.82	0.808	96
	Total	2.48	0.948	174
Radiology image HN	11-22	2.08	0.879	78
	23-34	3.22	0.771	96
	Total	2.71	0.997	174

Mauchly's test of sphericity is not significant so the assumptions are met. Tests of within-subjects effects and contrasts show that:

- There was a highly significant difference in the above question-types ($F(2, 344) = 12.24$, $p < .001$), partial eta squared = .066 (indicates medium effect size).
- There was a highly significant difference in the students' performance on three regional anatomy questions ($F(2, 344) = 67.78$, $p < .001$), partial eta squared = .283 (indicates large effect size).
- The interaction between the question-types and regional anatomy questions was significant ($F(4, 688) = 29.37$, $p < .001$), partial eta squared = .146 (indicates large effect size).

The analysis shows significant difference in performance of students:

- on questions-types:
  - with anatomical images better than no images, partial eta squared = .116 (large effect size)
  - with radiology images better than no images, partial et squared = .070 (medium effect size)
- on high low performing students, partial eta squared = .651 (very large effect size)

- on regional anatomy questions (see table 18):
    - with limb questions better than torso questions, partial eta squared = .325 (large effect size)
    - with limb questions better than HN questions, partial eta squared = .392 (large effect size)
- the interaction between the question types and regional anatomy questions (shown below).

This means students performed better on limbs questions as compared to torso and HN brain questions as shown below in table 18.

Table 18: Pairwise comparisons of questions on regional anatomy

Pairwise Comparisons						
Regional anatomy	Regional anatomy	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	.512*	.056	.000	.376	.649
	3	.598*	.057	.000	.461	.735
2	1	-.512*	.056	.000	-.649	-.376
	3	.085	.053	.337	-.044	.215
3	1	-.598*	.057	.000	-.735	-.461
	2	-.085	.053	.337	-.215	.044

As regional anatomy is also interacting at different levels with question-types, there was another level of complexity added to my research question on the students' performance on questions with and without anatomical and radiological images.

Table 19: Interactions between question types and regional anatomy

Question types * Regional anatomy				
Question types and regional anatomy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No image limbs	2.699	.068	2.565	2.833
No image torso	2.532	.069	2.397	2.667

No image HN	1.964	.074	1.818	2.111
Anatomical image limbs	3.337	.050	3.238	3.435
Anatomical image torso	2.243	.060	2.125	2.361
Anatomical image HN	2.351	.066	2.221	2.480
Radiology image limbs	2.720	.067	2.587	2.853
Radiology image torso	2.444	.066	2.312	2.575
Radiology image HN	2.648	.063	2.524	2.771

This means:

In limbs anatomy based questions, performance on:

- anatomical images was significantly better than no images questions
- radiology images was significantly better than no images questions

In torso anatomy based questions, performance on no images was significantly better than anatomical images

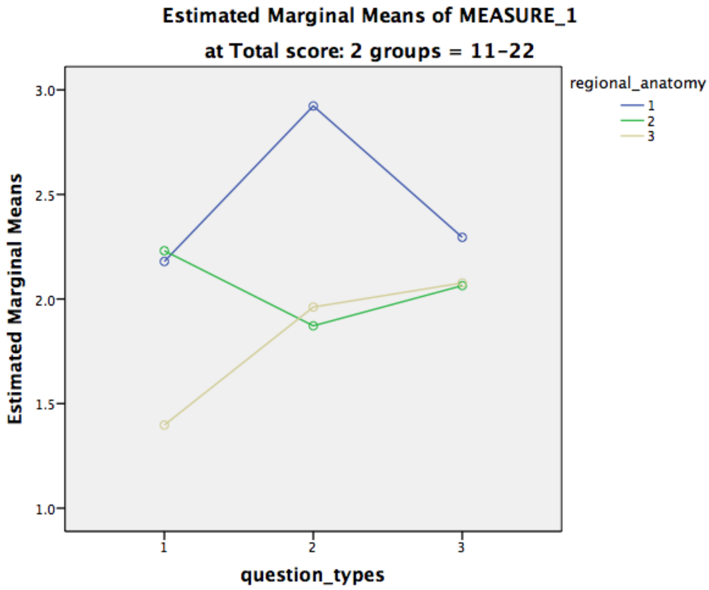
In HN anatomy based questions, performance on radiology images was significantly better than no images

In figure 17, for regional-anatomy, 1 (blue line) = limb questions, 2 (green line) = torso questions, 3 (yellow line) = head & neck questions

For question-types on x-axis: 1 = questions without images, 2 = questions with anatomical images, 3 = questions with radiological images

1st graph is of low performing students and 2nd graph is of high performing students

Figure 17: High and low performers' scores (as two separate figures) on question types and regional anatomy



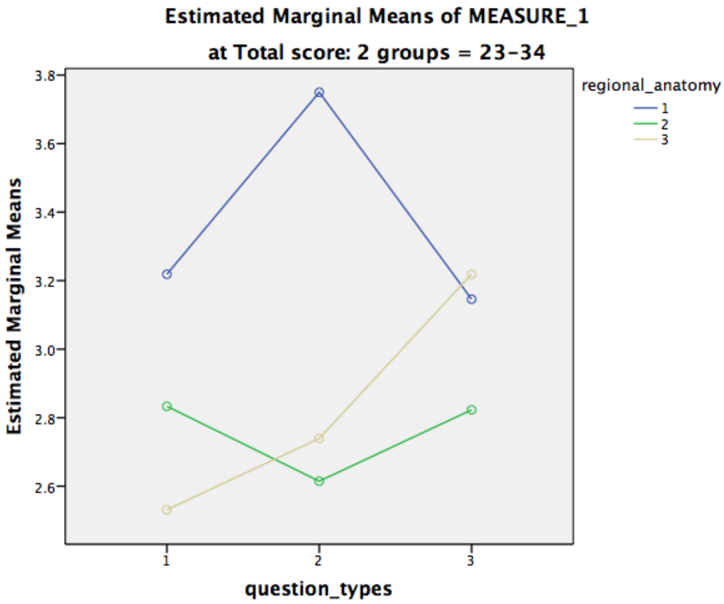


Figure 17 shows a significant difference in questions with and without anatomical and radiology images, three regional anatomy questions, and high and low performing students, i.e. the students' performance on questions with images was significantly better than on questions without images, the students' performance on limb anatomy questions was better than torso and head neck anatomy questions, and the performance of high and low performing students was significantly different. The interaction between the question-types and regional anatomy questions was significant. The performance on limb related anatomical image questions was significantly better than limb anatomy no-image questions, and limb anatomy radiology image questions performance was better than the limb anatomy no-image questions. The performance on torso anatomy no image questions was significantly better than the torso anatomy anatomical image questions. Moreover, the performance on HN anatomy radiology image questions was significantly better than HN anatomy no-image questions.

For 7th general linear model (images-questions, questions-subtypes and regional anatomy):

- Within-subjects' factors (IVs):
  - Question types (two types with images)
    - 1 = questions with anatomical images
    - 2 = questions with radiology images
  - Questions-subtypes
    - 1 = questions referring to bones
    - 2 = questions referring to soft tissue
  - Regional anatomy
    - 1 = limb questions
    - 2 = torso questions
    - 3 = head neck brain and neuroanatomy questions (HN)

The variables were computed as shown in table 20.

Between-subjects' factors: low and high performing students' scores

- 1 = students with scores between 11 and 22
- 2 = students with scores between 23 and 34

Table 20: Within-subjects factors for image questions, subtypes and regional anatomy

Within-Subjects Factors			
Image questions	subtypes	Regional anatomy	Dependent Variable
1	1	1	Anatomical image bones limbs
		2	Anatomical image bones torso
		3	Anatomical image bones HN
	2	1	Anatomical image soft tissue limbs
		2	Anatomical image soft tissue torso
		3	Anatomical image soft tissue HN
2	1	1	Radiology image bones limbs
		2	Radiology image bones torso
		3	Radiology image bone HN
	2	1	Radiology image soft tissue limbs
		2	Radiology image soft tissue torso
		3	Radiology image soft tissue HN

Within-subject factors were computed as follows:

- Anatomical_image_bones_limbs = TR5 + TR6 (2 questions)
- Anatomical_image_bones_ torso = TR29 + TR30 (2 questions)
- Anatomical_image_bones_ HN = TR18+ TR19 (2 questions)

- Anatomical_image_soft-tissue_limbs = TR7 + TR8 (2 questions)
- Anatomical_image_soft-tissue_ torso = TR31 + TR32 (2 questions)
- Anatomical_image_soft-tissue_ HN = TR17+ TR20 (2 questions)

- Radiology_image_bone_limbs = TR9 + TR10 (2 questions)
- Radiology_image_bone_torso = TR33 + TR34 (2 questions)
- Radiology_image_bone_ HN= TR21 + TR22 (2 questions)

- Radiology_image_soft-tissue_limbs = TR11 + TR12 (2 questions)
- Radiology_image_soft-tissue_ torso = TR35 + TR36 (2 questions)

- Radiology_image_soft-tissue_ HN = TR23 + TR24 (2 questions)

Descriptive data shows the means and standard deviations of low and high performing students on image questions, questions subtypes and regional anatomy. See table 21

Table 21: Means and standard deviations of low and high performing students on image questions, questions subtypes and regional anatomy

Descriptive Statistics				
	Total score: 2 groups	Mean	Std. Deviation	N
Anatomical image bones limbs	11-22	1.51	.597	78
	23-34	1.91	.293	96
	Total	1.73	.495	174
Anatomical image bones torso	11-22	1.27	.617	78
	23-34	1.36	.564	96
	Total	1.32	.589	174
Anatomical image bones HN	11-22	1.51	.679	78
	23-34	1.89	.320	96
	Total	1.72	.544	174
Anatomical image soft tissue limbs	11-22	1.41	.653	78
	23-34	1.84	.365	96
	Total	1.65	.557	174
Anatomical image soft tissue torso	11-22	.60	.566	78
	23-34	1.25	.665	96
	Total	.96	.700	174
Anatomical image soft tissue HN	11-22	.45	.617	78
	23-34	.85	.696	96
	Total	.67	.690	174
Radiology image bones limbs	11-22	1.38	.608	78
	23-34	1.74	.464	96
	Total	1.58	.561	174
Radiology image bones torso	11-22	1.22	.714	78
	23-34	1.74	.508	96
	Total	1.51	.661	174
Radiology image bone HN	11-22	1.18	.659	78
	23-34	1.79	.433	96
	Total	1.52	.624	174
Radiology image soft tissue limbs	11-22	.91	.687	78
	23-34	1.41	.689	96
	Total	1.18	.730	174
Radiology image soft tissue torso	11-22	.85	.666	78
	23-34	1.08	.691	96
	Total	.98	.688	174
Radiology image soft tissue HN	11-22	.90	.594	78
	23-34	1.43	.594	96
	Total	1.19	.648	174

Mauchly’s test of sphericity is not significant so the assumptions are met.

Tests of within-subjects’ effects and contrasts showed that there was a significant difference in the students’ performance related to:

- Question subtypes (F (1, 172) = 277.30, $p < .001$), partial eta squared = .617 (indicates very large effect size) i.e. questions referring to bone better than questions referring to soft tissue.
- regional anatomy questions (F (2, 344) = 67.78, $p < .001$), partial eta squared = .283 (indicates large effect size)
    - with limb questions better than torso questions, partial eta squared = .325 (large effect size)
    - with limb questions better than HN questions, partial eta squared = .392 (large effect size)
- The interaction between image questions, question subtypes and low-high performing students (F (1, 172) = 7.08, $p<.05$), partial eta squared = .040 (small effect size) – image type 2-1 with subtypes 2-1(F (1, 172) = 7.09, $p<.05$), partial eta squared = .040 (small effect size) i.e. low and high performing students better on image referring to bones than soft tissue.
- the interaction between image-questions and regional anatomy questions (F (2, 344) = 34.65, $p < .001$) partial eta squared = .168 (large effect size)

Table 22: Interactions between images questions and regional anatomy

Image questions * regional anatomy				
Image questions and regional anatomy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Anatomical image limbs	1.668	.025	1.619	1.717
Anatomical image torso	1.122	.030	1.063	1.180
Anatomical image HN	1.175	.033	1.111	1.240
Radiology image limbs	1.360	.034	1.294	1.427
Radiology image torso	1.222	.033	1.156	1.287
Radiology image HN	1.324	.031	1.262	1.386

- the interaction between the question-subtypes and regional anatomy questions ($F(2, 344) = 22.87$, $p < .001$), partial eta squared = .117 (large effect size) –
    - subtypes 2-1 with regional anatomy 2-1 ($F(1, 172) = 9.28$, $p<.05$), partial eta squared = .051 (small effect size), i.e. performance on limbs anatomy was better on bones than soft tissue, and performance on torso anatomy was better on bones than soft tissue.
    - subtypes 2-1 with regional anatomy and 3-1($F(1, 172) = 54.66$, $p<.001$), partial eta squared = .241(large effect size), i.e. performance on HN anatomy was better on bones than soft tissue.

Table 23: Interactions between subtypes and regional anatomy

Subtypes * regional anatomy				
Subtypes and regional anatomy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Bone limbs	1.636	.027	1.582	1.690
Bone torso	1.398	.033	1.333	1.462
Bone HN	1.592	.028	1.536	1.648
Soft tissue limbs	1.393	.034	1.325	1.460
Soft tissue torso	.946	.038	.871	1.020
Soft tissue HN	.907	.034	.841	.973

- the interaction between the image questions, question subtypes and regional anatomy questions (F (2, 344) = 40.62, p < .001), partial eta squared = .191 (large effect size) – subtypes 2-1 with regional anatomy 3-1 (F (1, 172) = 79, p<.001), partial eta squared = .315 (large effect size)
- the interaction between images questions, subtypes, regional anatomy and low-high performing students (F (2, 344) = 7.93, (p<.001), partial eta squared = .044 (small effect size) – subtypes 2-1 with regional anatomy 2-1 (F (1, 172) = 14.58, p<.001), partial eta squared = .078 (medium effect size)

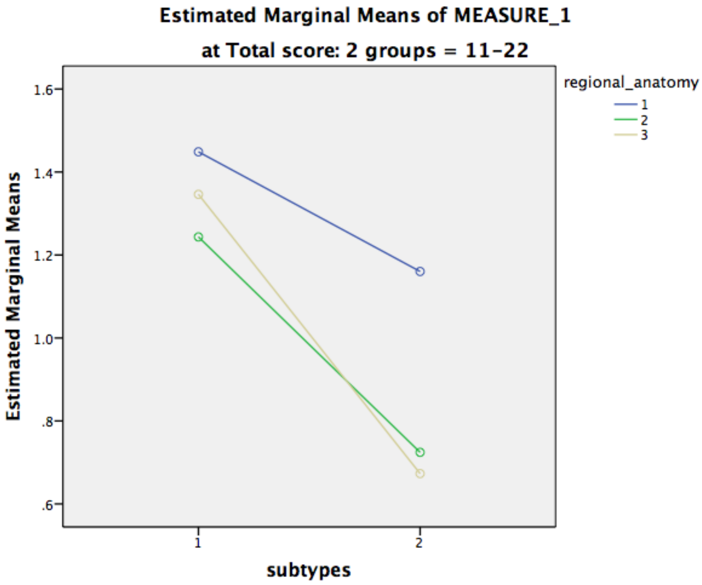
As image questions, question subtypes, regional anatomy questions and low-high performing students also interact, there is another level of complexity added to the analysis.

In figure 18, for regional-anatomy: 1 (blue line) = limb questions, 2 (green line) = torso questions, 3 (yellow line) = Head & neck questions.

For subtypes, 1 = questions referring to bones, 2 = questions referring to soft tissue (x-axis)

1st graph is of low performing students and 2nd graph is of high performing students

Figure 18: High and low performers' scores (as two separate figures) on questions subtypes on three anatomical regions



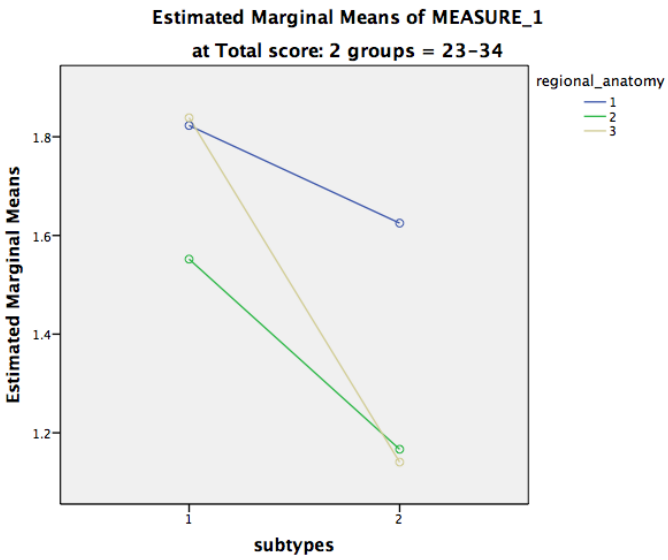


Figure 18 shows a significant difference in the above question subtypes, three regional anatomy questions, and high and low performing students. The interaction between the image questions, the above question subtypes and low-high performing students was significant, i.e. the students' performance on image questions referring to bones was significantly better than image-questions referring to soft tissues; the students' performance on limb anatomy questions was better than torso and head neck anatomy questions; and the performance of high and low performing students was significantly different. Moreover, the interaction between image questions and regional anatomy questions was significant. The interaction between question subtypes and regional anatomy questions was significant, i.e. in all three regions, the performance on image questions referring to bones was better than questions referring to soft tissues. Furthermore, the interaction between the image questions, the above question subtypes and regional anatomy questions was significant.

Investigating students' performance on the test with regard to the questionnaire variables

Time taken to complete the test

The minimum time taken to complete the test was 11 minutes 28 seconds, and the maximum was 1 hour 17 minutes 56 seconds (with mean of 30 minutes 17 seconds and standard deviation of 11 minutes 58 seconds)

The time variable was grouped into students needing under (and including) 36 minutes, and those needing over 36 minutes. This was done because usually 1 or 1.5 minute is assumed to be adequate to complete a question in anatomy examinations. In this test, 75.3% students were in the first group, while 24.7% students took above 36 minutes to complete the test.

The correlation between total time taken and total scores was not significant, therefore it was not analysed further.

Demographics characteristics (Medical schools, the anatomical resources used, sex, age distribution and training level):

Medical schools

ANOVA was carried out between total score and medical schools. See table 24 for descriptive statistics.

Table 24: Descriptive data for number of students participated from each medical school

Descriptive statistics				
Total Correct				
	N	Mean scores	Std. Deviation	Std. Error
School B	3	18.33	2.517	1.453
School H	13	19.23	3.059	.848
School K	121	24.50	4.850	.441
School P	12	21.33	2.570	.742
School S	17	20.88	4.781	1.160
School E	8	22.75	3.882	1.373
Total	174	23.34	4.869	.369

69.5% of data belonged to School K. Therefore, further analysis was not carried out.

Anatomical resources used to teach anatomy

Test was carried out between total score and anatomical resources used. See table 25 for descriptive statistics.

Table 25: Descriptive data for anatomical resources used to teach anatomy

Descriptive statistics				
Total Correct				
	N	Mean scores	Std. Deviation	Std. Error
Dissection of human cadavers only	4	21.50	10.472	5.236
Prosections (dissected body parts) only	7	20.43	3.359	1.270
Radiology images only	19	21.84	3.219	.739
All of the above	109	24.27	4.906	.470
Dissection of human cadavers and radiology images	2	21.50	7.778	5.500
Dissection of human cadavers and prosections	23	23.30	4.446	.927
Prosections and radiology images	10	19.40	2.413	.763
Total	174	23.34	4.869	.369

62.6% of students were taught anatomy in their schools through all the resources (dissection of human cadavers, prosections and radiology images).

Figure 19: Total scores and how anatomy has been taught in your school

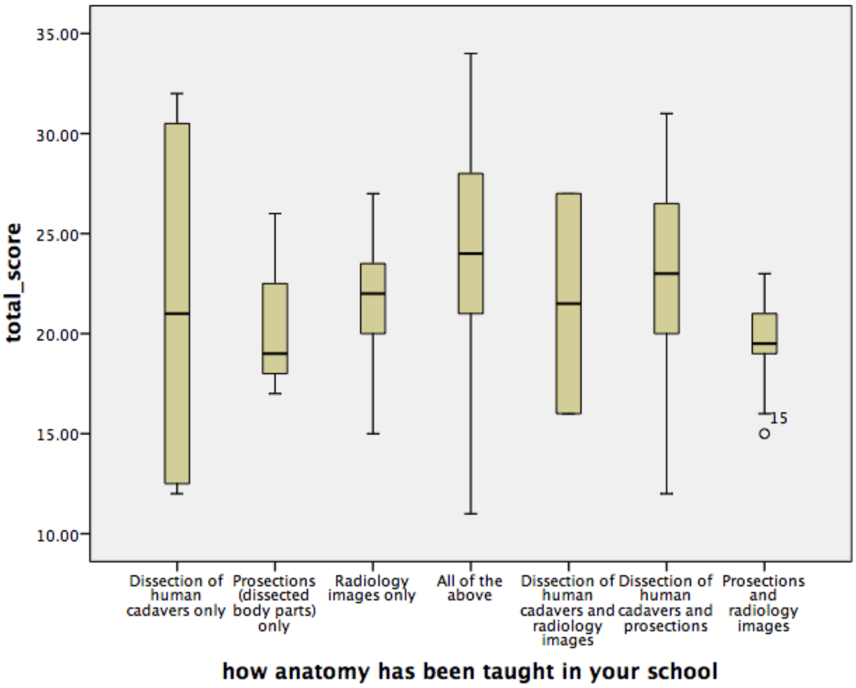


Table 26: ANOVA for total scores and how anatomy has been taught in your school

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	371.016	6	61.836	2.768	.014
Within Groups	3730.295	167	22.337		
Total	4101.310	173			

Further Bonferroni test was carried out to determine which anatomical resources used for teaching made a significant difference:

- All of the above resources were better than prosections and radiology images (mean difference 4.866, p = .045)

The data were variably distributed.

Sex of the students

Test was carried out between total score and sex of the students. See table 27 for descriptive statistics.

Table 27: Descriptive data for students' sex

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Female	96	22.84	4.671	.477
Male	78	23.96	5.064	.573
Total	174	23.34	4.869	.369

Figure 20: Total scores and students' sex

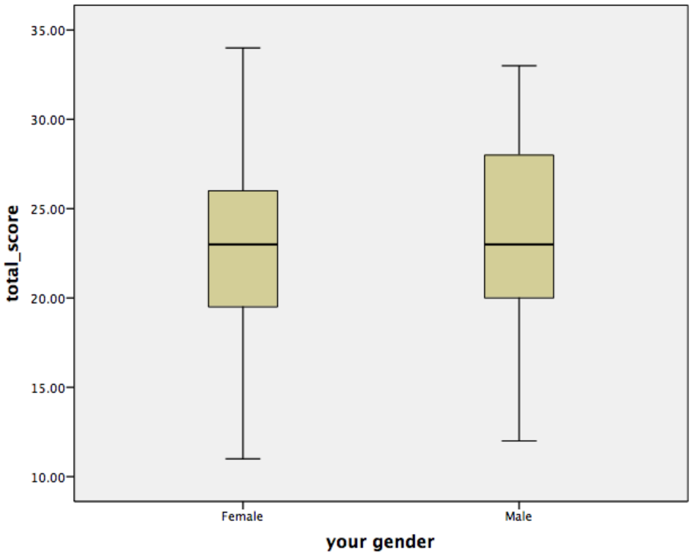


Table 28: ANOVA for total scores and students' sex

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.769	1	53.769	2.285	.132
Within Groups	4047.541	172	23.532		
Total	4101.310	173			

The relationship between total score and sex of the students was not significant.

Age range

Most students (69%) were of age group 19-21. Test was carried out between total score and the students' age range. See table 29 for descriptive statistics.

Table 29: Descriptive data for age range

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
16-18	5	24.40	5.595	2.502
19-21	120	22.56	4.570	.417
22-24	30	25.40	5.354	.977
25-27	10	23.40	4.402	1.392
28-30	4	28.00	2.160	1.080
31-33	2	25.50	.707	.500
34 or above	3	24.67	9.018	5.207
Total	174	23.34	4.869	.369

69% of data belonged to age range 19-21. Further analysis was not carried because the data were inconsistently variable.

The Students' Training Level

Test was carried out between total score and the students' training level. See table 30 for descriptive statistics.

Table 30: Descriptive data for students' training level

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
End of 2nd year student (MBBS 5 - Undergraduate stream)	150	22.74	4.636	.379
End of 1st year student (MBBS 4 - Graduate entry programme)	24	27.13	4.665	.952
Total	174	23.34	4.869	.369

As the distribution of data was not consistent, i.e. 86.2% students were at the end of their 2nd year (undergraduate stream), this variable was not investigated further.

<u>Preferences, views and experience (most likely prospective career, learning and assessment preferences) and views:</u>

Most likely prospective career

Test was carried out between total scores and most likely prospective career. See table 31 for descriptive statistics.

Table 31: Descriptive data for most likely prospective career

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Non-surgical	68	22.19	4.723	.573
Surgical	35	24.63	4.124	.697
Don't know	71	23.82	5.161	.613
Total	174	23.34	4.869	.369

In the test, 40.8% of students selected "don't know" option followed by "non-surgical" option (39.1%) and then "surgical" (20.1%) for their most likely prospective career. As a high percentage of students “didn’t know” about their prospective career, this was not investigated further.

Find/found hands-on cadaveric dissection an effective way of learning

Test was carried out between total score and students who find/found cadaveric dissection an effective way of learning. See table 32 for descriptive statistics.

Table 32: Descriptive data for hands-on cadaveric dissection an effective way of learning

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	114	24.10	5.026	.471
Unsure	22	23.64	5.260	1.122
Disagree	6	21.17	3.920	1.600
Not done	32	20.88	3.108	.549
Total	174	23.34	4.869	.369

A total of 65.5% agreed that hands-on cadaveric dissection is an effective way of learning.

Figure 21: Total scores and students who find/found cadaveric dissection an effective way of learning

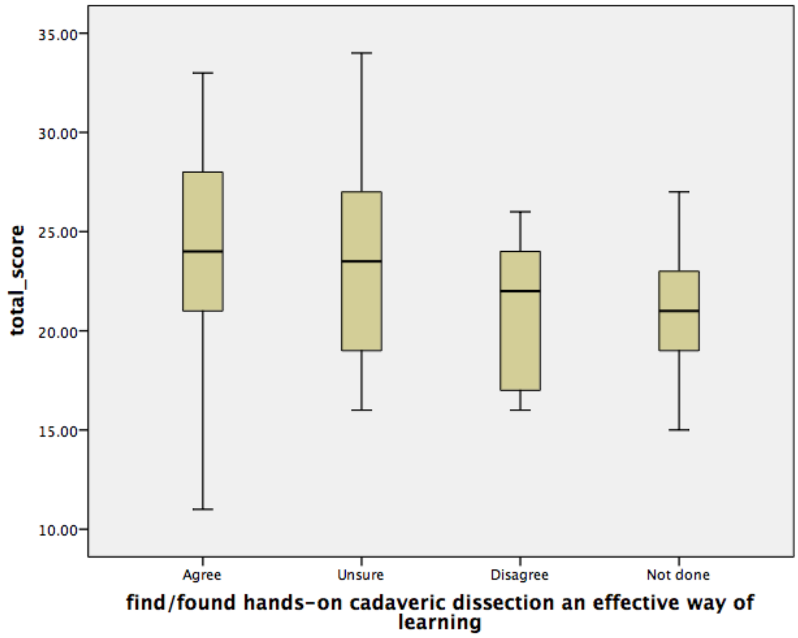


Table 33: ANOVA for total scores and students who find/found cadaveric dissection an effective way of learning

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	289.948	3	96.649	4.311	.006
Within Groups	3811.363	170	22.420		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether views on finding cadaveric dissection an effective way of learning made a significant difference to their scores:

- Students who agreed scored better than those who have not done cadaveric dissection (mean difference = 3.221, p = .005).

Find/found prosections (dissected body parts) an effective way of learning

Test was carried out between total score and students who find/found prosections (dissected body parts) an effective way of learning. See table 34 for descriptive statistics.

Table 34: Descriptive data for prosections (dissected body parts) an effective way of learning

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	125	23.89	5.132	.459
Unsure	24	22.79	4.107	.838
Disagree	7	20.14	4.375	1.654
Not used	18	21.56	3.053	.720
Total	174	23.34	4.869	.369

A total of 71.8% of students agreed that using prosections is an effective way of learning.

Figure 22: Total scores and students who find/found prosections as effective way of learning

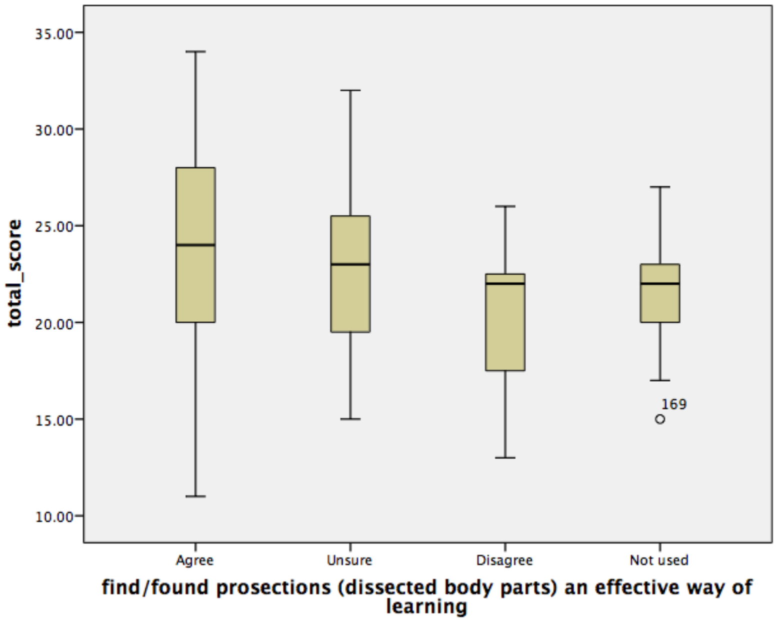


Table 35: ANOVA for total scores and students who find/found prosections as effective way of learning

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	173.618	3	57.873	2.505	.061
Within Groups	3927.692	170	23.104		
Total	4101.310	173			

This relationship was not significant.

Find/found cadaveric photographs an effective way of learning anatomy

Test was carried out between total score and the students' who find/found cadaveric photographs an effective way of learning anatomy. See table 36 for descriptive statistics.

Table 36: Descriptive data for students who find/found cadaveric photographs an effective way of learning anatomy

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	59	22.73	5.010	.652
Unsure	62	23.82	4.579	.582
Disagree	34	23.59	5.263	.903
Not used	19	23.26	4.794	1.100
Total	174	23.34	4.869	.369

A total of 35.6% of students were unsure and 33.9% agreed that cadaveric photographs represent an effective way of learning anatomy.

Figure 23: Total scores and students who find/found cadaveric photographs an effective way of learning anatomy

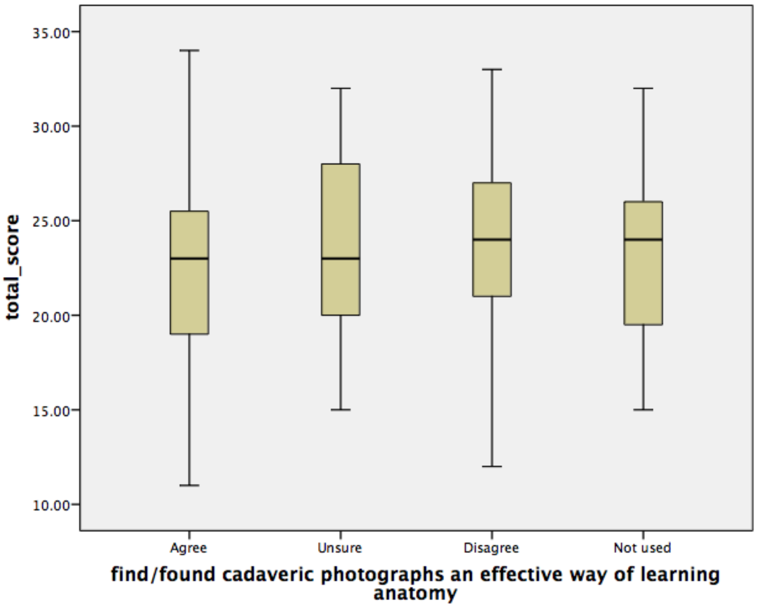


Table 37: ANOVA for total scores and students who find/found cadaveric photographs an effective way of learning anatomy

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	38.681	3	12.894	.540	.656
Within Groups	4062.629	170	23.898		
Total	4101.310	173			

This relationship was not significant.

Find/found radiological images an effective way of learning anatomy

Test was carried out between total scores and the students' who find/found radiological images an effective way of learning anatomy. See table RT38 for descriptive statistics.

Table 38: Descriptive data for radiological images an effective way of learning anatomy

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	105	23.80	4.919	.480
Unsure	38	23.34	5.026	.815
Disagree	23	20.65	3.868	.807
Not used	8	25.13	3.907	1.381
Total	174	23.34	4.869	.369

A total of 60.3% of students agreed that using radiological images is an effective way of learning anatomy.

Figure 24: Total scores and students who find/found radiological images an effective way of learning anatomy

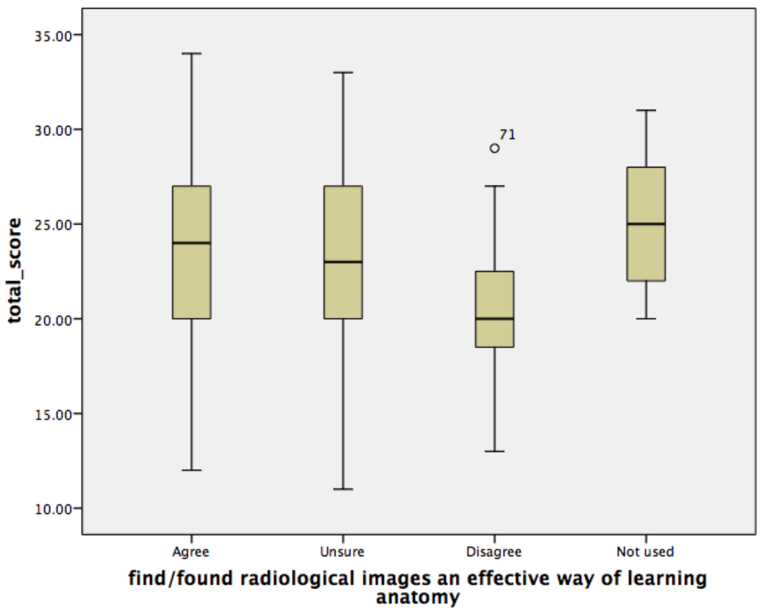


Table 39: ANOVA for total scores and students who find/found radiological images an effective way of learning anatomy

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	213.865	3	71.288	3.117	.028
Within Groups	3887.445	170	22.867		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether their views on learning from radiology images made a significant difference to their scores:

- Students who agreed scored better than those who disagreed (mean difference 3.148, p = .029).

Hence, the data showed that students' views, on learning preferences, inclining towards cadaveric dissection and radiology images, have significantly affected their scores on the test.

Furthermore, a few related questions were asked in the questionnaire to check the students' preferences on the assessment system; and to investigate if their preferences have any effect on their performance.

Cadaveric photographs are more effective for examinations than prosections and dissected material

Test was carried between total scores and their views on cadaveric photographs being more effective for examinations. See table 40 for descriptive statistics.

Table 40: Descriptive data for cadaveric photographs being more effective for examinations than prosections and dissected material

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	11	23.73	4.197	1.266
Unsure	66	22.02	5.146	.633
Disagree	97	24.21	4.580	.465
Total	174	23.34	4.869	.369

A total of 55.7% of students disagreed and 37.9% were unsure if cadaveric photographs are more effective for examination than prosections and dissected material.

Figure 25: Total scores and students' views on cadaveric photographs being more effective for examinations than prosections and dissected material

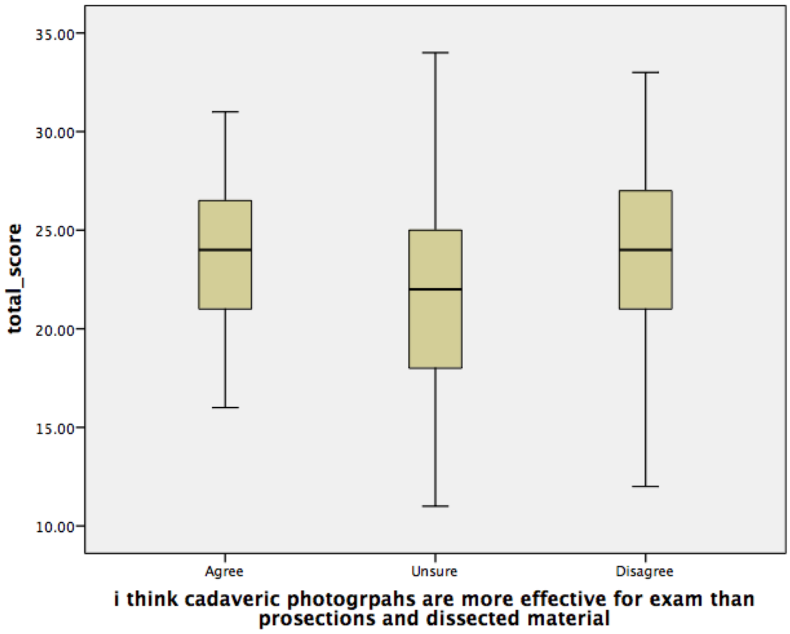


Table 41: ANOVA for total scores and students' views on cadaveric photographs being more effective for examinations than prosections and dissected material

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	190.267	2	95.134	4.159	.017
Within Groups	3911.043	171	22.872		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether views on use of cadaveric photographs as effective resource in examinations made a significant difference on the students' scores:

- Students who disagreed scored better than those who were unsure (mean difference = 2.191, p = .014).

Think radiology images are more effective for examination than prosections and dissected material

Test was carried between the total scores and the students' views on radiology images as effective resource for examination. See table 42 for descriptive statistics.

Table 42: Descriptive data for students who think radiology images are more effective for examination than prosections and dissected material

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	15	23.53	4.868	1.257
Unsure	61	22.97	4.810	.616
Disagree	98	23.55	4.941	.499
Total	174	23.34	4.869	.369

A total of 56.3% of students disagreed and 35.1% were unsure if radiological images are more effective for examination than prosections and dissected material.

Figure 26: Total scores and students' thoughts on radiology images as more effective for examination than prosections and dissected material

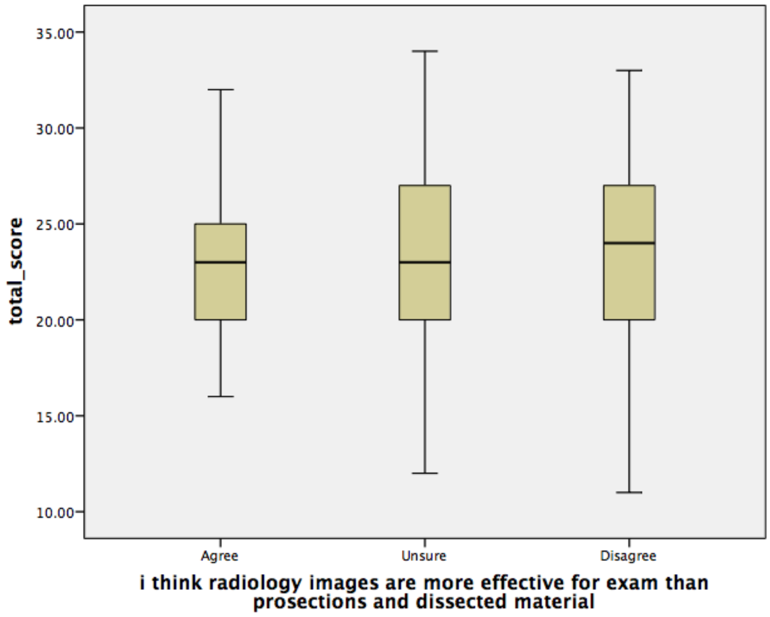


Table 43: ANOVA for total scores and students' thoughts on radiology images as more effective for examination than prosections and dissected material

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.398	2	6.699	.280	.756
Within Groups	4087.913	171	23.906		
Total	4101.310	173			

This relationship was not significant.

Furthermore, a question was included to see how students perceive clinically relevant anatomy (applied, living and surface anatomy).

Find/found clinically relevant anatomy learning (applied, living/surface) an effective way of learning anatomy

Test was carried out between the total scores and the students' thoughts on clinically relevant anatomy learning as effective way of learning. See table 44 for descriptive statistics.

Table 44: Descriptive data for clinically relevant anatomy learning (applied, living/surface) an effective way of learning anatomy

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	149	23.62	4.942	.405
Unsure	20	20.85	3.937	.880
Disagree	5	25.20	3.114	1.393
Total	174	23.34	4.869	.369

A high percentage of students (85.6%) were in favour of clinically relevant anatomy.

Figure 27: Total scores and students' views on clinically relevant anatomy learning (applied/living/surface) an effective way of learning anatomy

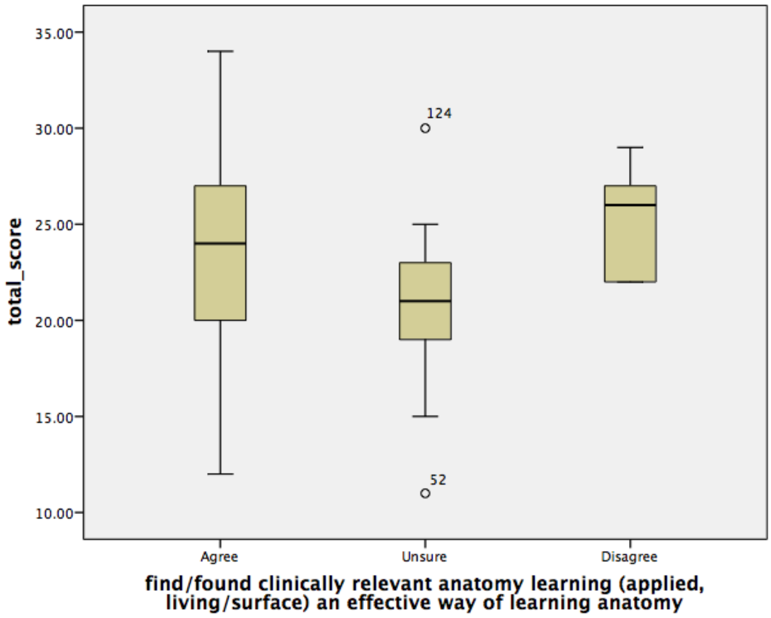


Table 45: ANOVA for total scores and students' views on clinically relevant anatomy learning (applied/living/surface) an effective way of learning anatomy

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	152.766	2	76.383	3.308	.039
Within Groups	3948.545	171	23.091		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether students' views on clinically relevant anatomy learning being an effective way of learning anatomy made a significant difference to their scores:

- Students who agreed scored better than those who were unsure but this was on the borderline (mean difference = 2.767, p = .050)

Believe anatomy knowledge should be tested with clinical knowledge

Test was carried out between total scores and students who believe that anatomy knowledge should be tested with clinical knowledge. See table 46 for descriptive statistics.

Table 46: Descriptive data for anatomy knowledge should be tested with clinical knowledge

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	146	23.48	4.739	.392
Unsure	26	22.88	5.638	1.106
Disagree	2	19.50	3.536	2.500
Total	174	23.34	4.869	.369

It was found that not only students preferred clinically relevant anatomy learning but also a high percentage (83.9%) preferred anatomy knowledge to be tested with clinical knowledge.

Figure 28: Total scores and students who believe that anatomy knowledge should be tested with clinical knowledge

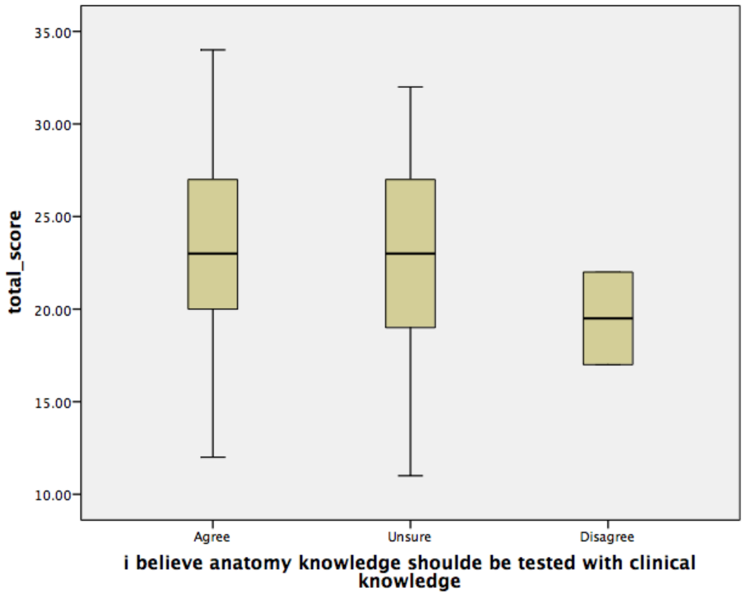


Table 47: ANOVA for total scores and students who believe that anatomy knowledge should be tested with clinical knowledge

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	37.718	2	18.859	.794	.454
Within Groups	4063.592	171	23.764		
Total	4101.310	173			

This relationship was not significant.

Think online anatomy examination are more effective than practical examination in the dissecting room

Along with their preferences on the style of the assessment system, their preference on the technique to deliver these examinations was also investigated.

Test was carried out between the total scores and the students' views on online anatomy examinations. See table 48 for descriptive statistics.

Table 48: Descriptive data for online anatomy examination are more effective than practical examination in the dissecting room

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	20	21.15	5.214	1.166
Unsure	21	24.52	4.718	1.030
Disagree	25	25.76	4.807	.961
not done online exam	69	22.25	4.587	.552
Not done practical exam	39	24.23	4.498	.720
Total	174	23.34	4.869	.369

It was found that 39.7% students had not taken an online examination, and 22.4% students had not taken a practical examination. As this variable was inconsistently distributed, it was not investigated further.

Main motivation for learning anatomy is to pass the examination

Assessment is often known to motivate strategic learners, and this question was added to find out the students' views.

Test was carried out between the total scores and the students' motivation to learn anatomy for passing examinations. See table 49 for descriptive statistics.

Table 49: Descriptive data for main motivation for learning anatomy is to pass the examination

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	81	22.25	4.724	.525
Unsure	23	23.17	5.024	1.048
Disagree	70	24.67	4.723	.565
Total	174	23.34	4.869	.369

The results were not discrete as 46.6% students agreed that their main motivation for learning anatomy is to pass the examination whereas 40.2% disagreed.

Figure 29: Total scores and students' motivation for learning anatomy is to pass the examination

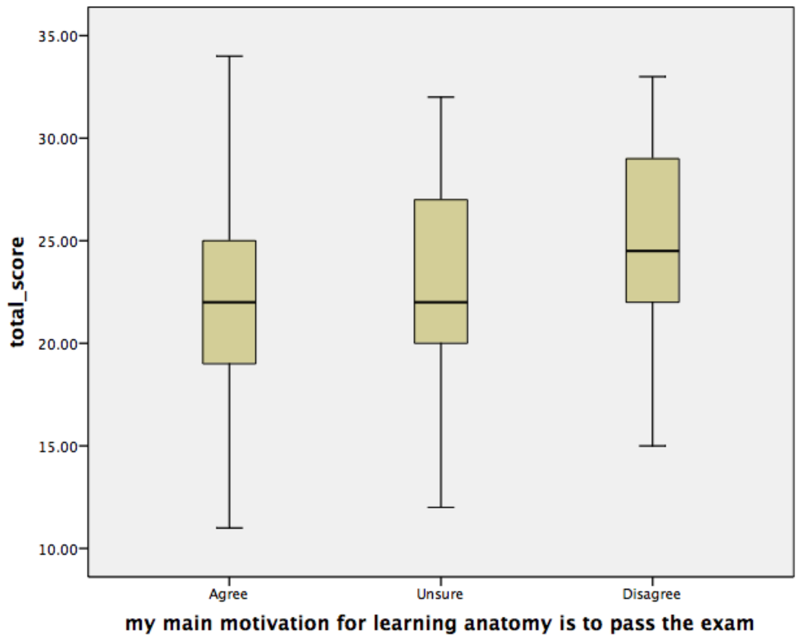


Table 50: ANOVA for total scores and students' motivation for learning anatomy is to pass the examination

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	221.501	2	110.751	4.881	.009
Within Groups	3879.809	171	22.689		
Total	4101.310	173			

Further Bonferroni test was carried out to determine which their motivation had made a significant difference to their scores:

- Students who disagreed scored better than those who agreed (mean difference = 2.425, p = .006).

Think it requires deep understanding to answer questions with prosections and dissected material

Furthermore, the students' views on relationship between the use of particular resources (prosections and dissected material, cadaveric photographs and radiological images) and deep understanding to answer a question were investigated.

Test was carried out between total scores and their views that deep understanding is required for questions with prosections. See table 51 for descriptive statistics.

Table 51: Descriptive data for deep understanding to answer questions with prosections and dissected material

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	115	24.00	4.733	.441
Unsure	42	22.95	4.844	.747
Disagree	17	19.88	4.512	1.094
Total	174	23.34	4.869	.369

A total of 66.1% agreed that it requires deep understanding to answer question with prosections and dissected material.

Figure 30: Total scores and students' views that it requires deep understanding to answer questions with prosections and dissected material

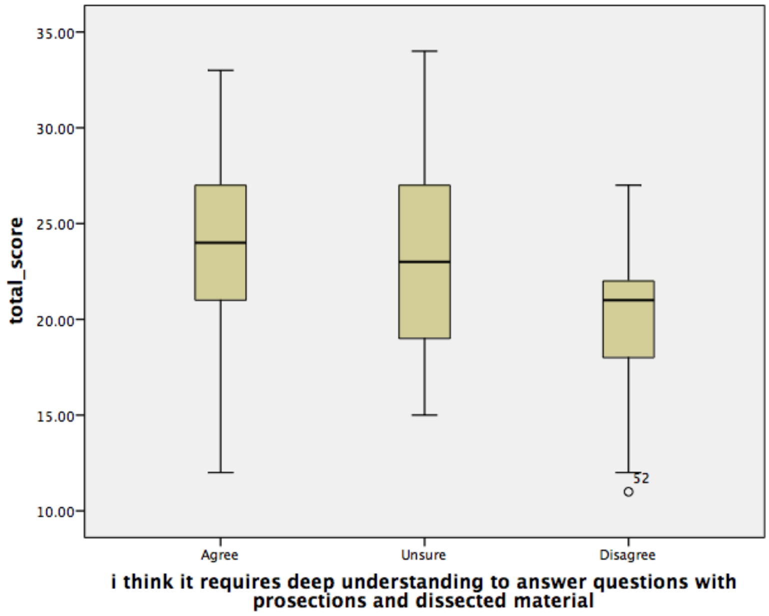


Table 52: ANOVA for total scores and students' views that it requires deep understanding to answer questions with prosections and dissected material

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	259.641	2	129.820	5.779	.004
Within Groups	3841.669	171	22.466		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether their views that deep understanding is required to answer questions with prosections had made a significant difference to their scores:

- Students who agreed scored better than those who disagreed (mean difference = 4.118, p = .003).

Think it requires deep understanding to answer questions with cadaveric photographs

Test was carried out between total scores and their views that deep understanding is required for questions with cadaveric photographs. See table 53 for descriptive statistics.

Table 53: Descriptive data for deep understanding to answer questions with cadaveric photographs

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	102	23.52	4.641	.460
Unsure	62	23.71	5.068	.644
Disagree	10	19.30	4.523	1.430
Total	174	23.34	4.869	.369

A total of 58.6% agreed but 35.6% of students were unsure that it requires deep understanding to answer questions with cadaveric photographs.

Figure 31: Total scores and students' views that it requires deep understanding to answer questions with cadaveric photographs

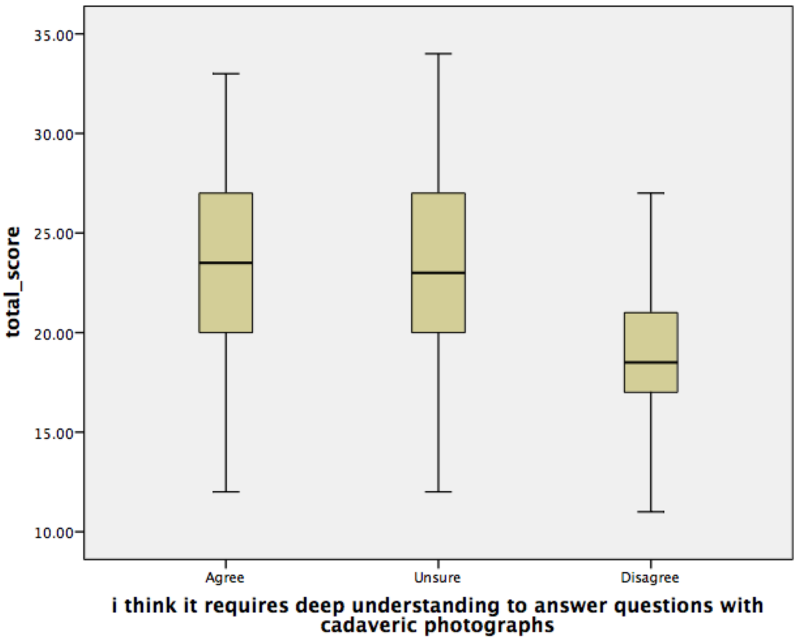


Table 54: ANOVA for total scores and students' views that it requires deep understanding to answer questions with cadaveric photographs

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	174.975	2	87.488	3.810	.024
Within Groups	3926.335	171	22.961		
Total	4101.310	173			

Further Bonferroni test was carried out to determine whether their views that deep understanding is required to answer questions with cadaveric photographs had made a significant difference to their scores:

- Students who agreed scored better than those who disagreed (mean difference = 4.220, p = .026).
- Students who are unsure scored better than those who disagreed (mean difference = 4.410, p = .023).

Think it requires deep understanding to answer questions with radiological images

Test was carried out between total scores and their views that deep understanding is required for questions with radiological images. See table 55 for descriptive statistics.

Table 55: Descriptive data for deep understanding to answer questions with radiological images

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	122	23.59	4.770	.432
Unsure	35	23.66	5.368	.907
Disagree	17	20.94	4.023	.976
Total	174	23.34	4.869	.369

A total of 70.1% of students agreed that it requires deep understanding to answer questions with radiological images.

Figure 32: Total scores and students' views that it requires deep understanding to answer questions with radiological images

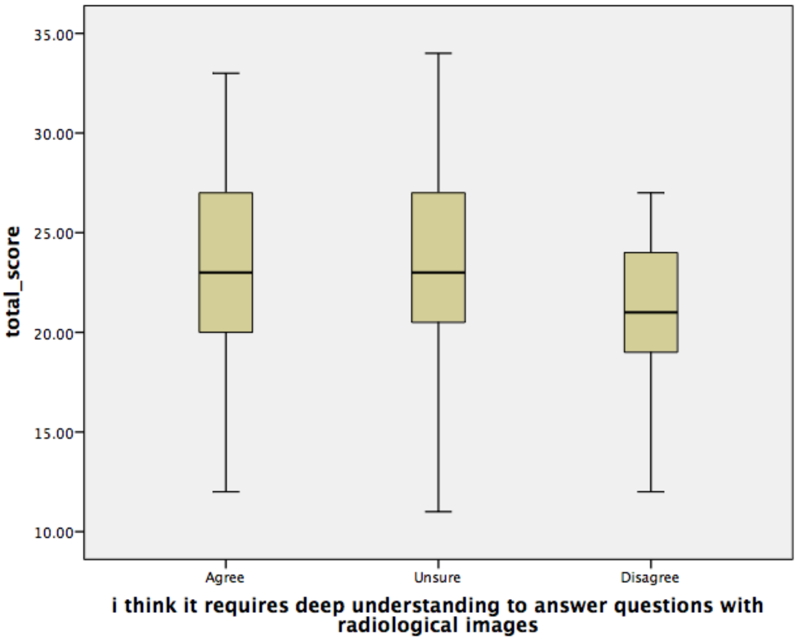


Table 56: ANOVA for total scores and students' views that it requires deep understanding to answer questions with radiological images

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	108.975	2	54.488	2.334	.100
Within Groups	3992.335	171	23.347		
Total	4101.310	173			

This relationship was not significant.

Participation in voluntary demonstrating programme:

Involved in demonstrating anatomy

In the study, 97.1% of students were not involved formally in demonstrating anatomy.

Test was carried out between total scores and their participation in voluntary demonstrating programme. See table 57 for descriptive statistics.

Table 57: Descriptive data for demonstrating anatomy

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Yes	5	24.60	4.450	1.990
No	169	23.31	4.888	.376
Total	174	23.34	4.869	.369

As 97.1% of students have never been involved in formal demonstrating, this variable was not analysed further.

Think anatomy demonstrating helped me to learn anatomy

A further question was included in the questionnaire to investigate if informal demonstrating anatomy had helped students to learn anatomy. A total of 55.7% agreed that demonstrating helped them to learn anatomy.

Test was carried out between total scores and their thinking that anatomy demonstrating helped them learn anatomy. See table 58 for descriptive statistics.

Table 58: Descriptive data for thinking anatomy demonstrating helped to learn anatomy

Descriptive statistics				
Total Correct				
	N	Mean	Std. Deviation	Std. Error
Agree	97	23.93	4.988	.506
Unsure	18	21.94	5.023	1.184
Disagree	7	21.29	1.890	.714
Not done	52	23.02	4.767	.661
Total	174	23.34	4.869	.369

Figure 33: Total scores and students' thoughts that anatomy demonstrating helped them to learn anatomy

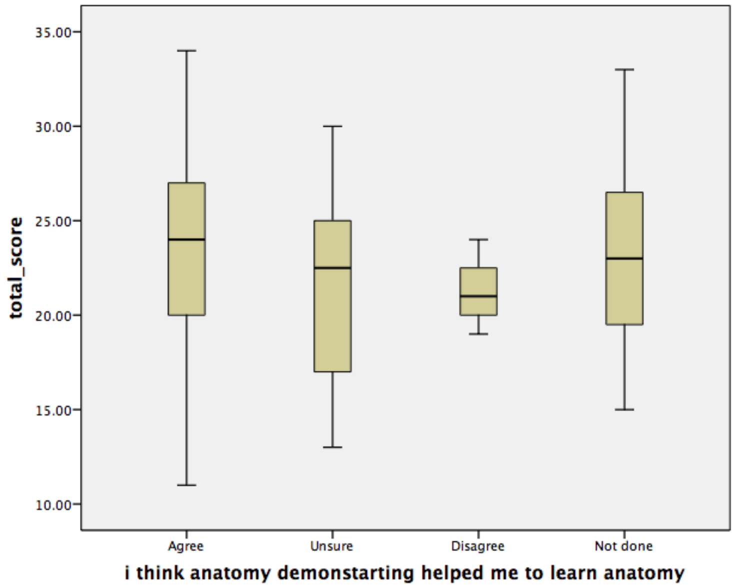


Table 59: ANOVA for total scores and students' thoughts that anatomy demonstrating helped them to learn anatomy

ANOVA					
Total Correct					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	103.462	3	34.487	1.466	.225
Within Groups	3997.849	170	23.517		
Total	4101.310	173			

This relationship was not significant.

Thus the students' performance significantly varies with regard to the following variables:

- Anatomical resources used to teach anatomy
    - Students who used all of the anatomical resources scored significantly better than those who used prosections and radiology images (mean difference 4.866, p = .045)
- Find/found hands-on cadaveric dissection an effective way of learning

    - Students who agreed that cadaveric dissection was an effective way of learning scored better than those who had not done cadaveric dissection (mean difference = 3.221, p = .005).
- Find/found radiological images an effective way of learning anatomy
    - Students who agreed that radiological images are effective way of learning anatomy scored better than those who disagreed (mean difference 3.148, p = .029).
- Cadaveric photographs are more effective for examinations than prosections and dissected material
    - Students who disagreed that cadaveric photographs are more effective for examination than prosections and dissected material scored better than those who were unsure (mean difference = 2.191, p = .014).
- Main motivation for learning anatomy is to pass the examination
    - Students who disagreed that their main motivation for learning anatomy is to pass the examination scored better than those who agreed (mean difference = 2.425, p = .006).
- Think it requires deep understanding to answer questions with prosections and dissected material
    - Students who agreed that it requires deep understanding to answer question with prosections and dissected material scored better than those who disagreed (mean difference = 4.118, p = .003).
- Think it requires deep understanding to answer questions with cadaveric photographs
    - Students who agreed that it requires deep understanding to answer questions with cadaveric photographs scored better than those who disagreed (mean difference = 4.220, p = .026).

  - Students who are unsure that it requires deep understanding to answer questions with cadaveric photographs scored better than those who disagreed (mean difference = 4.410, p = .023).

<u>Investigating the students' performance on question-types with regard to the questionnaire variables:</u>

Repeated measures ANOVA was carried out to investigate if questionnaire variables affected the students' performance on questions with and without anatomical and radiology images.

- Within-subjects' factors (IVs):
  - Question types
    - 1 = questions with no images
    - 2 = questions with anatomical images
    - 3 = questions with radiological images
- Between-subjects' factors: low and high performing students' scores
  - 1 = students with scores between 11 and 22
  - 2 = students with scores between 23 and 34
- Covariates – questionnaire variables

Mauchly's test of sphericity is not significant so the assumptions are met.

Tests of within-subjects' effects and contrasts show that there is significant difference in interaction between question types and the following variables ($p<.05$):

- QR6 (how anatomy has been taught), p = .024, partial eta squared = .023 - (level 1-2 p = .013, partial eta squared = .038; level 1-3 = p = .034, partial eta squared = .028). This means the resources used to teach anatomy made a significant difference to the students' performance on questions with no image and anatomical images, and questions with no images and radiology images. Students performed better on questions with anatomical images than no images, and radiology images than no images (small effect size in both cases).

- QR14 (I believe anatomy should be tested with clinical relevance), $p = .020$, partial eta squared = .025 - (level 1-2 $p = .021$, partial eta squared = .033) (level 2-3 $p = .006$, partial eta squared = .047). This means students' beliefs in being tested with clinically oriented anatomy questions had a significant difference to their performance on questions with no image and anatomical images, and questions with anatomical images and radiology images. Students performed better on questions with anatomical images than no images, and anatomical images than radiology images (small effect size in both cases).
- QR21 (I think it requires deep understanding to answer questions with radiology images), $p = .029$, partial eta squared = .022 - (level 2 and 3 $p = .005$, partial eta squared = .050. This means their belief that it requires deep understanding to interpret radiology images made a significant difference to their performance on questions with anatomical images and radiology images. Students performed better on questions with anatomical images than radiology images (small effect size).

Summary

Consequently, from the test data, a highly significant difference ($p< .001$) in performance was seen between easy and difficult questions; high and low performing students; questions with and without images (two question types and three question types); questions subtypes referring to bones and soft-tissues; and regional anatomy questions. Moreover, significant interactions were seen between three question types and easy-difficult questions and high-low performing students; image questions, question subtypes and high-low performing students; three question types and regional anatomy; and image questions, question subtypes and regional anatomy questions.

From the questionnaire data, a relatively less significant ($p < .05$) performance on the overall test was seen where all anatomical resources were used to teach anatomy; with students who agreed that cadaveric dissection, radiological images are effective ways of learning anatomy; with students who agreed that deep understanding is required to answer questions with cadaveric photographs; and with students who disagreed that cadaveric photographs are more effective for examinations than prosections and dissection material, and students who disagreed that their main motivation for learning anatomy is to pass the examination.

A less significant interaction ($p < .05$) was seen between three question-types and the resources used to teach anatomy, the students' beliefs in being tested with clinically oriented anatomy questions, and the students' beliefs that deep understanding is required to interpret radiology images.

Furthermore, feedback of the students is used to illustrate the quantitative data, and this is discussed with the findings in the next chapter.

Chapter 5 - Discussion

This study is first of its kind to statistically investigate the performance of medical students on the presence and absence of images in clinically-oriented anatomy questions, along with their views and preferences.

In line with cognitive theories of multimedia learning (Sweller et al. 1998; Mayer 2009; Schnotz, Kurschner, 2008), learning is a cognitive process which cannot be directly observed but it can be inferred through the performance on a task. In this study, the students' performance on the test is considered as a representation of their mental models, and it is believed that their scores in the test demonstrate their knowledge of anatomy. Here mental model refers to an internal representation derived from their pre-existing knowledge, along with the ability to integrate internal representation with external text/visuals available in the test.

Although anatomy is integrated throughout the medical curriculum, it is only explicitly taught in years 1 and 2 in most medical schools. This means students in medical schools in the UK are required to cover the learning objectives suggested by the Anatomical Society of Great Britain and Ireland (McHanwell et al., 2007) in those two years, and, implicitly, build the knowledge further in their clinical years. Therefore, there is an expectation of having appropriate mental images to understand anatomical text and images.

Moreover, a variety of visuals are used in anatomy teaching across medical schools in the UK and the world. Owing to authenticity and face validity of a number of anatomical resources (as discussed by Gunderman, 2008; Sugand et al., 2010), this does not only put pressure on students to gradually make mental images of relevant text and visuals but also to be adequately capable to interpret a variety of visuals used in anatomy and clinical settings. For example, they may have mainly used three dimensional cadaveric resources to understand a concept during their formal learning but may be required to exhibit the knowledge through a question based on an X-ray. This transition is somehow expected to occur without standardised formal training. From my experience, this issue is required to be addressed and researched, especially, with regard to our expectation for students to be able to translate grey scale radiology images and cadaveric images, and understand textual information without images. This underlines a pre-training principle of cognitive theory of multimedia learning (Mayer, 2009), which states people learn well from external representations (text and visuals) when they receive pre-training in the names and characteristics of their key elements.

Furthermore, various types of anatomy examination tools are used across the world (Chirculescu et al., 2007; Schubert et al., 2009; Inuwa et al., 2011; Yaqinuddin et al., 2013; Smith and McManus, 2015) and this seems to depend on the feasibility measures, especially in relation to practical anatomy assessments conducted in a dissecting room environment (Rowland et al., 2011; codes of practice by Human Tissue Act, 2004). On the practical side, procurement of cadaveric resources has been a problem for many medical schools all over the world, and it is often related to religious beliefs with regard to cremation or burial of dead bodies (Richardson, 1988; Inuwa et al. 2011). In addition, it is not feasible for all schools to conduct these assessments in pre-clinical years with cadaveric resources because of limitations of time and labour that are essential in assembling and dissembling these examinations.

Considering the advantage of the three-dimensional aspect of these specimens in examinations conducted in a dissecting room environment, where students are able to manoeuvre around the specimens to have a better view, there is a definite loss on the spatial front in less sophisticated online examinations. However, practical examinations are not spared of disadvantages on the spatial front, i.e. mostly prosections (separated body parts) rather than fully dissected cadavers are used in these tests, and thus prosections often fall short in relation to the continuity and the spatial aspects. On the other hand, although cadaveric resources are known to be authentic to facilitate anatomy learning (Smith and Mathias, 2011), it could be argued that they lack face validity because in the clinical years, medical students either perform physical examinations on patients or interpret radiological images. Moreover, conducting these examinations in a dissecting room environment and making the students move to consecutive stations on the sound of a buzzer (as often conducted) may add a considerable extraneous effort (Sweller, 1994; suggested by many colleagues and one of the interviewees in my previous study), especially when assessing more than identification skills.

As the emphasis of this study is on the students' performance in an online test, I will discuss my findings in the light of literature that investigates the effect of images in clinically oriented anatomy assessment conditions. In anatomy, images are known to provide extra/more information than merely complementing the text; therefore, these have been referred as supplementary in this study (Schnotz, 2002). Here interpretation of an image refers to understanding additional information provided through the image along with integrating appropriate schemas to comprehend it (combination of bottom-up and top-down approaches, and cognitive theories).

Answering the research questions

The investigation was carried out to answer the following research questions:

Do students' (2^{nd} year medical students) scores vary depending on whether they receive clinically-oriented anatomy questions with

- no image *or*
- with various images: *either*
  - with anatomical images (cadaveric and clinical findings images) *or*
  - with radiology images (X-ray, CT scans, MRIs)?

Sub-questions:

To investigate whether the following influence their performance:

1. Students' characteristics
2. Students' preferences and experiences
3. Students' participation in anatomy demonstrating activities

High-low performing students and easy-difficult questions

Before discussing the above research questions, it is important to consider the level of difficulty of the questions and the level of competence of the students. Literature has addressed the difference in performance of students depending on their competence and how this varies the level of difficulty of a question (Swanson et al., 2006). During the study design, the level of difficulty of the questions was confirmed by seven anatomy demonstrators (currently working as surgical and radiological trainees and registrars) and one highly experienced surgeon and anatomist. The Angoff method was chosen to define questions' difficulty level, and each member was requested to estimate the proportion of borderline candidates likely to respond to each question correctly. For each question, an average was calculated and 56% was decided as the cut-off score. The questions, which were put through an Angoff process and acquired “56 above” were grouped as easy questions, while “56 and below” were grouped as difficult questions.

For the students’ level of competence, the ones who answered 11-22 questions correctly were grouped as low performing students, and those who answered 23-34 questions correctly were grouped as high performing students (range of questions answered correctly was between 11 and 34).

The results of first ANOVA show the performance of students on easy and difficult questions, and high and low performing students were highly significant with large effect size (Figure 12).

Performance on questions with and without images

With reference to the Angoff method, the three question types (with no images, with anatomical images and with radiological images) had comparable number of easy and difficult questions. In no image question type, there were six questions in each of the easy and difficult categories. In anatomical image question type, there were seven questions in easy category, and five questions in difficult category. In radiology image question type, there were five questions in easy category, and seven questions in difficult category.

The second ANOVA shows that the students’ performance significantly varied on questions with no-images, anatomical images and radiology images; namely, students scored significantly higher on questions with images as compared to no images (as shown in figure 13 and 14 – medium effect size). This is in concordance with cognitive theory of multimedia learning that suggests that people learn better from words and images than from words or images alone (Mayer, 2005, 2009). However, although visuals are known to facilitate learning, there is plenty of literature that emphasise both cognitive benefits and costs of visuals (Crisp and Sweiry, 2006; Berends and Van Lieshout, 2009; Vorstenbosch et al., 2013, 2014). The better performance of students on questions with images may imply their ability to interpret visuals successfully. On the other hand, it may suggest that second year students lack deep understanding of the topic; and therefore, were unable to form appropriate mental models without images and thus did not perform as well on questions without images (as recommended by Mayer 2005; Regehr and Norman, 1996; Sweller 1994). Both of the above factors support Vorstenbosch et al's. (2014) study which proposed that visual options may promote cueing.

Further analysis showed various interactions in the above findings (Figure 15). A highly significant difference in performance was seen in easy and difficult questions. The interaction between question-types (no images versus anatomical image questions, no image versus radiological image questions) and question difficulty (easy versus difficult questions) was highly significant with very large effect size. This suggested a significant difference in easy and difficult no-image questions versus anatomical image questions, and easy and difficult no-image questions versus radiology image questions. The interaction between question-types, question difficulty and high-low performing students was also significant but with small effect size, which means there was a significant difference between high and low performing students; however, this significance was relatively low. Thus other variables such as question difficulty and low-high performing students affected their scores on various question types. Vorstenbosch et al's. (2013) study showed that the use of different types of images within multiple-choice format does not lead to predictable effect but instead may have variable effects on individual items. In their study, fourteen items were more difficult with labelled images as opposed to textual options, while ten items were easier. In a study by Hunt (1978), the effect of radiological images in multiple choice questions was also not consistent; fourty-three items were found difficult with the inclusion of an image, eighteen were easier for the students to answer correctly, and the remaining nine items showed no difference between the two groups. Overall, the students performed badly on questions with original images or radiological images as opposed to those with written description, and it was deemed to be due to the complexity of interpreting a complex task on a grey scale image. However, in Holland et al.'s (2015) study, item analysis of three consecutive years of histology MCQ examinations were analysed and the mean values showed no significant difference in item discrimination or difficulty with and without inclusion of an image.

Although the main focus of the study was to investigate the role of visuals on the students' performance, there are more layers of information present in these questions, which is discussed further.

With regard to distribution of questions with and without images, twelve were based no-images, twelve on anatomical images and twelve on radiological images. Within the image questions, six questions referred to bony structures while the other six referred to soft tissues. This was done because the structures in these images, especially radiological images, are not homogenous i.e. bones appear different to soft tissue. So, along with investigating the performance of students on questions with and without images, their performance on the deep component indicated (bone or soft tissue) in an image, was also investigated.

Moreover, questions were equally distributed with regard to regional anatomy, to blueprint the test with the learning objectives suggested by the Anatomical Society of Great Britain and Ireland (McHanwell et al. 2007); i.e. twelve questions were based on the limbs anatomy (both upper and lower limbs), twelve on the anatomy of the torso and twelve on the head & neck neuroanatomy.

Before further discussing the role of images in these questions, I would like to address the students' feedback collated on the tool. These statements of feedback are used to illustrate the quantitative data.

Students' feedback (qualitative data)

There were a range of views towards the use of images in anatomy questions. See the comments below:

> *"I prefer questions with images, as it makes it much easier to visualise the problem (even if the written description is excellent) and recall the anatomy involved. A good example of this is when talking about facial and trigeminal nerve problems, it is not difficult to imagine areas of the face but having a photograph to look at makes it much easier to map the anatomy from the dissecting room and pictures from textbooks onto the clinical problem posed by the question".*

> *"The advantages of using images that I found was I could imagine it or answer the question more easily as there was a visual cue right in front of me, e.g. muscles of the face. Images had made a difference in the test as I could see something. The thought process of answering the question felt easier as I didn't have to imagine myself but could work with the image that was there, if that makes sense?"*

> *"I greatly preferred the questions with images, as I found them more interactive; they required more thought and application of anatomy knowledge in a practical way. Of the type of images, I preferred the clinical*

findings images and radiographic ones over the cadaveric. This was because they were more akin to the types of images we will face in the future".

Most students preferred questions with images; and, according to their views, images made it much easier for them to visualise a problem and thus answer a question. This is in concordance with studies by Levie and Lentz (1982), Winn (1989), Peeck, (1993), Carney and Levin (2002) which emphasised the role of visuals in simplifying the text and making the abstract more concrete. One student preferred image-questions as these encouraged his/her thought processing i.e. these questions stimulated his/her pre-existing knowledge of anatomy to answer questions with images. Although it is a single student view, this supports the supplementary role of anatomical images to the text (Schnotz, 2002), and combination of bottom-up and top-down approaches (Neisser, 1976) to answer questions with these supplementary images. A number of students preferred radiological and clinical images because these have better face validity and are frequently seen in clinical settings. This highlights students' understanding of the importance of clinically relevant anatomy (Miller, 1990; Bloom, 1956; Sood and Singh, 2012).

However, the inclination towards questions with images was not consistently positive among the students. See the comments below:

"I think text-only questions are good at testing pure academic knowledge that sometimes just needs to be memorised, and so knowing that these will be in an exam encourages you to learn it, rather than relying on the visual prompts of image-based questions".

"I suppose the issue of orientation applies to all image-based questions. Ensuring that the image is sufficiently labelled so that the student can understand the question, without making it too easy to answer. I can't remember if it was this test or other anatomy exams, but I have experienced questions in which more than one answer could be right due to the vagueness of the question or the image used".

"I appreciate its usefulness now that I have done it, but at first I did find the image questions the hardest, as it is a learning style that we haven't been exposed to before, and it can take some time to orientate yourself, especially if, like me, you were previously only used to clear-cut diagrams".

One of the students suggested that text-only questions in a test encourage him/her to understand the topic in depth, rather than relying on the visual stimuli in questions with images. A few students emphasised the importance of pre-existing knowledge to orient and make sense of these images. This supports the notion of acquiring appropriate mental models to deal with questions with and without images (Pollitt and Ahmend 1999; Schnotz, 1993). It also raises the question of whether the images used in examination conditions should be similar to the ones used during teaching or not. Some studies argue that the use of familiar images used in teaching may be reassuring but this may promote a positive cueing effect (Crisp and Sweiry, 2006). This emphasises Vorstenbosch et al's. (2014) findings that text-only questions promote internal visualisation of answers; whereas visual options promote cueing and the ability to interpret visual information. Moreover, in an assessment setting, it may have a wider impact because assessments are known to facilitate future learning patterns (Schuwirth and Van der Vleuten, 2006). However, it is concerning that even publications from the two North American medical licensing institutions give no guidance with regard to the use of images in multiple-choice format (Case and Swanson, 2002; Wood et al., 2004).

One of the students found the image-questions difficult to interpret because he/she was used to clear-cut diagrams and thus it took extra effort for interpreting the images used. This supports the notion of cognitive costs of an image (Schnotz and Bannert, 2003) and also emphasise the importance of combination of top-down and bottom-up approaches (as suggested by Neisser's cyclic approach, 1976) and neuroscientific understanding of visual pathways (Crossman and Neary, 2014). It addresses how an unknown external visual might interfere with or support the translation of an internal depictive representation. It seems particular images like cartoon images may have their value early on in the course to help build concepts; however, for the understanding of concepts in depth, accurate external representations (authentic images with face validity) of anatomy are important.

Performance on questions with anatomical and radiological images

Further analysis was carried out to see if the use of images regardless of their types had a consistent effect on the students' scores. No significant difference was seen in the students' mean scores on questions with anatomical and radiology images in the study. This showed that the type of images does not make a significant difference to the performance of students. This is in concordance with Khalil et al's. (2005), Schubert et al's., (2009) and Inuwa et al's., (2011, 2012) studies that showed no significant differences in mean scores in various imagery strategies for the "immediate recall of anatomical information". However, in Crisp and Sweiry's (2006) study that investigated the effects of various visuals on students' performance and showed that differences in the images significantly affected marks of one question and had smaller effects on marks and the nature of answers with some of the others. Moreover, in Berends and Van Lieshout's (2009) study the presence of images increased item difficulty and slowed down the speed at which students were able to process information.

Considering the authenticity of cadaveric resources (Smith and Mathias, 2011) and face validity of radiology and other clinical images (Dettmer et al., 2013), it is important to incorporate these visual resources appropriately in anatomy examinations to test students understanding of this multifaceted subject.

Students' feedback (qualitative data)

Some of the comments on questions with anatomical and radiological images are shown below:

> *"Radiological images are hard to decipher! In my experience, they seem a lot harder than they actually are (maybe because the test setters are quite kind in the questions they ask at my stage of the medical degree!). However, it is high valuable to be tested on them as they are so relevant to the teaching (and quizzing) we will have in the hospital. I especially like getting radiological questions correct as it makes me think I really understand the material and have properly learned and understood the anatomy to be able to pick it out from a black and white image".*

> *"In regards to radiology, the x-rays I found easier to answer again as I could see what you were pointing to instead of just asking what is at that location. CT images I found confusing but I am sure that may be due to my little experience of seeing them as much as x-rays".*
>
> *"Radiology I think is important but I personally struggle with them mainly because they haven't been taught to us in much detail nor particularly well, at least so far".*

“The radiography ones were in my opinion the best for the reason above and because the different imaging techniques in the different pictures - CT, MRI, X-RAY - that helped me to appreciate the uses of each in different clinical scenarios.”

“Radiological images can be a little abstract to get your head around and so depend on the student’s background. For a class of all medics, their experience can be assumed, but students from other disciplines may not have had the same exposure”.

Besides, most students at this stage in their degree are expected to have seen cadaveric and normal radiological images. The students’ comments above highlight the difficulty of interpreting radiological images. It could be as a result of lack of integration of radiology within anatomy teaching in pre-clinical years in many schools. Therefore, students either depend on radiology-dedicated classes, if available in the formal curriculum, or on self-directed internet search for this information (Dettmer et al., 2013). However, it was good to know that students understand the clinical value of radiographic images; especially as radiology is not a formal part of their pre-clinical years and only taught on an ad hoc basis with other subjects (Dettmer et al., 2013). One of the students expressed excitement when getting a radiological question correct because according to the individual, getting a radiological question correct means one understands the material in depth. A couple of students found CT scans quite challenging to interpret because of lack of exposure to and experience of these scans in their years 1 and 2. It has been pointed out in the literature that the use of medical images at different stages of a medicine degree is important for predictive validity of anatomy (Lufler et al 2012).

Along with radiology images, images of clinical findings were found useful. See the comment below:

> *"Photographs of clinical findings are very useful. The test used excellent, clear photographs, which left no doubt about the sign/ condition that was being questioned".*

Like radiological and clinical findings images, cadaveric images were also valued by many students but they found them difficult to orient in order to answer a question correctly. This may be as a consequence of the time lapse since they had used cadaveric material for learning and/or done any dissections themselves; some did confess their lack of knowledge to be the reason for not being able to answer questions with cadaveric images correctly. One student noted that cadaveric images require a much deeper understanding in order to identify structures. Another expressed that cadaveric image questions are good at testing the direct physical appearance of anatomical structures, as they appear to the naked eye. Similar points were raised by other students stating that it was difficult to understand three dimensional structures from pictures of prosections.

See the students' comments on cadaveric images below:

> *"Cadaveric images can be a bit harder to figure out, especially as some time had passed since I had last used cadaveric material to study from/ done any dissecting myself up the time I sat the test. However, I thought*

that the cadaveric images used in the test were excellent and added a great deal of clarity to the questions; they were more useful that any clinical photographs would have been".

"I think cadaveric images are a good choice because they require a much deeper understanding in order to identify structures. The only issue was sometimes it is hard to see where the arrow is pointing, especially in poorly lit images. Stil,l I think cadavers show much better applied knowledge than simplified diagrams".

"The cadaveric ones were still useful in testing anatomy knowledge, but sometimes I found them more challenging as I had difficulty identifying structures - this possibly is saying more about the lack of my anatomy knowledge in certain areas than of the usefulness of the question".

However, a general issue with cadaveric image appearance and quality was noticed in the students' comments. It seems that making sense of three dimensional objects in two dimensional images requires effort and this transition is not as smooth as it is assumed by the experts in the field. Moreover, lack of contrast in cadaveric images as compared to cartoon images seem to affect the capacity of students to deal with these questions (see comment below).

"I think the issue with cadaveric images in an exam is always going to be the contrast. Even in the dissection room with a specimen, despite having a good knowledge of anatomy, it can be difficult to distinguish very similar looking structures due to the generalised brownish/grey appearance of almost everything. I can imagine students from institutions which do not

have cadaveric learning would find these questions difficult and unfamiliar, and so would be at a disadvantage if the same exam was used to test students from across multiple institutions".

This issue will need further research to understand it and suggest ways to resolve it. Like other images, there was a common theme of having appropriate mental models to understand various types of images used in anatomy examinations.

Some specific comments were also made on having sufficient previous knowledge to interpret these images:

"I absolutely hated the questions with the layers of muscles/fascia but again, mainly because we haven't been taught layers that well or types of fascia. Again, I think they're very good questions for understanding though."

"Having a grasp of all these methods really aids in producing the 3D mental imagery required to completely understand anatomy. It also allows you to understand the anatomical relationships with neighbouring structures".

Further to the above comment, it was insightful to see students' understanding between their learning processes and assessments.

Performance on questions referring to bones and soft tissue

The next level of analysis was carried out to investigate if a deep component (bones and soft tissue) indicated on anatomical and radiology images have any effect on the students' scores. As explained in the material and method section, these categories were developed because anatomical and radiology images are not homogenous images i.e. bones appear different to soft tissue in these images. This is especially true with radiological image modalities because in these images, rays or sound waves get absorbed and/or reflected back differently from bones and soft tissue and thus they appear brighter or darker depending upon the density of the structure.

It was found that the structures (bones and soft-tissue) indicated in anatomical and radiology images have a highly significant effect on the students' performance with large effect size (shown in figure 16). Moreover, a significant interaction between question types (anatomical and radiological images), question subtypes (referring bones and soft-tissues) and high-low performing students was seen. It was found that both high and low performing groups performed better on image-questions indicating bones as compared to soft-tissue (as shown in figure 16 and table 14), but the significance was relatively low with small effect size. This may be as a result of inadequacy to process relatively more layers of information to answer image questions indicating soft-tissues as opposed to bones. Layers of information refer to the detail that images with soft-tissue often have as a consequence of the combination of inter-related structures; muscles, nerves, arteries, veins and bones are all present in an image. Furthermore, with regard to cadaveric images, these usually focus on a part of body (separated body part), which requires a number of processes; such as orientation of an image, followed by making sense of all neighbouring structures and the spatial understanding of these structures. This seems to get relatively more complex in images with structures of different densities as opposed to images with only bones.

Building on Alternative Multimedia Learning Theory

According to alternative multimedia theory, along with cognitive benefits (Mayer, 2009), visuals also have cognitive costs (Schnotz and Bannert, 2003; Schnotz and Kurschner, 2008; Schnotz and Baadte, 2015). In line with the theory, interpreting an image is a matter of complex interactions between a numbers of factors; i.e. perceptual surface structure, semantic deep structure, image familiarity, sufficient internal representations, and orchestration between existing mental models with external representation (the combination of text and image in this study).

Interestingly, the above part of the analysis adds another element to the theory of alternative multimedia learning, relating to assessments with supplementary images (such as anatomy assessments). Although the students' scores did not differ significantly between questions with anatomical images or radiology images in this study, the deep component of non-homogeneous structures (bones or soft-tissue) within these images has made a significant difference to the students' scores. Thus, significant interactions were seen between question types, question difficulty and high-low performing students, and image questions, question subtypes and high-low performing students. Consequently, along with the use of valid and authentic images in SBA-type assessments, it is important to have a spread of questions indicating deep components (both bones and soft-tissues) within various types of images for construct and content validity of a test instrument involving supplementary images.

Furthermore, it can be seen that low-performing students seem to find it difficult to interpret soft-tissue (layers of information) as opposed to high performing students in questions with anatomical and radiological images. It may be because relatively better schematic networks are required to orient and spatially understand an image with layers of information (as in soft tissue images) as opposed to images representing one structure (bony images) in the image questions used in the study.

Performance on questions based on different regions of anatomy (limbs, torso, head & neck neuroanatomy)

This investigation was done to explore if regional anatomy makes a difference to students' performance on questions with and without various images. A highly significant difference in the students' performance on regional anatomy questions was observed i.e. the students performed better on limb anatomy questions as opposed to torso anatomy questions and better than on head neck neuroanatomy questions with a large effect size (see figure 17). Further, the interaction between question types and regional anatomy was significant; i.e. in limbs anatomy based questions, performance on anatomical and radiology images was significantly better than in no-image questions, and in head and neck neuroanatomy anatomy questions, performance on radiology images was significantly better than with no images. This supports the cognitive theory of multimedia learning (Mayer, 2009). Interestingly, the above theory was not consistent with the performance on torso anatomy questions. In these questions, performance on no-image questions was found to be significantly better than those with anatomical images. In torso questions, anatomical images seemed to interfere and had a negative impact on the students' performance, as opposed to no images, anatomical and radiology image questions.

Low performance on anatomical images in torso and head neck neuroanatomy questions may be because typical cadaveric images are zoomed-in images of dissected parts, which occasionally makes them difficult for orienting and interpreting. Although prosections are used in many medical schools, there is an inclination towards online assessments as more convenient than the practical examinations; so, two-dimensional (2D) images of these three-dimensional (3D) prosections are used in assessments. This highlights the issue of translating 3D resource information to 2D images (Miller, 2000). The importance of spatial skills in teaching and learning anatomical knowledge has been raised in the literature (Hegarty et al., 2009), though mainly in dental studies.

In the questionnaire study, a relatively high percentage of students disagreed that cadaveric photographs are effective for examinations, which raises the issues about interpreting 2D structures (cadaveric photographs) when 3D structures (cadaveric material) are used during teaching. However, several studies have shown no significant difference in the students' performance on factual questions that test the recall of knowledge, with various 2D and 3D visuals resources (Khalil et al., 2005; Schubert et al., 2009; Inuwa et al., 2011, 2012). These studies compared students' performance on 3D and 2D image strategies on factually based questions. Although the assessment tool in this study was online and no such comparison was made between 2D online and 3D cadaveric resources in practical set-up, this transition of learning from 3D resources to interpreting 2D images requires further empirical research.

See comments below:

> *"This experience has highlighted how little anatomy is taught at my medical school and how, when presented with an image of a cadaver we are stumped. Anatomy at my medical school is primarily taught with coloured images, models and living anatomy. When the colour is taken away and we are presented with surgical or cadaveric dissection images we are left at a loss as to how to identify structures".*

“very difficult to understand 3D structures from pictures of prosections”.

This resonates with the notion of cognitive costs of images; especially so in these supplementary images (Schnotz and Bannert, 2003). This reinforces the assumption that images do not always have a positive impact on the performance of students: their inclusion may cause interference, and the degree and quality of impact is dependent on other factors like competence of the students, image types, image subtypes, and regional anatomy (as internal validity criteria suggested by Campbell and Stanley, 1963). Moreover, it is likely to be a difficult task to interpret cadaveric images for students who have not been exposed to these resources during formal anatomy teaching because of lack of pre-existing schemas required to facilitate their understanding. This applies to a number of medical schools which rely on coloured cartoon diagrams instead of cadaveric specimens. Furthermore, for those who have used these resources for learning, it is quite possible they do not have enough mental representations to be able to translate 3D information in order to interpreting 2D images.

Students’ feedback (qualitative data)

Formal teaching and knowledge gained play an important part in the performance of students. In the feedback section, one of the students wrote about the lack of cadaveric anatomy taught in his/her medical school, which made it difficult to make sense of a cadaveric image. Although it is only one student’s view, this gap has been identified in the literature, explaining poor performance in anatomy examinations, and thus affecting predictive validity of anatomy (Drake et al., 2009; Gogalniceanu et al., 2009; Standring and Larvin, 2011). Standardisation for assessing anatomy has been suggested as a plausible solution in the literature

(Rowland et al., 2011) and it may also be a way to subsequently ensure standardisation of teaching of anatomy across the board. This question on anatomy examination standardization was put to one of my interviewees (a senior anatomist) during my previous research (MoE2). Unexpectedly, the standardisation of anatomy assessments was not supported by the senior anatomist, and this was justified on the basis of difference in needs and limitations of each institution. However, she acknowledged that the existing assessment tools are not powerful enough to prepare students for future practice, so the need for a standard curriculum instead was suggested.

As a practitioner, I appreciate the limitations in the area, especially with the accessibility of cadaveric resources; however, as we are all working towards building a good foundation for our future doctors - I believe we require some standardized guidelines for testing their knowledge so they are able to cope with the system. Moreover, as assessments are known to facilitate students' learning patterns (Schuwirth and Van der Vleuten, 2006; Larsen et al., 2008, 2009), inconsistency in the assessment system could lead to variation in the competence of future doctors. Older (2004) and Rowland et al., (2011), showed concerns about varying anatomy assessment systems in UK medical schools and recommended that any significant change should ensure development through validated standardised programmes. These views regarding standardisation of anatomy examinations are in concordance with the views of another academic whom I interviewed in my MoE2 research.

Along with radiology discordance, the comment below also emphasises the lack of clarity in learning objectives amongst institutions' learning objectives (Rowland et al., 2011 and Older 2004), and the resulting variations in teaching and assessment styles across the institutions. In views of Older (2004), the changes in the curriculum of undergraduate medical schools in the UK have been implemented without any rigorous research and agreement of national core curriculum, and this has affected the proficiency of future doctors. According to an interviewee (a senior anatomist) in my MoE2 study, the reduced share of anatomy in the curriculum is a practical compromise and a real test to deliver the best in their share of time. Therefore, there are issues around consistency of the anatomy curriculum in the UK, which along with experts are also recognised by students too, along with experts.

See comment below:

> *"It was a very helpful tool for anatomy revision. However, as it was not geared towards the anatomy syllabus of one particular university, naturally there were some parts that I wasn't familiar with. In my case I wasn't used to the radiology images and found those more challenging to tackle. All in all, it was great fun to see the clinical side of anatomy, would definitely love to have more of these questions/quizzes!"*

Clinically-oriented questions

Although discussion on clinically relevant anatomy questions is not the focus of the study, I would like to address the inclination towards clinically relevant teaching and assessment in the literature as a consequence of evolution of the medical field (Smith and McManus, 2015; Yaqinuddin et al. 2013). In the assessment literature, these types of tests are habitually characterised according to Miller's pyramid or Bloom's taxonomy (Miller, 1990; Bloom, 1956) in order to focus on propositional as well as functional knowledge. Moreover, there have been a number of studies in support of clinically oriented anatomy teaching and demanding retrieval of knowledge in contextually rich assessments (Sood and Singh, 2012; Yaqinuddin et al., 2013; Smith and McManus, 2015; Ikah DSK et al., 2015).

In the study, the students also expressed their views on the clinically-oriented anatomy test. These data are presented to show whether their views are in agreement or disagreement with experts' views available in the literature. Unanimously, students valued the importance of clinically relevant anatomy questions to test their integrated knowledge of anatomy.

Some of the comments are as follows:

> *"Really an excellent test. What made it better than most examinations of medical knowledge was that sort of 'extra step' you had in many questions. For example, instead of asking simply what innervated the upper larynx, you asked what might cause a cough reflex there. We learn so much of our course through text that when I get to a question about, say, the lumbar puncture layers, I'm made to look deep into my knowledge of the structure and use many of those text based facts I know to answer the single question. Standard examination questions often do not do this and rather rely on us to just remember single sentence obscurities from lectures to assess our depth of knowledge. Thank you very much and I hope my results are useful!"*

> *"Clinically relevant questions that made you think logically about multiple levels of anatomy and anatomical functions to form an answer. Online platform worked well".*

"It was quite a hard exam. It's made me realise how I focus a lot more on learning the anatomy itself than the clinical aspects. I tend to learn them separately; however I now think it would be a better style of learning if I learnt them together. A very useful exam, nonetheless. Easy to understand and use".

"The questions provided a good practical application of anatomy. However, during our teaching the practical side has not been emphasised as much as opposed to learning the theory; hence making the 'jump' was something quite difficult - especially as we are taught with some radiology images but not many. This left me being unable to work out which side of the body was shown or which ligament etc. although I knew the knowledge".

"The clinical setting of the questions was particularly good because it gave meaning to the questions as opposed to the feeling of "random anatomy fact checking" that I get with some questions and exams".

The comments above emphasised the consensus between students and experts about anatomical knowledge being delivered and tested in a clinical context. However, some found this transition overly challenging, which may be as a consequence of a lack of adequate mental models and/or relatively less exposure to clinical and applied anatomy.

With regard to making these questions clinically relevant (as recommended by Yaqinuddin 2013; Smith and Mathias, 2015), it is important to understand the issues described in the literature with regard to adding clinical vignettes, especially in the practical set-ups (examinations conducted in a dissecting room environment). In my Methods of Enquiry (MoE2) essay, one of the interviewees (a senior academic) expressed concern about having clinical scenarios/vignettes in practical examinations. He coined it as a “clinical conundrum” that could disadvantage students’ performance in a practical set-up. As there are a couple of extra steps involved in these examinations as opposed to online ones, i.e. students moving to the subsequent stations on the sound of the buzzer, and stabilising themselves on each station prior to attempting each question - this may have a negative impact on the students’ performance. Thus, the interviewee preferred the clinical relevance in teaching anatomy; however, he believed the integrated knowledge of applied anatomy should only be assessed where students can sit comfortably and concentrate. Considering the affirmed importance of teaching and assessing applied knowledge of anatomy, and limitations of administering it in a practical set-up, various online tests are used throughout the world (Karay et al., 2012; Orsbon et al., 2014).

In medical schools with no policies on voluntary body donations, lack of cadaveric resources and/or lack of manpower to keep these resources viable; especially in some countries where body donation is not fully accepted in the culture, there is an obvious apprehension about it (Orsbon et al., 2014; Inuwa et al., 2011, 2012), and thus the inclination is towards online resources for teaching and assessing anatomy. Besides, for those who are more inclined to test applied knowledge of anatomy along with factual knowledge rather than testing these components separately, an online platform with appropriate use of authentic and valid images may be an option. Additionally, the online resources could prove useful in providing immediate constructive feedback to the students (Krippendorf et al., 2008). These are less labour-intensive as opposed to practical examinations, and also relatively easy to administer, mark and analyse (Morris and Chirculescu, 2007). However, online examinations lose the authenticity of involving real cadaveric resources, which may be important for testing their topographical anatomical knowledge in pre-clinical years at a stage when students are still learning to make mental models of three dimensional structures. These online systems are more likely to be useful for students who have some mental models, and are able to translate three-dimensional knowledge gained in order to read two-dimensional images.

Role of feedback in the system

Providing feedback to the students was not the focus of the study but it was added to facilitate their future learning, and to show gratitude to the students who voluntarily participated in the test. The comments received from the students are consistent with the interactivity principle of cognitive theory of multimedia learning (CTML) theory, which suggests that interactivity, for example, learners' control, guidance and feedback improve learning transfer and performance (Mayer, 2009).

See participants' comments below:

> *"The most useful part of the quiz was definitely the feedback section at the end. Having such a detailed explanation of not only the reasons for which the right answer is correct, but also explaining the anatomy relevant to the other options has really helped me out in revision and also made me look at areas I might otherwise have not thought to study".*

> *"To improve the test, some of the explanations could have been expanded to include why some of the other answers were not correct. I think (but can't be sure!) that this was done for some questions but not all. For example, a question about a symptom experienced in the hand, related to a nerve in the arm: if the answer includes radial, ulnar, and*

median nerves, it might be useful to have a brief overview of the symptoms associated with each nerve. When I make an attempt at a question and I get it wrong, it's good to know the true answer to that question, but it's also good to know why I thought my answer was correct, and refresh my memory there and then of the correct question (/correct symptom) for my answer, if that makes any sense!"

"Furthermore, I liked the clinical context of many of the questions. Even if I didn't get them right the first time, I learnt a lot from the feedback at the end".

Thus students found it extremely important to know why an incorrect answer was incorrect along with the explanation for an answer they answered correctly. They also commented on the usefulness of immediate and detailed feedback. Feedback is an important element for continuous learning (Larsen et al., 2008), and although assessment is known to facilitate learning, feedback essentially is known as a tangible component and a scaffold to drive future learning (Black and Wiliam, 1998a).

Discussing students' performance on the test with regard to the questionnaire variables

Analysis of variance (ANOVA) was conducted using the questionnaire's elements/questions as covariates to investigate if these factors have any relationship with the students' performance on the test.

A significant interaction was found between the students' performance and the following variables:

- All anatomical resources used to teach anatomy

- Agreed that finding hands-on cadaveric dissection an effective way of learning
- Agreed that finding radiological images an effective way of learning anatomy
- Disagreed that cadaveric photographs are more effective for examinations than prosections and dissected material
- Disagreed that their main motivation for learning anatomy is to pass the examination
- Agreed that deep understanding is required to answer questions
  - with prosections and dissected material
  - with cadaveric photographs

These variables are discussed in the next section.

<u>Anatomical resources used to teach anatomy</u>

Analysis revealed a majority of students (62.6%) used all three types of anatomical resources (dissection of human cadavers, prosections and radiology images), and there was a significant difference in the scores of students between groups; i.e. students who have been taught anatomy with “all the above resources” in their schools performed significantly better than those who were taught with “prosections and radiology images” ($p < .05$). The literature has emphasised the importance of various resources in teaching anatomy (McHanwell et al., 2007).
Smith and Mathias (2011) have suggested the importance of visualization of three-dimensional anatomy for students and clinicians. Moreover, with the advent of radiology in the field of medicine, there is much emphasis on images seen in clinical settings (radiology images and clinical findings) (Phillips et al., 2013; Dettmer et al., 2013).

Learning preferences

Learning preference concerns the representation of information one prefers during learning (Vorstenbosch et al. 2014). Some learners prefer reading texts, others favour listening to an explanation, and some choose visual information from images. Learning preference influences the cognitive processes used during examination (Mayer and Massa, 2003; Leite et al., 2010). There are a number of studies on learning preferences and styles. Studies by Fleming et al. (2011) and Wilkinson et al. (2014) do not support a correlation between learning styles and students' performance. Mahony et al. (2016) reported that individual learning styles contribute little to academic performance. However, as this is not the focus of the study, I have not discussed the literature on learning styles. A few anatomy-related questions on individual learning preferences were added in the questionnaire to investigate if their preferences in learning (hands-on cadaveric dissection, prosections, cadaveric photographs and/or radiological images) affect their overall performance on the test. As "not done", "not used" options were included considering not all students have experienced all the above resources, these results were relatively weak. Out of all the variables, the ones that made a significant difference ($p < .05$) to the students' scores on the tests were those where students "agreed" that hands-on cadaveric dissection was more effective way of learning than those where they stated to have "not done cadaveric dissection". Those who "agreed" that radiological images are an effective way of learning anatomy scored significantly more than those who "disagreed". This finding is in agreement with the literature that emphasise hands-on cadaveric dissection as an authentic skill for learning anatomy (Smith and Mathias, 2010). Radiology images are widely known to enhance quality and efficiency of anatomy teaching because of their relatively higher face validity in clinical settings (Grignon et al., 2016).

With regard to photographic representation of cadaveric material, it is important to conduct further research in the area because these do not seem to serve a similar purpose as cadaveric material because of their 2D representation as compared to 3D cadaveric material (Hegarty et al., 2009). Prosections have been considered as a good learning resource possibly because these provide three-dimensional knowledge with relatively less hassle (dissected specimens) as opposed to whole cadavers. These resources are frequently available in anatomy dissecting rooms for revision sessions conducted out of scheduled teaching hours and thus more convenient for access as opposed to the dissection of cadavers.

Assessment preferences

The assessment preferences of the students were considered in the context of investigating whether their preferences in examinations (via cadaveric photographs, prosections and dissected material or radiological images) affect their overall performance on the test. Out of the variables, the ones that made a significant difference to students' scores on the test were the ones showing that those who "disagreed" that cadaveric photographs are an effective resource for examination scored better than those who were "unsure". The dissected material and their images seem to play different roles according to the students' views. It may be that two-dimensional photographs of the cadaveric material make it difficult to orient and interpret and/or students lack appropriate mental models to translate the information (Miller, 2000). It reiterates that appropriate training to progress through this transition of making sense of three-dimensionality of structures through two-dimensional images is important to consider in the curriculum.

Assessment as main motivator

According to Smith and Mathias (2010) deep approaches to learning are especially encouraged by assessment methods and teaching practices which aim at deep learning and conceptual understanding. As assessment is often known as the main motivation for learning for strategic learners (Standring and Larvin, 2011; Reid et al., 2007; Black and Wiliam, 1998a), this question was added to investigate the students' views. It has been known that students take different approaches to learning: deep (understanding material), surface (memorising details), and strategic (motivated by assessments) (Smith and Mathias, 2010). Considering the demands of the medical curriculum, students often learn strategically, and the assessment system works as a key component in facilitating their learning (Moxham et al., 2011).

Students often learn about the importance of a subject from their seniors and mentors, and there is ample literature that emphasises the importance of anatomy in medicine and the disastrous consequences of lack of anatomical knowledge in medicine and surgery (Rainsbury et al., 2007; Standring and Larvin, 2011; Cooper and Gray, 2014). Interestingly, further investigation showed that the groups of students who "disagreed" that their motivation to learn anatomy is to pass the examination scored significantly higher than those who "agreed". It was reassuring to see that, although a slightly higher percentage of students agreed that their main motivation for learning anatomy is to pass the examination, the students who performed significantly better were those who disagreed. This emphasises that not all high performing students are strategic learners.

Visual resources and deep understanding

Furthermore, the students' views on relationships between the use of particular resources, (prosections and dissected material, cadaveric photographs or radiological images), and deep understanding to answer a question correctly were acquired. Out of the variables, the group that "agreed" that deep understanding to answer questions with prosections and dissected material scored significantly higher than the "disagree" group. Moreover, the group "agreeing" that deep understanding is required to answer questions with cadaveric photographs scored significantly higher than those who "disagreed", and those who are "unsure" scored higher than those who "disagreed". These findings highlight the need for research in the area of learning from 3D and 2D resources, and translating 3D resource information to 2D images (Miller, 2000).

Other variables with no significant relation to students' performance

The questionnaire elements which did not make a significant difference to the students' scores are discussed below:

Time taken to complete the test

Although this is not the focus of the study, the time taken to attempt each question by each student was collected. This feature was built in the software used. These data were analysed to see how much time students took to complete 36 applied anatomy questions. In practical tests often 1 minute (Shaibah and Van der Vleuten, 2013; Smith and McManus, 2015) or 1.5 minute (Khalil et al. 2005; Yaqinuddin et al., 2013) is assumed to be sufficient to attempt each question or station. In the study, although 1 hour 30 minutes were given, it was thought that most students would finish the test in 36 minutes. The tool had a participant information sheet, consent form and questionnaire prior to the actual test; however, and the clock started only on their confirmation to start the actual test. The average time required to complete the test was 30 minutes 17 seconds with a standard deviation of 11 minutes 58 seconds. The minimum time a student took to complete the test was 11 minutes and 28 seconds. The maximum time taken was 1 hour 17 minutes and 56 seconds. Considering the variation in the time taken by the group to complete the test, it is important to note that allocating a set time of 1 or 1.5 minutes to each question does defeat the assumption that individuals' problem solving processes are idiosyncratic. Students may need more or less time to deal with a question, depending on their level of competence and their pace of reading and trying to solve a problem. Zhang et al., (2013) showed that there is no significant difference between a timed and untimed steeplechase examination, and Smith and McManus (2015) suggested the timing as more of a practical consideration than a cognitive one. The time taken by an individual to solve a problem could be considered as a cognitive issue. People's ability and the time needed to solve a problem are idiosyncratic (Sweller, 1994). Therefore, the administration of practical examinations that do not provide the flexibility to move back and forth between questions may put some students at a disadvantage. Unlike practical examinations, online examinations can be easily designed to provide the flexibility to go back and forth, re-think, and students can also take longer or shorter time on a question as per their requirement (Paalman 2000). This is in concordance with the interactivity principle of CTML theory (Mayer, 2009). The correlation between total time taken and total scores was not significant, therefore it was not analysed further.

Medical Schools

This variable was added because medical schools in the UK utilise either all or some combinations of available anatomical resources; i.e. dissections (dissecting cadavers), prosections (pre-dissected body parts) and radiological images

(Rowland et al, 2011). Out of six schools included in the study, three used prosections and radiological images, two used radiological images only, and one had dissections, prosections and radiological images for teaching anatomy. The data distribution was inconsistent because a high percentage of participants belonged to one medical school, and thus further analysis was not carried out.

Sex of the students

This analysis was carried out mainly because in my previous study (Institutional focused study), a positive trend in relation to sex was seen. In the study, the adjusted mean difference between male and female students' scores in the anatomy practical examination was 4.4 marks, with the mean females mark being higher (Sagoo et al., 2016). The students' sex having an effect on anatomical learning has been studied by Hisley et al. (2008) in a study comparing dissection with digital software based dissection, and similar findings were noticed; i.e. females performed significantly better than male students. However, in the present study, no significant difference was found in the performance of male and female students on the test.

Age Range

In the study, age was included to investigate if it had any effect on student performance. It was assumed that graduate students (relatively older students) would attain higher marks because they are likely to have a better idea of how to learn. However, in this study, further analysis was not conducted because the distribution of students in different age groups was highly inconsistent with a majority of students between age range19-21.

Students' Training Level

Most medical schools enrol students on their MBBS course via two streams; an undergraduate stream of students (school-leavers) and a graduate stream of students (who already have a graduate degree in science or non-science background). These students are required to achieve high marks in UKCAS or GAMSAT tests to enter the MBBS course. In this study, as a high percentage of students were undergraduate, this variable was not investigated further.

Most Likely Prospective Career

As anatomy is considered as the foundation subject for medicine and surgery (Gunderman and Wilson, 2005; Sugand et al., 2010; Benninger et al., 2014), it was thought that students who are interested in "surgical field" as their most likely prospective career may perform better in the test. However, 40.8% of students selected the "don't know" option followed by "non-surgical" and then "surgical" option for their most likely prospective career. As they are in early stages of their career, it is understandable that they are still exploring the disciplines and may not have made a prospective choice of the field they would like to gain expertise in. Further analysis was not conducted because the data were inconsistently distributed.

Clinically Relevant Anatomy Learning and Testing

Questions were included to see how students perceive clinically relevant anatomy (applied, living and surface anatomy) for learning and assessment. As anticipated, a high percentage of students were in favour of clinically relevant anatomy. This is in line with experts' views on integration of anatomy with clinical disciplines (McHanwell et al., 2007; Moxham et al., 2011). It was found not only that the students' preferred clinically relevant anatomy learning, but 83.9% would like their anatomy knowledge to be tested with clinical knowledge. Further analysis found a significant difference between groups that found clinically relevant anatomy learning an effective way of learning anatomy; however, the distinction did not had a clear significance on performance. The analysis of the relationship between the students’ scores and beliefs that anatomy knowledge should be tested with clinical knowledge, was not statistically significant.

Assessment Techniques

Along with the students’ preferences on the style of the assessment, their preferences on the technique to deliver these type of examinations was also investigated. It is known from the literature that there is no consensus on a preferred approach to anatomy assessments; therefore, many universities follow traditions and rely on convenience and beliefs with regard to the use of anatomical resources in assessments (Rowland et al, 2011). The questionnaire data gathered from the study resonates with the assessment issues highlighted in the literature. In the study, 39.7% students had never taken an online examination and 22.4% students had not taken a practical examination. The rest of the data were inadequate to investigate any further.

Participation in an Anatomy Demonstrating Programme

In the literature, anatomy demonstrating is defined as the ability of peer tutors and tutees to communicate effectively, to enhance the learning of anatomy, and it is recognised as a valuable approach for learning (Youdas et al., 2008; GMC, 2009). In some cases, tutoring involves experienced students at more advanced stages of their training acting as tutors and hence the term “near-peer teaching” is used in such context (Bulte et al., 2007). “Peer teaching” is conducted usually by peers of similar age or level of learning, and who are therefore relatively inexperienced. In anatomy, near-peer teaching and peer teaching are widely used, and it is probably best exemplified by the anatomy demonstrators employed extensively across medical schools in the UK (Houwink et al., 2004). To see any relationship between the students’ performance and their anatomy demonstrating experience, two questions were included in the questionnaire. In the present study, a very high percentage of students had never been involved in demonstrating anatomy formally. This may be because very few schools give a formal opportunity to the senior students to teach their juniors. Furthermore, another question was included to find out if informal anatomy demonstrating helped their anatomy learning. More than half of them agreed that demonstrating informally helped them to learn anatomy. This opens up a wide area of research on what role these formal (near-peer teaching) and informal (peer teaching) demonstrations play in teaching, learning and mentorship; however, it is not in the scope of this study to discuss it. Further analysis was not conducted as most of the students never taught anatomy formally. No significant difference was seen in students' the performance of those who informally taught anatomy.

Discussing students' performance on question-types in the test with regard to the questionnaire variables

Further ANOVA was carried out to investigate the students' performance on questions with (anatomical and radiology images) and without images in relation to the questionnaire variables (as covariates). It was found that resources used to teach anatomy made a significant difference ($p < .05$) to the students' performance with small effect size i.e. the students performed better on questions with anatomical images than no images, and radiology images than no images. Moreover, the students' views on being tested with clinically oriented anatomy questions made a significant difference to their performance with small effect size i.e. students performed better on questions with anatomical images than no images, and anatomical images rather than radiology images. Furthermore, students who believed that it requires deep understanding to interpret radiology images had a significant difference to their performance with small effect size; i.e. students performed better on questions with anatomical images than radiology images. This is in concordance with the cognitive theory of multimedia learning

(Mayer, 2009) as well as the alternative theory of multimedia learning that refers to the notion of both benefits and costs of different types of images (Schnotz and Kurschner, 2008; Schnotz and Baadte, 2015).

Hence, the findings suggest that the students' performance on clinically-oriented anatomy questions with and without images is dependent on an intricate network of factors; including external and internal representations, the level of the students' competence, and the deep component of an image, question difficulty, the context of the question (regional anatomy) and individual preferences.

Limitations of the study

Some of the similar studies have used two separate groups as control and tests for assessing the students' performance (Holland et al., 2015; Crisp and Sweiry, 2006; Inuwa et al., 2011, 2012). In this case, it was not possible to assess one group with questions with images, and the other group with the same questions without images. This is because of the limitation of having only a single contact window with medical students of six medical schools and the responsibility of providing consistent revision material (test) to all the participants. However, to avoid any pitfalls, the study was designed in a particular way to have the same group of students acting as a control group while attempting questions with only textual material, and as a test group while attempting questions with textual material and images.

As it was a multi-institutional study, it was not feasible to know if the students had come across the same images and text (used in the test) during their teaching which may have had an impact on their results because of a cueing effect. The images and text used were within the context of the objectives suggested by the Anatomical Society of Great Britain and Ireland (McHanwell et al., 2007). There was a possibility of a cueing effect if they had previously seen the same visual and textual material. As various types of visuals are used for teaching anatomy, in order to see relationship between performance and visuals used during learning, the questionnaire was involved in this study to investigate whether use of certain types of visuals in teaching has any impact on participants' performance. As most students have been taught anatomy with "all the above resources" in their schools, this variable was not tested entirely. Nevertheless, it may had been technically problematic to carry out such an analysis because of differences in student cohorts, teaching styles and use of visuals in different medical schools.

Furthermore, this study was designed to only assess cognitive skills and not psychomotor and affective skills. Testing other components would require integrating this type of assessment with skills-tests through applied anatomy Objective Structured Practical Examination (OSCEs) stations.

A number of other issues such as unclear wording of some questions, unclear images, issues with loading images on the feedback page, and issues with smoothly running the tool on private browsers etc. were recognised at the piloting stage, and these were all corrected before releasing the tool for collecting the data for this study.

<u>Further investigation</u>

For the future study I am keen to explore further in the area of visuals; mainly cross-sectional images (anatomical and radiological). I have recently come across very interesting literature on the role of spatial skills in making mental models of these images in anatomy learning and assessments. In the literature there have been debates in the area of alternative models of individual differences in spatial performance. It has been argued as an "innate ability" by some (Curtis et al, 2007;

Sandow et al., 2002) and a “skill” by others which assumes that skill acquisition is essentially a matter of practice (Gawande, 2002). It is also considered possible that spatial ability predisposes one to do well in medical training while being developed further during this training, and it depends on the level of expertise (Ackerman, 1988). Currently, spatial skills testing is part of the selection process for students on dentistry courses in North America (Ranney et al., 2005). Like dentistry, anatomy is a multifaceted hands-on subject, and medical students learn anatomy through dissections done during classes, dissected prosections, living anatomy sessions and relevant radiology stations set up in the dissecting room. Therefore, I am keen to invest in the area by designing a new study to investigate whether interpreting the appearance of a cross-sectional image of an unfamiliar anatomical object depends on spatial ability. Furthermore, do spatial skills enhance learning of anatomy?

Chapter 6 - Conclusion

The initial analysis showed a significant difference in the students' performance on clinically oriented anatomy questions with and without images; however, no significant difference in performance was seen between the questions with anatomical and radiological images. The performance on internal representations integrated with authentic and valid images was better than the performance on only internal representations in clinically oriented anatomy text. Although this study was based on an assessment of learning, and not-learning, the initial findings are in concordance with the central idea of the cognitive theory of multimedia learning that text and images together are better than text or images alone (Mayer 2009). However, these findings are dependent on a number of factors such as the level of question-difficulty, high and low performing students, regional anatomy and interactions between these factors.

As anatomical and radiological images form a crucial component in multifaceted anatomy, and are intrinsic and have built-in (supplementary) meaning, students are required to have appropriate mental models to interpret these images in a context, i.e. one needs the ability to integrate internal representations with external representations to comprehend these images. Without appropriate internal representation, these images may increase the difficulty of a question. Further analysis showed the deep component of an image (indicating bones or soft-tissues) in anatomical and radiology images affects the performance of students significantly i.e. they performed better on questions referring to bones than soft-tissues. This supports the concept of cognitive benefits and costs of supplementary images (Schnotz and Bannert, 2003). This finding on difference in performance on questions referring to bone or soft-tissue regardless of the image-type can be employed to further develop Alternative Multimedia Learning Theory for assessments involving supplementary images; that is, along with surface and deep structure of an image, the deep components within non-homogeneous supplementary images makes a significant difference to the students' scores, especially in a subject like human anatomy. Therefore, along with the use of valid and authentic images in assessments, it is important to assess students on deep components of these images for construct and content validity of the test instrument.

This reinforces the conclusion that images do not have a consistent impact on students' performance. It depends on a more intricate network of interpretation of perceptual surface structure and semantic deep structure of an image, image familiarity, sufficient internal representations, orchestration between existing mental models with external representation, students' level of competence, question-difficulty, and deep components indicated on an image, and the subject area/context (Campbell and Stanley, 1963; Mayer, 2009).

Academic and professional contribution

This study supports the current teaching styles (in applied and clinical anatomy) by incorporating clinical scenarios in the questions and providing a clinical context to anatomy questions (Yaqinuddin et al. 2013; Miller, 1990; Bloom, 1956). Moreover, the use of authentic and valid images tests the robustness of inter-connections between external and internal representations; and questions with no images test mental models without the support or interference of external visual representations (images). Furthermore, the study provides a tool to assess the multifaceted nature of anatomy outside the dissecting room environment that has a potential to reduce the dissonance phase in assessment of anatomy in "knows" and "knows how" levels in Miller's Pyramid and Bloom's taxonomy level 2 and 3 (Miller, 1990; Bloom 1956), and fulfils the objectives set by the Anatomical Society of Great Britain and Ireland and General Medical Council's "*Tomorrow's Doctors"*. It suggests an alternative method for testing applied knowledge of anatomy of "advanced beginners" (Dreyfus 2004) without the disadvantage of practical examinations, which limits students to answer a question in a set time of 1 to 1.5 minutes. Unlike online examinations, practical examinations may lead to physical fatigue as students are continuously moving to subsequent stations on the sound of the buzzer, and stabilising themselves on each station prior to attempting each question, which may have a negative impact on their performance. Moreover, these examinations defeat the assumption of idiosyncrasy of individuals in solving problems - as each individual, each question and interactions between them are different, practical examinations set-up can put students with varying levels of competence into a disadvantage, and thus makes the test unreliable.

The principal implication of the findings of this study is that incorporating images impacts on students' performance on applied anatomy assessments, and teachers and examiners ought to take this into account in designing these assessments and interpreting the results. Along with aligning these assessments with the learning objectives and teaching styles, the blueprinting of these examinations should involve effective use of authentic and valid images aimed at an appropriate level for these advanced beginners.

As the type of image did not affect the students' performance in the study, it is important to have a mixture of authentic and valid images to appropriately test anatomical knowledge. Furthermore, the recommendation is that the students' results must be analysed separately to see whether presence or absence of images have any effect on their performance; as well as if they are able to cope equally well in questions with cadaveric, clinical findings and radiological images.

Moreover, the deep component of the images seems to play a significant role; therefore, questions referring to bones and soft tissues should be one of the criterion for blueprinting, and the analysis of results should take the students' performance on these supplementary and non-homogenous images into consideration.

Hence, this study enhances the existing and frequently used anatomy assessment style with regard to assessing applied anatomical knowledge and the use of relevant images in SBA type online assessment. It suggests that the existing anatomy assessment system will benefit by further refining the blueprinting process of single-best-answers questions in anatomy with presence and absence of various types of images and their deep component, and will benefit from the results' analysis to help future learning patterns of these advanced beginners.

<u>References</u>

Ackerman, P. L. (1988) ‘Determinants of individual differences during skill acquisition: cognitive abilities and information processing’, *Journal of Experimental Psychology*, General, 117, pp 288−318

Ahmed, A. and Pollitt, A. (2007) 'Assessment in Education: principles, policy and practice'. *Assessment in Education*, vol. 14, no. 2, pp. 201-232

Ahmed, K., Rowland, S., Patel, V., Khan, R.S., Ashrafian, H., Davies, D.C., Darzi, A., Athanasiou, T., Paraskeva, P.A. (2010) ‘Is the structure of anatomy curriculum adequate for safe medical practice’? Surgeon, vol. 8, pp. 318–324

Ainsworth, S. (1999) ‘The functions of multiple representations’, *Computer & Education*, vol. 33, pp. 131–152

Anderson, J. (1979) ‘For Multiple Choice Questions’, *Medical Teacher*, vol. 1, no. 1, pp. 37-42

Angoff, W.H. (1971) ‘Scales, norms and equivalent scores’, in Thorndike, R.L., *Educational Measurement*. 2nd Ed. Washington, DC: American Council on Education, pp. 508-600

Baddeley, A.D. (1992) ‘*Working memory’*, Science, 255, pp. 556-559

Baartman, L.K.J., Bastiaens, T.J., Kirschner, P.A. and Van der Vleuten, C.R.M. (2006) 'The wheel of competency assessment: presenting quality criteria for competency assessment programs', *Studies in Educational Evaluation*, vol. 32, pp. 153-170

Beagle, J.M. (2009) *Pictures versus text: Modality effects across three levels of learning and study time*, Purdue University: West Lafayette, IN. Doctorate of Philosophy Dissertation, p. 314

Benninger, B., Matsler, N., Delamarter, T. (2014) 'Classic versus millennial medical lab anatomy', *Clinical Anatomy*, vol. 27, pp. 988–993

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

Berends, I.E., Van Lieshout, E.C. (2009) 'The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load', *Learning and Instruction*, vol. 19, pp. 345–53

Biedermann, I. (1981) On the semantics of a glance at a scene, in: Kubovy, M. and Pomerantz, J. R. (Eds) *Perceptual organization* (Hillsdale, NJ, Erlbaum)

Biggs, J. (1996) 'Enhancing teaching through constructive alignment', *Higher Education*, vol. 32, pp. 347-364

Biggs, J. (2003) 'Aligning teaching and assessing to course objectives', *Teaching Learning in Higher Education: New Trends Innovations*, vol. 2, pp. 13–17

Biggs, J.B. and Tang, C. (2011) *Teaching for Quality Learning at University*. 4th Ed. Maidenhead, UK, Open University Press

Black, P. and Wiliam, D. (1998a) 'Assessment and classroom learning: Principles, policy and practice', *Assessment in Education*, vol. 5, no. 1, pp. 1-78

Bloodgood, R.A., Ogilvie, R.W. (2006) 'Trends in histology laboratory teaching in United States medical schools', *The Anatomical Record B New Anatomy*, 289, pp. 169–175

Bloom, B. S. (1956) *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. 1st Ed. New York: David McKay Co. 201 p.

Boshuizen, H.P.A., Schmidt, H.G. (1992) On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices, Cognitive science, vol. 16 (2), pp. 153-184

Brainerd, C.J., Desrochers, A., Howe, M.L. (1981) 'Stages-of-learning analysis of picture-word effects in associative memory', *Journal of Experimental Psychology Human Learning and Memory*, vol. 7, pp. 1–14

Brenner, E., Chirculescu, A.R.M., Reblet, C., Smith, C. (2015) 'Assessment in anatomy', European Journal of Anatomy, vol. 19, no. 1, pp. 105-124

BERA (2011) Ethical Guidelines for Educational Research, London: BERA

Bulte, C., Betts, A., Garner, K., Durning, S. (2007) ‘Student teaching: Views of student near-peer teachers and learners’, Medical Teacher, 29, pp. 583–590

Campbell, D. T., Stanley, J. C. (1963) ‘Experimental and quasi-experimental designs for research on teaching’. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally

Carney, R., Levin, J. (2002) ‘Pictorial illustrations still improve students' learning from text’, *Educational Psychology Review*, vol. 14, pp. 5–26

Case, S.M. and Swanson, D.B. (2002) *Constructing Written Test Questions for the Basic and Clinical Sciences*. 3rd Ed. Philadelphia, PA: National Board of Medical Examiners

Chakravaraty, M., Latif, N.A., Abu-Hijlah, M.F., Osman, M., Dharap, A.S., Ganguly, P.K. (2005) ‘Assessment of anatomy in a problem-based medical curriculum’, *Clinical Anatomy*, vol. 18, pp. 131-136

Chandler, P. and Sweller, J. (1991) ‘Cognitive load theory and the format of instruction’, *Cognition and Instruction*, vol. 8, no. 4, pp. 293-332

Chen, R., Grierson, L.E., Norman, G.R. (2015) ‘Evaluating the impact of high-and low- fidelity instruction in the development of auscultation skills’, *Medical Education*, 49, pp. 276–285

Chirculescu, A.R.M., Chirculescu, M., Moris, J, F. (2007) 'Anatomical teaching from medical students from the perspective of European Union enlargement'. *European Journal of Anatomy*, vol. 11, pp. 63-65

Cook, M.P. (2006) 'Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles', *Science Education*, 90, pp. 1073–1091

Cooper, C., Gray, L.A. (2014). Lack of anatomy training could lead to shortage of surgeons. The Independent. Available at: http://www.independent.co.uk/life-style/health-and-families/health-news/lack-of-anatomy-training-could-lead-to-shortage-of-surgeons-9570684.html

Clark, J. M., and Paivio, A. (1991) 'Dual coding theory and education'. *Educational Psychology Review*, vol. 3, pp. 149-210

Clough, R.W., Lehr, R.P. (1996) 'Testing knowledge of human gross anatomy in medical school: an applied contextual-learning theory method', *Clinical Anatomy*, vol. 9, no. 4, pp. 263-268

Crisp, V.M. and Sweiry, E. (2006) 'Can a picture ruin a thousand words? The effects of visual resources in exam questions', *Educational Research*, vol. 48, pp. 139-144

Cronbach, L., & Meehl, P. (1955) 'Construct validity in psychological tests', *Psychological Bulletin, 52, pp.* 281–302

Crossman, A.R., Neary, D. (2014) Neuroanatomy: An Illustrated Colour Text, London, UK, Churchill Livingstone

Crotty, M. (1998) 'The Foundations of Social Research: Meaning and Perspective in the Research Process', London: Sage, Chapter 1, 'Introduction: The Research Process', pp. 1-17

Curtis, D. A., Lind, S. L., Plesh, O., Finzen, F. C. (2007) 'Correlation of admissions criteria with academic performance in dental students', Journal of Dental Education, 71, pp. 1314−1321.

De Gruijter, D.N.M. (1985) Compromise models for establishing examination standards, Journal of Educational measurement, vol. 22 (4), pp. 263-269

De Jong, T. (2010) 'Cognitive load theory, Educational Research, and Instructional Design: Some food for thought', *Instructional Science*, vol. 38, pp. 105-134

DeLeeuw, K.E., Mayer, R.E. (2008) 'A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load', *Journal of Educational Psychology*, 100, pp. 223-234

Dettmer, S., Schmiedl, A., Meyer, S., Glesemann, A., Pabst, R., Weidemann, J., Wacker, F.K., Kirchhoff, T. (2013) 'Radiological Anatomy – Evaluation of Integrative Education in Radiology', *Radiological Anatomy*, 185, pp. 838-843

Derrick, J., Ecclestone, K., Gawn, J. (2009) *Formative assessment in adult literacy, language and numeracy,* Leicester (NIACE)

Dijkstra, J., Van der Vleuten, C.P.M., Schuwirth, L.W.T. (2010) ‘A new framework for designing programmes of assessment’, *Advances in Health Sciences Education*, vol. 15, pp. 379-393

DiLullo, C., Morris, H.J., Kriebel, R.M. (2009) ‘Clinical competencies and the basic sciences: An online case tutorial paradigm for delivery of integrated clinical and basic science content’, *Anatomical Sciences Education, vol.* 2, pp. 238–243

Downing SM. (2003) 'Validity: on meaningful interpretation of assessment data', *Medical Education*, vol. 3, no. 9, pp. 830-837

Drake, R.L., McBride, J.M., Lachman, N., Pawlina, W. (2009) ‘Medical education in the anatomical sciences: The winds of change continue to blow’, Anatomical Sciences Education, vol. 2, pp. 253–259

Dreyfus, S.E. (2004) ‘The five-stage model of adult skill acquisition’, *Bulletin of science, Technology and society*, vol. 24, no. 3, pp. 1177-1181

Ebel R.L. (1983) ‘The practical validation of tests of ability. Educational Measurement: issues and practice’, vol. 2 (2), 7-10

Elkins, J. (1998) On pictures and the words that fail them (Cambridge, Cambridge University Press)

Engelhardt, P.V. (2009) ‘An Introduction to Classical Test Theory as Applied to Conceptual Multiple-choice Tests’, C Henderson and KA Harper, editors. Getting Started in PER. College Park: *American Association of Physics Teachers; Reviews in PER*, 2

Epstein, R.M., Hundert, E.M. (2002) ‘Defining and assessing professional competence’, *American Medical Association*, 287 (2)

Evans, D.J., Cuffe, T. (2009) ‘Near-peer teaching an anatomy: an approach for deeper learning’, *Anatomical Sciences Education*, vol. 2, no. 5, pp. 227-233

Fabrizio, P.A. (2013) ‘Oral anatomy laboratory examinations in a physical therapy program’, *Anatomical Science Education*, vol. 6, pp. 271-276

Fisher-Hoch, H., Hughes, S., Bramley, T. (1997) What makes GCSE examination questions difficult? Outcomes of manipulating difficulty of GCSE questions: a paper presented at the British Educational Research Association Annual Conference (Cambridge, University of Cambridge Local Examinations Syndicate). Available online at: http://www.cambridgeassessment.org.uk/ research/confproceedingsetc/ given a date retrieved?

Fitzgerald, J.E., White, M.J., Tang, S.W., Maxwell-Armstrong, C.A., James, D.K. (2009) ‘Response to letter: Are we teaching sufficient anatomy at medical school? The opinions of newly qualified doctors’, *Clinical Anatomy*, vol. 22, pp. 288–289

Fleming, S. McKee, G., Huntley-Moore, S. (2011) ‘Undergraduate nursing students’ learning styles: a longitudinal study’, Nurse Education Today, vol. 31 (5), pp. 444-449

Fraher, J.P., Evans, D.J.R. (2009) ‘Training Tomorrow's Anatomists Today: A Partnership Approach’, Anatomical Sciences Education, vol. 2, pp. 119-125

Gawande, A. (2002, January 28) ‘The learning curve’, The New Yorker (pp. 52−61)

Gibson, J. J. (1979) ‘The ecological approach to visual perception’. Boston: Houghton Mifflin Company.

Ginns, P. (2005) ‘Meta-analysis of the modality effect’, Learning and Instruction, vol. 15, pp.313– 331

GMC. (2009) *General Medical Council. Tomorrow’s Doctors. Outcomes and Standards, for Undergraduate Medical Education*, Regulating Doctors, Ensuring Good Medical Practice. 3rd Ed. London, UK: General Medical Council. 104 p.

Gogalniceanu, P., O’Connor, E.F., Raftery, A. (2009) ‘Undergraduate anatomy teaching in the UK’, Annals of The Royal College of Surgeons of England, 91, pp. S102–S106

Gregory, R. L. (1973) ‘*Eye and brain’*. New York: McGraw–Hill

Grignon, B., Oldrini, G., Walter, F. (2016) ‘Teaching medical anatomy: What is the role of imaging today?’ *Surgical and Radiologic Anatomy,* 38, pp. 253–260

Gunderman, R.B. (2008) ‘Giving ourselves: The Ethics of Anatomical Donation’, *Anatomical Sciences Education*, vol. 1, pp. 217-219

Gunderman, R.B., Wilson, P.K. (2005) ‘Exploring the human interior: the roles of cadaver dissection and radiologic imaging in teaching anatomy’. *Academic Med*icine, vol. 80, pp. 745–749

Hegarty, M., Keehner, M., Khooshabeh, P., Montello, D.R. (2009) ‘How spatial abilities enhance, and are enhanced by, dental education’, *Learning and Individual Differences*, vol.19, pp. 61–70

Haladyna, T.M., Rodriguez, M.C. (2014) *Developing and validating test items*, Routledge, Taylor and Francis group, New York and London

Hisley, K.C., Anderson, L.D., Smith, S.E., Kavic, S.M., Tracey, J.K. (2008) ‘Coupled Physical and Digital Cadaver Dissection followed by a Visual Text Protocol provides insights into the nature of Anatomical knowledge and its evaluation’, Anatomical Sciences Education, vol. 1, pp. 27–40

Holland, J., O’Sullivan, R., Arnett, R. (2015) ‘Is a picture worth a thousand words: an analysis of the difficulty and discrimination parameters of illustrated vs. text-alone vignettes in histology multiple choice questions’, *BMC Medical Education*, vol. 15, p. 184

Houwink, A.P., Kurup, A.N., Kollars, J.P., Kral Kollars, C.A., Carmichael, S.W., Pawlina, W. (2004) ‘Help of third-year medical students decreases first-year medical students' negative psychological reactions on the first day of gross anatomy dissection', Clinical Anatomy, 17, pp. 328–333

Human Tissue Act, (2004) Available at:
http://www.legislation.gov.uk/ukpga/2004/30/pdfs/ukpga_20040030_en.pdf
Hunt, D.R. (1978) 'Illustrated multiple choice examinations', *Medical Education*, vol. 12, pp. 417-420

Ikah, D.S.K., Finn, G.M., Swamy, M., White, P.M., McLachlan, J.C. (2015) 'Clinical vignettes improve performance in Anatomy Practical Assessment', *Anatomical Sciences Education*, vol. 8, pp. 221-229

Inuwa, I.M., Taranikanti, V., Al-Rawahy. M., Habbal.O. (2011) 'Perceptions and attitudes of students towards two methods of assessing practical anatomy knowledge', *Sultan Qaboos University Medical Journal*, vol. 11, pp. 383–390

Inuwa, I.M., Taranikanti, V., Al-Rawahy, M., & Habbal, O. (2012) 'Anatomy practical examinations: How does student performance on computerized evaluation compare with the traditional format'? *Anatomical Sciences Education*, vol. 5, no. 1, pp. 27-32

Intercollegiate Committee for Basic Surgical Examinations 2012/13 Annual Report. Available at:
http://www.intercollegiatemrcs.org.uk/new/pdf/Annual_Report_12_13.pdf

Johnson-Laird, P. (1981) Mental models of meaning, in: A. K. Joshi, B. L. Webber & I. A. Sag (Eds) Elements of discourse understanding (Cambridge, Cambridge University Press)

Kalyuga, S., and Sweller, J. (2004) 'Measuring knowledge to optimise cognitive load factors during instruction'. *Journal of Educational Psychology*, 96, pp. 558-568

Karay, Y., Schauber, S.K., Stosch, C., Schuettpelz-Brauns, K. (2012) ‘Can computer-based assessment enhance the acceptance of formative multiple choice exams’? A utility analysis, Medical Teacher, vol. 34 (4), pp. 292-296

Khalil, M.K., Johnson, T.E., Lamar, C.H. (2005) ‘Comparison of computer-based and paper-based imagery strategies in learning anatomy’. *Clinical Anatomy*, vol. 18, pp. 457-464

Knauff, M., Johnson-Laird, P.N. (2002) Visual imagery can impede reasoning, Memory and Cognition, 30 (3), 363-371

Krippendorf, B.B., Bolender, D.L., Kolesari, G.L. (2008) ‘Computerised grading of anatomy laboratory practical examinations’, *Anatomical Science Education*, vol. 1, pp. 220-223

Kulhavy, R.W. Stock, W.A., Caterino, L.C. (1994) Reference maps as a framework for remembering text. In W.Schnotz and R.W. Kulhavy (Eds.), Comprehension of graphics (pp. 153-162). Amsterdam: Elsevier Science.

Larsen, D.P., Butler, A.C., Roediger, H.L. (2008) 'Test-enhanced learning in Medical Education', *Medical Education*, vol. 42, pp. 959-966

Larsen, D.P., Butler, A.C., Roediger, H.L. (2009) 'Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial', *Medical Education*, vol. 43, pp. 1174–1181

Leite, W.L., Svinicki, M., Shi, Y. (2010) 'Attempted validation of the scores of the VARK: Learning styles inventory with multitrait-multimethod confirmatory factor analysis models', *Educational and Psychological Measurement*, 70, pp. 323–339

Leonard R.J., Acland, R., Aguyr, A., Blevins, C.E., Cahill, D.R., Collins, J.D., Dalley, A.F. I.I., Dolph, J., Hagedoorn, J.P., Hoos, P.C., (1996) 'A clinical anatomy curriculum for the medical student of the 21st century: Gross anatomy', Clinical Anatomy, vol. 9, pp. 71–99

Levie, W.H., Lentz, R. (1982) 'Effects of text illustrations: a review of research', *Educational Communication and Technology Journal*, vol. 30, pp. 195–232

Liddell, K. and Hall, A. (2005) 'Beyond Bristol and Alder Hey: the Future Regulation of Human Tissue'. *Medical Law Review*, vol. 13, no. 2, pp. 170-223

Louw, G., Eizenberg, N., Carmichael, S.W. (2009) The place of anatomy in medical education: AMEE Guide no 41, *Medical Teacher*, vol. 31, no. 5, pp. 373-386

Lowe, R.W. (1996) Animation and learning: selective processing of information in dynamic graphics', Learning and instruction, vol. 13, pp. 157-176

Lufler, R.S., Zumwalt, A.C., Romney, C.A., Hoagland, T.M. (2012) 'Effect of visual-spatial ability on medical students' performance in a gross anatomy course', *Anatomical Sciences Education*, vol. 5, pp. 3–9

Lukic, I.K., Gluncic, V., Katavic, V., Petanjek, Z., Jalsovec, D., Marusic, A. (2001) 'Weekly quizzes in extended-matching format as a mean of monitoring students' progress in gross anatomy', Annals of Anatomy, 183, pp. 575-579

Mahony, S.M., Sbayeh, A., Horgan, M., O' Flynn, S. O' Tuathaigh, C.M.P. (2016) Association between learning style preferences and anatomy assessment outcomes in graduate-entry and undergraduate medical students, Anatomical Sciences Education, vol. 9 (4), pp. 391-399

Mathai, S., Ramadas, J. (2009) 'Visuals and visualisation of human body systems', *International Journal of Science Education*, vol. 31, pp. 439–458

Mayer, R.E. (2005b) 'Introduction to multimedia learning', in Mayer, R.E. (Ed.), *The Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press

Mayer, R.E. (2009) Multimedia learning (2nd ed). New York: Cambridge University Press

Mayer, R.E. (2010) ‘Applying the science of learning to medical education’, *Medical Education*, vol. 44, pp. 543–549

Mayer, R.E., Massa, L.J. (2003) ‘Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference’, *Journal of Educational Psychology*, 95, pp. 833– 841

Mayer, R.E., Moreno, R. (2003) ‘Nine ways to reduce cognitive load in multimedia learning’, *Educational Psychology*, vol. 38, pp. 43–52

McHanwell, S., Atkinson, M., Davies, D.C., Dyball, R., Morris, J., Ockleford, C., Parkin, I., Standring, S., Whiten, S., Wilton, J. (2007) ‘A core syllabus in anatomy - Adding common sense to need to know’, *European Journal of Anatomy*, vol. 11, pp. 3-18

McWhorter, D.L. and Forester, J.P. (2004) ‘Effects of an alternate dissection schedule on gross anatomy laboratory practical performances’, *Clinical Anatomy, vol.* 17, pp.144-148

Menzes, R.G., Nayak, V.C., Binu, V.S., Kanchan T., Rao, P.P.J., Baral, P., Lobo, S.W. (2011) ‘Objective Structural Practical Examination (OSPE) in forensic medicine: Students' point of view’, *Journal of Forensic and Legal Medicine*, vol. 18, pp. 347-349

Messick, S. (1994) ‘The interplay of evidence and consequences in the validation of performance assessments’. *Educational Researcher*, vol. 23, pp. 13-23

Meyer, A.J., Innes, S.I., Stomski, N.J., Armson, A.J. (2016) ‘Student performance on practical gross anatomy examinations is not affected by assessment modality’, American Sciences Education, vol.9 (2), pp. 111-120

Miller, G.E. (1990) 'The assessment of clinical skills/competence/performance’, *Academic Medicine*, vol. 65, no. 9, pp. 63–67

Miller, R. (2000) ‘Approaches to learning spatial relationships in gross anatomy: perspective from wider principles of learning’, *Clinical Anatomy*, Medical Education, vol. 13, pp. 439-443

Molyneux, C., Robson, L. (2012) ‘An online image-based examination for the assessment of clinical anatomy’, Journal of Anatomy, vol. 221 (1), pp. 95-95

Morris, J. F., and Chirculescu, A. R. M. (2007) ‘Structure and assessment of a short intense clinical anatomy course shortly before clinical studies’, *Clinical Anatomy*, vol. 11(Supplement 1), pp. 95-98

Moxham, B.J., Shaw, H., Crowson, R., Plaisant, O. (2011) 'The future of clinical anatomy'. *European Journal of Anatomy*, vol. 15, no. 1, pp. 29-46

Moqattash, S., Harris, P.F., Gumaa, K.A., Abu-Hijleh, MF. (1995) Assessment of basic medical sciences in an integrated system-based curriculum’, Clinical Anatomy, vol. 8 (2), pp. 139-147

Nayar, U., Malik, S.L., Bijlani, R.L. (1986) ‘Objective Structural Practical Examination: A new concept in assessment of laboratory exercises in preclinical sciences’, Medical Education, vol. 20, pp. 204-209

Neisser, U. (1976) '*Cognition and reality'*. San Francisco: W. H. Freeman

Nicholson, D.T., Chalk, C., Funnell, W.R., Daniel, S.J (2006) Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model', Medical Education, vol. 40 (1), pp. 1081-1087

O'Brien, K.E., Cannarozzi, M.L., Torre, D.M., Mechaber, A.J., Durning, S.J. (2008) 'Training and assessment of CXR/basic radiology interpretation skills: results from the 2005 CDIM survey', *Teaching and Learning in Medicine*, 20, pp. 157–162

Older, J. (2004) 'Anatomy: A must for teaching the next generation', *Surgical Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*, pp. 79-90

Ollerenshaw, A., Aidman, E. & Kidd, G. (1997) 'Is an illustration always worth ten thousand words? Effects of prior knowledge, learning style and multimedia illustrations on text comprehension', *International Journal of Instructional Media*, vol. 24, no. 3, 227–238

Orsbon, C.P., Kaiser, R.S., Ross, C.F. (2014) 'Physician opinions about an anatomy core curriculum: a case for medical imaging and vertical integration', *Anatomical Sciences Education*, vol. 7, no. 4, pp. 251-261

Paalman, M.H. (2000) 'Why teach anatomy? Anatomists respond', The Anatomical Record, vol. 261 (1), pp. 1-2

Page, G and Bordage, G. (1995) 'The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills', *Academic Medicine*, vol. 70, pp. 104-110

Palmer, E.J., Devitt, P.G. (2007) 'Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions'? *BMC Medical Education*, vol. 7, p. 49

Papa, V., Vaccarezza, M. (2013) ‘Teaching anatomy in the XXI century: new aspects and pitfalls’, The Scientific World Journal, http://dx.doi.org/10.1155/2013/310348

Papachristodoulou, D. (2010) 'Learning experiences and assessment in the first 2 years of the medical course at King's College London School of Medicine', *The Keio Journal of Medicine*, 59, no. 4, pp. 140-145

Paivio, A. (1986) *Mental representations: A dual coding approach*, Oxford, England: Oxford University Press

Pawlina, W. (2009) ‘Basic sciences in medical education: Why? How? When? Where’? *Medical Teacher,* 31, pp. 787–789

Peeck, J. (1974) ‘Retention of pictorial and verbal content of a text with illustrations,’ *Journal of Educational Psychology,* 66 (6), pp. 880-888

Peeck, J. (1987) ‘The role of illustrations in processing and remembering illustrated text’, in: Willows, D. M. and Houghton, H. A. (Eds) *The psychology of illustration*, vol. 1: Basic research (New York, Springer-Verlag)

Peeck, J. (1993) ‘Increasing picture effects in learning from illustrated text’, *Learning and Instruction*, vol. 3, pp. 227–238

Peel, S. (1998) ‘An innovative problem-solving assessment for groups of first-year medical undergraduates’, Think Tanks, *Medical Education*, vol. 32, no. 1, pp. 35-39

Philip, C.T., Unruh, K.P., Lachman, N., Pawlina, W. (2008) ‘An explorative learning approach to teaching clinical anatomy using student generated content’, Anatomical Sciences Education, vol. 1, pp. 106–110

Phillips, A.W., Smith, S.G., Straus, C.M. (2013) ‘The role of radiology in preclinical anatomy: a critical review of the past, present, and future’, *Academic Radiology*, vol. 20, pp. 297-304

Pollitt, A. & Ahmed, A. (1999) ‘A new model of the question answering process’. Paper presented at the International Association for Educational Assessment Annual Conference (Cambridge, University of Cambridge Local Examinations Syndicate). Available online at: http://www.cambridgeassessment.org.uk/research/confproceedingsetc/

Raftery A. (2006) ‘Anatomy teaching in the UK,’ Surgery, 25 (1), pp. 1–2

Rainsbury, D., Barbour, A., Mahadevan, V. (2007) 'Anatomy Teaching – The cruellest cut of all', *Annals of the Royal College of Surgeons of England,* 89, pp. 196-197

Ranney, R. R., Wilson, M. B., & Bennett, R. B. (2005) 'Evaluation of applicants to predoctoral dental education programs: Review of the literature', Journal of Dental Education, 69, pp. 1095−1106

Regehr, G., Norman, G.R. (1996) 'Issues in cognitive psychology: implications for professional education', *Academic Medicine*, 71, p. 9

Reid, W.A., Duvall, E., Evans, P. (2007) 'Relationship between assessment results and approaches to learning and studying in year two medical students', *Medical Education*, 41, pp. 754-762

Robson, C. (2011) *Real World Research*, 3rd edition, Chichester: John Wiley and Sons Ltd

Richardson, R. (1988) 'Death, Dissection and the Destitute: The Politics of the Corpse in Pre-Victorian Britain', London and New York: Routledge

Rowland, S., Ahmed, K., Davies, D.C., Ashrafian, H., Patel, V., Darzi, A., Paraskeva, P.A., Athanasiou, T. (2011) 'Assessment of anatomical knowledge for clinical practice: perceptions of clinicians and students', *Surgical and Radiologic Anatomy*, vol. 33, pp. 263-269

Royal, K.D., Gilliland, K.O., Kernick, E.T. (2014) 'Using Rasch Measurement to score, evaluate, and improve examinations in an anatomy course', *Anatomical Sciences Education,* vol. 7, pp. 450-460

Sagoo, M.G., Smith, C.F., Gosden, E. (2016) 'Assessment of anatomical knowledge by practical examinations: The effect of question design on student performance', *Anatomical Sciences Education,* vol. 9 (5), pp. 446-452

Sandow, P. L., Jones, A. C., Peek, C. W., Courts, F. J., Watson, R. E. (2002) 'Correlation of admission criteria with dental school performance and attrition', Journal of Dental Education, 66, pp. 385−392

Schmidt, H.G., Rikers, R.M. (2007) How expertise develops in medicine: knowledge encapsulation and illness script formation,' Medical Education, 41 (12), pp. 1133-1139

Schnotz, W. (1993) 'Some remarks on the complementary on the relation of dual coding and mental models in graphics comprehension', *Learning and Instruction*, vol. 3, pp. 247-249

Schnotz, W. (2002) 'Towards an integrated view of learning from text and visual displays', *Educational Psychology Review*, vol. 14, no. 1, pp. 101-120

Schnotz, W., (2005) 'An integrated model of text and picture comprehension', in Mayer, R.E. (Ed.), *The Cambridge Handbook of Multimedia Learning*, 1st Ed. New York: Cambridge University Press, pp. 49-69

Schnotz, W. Baadte, C. (2015) ‘Surface and deep structures in graphics comprehension’, *Memory Cognition*, vol. 43, pp. 605-618

Schnotz, W., Bannert, M. (2003) ‘Construction and interference in learning from multiple representation’, *Learning Instruction*, vol. 13, pp. 141-156

Schnotz, W., Kurschner, C. (2008) ‘External and internal representations in the acquisition and use of knowledge: visualization effects on mental model construction’, *Instructional Science*, vol. 36, pp. 175-190

Schoeman, S. and Chandratilake, M. (2012) ‘The anatomy competence score - a new marker for anatomical ability’, *Anatomical Science Education*, vol. 5, pp. 33-40

Schubert, S., Schnabel, K.P., Winkelmann A., (2009) ‘Assessment of spatial anatomical knowledge with a 'three-dimensional multiple choice test' (3D-MC)’, *Medical Teacher*, vol. 31, pp. 13-17

Schuwirth, L.W.T., Van der Vleuten, C.P.M., Donkers, H.H. (1996) ‘A closer look at cueing effects in multiple-choice questions,’ Medical Education, 30 (1), pp. 44-49

Schuwirth, L.W.T, Verheggen, M., Van der Vleuten, C., Boshuizen, H., Dinant, G. (2001) ‘Do short cases elicit different thinking processes than factual knowledge questions do’? *Medical Education*, vol. 35, pp. 348–356

Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2003) 'The use of clinical simulations in assessment', *Medical Education*, vol. 37 (supplementary 1), pp. 65-71

Schuwirth, L.W.T., Van der Vleuten, C.P.M. (2004) ‘Changing education, changing assessment, changing research’? *Medical Education,* vol. 38, pp. 805-812

Schuwirth, L.W.T., Van der Vleuten, C.P.M. (2004) ‘Different written assessment methods: What can be said about their strengths and weaknesses’? *Medical Education*, vol. 38, pp. 974–979

Schuwirth, LW. & Van der Vleuten, CPM. (2006) ‘How to design a useful test: the principles of assessment’, A*ssociation for the study of Medical Education, Medical Education*: Edinburgh, p. 1-7

Schuwirth, L.W.T and Van der Vleuten, C.P.M. (2011) ‘Programmatic assessment: From assessment of learning to assessment for learning’, *Medical Teacher*, vol. 33, pp. 478-485

Severo, M. and Tavares, M.A.F. (2010) ‘Meta-Evaluation in Clinical Anatomy: A practical application of item response theory in multiple choice examinations’, *American Association of Anatomists*, vol. 3, pp. 17-24

Shaibah, H.S., Van der Vleuten, C.P.M. (2013) 'The validity of multiple-choice practical examinations as an alternative to traditional free response examination formats in gross anatomy', *Anatomical Sciences Education*, vol. 6, pp. 149-156

Smith, C. F., & Mathias, H. S. (2010) ‘Medical students’ approaches to learning anatomy: students' experiences and relations to the learning environment’, *Clinical Anatomy*, vol. 23, no. 1, pp. 106-114

Smith, C.F., Mathias, H.S. (2011) ‘What impact does anatomy education have on clinical practice’? *Clinical Anatomy*, vol. 24, pp. 113–119

Smith, C.F., McManus, B. (2015) ‘The integrated anatomy paper: a robust assessment method for anatomy education today’, *Anatomical Science Education*, vol. 8, no. 1, pp. 63-73

Smith, C.F., Finn, G.M., Stewart, J., Atkinson, M.A., Davies, D.C., Dyball, R., Morris, J., Ockleford, C., Parkin, I., Standring. S., Whiten. S., Wilton. J., McHanwell, S. (2016) ‘The Anatomical Society core regional anatomy syllabus for undergraduate medicine’. *Journal of Anatomy*, vol. 228, pp. 2–14

Smyth, A. and Holian, R. (2008) 'Credibility issues in research in research from within organisations', in Sikes, P. and Potts, A. (ed.), *Researching Education from the Inside*, London and New York: Routledge, Taylor and Francis Group

Sood, R., Singh T. (2012) ‘Assessment in medical education: Evolving perspectives and contemporary trends’, *The National Medical Journal of India*, vol. 25, p. 6

Sorden, S.D. (2005) ‘A cognitive approach to instructional design for multimedia learning’, *Educational Psychology Review*, vol. 19, pp. 469-508

Standring, S. and Larvin, M (2011) ‘Anatomy teaching: the cruellest cut - a progress report ‘, *Annals of the Royal College of Surgeons of England*, vol. 93, pp. 16-18

Standing, L., Smith, P. (1975) ‘Verbal-pictorial transformation in recognition memory’, Canadian Journal of Experimental Psychology, 29, pp. 316–326

Stewart, J. H., Van Kirk, J., Rowell, R. (1979) ‘Concept maps: a tool for use in biology teaching’, *American Biology Teacher*, vol. 41, pp. 171–175

Sugand, K., Abrahams, P., Khurana, A. (2010) ‘The anatomy of anatomy: a review of its modernization’, *Anatomical Sciences Education*, vol. 3, pp. 83-93

Swanson, D.B., Norcini, J.J., Grosso, L.J. (2006) 'Assessment of clinical competence: written and computer-based simulations', *Assessment and Evaluation in Higher Education*, vol. 12, no. 3, pp. 220-246

Sweller, J., Chandler, P., Tierney, P. & Cooper, M. (1990) ‘Cognitive load as a factor in the structure of technical material’, *Journal of Experimental Psychology*, General, 119, pp. 176-192

Sweller, J. (1994) 'Cognitive load theory, learning difficulty, and instructional design, *Learning and instruction*, vol. 4, pp. 295-312

Sweller, J., Van Merrienboer, J. J. G., Paas, F. (1998) ‘Cognitive architecture and instructional design’. *Educational Psychology Review*, vol. 10, no. 3, pp. 251-296

Van der Vleuten, C.P.M. and Schuwirth, L.W.T. (2005) 'Assessing professional competence: From methods to programmes', *Medical Education*, vol. 39, pp. 309-317

Van der Vleuten, C.P.M., Schuwirth, L.W.T., Scheele, F., Driessen, E.W. (2010) 'The assessment of professional competence: building blocks for theory development'. *Best practice & Research Clinical Obsteterics and Gynaecology*, vol. 24, pp. 703-719

Van Merrienboer, J.J.G. and Sweller, J. (2010) 'Cognitive load theory in health professional education: design principles and strategies' *Medical Education*, vol. 44, pp. 85-93

Vasan, N.S., DeFouw, D.O., Holland, B.K. (2008) 'Modified use of team-based learning for effective delivery of medical gross anatomy and embryology', *Anatomical Sciences Education, vol.* 1, pp. 3–9

Vasan, N.S., DeFouw, D.O., Compton, S. (2009) A survey of student perceptions of team-based learning in anatomy curriculum: Favorable views unrelated to grades', *Anatomical Sciences Education, vol.* 2, pp. 150–155

Vasan, N.S., DeFouw, D.O., Compton, S. (2011) 'Team-based learning in anatomy: An efficient, effective, and economical strategy', *Anatomical Sciences Education, vol.* 4, pp. 333–339

Vorstenbosch, M.A.T.A., Bouter, S.T., Van der Hurk, M.M., Kooloos, J.G.M., Bolhuis, S.M., Laan. R.F.J.M. (2014) 'Exploring the validity of assessment in anatomy: do images influence cognitive processes used in answering extended matching questions'? *American Association of Anatomists*, vol. 7, no. 2, pp. 107-116

Vorstenbosch, M.A.T.A., Klaassen, T.P., Kooloos, J.G., Bolhuis, S.M., Laan, R.F. (2013) 'Do images influence assessment in anatomy? Exploring the effect of images on item difficulty and item discrimination', *Anatomical Sciences Education*, vol. 6, pp. 29–41

Vorstenbosch, M.A.T.A., Kooloos, J.G., Bolhuis, S.M., Laan, R.F. (2016) 'An investigation of anatomical competence in junior medical doctors'. *Anatomical Sciences Education*, vol. 9, pp. 8–17

Wass, V., Van der Vleuten, C. V. D., Shatzer, J., & Jones, R. (2001) 'Assessment of clinical competence', *Medical Education Quartet,* The Lancet, 357, pp. 945-949

Weidenmann, B. (1989) 'When good pictures fail: an information-processing approach to the effect of illustrations', in: Mandl, H. and Levin, J. R. (Eds) *Knowledge acquisition from text and pictures*: Advances in Psychology 58 (Amsterdam, Elsevier)

Wellington, J. (2000) *Educational Research: contemporary issues and practical approaches*, Continuum International Publishing Group, Great Britain

WFME. (2007) WFME Global Standards for Quality Improvement in Medical Education - European Specifications. Denmark

Winn, W. (1987) ‘Charts, graphs, and diagrams in educational materials’, in: Willows, D. M. and Houghton, H. A. (Eds) The psychology of illustration, vol. 1: Basic research (New York, SpringerVerlag)

Winn, W. (1989) ‘The design and use of instructional graphics’, in: Mandl, H. and Levin, J. R. (Eds) Knowledge acquisition from text and pictures: Advances in Psychology 58 (Amsterdam, Elsevier)

Winn, W. D. (1994) ‘Contributions of perceptual and cognitive processes to the comprehension of graphics’. In Schnotz, W., and Kulhavy, R. (eds.), *Comprehension of Graphics*, Elsevier, Amsterdam, pp. 3–27

White, S., Sykes, A. (2012) ‘Evaluation of a blended learning approach used in an anatomy and physiology module for pre-registration in healthcare students’. The fourth international conference on mobile, hybrid, and on-line learning, Iaria, pp. 1-9

Wilkinson, T., Boohan, M., Stevenson, M. (2014) ‘Does learning style influence academic performance in different forms of assessment?’ *Journal of Anatomy,* 224, pp. 304–308

Witteman, M.J., Segers, E. (2010) ‘The modality effect tested in children in a user paced multimedia environment’, *Journal of Computer Assisted Learning*, 26, pp. 132–142

Wood, T., Cole, G., Lee, C. (2004) ‘Developing multiple choice questions for the RCPSC certification examinations. The Royal College of Physicians and Surgeons of Canada, Office of Education

Yaqinuddin, A., Zafar, M., Ikram, M.F., Ganguly, P. (2013) ‘What is an Objective Structural Practical Examination in Anatomy?’ *Anatomical Sciences Education*, vol. 6, pp. 125-133

Yiou, R., Goodenough, D. (2006) ‘Applying problem-based learning to the teaching of anatomy: The example of Harvard Medical School’, *Surgical and Radiologic Anatomy,* 28, pp. 189–194

Youdas, J.W., Hoffarth, B.L., Kohlwey, S.R., Kramer, C.M., Petro, J.L. (2008) ‘Peer teaching among physical therapy students during human gross anatomy: Perceptions of peer teachers and students’, Anatomical Sciences Education, vol. 1, pp. 199–206

Zhang, G., Fenderson, B.A., Schmidt, R.R., Veloski, J.J. (2013) ‘Equivalence of students’ scores on timed and untimed anatomy practical examinations’, Anatomical Sciences Education, vol. 6 (5), pp. 281-285

Appendix

Email sent to the students

Dear Students,

Fantastic opportunity to test your applied anatomy knowledge, especially worth doing it before your exams!

These questions are reviewed by the legend himself, Professor Harold Ellis! On top of that at the end of the tool you will receive an elaborate feedback on each question - so double benefit!!

This tool would be mutually beneficial as you can test your knowledge and learn from it, and it forms a part of my doctorate project. My name is Mandeep Gill Sagoo and I work as an Anatomy Demonstrator at King's College London and a doctorate student at UCL, Institute of Education.

To access the questions, please use your medical school email address (gmails and hotmails will not work!). Then please fill in the consent form (1 question) and the questionnaire (22 questions).

After this you could take the test (36 applied anatomy questions).

The test will take 45 minutes of your time but we have given you 1 hour 30 minutes to complete the test. After this time the test will expire. So please start the test when you are ready.

On completion and submission of the test, you will receive the results and an elaborate feedback on each question. To see the feedback please make sure to expand each question on the page.

All student scores and details will remain anonymous so no need to worry!

Click on the link below and Start now!

www.myanatomygrowth.com

All the Best!

<u>Consent form</u>

By agreeing to participate in this study designed to assess the relationship between the design of anatomy questions and their effect on your performance, you understand that:

i) My test results will be used for data analysis

ii) My questionnaire results will be linked to my test results for data analysis

All the information will be kept strictly anonymous and you have the right to withdraw from the study process at any time. This tool has no connection with your formal examinations, and your participation or withdrawal from this study will have no impact on your future exanimations or training.

Please choose one of the options.

- Yes - I consent to take part
- No - I do not consent to take part

Thank-you!

Participation information sheet

INVITATION

I am *Ms. Mandeep Gill Sagoo*, an Anatomist at King's College, London and a doctorate student at the UCL Institute of Education. I would like to invite you to participate in this educational research project that forms a part of my doctorate research.

WHAT DO I HAVE TO DO?

Please complete the questionnaire provided and answer the questions on the tool.

WHAT IS THE PURPOSE OF THIS STUDY?

This study addresses the relationship between various designs of anatomy questions (without and with a number of visual resources used in contextually rich questions) and their effect on students' scores.

WHY AM I BEING CHOSEN?

As you are taking anatomy examination in the near future, it is mutually beneficial because it will enable you to see how you are doing. The correct answers and elaborate feedback will be made available to you to aid your revision and to express gratitude for your participation. Your results on the test and the information received from you on the questionnaire will help me to for my research.

DO I HAVE TO TAKE PART?

No - there is no obligation on you to take part. The tool is entirely voluntary, and you may withdraw at any time and it will not affect your training or any future examinations.

WILL CONFIDENTIALITY BE ENSURED?

Yes, confidentiality of your personal information is assured. Individual identities and identification factors will not be disclosed during analysis, reporting and dissemination. Your future progression will not be affected in anyway if you decide not to participate, or any of the answers that you provide if you do decide to participate.

HOW WILL BE THE DATA PROTECTED?

The data collected will be stored in compliance with the legal requirements of the Data Protection Act 1998.

ARE THERE ANY RISKS OR BENEFITS?

There are no risks or direct benefits to you. However, this study is part of a continual process of improving the assessment system and may benefit others in the future. This data will be analysed for my research and the anonymised information will be disseminated through conferences and publications.

WHO HAS REVIEWED THE STUDY?

The project is reviewed by my supervisors, Dr. Charlie Owen and Dr. Caroline Pelletier, and ethically approved by the Institute of Education, UCL.

CONTACT AND FURTHER INFORMATION?

If you would like any more information, please contact me: *mgsagoo@gmail.com*

The following websites may be of assistance: British Education Research Association http://www.bera.ac.uk/

THANK YOU FOR TAKING PART IN THIS STUDY

Questionnaire

1. Your Gender

- Female
- Male

2. Your Age range

- 16-18
- 19-21
- 22-24
- 25-27
- 28-30
- 31-33
- 34 or above

3. Your Training Level

- End of 2nd year student (MBBS 5 – Undergraduate stream)
- End of 1st year (MBBS 4 – Graduate entry programme)

4. Your Most Likely Prospective Career?

- Non-surgical
- Surgical
- Don't know

5. Which Medical School/University are you studying at?

- Barts and The London School of Medicine and Dentistry
- Brighton & Sussex Medical School
- Hull York Medical School

- Imperial College London
- King's College London
- Plymouth University Peninsula Schools of Medicine and Dentistry
- St. George's, University of London
- University College London
- University of Newcastle
- University of Birmingham
- University of Southampton
- University of Exeter

6. How anatomy has been taught in your medical school/university?

- Dissection of human cadavers only
- Prosections (dissected body parts) only
- Radiology images only
- All of the above
- Dissection of human cadavers and radiology images
- Prosections and radiology images

7. Have you been formally involved in demonstrating anatomy to junior students?

- Yes
- No

8. I find/found hands-on cadaveric dissection an effective way of learning anatomy.

- Agree
- Unsure
- Disagree

- Not done

9. I find/found prosections (dissected body parts) an effective way of learning anatomy.
   - Agree
   - Unsure
   - Disagree
   - Not used

10. I find/found cadaveric photographs an effective way of learning anatomy.
   - Agree
   - Unsure
   - Disagree
   - Not used

11. I find/found radiological images (x-rays, MRI, CT etc.) an effective way of learning anatomy.
   - Agree
   - Unsure
   - Disagree
   - Not used

12. I find/found clinically relevant anatomy learning (applied, living/surface anatomy) an effective way of learning anatomy.
   - Agree
   - Unsure
   - Disagree

13. My main motivation for learning anatomy is to pass the examination.

- Agree
- Unsure
- Disagree

14. I believe anatomy knowledge should be tested with clinical knowledge.

- Agree
- Unsure
- Disagree

15. I think anatomy demonstrating has helped me to learn anatomy.

- Agree
- Unsure
- Disagree
- Not done

16. I think online anatomy examinations are more effective than practical examinations (spotter tests) conducted in the dissecting room.

- Agree
- Unsure
- Disagree
- Not done online examinations
- Not done practical examinations

17. I think cadaveric photographs are more effective for examinations than prosections and dissected material.

- Agree
- Unsure
- Disagree

18. I think radiological images (x-rays, MRI, CTs etc.) are more effective for examinations than prosections and dissected material.

- Agree
- Unsure
- Disagree

19. I think it requires deep understanding of the subject to answer questions with prosections and dissected material.

- Agree
- Unsure
- Disagree

20. I think it requires deep understanding of the subject to answer questions with cadaveric photographs.

- Agree
- Unsure
- Disagree

21. I think it requires deep understanding of the subject to answer questions with radiological images.

- Agree

- Unsure
- Disagree

Applied anatomy test

1. A 52-year-old woman underwent a right-sided radical mastectomy (surgical removal of breast) and excision of all axillary lymph nodes on the affected side due to breast cancer. Postoperatively, the nurse noticed that she had “winging of the scapula".

Which of the following nerves damage has caused this condition?

- Axillary nerve
- Medial pectoral nerve
- Musculocutaneous nerve
- Long thoracic nerve *****true
- Lower subscapular nerve

2. A boxer sustained a right-sided brachial plexus injury as a result of a fight. Neurological assessment revealed that abduction cannot be initiated, but if the arm is helped through the first 15°of abduction, the patient can fully abduct the arm. From this amount of information and your knowledge of the brachial plexus, where would you expect the injury to be?

- Axillary nerve
- Long thoracic nerve
- Radial nerve

- Suprascapular nerve *****true
- Ulnar nerve

3. A young man involved in a head-on car collision hit his flexed knee on the dashboard of the car. He was later found to have a major instability of the knee, in that his tibia could be moved posteriorly relative to the femur.

What of the following ligaments was likely damaged?

- Anterior cruciate
- Deltoid
- Lateral collateral
- Medial collateral
- Posterior cruciate *****true

4. A 43-year-old woman presents with lumbar pain and sciatica (pain over the sciatic nerve distribution), with associated numbness over the lateral border of the right foot and ankle, along with an absent ipsilateral ankle jerk.

Which spinal nerve distribution does this represent?

- L4
- L5
- S1 *****true
- S2
- S3

5. A 12-year-old skater is brought into A&E after a bad fall onto his right hand. He has been found to have a displaced spiral fracture in the area pointed to on the diagram.

Which of the following nerves is most likely to be damaged in this case?

- Axillary nerve
- Long thoracic nerve
- Median nerve
- Radial nerve *****true
- Ulnar nerve

6. A patient with a fracture to the area indicated was treated with a plaster cast. A few days later he started to develop progressive numbness over the dorsum of the foot and weakness in dorsiflexion. The cast was quickly changed and the signs were attributed to nerve compression.

Which one of the following nerves is most likely to be compressed?

- Common fibular *****true
- Femoral
- Obturator
- Sciatic
- Tibial

7. A 20-year-old woman is found to have the finding shown in the image. Which one of the following nerves is most likely to be injured that explains the examination findings?

- Femoral nerve
- Inferior gluteal nerve
- Obturator nerve
- Sciatic nerve
- Superior gluteal nerve *****true

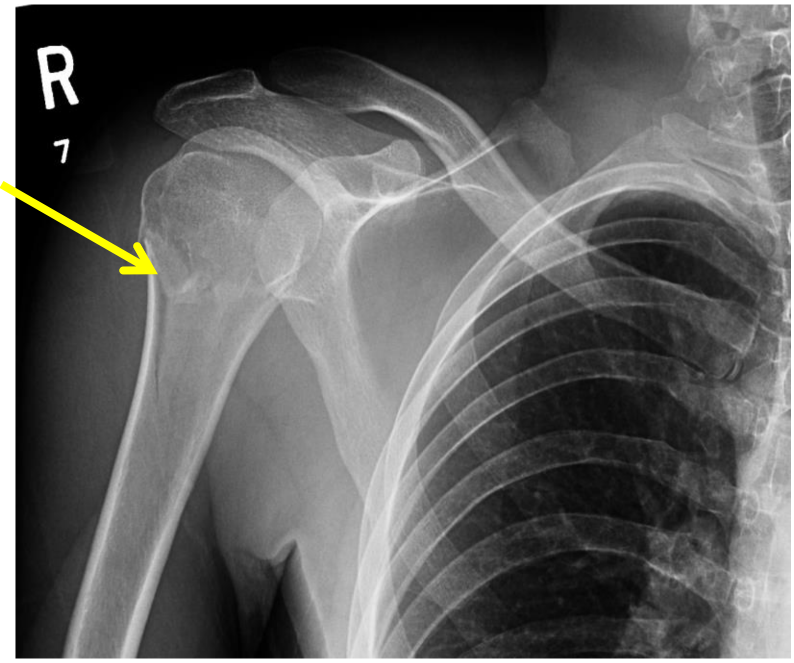
8. A 48-year-old woman presents with a 1-year history of numbness and tingling affecting the area shown in the image. On examination, she has signs of atrophy of the thenar eminence of the hand.

Compression of which structure is causing the patient's symptoms?

- Axillary nerve
- Median nerve *****true
- Radial nerve
- Thoracodorsal nerve
- Ulnar nerve

9. A front-seat passenger sustained a fracture of the area shown in the image as a result of hitting the dashboard with his shoulder during a high-speed head-on-collision.

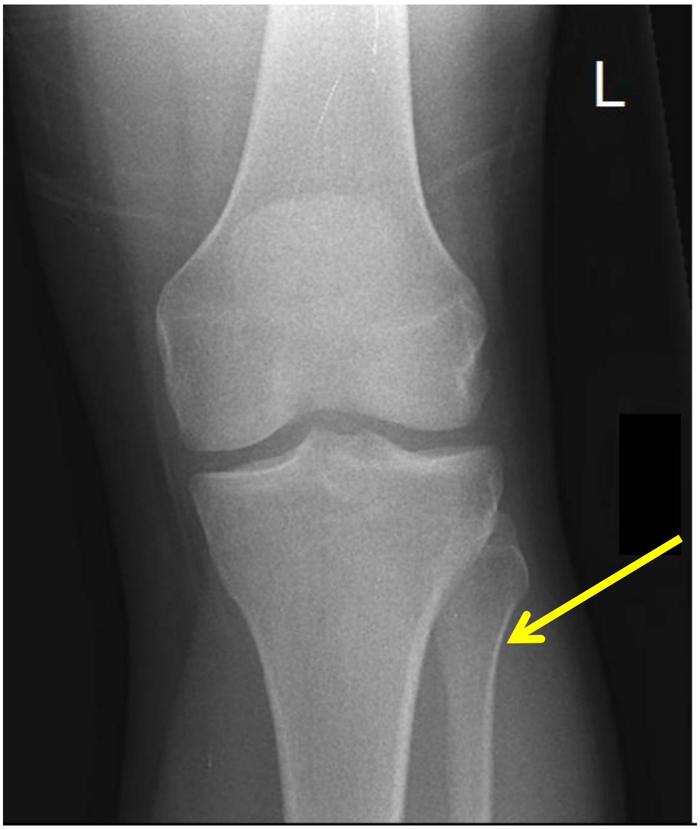
Innervation to which of the following muscles is most likely to be affected in this case?



- Biceps Brachii
- Brachialis
- Brachioradialis
- Deltoid *****true
- Teres major

10. A patient was brought to A&E with a fracture in the structure indicated on the image. After the bone healed, she had "foot drop", and so the foot flopped onto the ground during walking.

Paralysis of which of the following muscles would be associated in this case?



- Quadriceps femoris
- Flexor hallucis longus
- Popliteus
- Tibialis anterior *****true
- Tibialis posterior

11. A 60-year-old woman presents to her GP with knee instability, pain and swelling following an injury. There is an injury to the ligament indicated by the image.

Which of the following signs is most likely to be found on physical examination?

- Abnormal abduction of the knee
- Abnormal adduction of the knee
- Anterior displacement of the tibia on the femur *****true**
- Posterior displacement of the tibia on the femur
- Rotation of the tibia on the femur

12. In order to check the pulse of a teenager whose forearm is in a cast, the doctor presses his finger into the depth of the area marked.

The tendon lying immediately medial (ulnar) to the doctor's finger belongs to which of the following muscles?

- Brachioradialis
- Extensor carpi radialis brevis
- Extensor carpi radialis longus
- Extensor pollicis brevis
- Extensor pollicis longus *****true

13. During surgery, the surgeon decided to transect the anterior scalene muscle where it inserts on the first rib.

Which of the following structures is in contact with the anterior surface of the muscle that the surgeon must be careful of sparing?

- Inferior trunk of the brachial plexus
- Long thoracic nerve
- Phrenic nerve *****true
- Sympathetic trunk
- Vagus nerve

14. A patient is brought in A&E in respiratory distress. It is quickly decided to perform an emergency tracheostomy.

At what level could you rapidly create an airway below the vocal cords with a minimum danger of haemorrhage?

- Just below the cricoid cartilage
- Just above the jugular notch
- Just above the thyroid cartilage
- Just below the thyroid cartilage *****true
- Through the 3rd tracheal ring

15. A student came into A&E with photophobia, neck stiffness, high fever and a non-blanching rash. Tests reveal a high white blood cell count in her cerebrospinal fluid (CSF) taken from a lumbar puncture. She is diagnosed of bacterial meningitis. From which of the following structures/areas was the CSF taken?

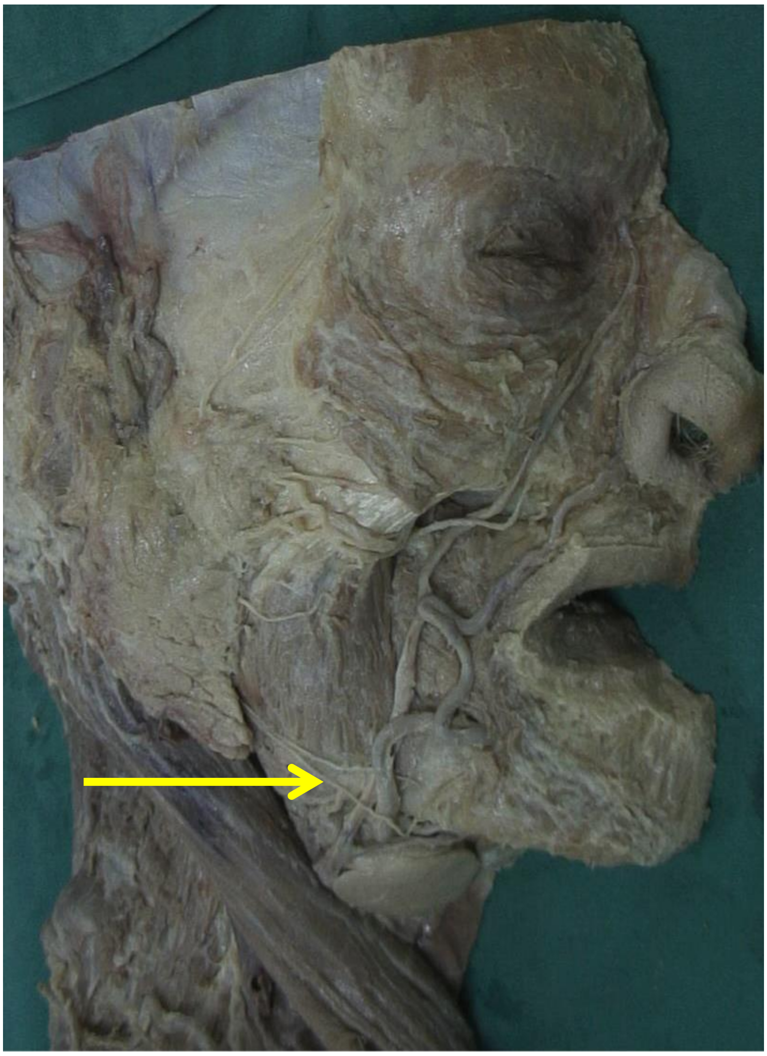
- Cavernous sinus
- Epidural space
- Subarachnoid space *****true
- Subdural space
- Verterbal venous plexus

16. A patient is known to have brain aneurysm presented with a sudden onset of "thunderclap" headaches, nausea and vomiting. He is diagnosed of subarachnoid haemorrhage. For reducing intracranial pressure, a lumbar puncture is done.

For this, the surgeon will have to pass through the skin, subcutaneous tissue, deep back muscles and then, in order, the:

- Anterior longitudinal ligament, ligamenta flava, epidural space, dura, subdural space, arachnoid, subarachnoid space
- Interspinal ligament, ligamenta flava, posterior longitudinal ligament, epidural space, dura, subdural space, arachnoid, subarachnoid space
- Intertransverse ligament, ligamentum flava, posterior longitudinal ligament, epidural space, dura, subarachnoid space, arachnoid
- Posterior longitudinal ligament, interspinal ligament, epidural space, dura, subdural space, arachnoid, subarachnoid space
- Supraspinal ligament, interspinal ligament, ligamenta flava, epidural space, dura, subdural space, arachnoid, subarachnoid space *****true

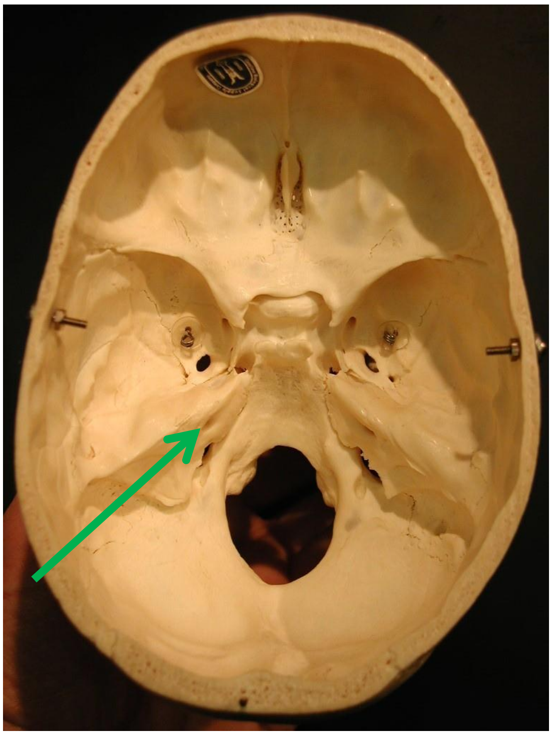
17. During a face lift operation, the plastic surgeon inadvertently cut the nerve indicated. Which of the following muscles would be paralyzed because of the injury?



- Buccinator
- Depressor anguli oris *****true
- Levator anguli oris
- Levator labii superioris
- Stylohyoid

18. A 64-year-old man was diagnosed with an acoustic neuroma (benign tumour of nerve cells) at the marked level in the image.

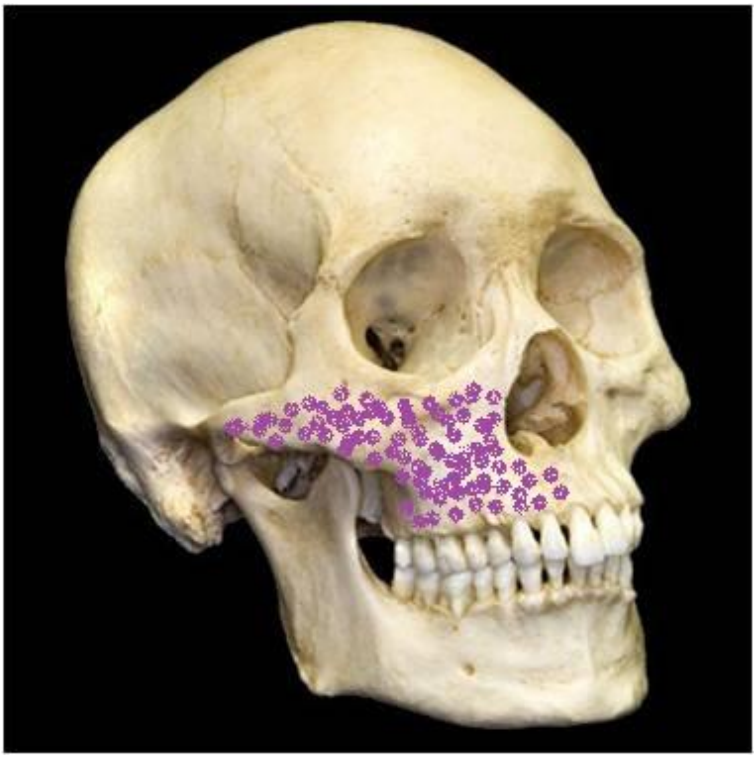
What other cranial nerve might also be affected since this nerve uses the same foramen?



- Abducens
- Facial *****true
- Glossopharyngeal
- Trigeminal
- Vagus

19. While recovering from multiple dental extractions, a patient experienced a radiating pain affecting the area indicated in the image.

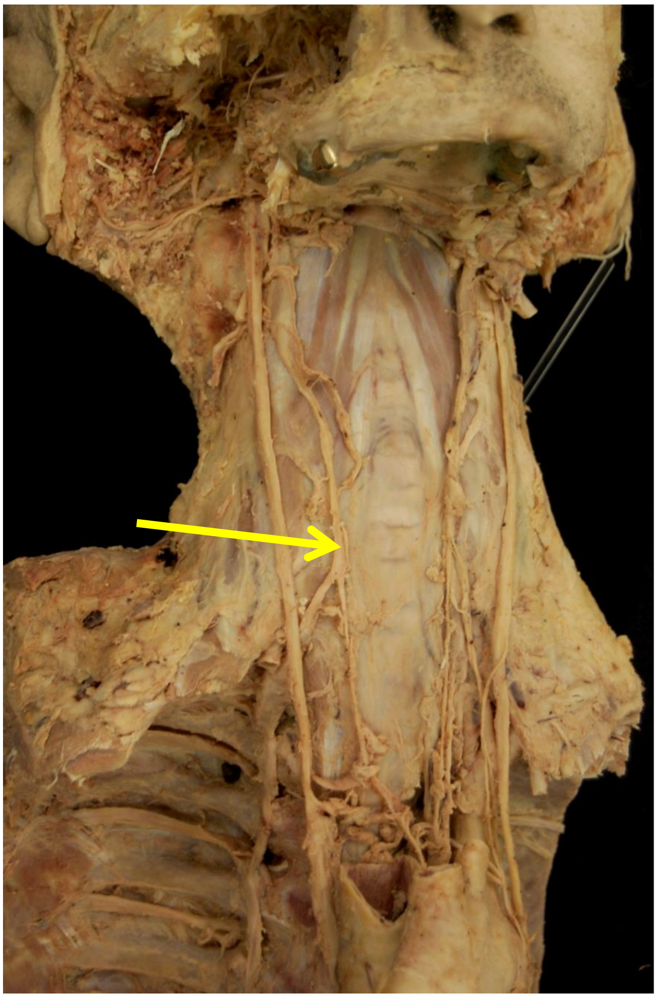
Which of the following nerves is involved?



- Facial
- Glossopharyngeal
- Ophthalmic division of trigeminal
- Mandibular division of trigeminal
- Maxillary division of trigeminal *****true

20. A 36-year-old man has a neck tumour which has compressed the structure indicated in the image.

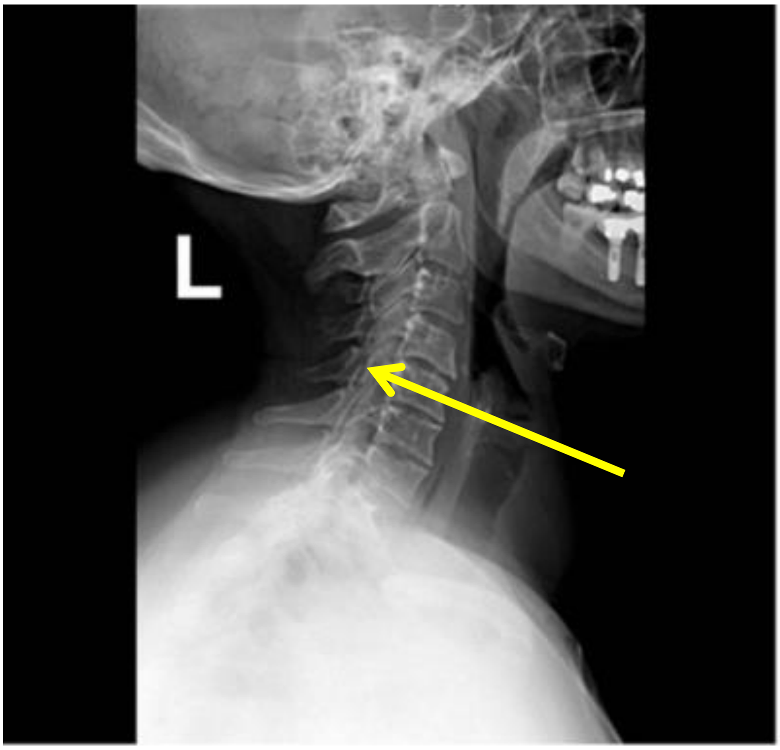
Which of the following physical signs would you expect in this case?



- Increased sweat secretion on the right side of the face
- Lateral deviation of the right eye
- Pale skin on the right side of his face
- Ptosis on the right eye *****true
- Pupil dilatation of the right eye

21. In a fall from a horse, a rider sustains a severe neck injury at the lower level of his neck. In addition to damage to the spinal cord, the transverse process of the vertebra marked in the image is fractured.

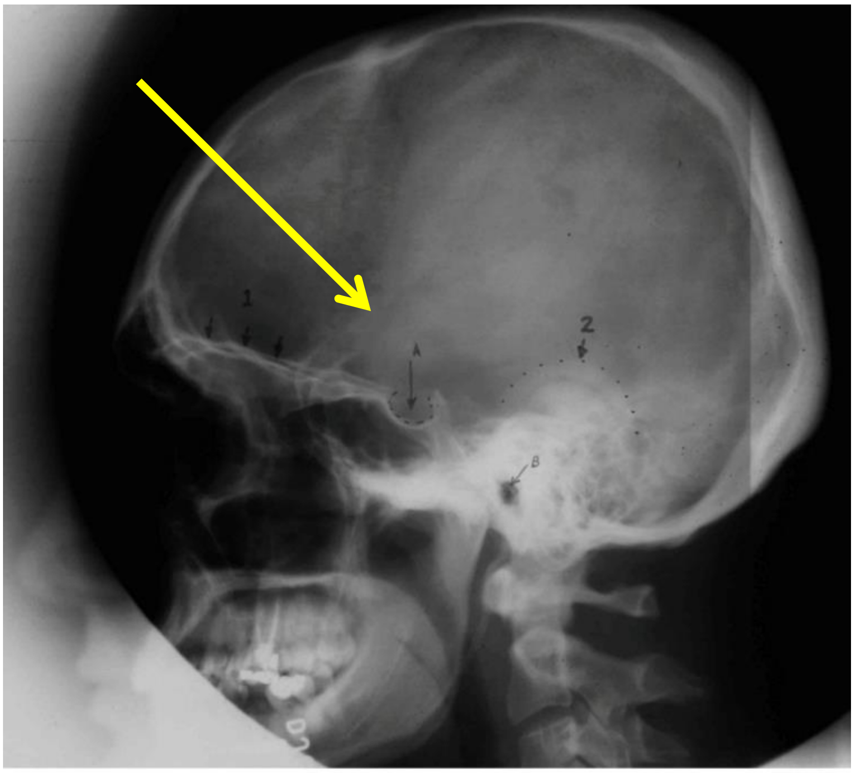
Which of the following arteries is endangered?



- Common carotid
- Costocervical
- Inferior thyroid
- Internal carotid
- Vertebral *****true

22. A man is hit by a baseball on the side of his head. He immediately loses consciousness, wakes up momentarily and then passes out. He is rushed to A&E and is immediately scanned. The scan shows a skull fracture at the site indicated in the image. He is rushed to surgery where he undergoes a Burr Hole surgery to relieve the pressure. After a few hours, he regains consciousness.

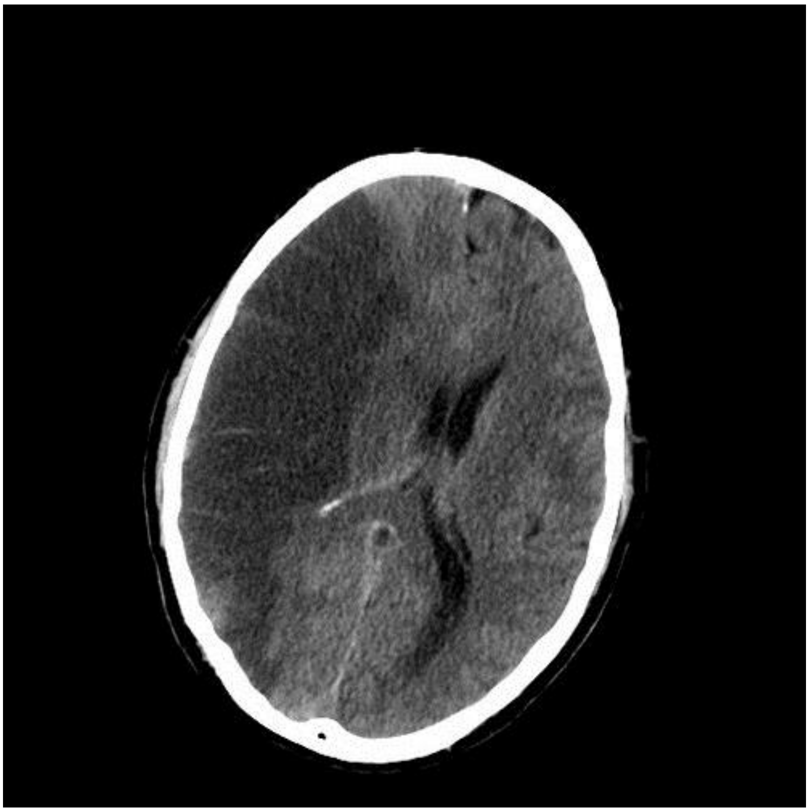
Which of the following best describes the haemorrhage from the fracture?



- Extradural *****true
- Intracerebral
- Subaponeurotic
- Subarachnoid
- Subdural

23. A 90-year-old man suffers a stroke resulting in left-sided paralysis. Computed tomography (CT) shows that the intracerebral haemorrhage has interrupted the blood supply to the area on the scan.

Which of the following vessels is affected?



- Anterior cerebral artery
- Middle cerebral artery *****true
- Middle meningeal artery
- Posterior cerebral artery
- Vertebral artery

24. A 32-year-old man presents with three months history of headaches, hearing loss, dizziness and other neurological deficit. A scan reveals a convexity meningioma as shown in the image.

Which of the following best describes the region of the body affected?

- Head, neck and tongue movements on the right side of the body *****true
- Head, neck and tongue movements on the left side of the body
- Leg and trunk sensations on the left side of the body
- Leg and foot movements on the left side of the body
- Leg and trunk sensations on the right side of the body

25. While observing a mastectomy (surgical removal of breast) on a 60-year-old female patient, you have been asked by the surgeon to help tie off the arteries that supply the medial side of the breast.

Which of the following arteries gives origin to these small branches?

- Internal thoracic *****true
- Musculophrenic
- Posterior intercostal
- Superior epigastric
- Thoracoacromial

26. A 78-year-old man suffers a myocardial infarction and is subsequently found to have a complication of complete heart block (that is, the right and left bundles of the conduction system have been damaged).

Which of the following arteries is most likely to be involved in this case?

- Acute marginal
- Anterior interventricular *****true
- Circumflex
- Obtuse marginal
- Posterior interventricular

27. A 45-year old woman presents with severe pain on the left side of her pelvis that radiates to the left upper medial thigh. She is diagnosed with endometriosis irritating the obturator nerve. The condition was treated surgically through a midline incision. Towards the end of the surgery the consultant asked a medical student to identify the layers of the abdominal wall (from inside to outside) that needed to be sutured.

Which of the following lists best describes these layers?

- Parietal peritoneum, visceral peritoneum, superficial fascia, skin
- Pelvic diaphragm, perineal membrane, perineal muscles, Colle's fascia, skin
- Peritoneum, external oblique, transverse abdominal muscle, superficial fascia, skin
- Peritoneum, internal oblique, transverse abdominal muscle, Scarpa's fascia, skin
- Peritoneum, linea alba, superficial fascia, skin *****true

28. A child presents in a paediatric ward with no cremasteric reflex (stroking of the upper medial thigh does not elicit testes retraction). Paralysis of which of the following muscles is mostly likely to be responsible for the lack of the reflex?

- Bulbospongiosus and cremaster
- Cremaster and dartos *****true
- Cremaster and internal urethral sphincter
- Cremaster and ischiocavernosus
- Detrusor and cremaster

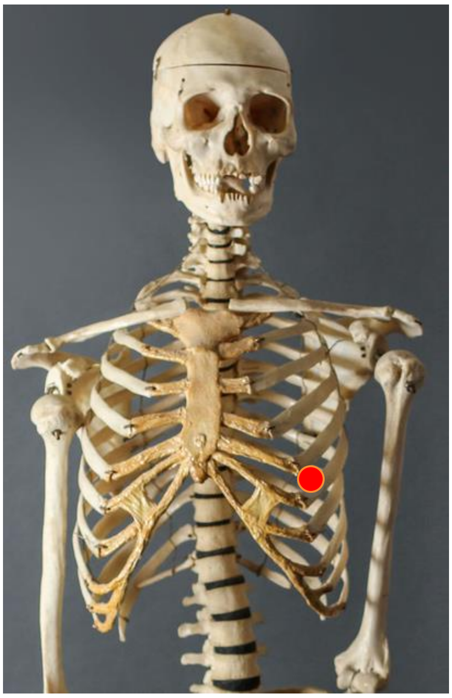
29. A 16-year-old lifts a large chest of drawers and as he lifts he feels a severe pain in the lower right quadrant of his abdomen. He finds that he can no longer lift without pain and the next day goes to see his GP. Surgery is indicated and during the surgery the surgeon corrects a sac projecting through the abdominal wall just above the structure joining point 1 and 2 in the image shown and lateral to the inferior epigastric vessels.

Which of the following is the cause of his condition?

- A congenital inguinal hernia
- A direct inguinal hernia
- A femoral hernia
- An incisional hernia
- An indirect inguinal hernia *****true

30. During examination of a 62-year-old man, you put your stethoscope on the area indicated (by red dot) in the image, and listen for a clearly audible pansystolic murmur. You hear it distinctly and know it must be associated with regurgitation of blood.

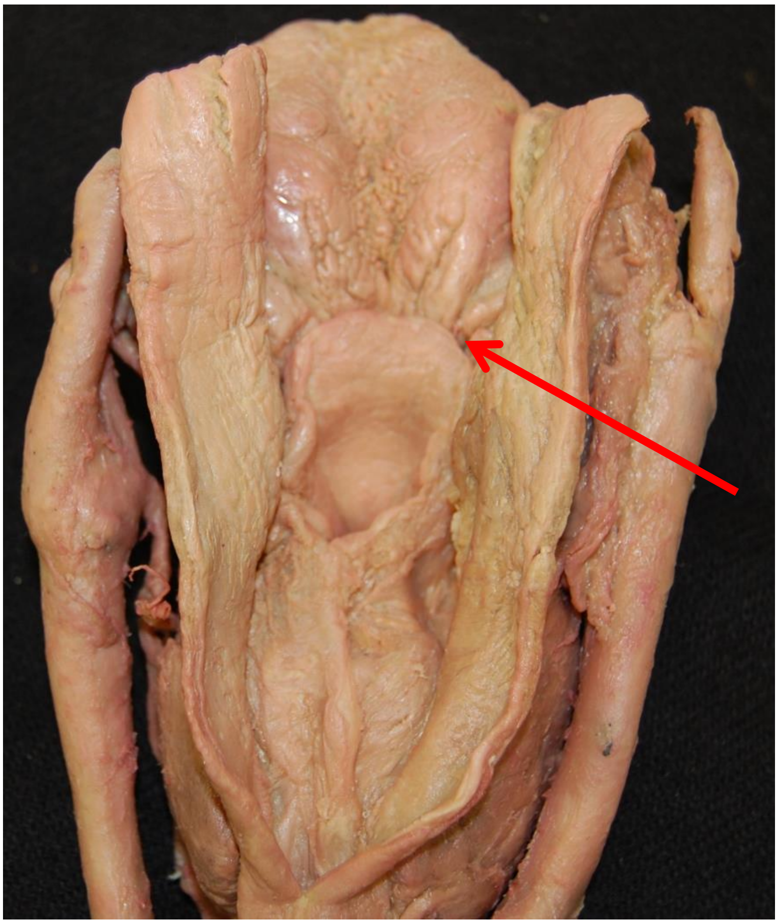
With the information provided, which one of the following heart valves is affected?



- Aortic
- Mitral *****true
- Pulmonary
- Semilunar
- Tricuspid

31. A 4-year-old girl is brought to A&E with a severe cough and discomfort in her throat. You are told by her mother that she had been playing with some beads and had apparently aspirated one. You noticed that the bead is stuck in the area marked in the image.

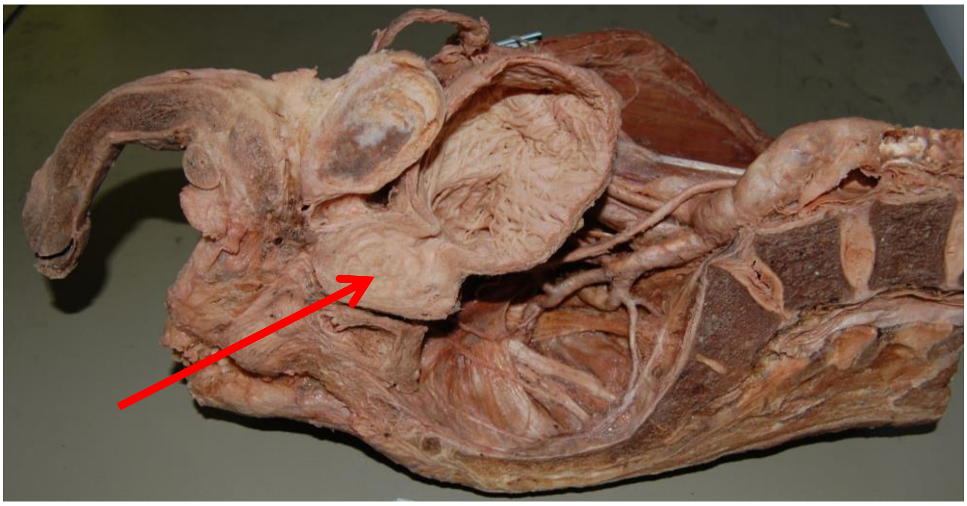
Irritation of which of the following nerves is mostly likely to cause the cough reflex?



- External laryngeal nerve
- Glossopharyngeal nerve
- Internal laryngeal nerve *****true
- Pharyngeal plexus
- Recurrent laryngeal nerve

32. A 69-year-old male is diagnosed with metastatic cancer. Thereafter a secondary malignant brain tumour is found on CT. However, despite best efforts the patient subsequently dies. An autopsy reveals tumour sites in the area indicated, the vertebral column and brain but no other organs.

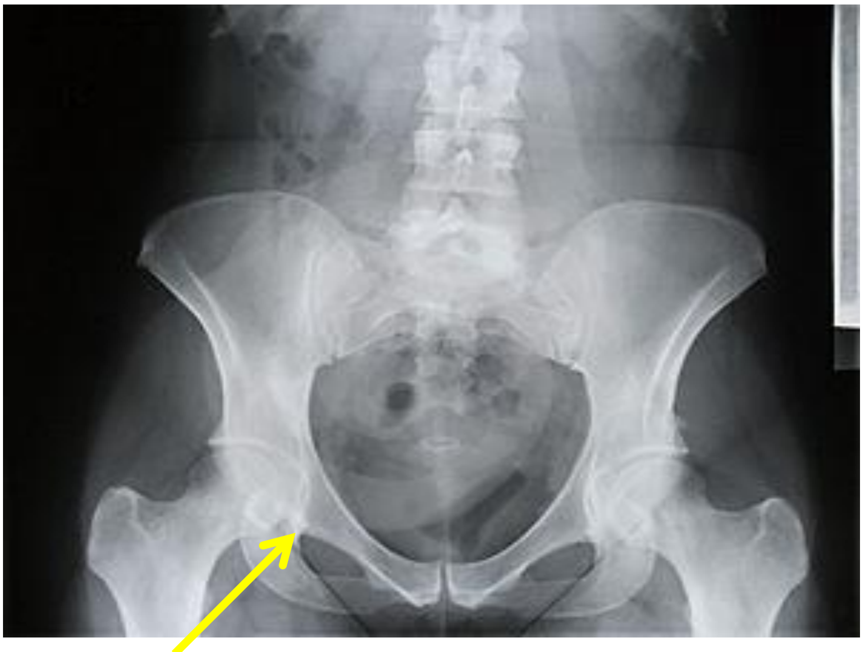
Which of the following structures caused the cancerous cells to reach the brain?



- Anterior spinal artery
- Azygos venous system
- Thoracic duct
- Vertebral artery
- Vertebral venous plexus *****true

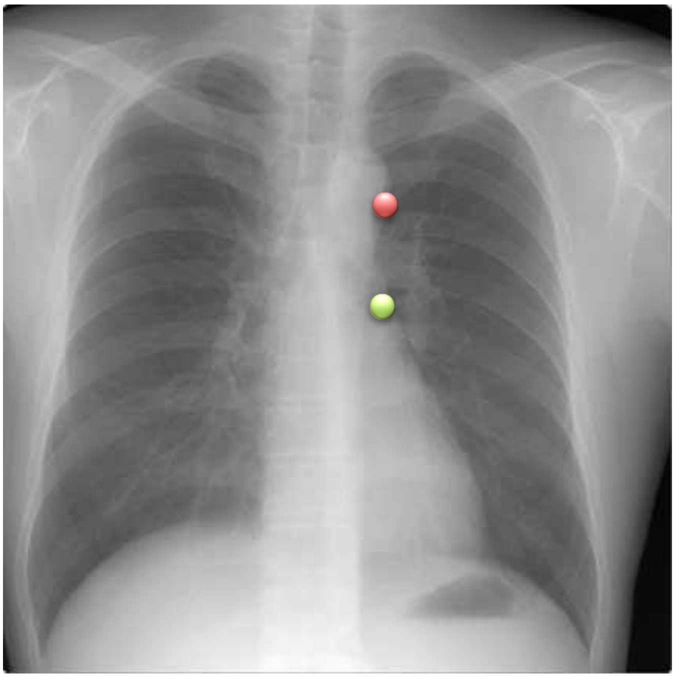
33. During childbirth a bilateral nerve block may be performed to provide anaesthesia to the majority of the perineum and the lower one fourth of the vagina. To do this the physician inserts a finger into the vagina and presses laterally to palpate the landmark indicated in the image.

Which of the following nerves is anaesthetised in this case?



- Femoral nerve
- Genitofemoral nerve
- Inferior gluteal nerve
- Obturator nerve
- Pudendal nerve *****true

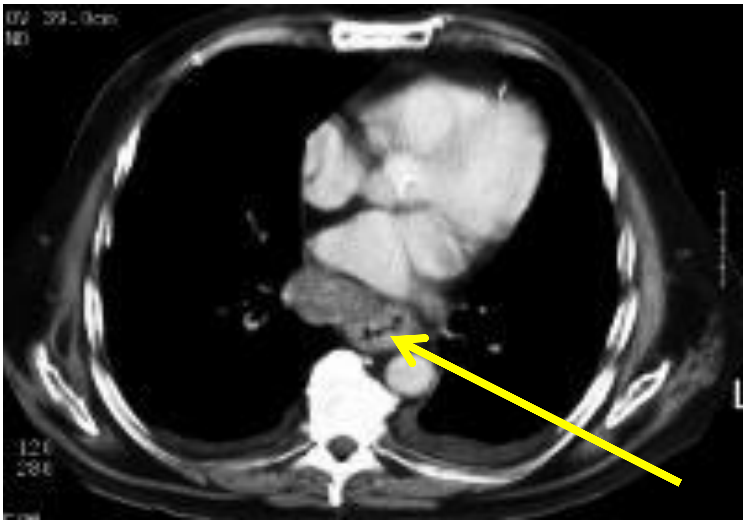
34. A 2nd-year medical student was doing her first physical exam. Which of the following heart sounds is she mostly likely to hear at the red dot and green dot in the image respectively?



- Aortic and pulmonary
- Mitral and aortic
- Aortic and tricuspid
- Pulmonary and tricuspid *****true
- Tricuspid and mitral

35. The structure indicated in the image was punctured from within. The patient subsequently developed an infection in the space around it.

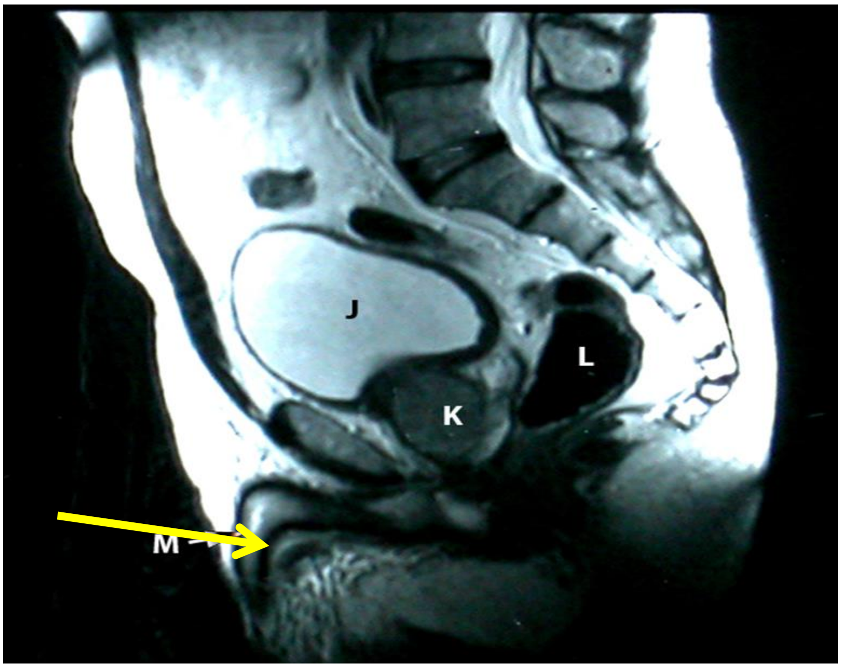
Which of the following best describes the space?



- Anterior mediastinum
- Middle mediastinum
- Pericardial cavity
- Pleural cavity
- Posterior mediastinum *****true

36. A 60-year-old man is brought to A&E after a forceful blow to his perineum subsequent to falling on a metal beam. An emergency MRI is requested. The structure marked by an arrow on the MRI is ruptured.

Which of the following best describes the site at which fluid (blood or urine) is most likely to accumulate?



- In the deep perineal pouch
- In the ischioanal fossa
- In the pararectal fossa
- In the rectovesical pouch
- In the superficial perineal pouch *****true

Ethics approvals

Direct Line: 020 8266 6397
Email:

30th March 2015

Dear Mandeep Gill Sagoo

PROJECT TITLE	Rethinking the assessment of applied anatomy knowledge of medical students: An investigation of the effect of visual resources, through contextually rich single best questions, on their performance and their views on anatomy
REC Reference	n/a
Reference	**SGREC15.0001**
Sponsor	Institute of Education, University of London
Principal Investigator (PI):	Mandeep Sagoo

Dear Mandeep

Research Ethics Committee Approval

Review
This study was approved by Charlie Owen (on 20/02/15) and by Sarah Crafter (on 24/03/15), both of the Institute of Education, University of London. Professor Nigel Brown, Director of the Institute of Medical & Biomedical Education and Pippa Tostevin (MBBS course director) University of London have confirmed their support for this study.

Ethical opinion
According to the SGREC process, this study was eligible for approval by the Joint Research and Enterprise Office on behalf of the Committee.

Therefore, a favourable ethical opinion of the above research has been given on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below. This approval is for 5 years from the date of this letter.

The list of documents reviewed and approved by the Research Ethics Committee (SGREC) under requirements of the Research Governance Framework are as follows:

Document	Version	Date
IOE Ethics form		Signed by Charlie Owen 20/02/2015 Signed by Sarah Crafter 24/03/2015
Consent form	1	February 16 2015
Participant Information Sheet	1	February 16 2015

After ethical review

Amendments

Amendments including changes to the study team, study design or supporting documents must be notified and approved by the committee before being implemented.

Annual Progress Report

You are required to submit an annual progress report within 30 days of the anniversary of your letter.

If your study is completed within a year of you first obtaining ethical approval, you should complete a declaration of the end of a study instead.

The annual progress report is a condition of you receiving ethical approval. If you do not submit an annual progress report, your study will be assumed to be closed, and ethical approval will be stopped.

Declaration of the End of a Study

You need to notify the Committee within 90 days of the termination of your study and within 15 days if you have terminated the study early.

All forms can be found on the SGREC website or from the SGREC coordinator (sgrec@sgul.ac.uk)

The SGREC website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

I wish you well in your research

[redacted]

Lucy H H Parker
Clinical Research Governance Manager
On behalf of Research Ethics Committee

CC Rachel Allen

1 Apr (4 days ago)

Dear Mandeep,

Confirming that approval to extend this project to …. student has been granted on the following conditions:

1. Students do not feel that they have to take part - and this should be clearly communicated.
2. The research occurs outside of core teaching time.
3. It is made clear how results of the formative assessment will be given to students and how long after the assessment? Also, will students receive just a score as feedback or can they go back and a refer to a master copy with answers and explanations? The latter would be preferred.
4. We couldn't see amongst the attachments the official ethics approval letter, please could we receive this.

Please could you provide a response to the above.

With best wishes,

…………………………………………..

Mandeep Gill Sagoo

Anatomy Demonstrator

Department of Biomedical Sciences

17 April 2015

Dear Ms Sagoo

RE. Rethinking the assessment of applied anatomy knowledge of medical students: An investigation of the effect of visual resources, through contextually rich single best questions, on their performance and their views on anatomy

Thank you for submitting an application together with the study protocol and associated material for review by the …. Ethics Committee. We noted existing ethical approval from the Ethics Committee of the Institute of education, University of London.

We are happy to approve your study and your intention to seek to recruit …. undergraduate students (MBBS) as participants.

Please consider these two comments from the Committee:

1. The participant information leaflet was commended for its clarity, but a small number of grammatical errors interfered slightly with its readability.

2. The participant information leaflet could suggest justifiably that the study might benefit participants by supporting their learning on the topics addressed.

Please ensure that the documents used in the study are equivalent to the attached referenced versions which you should retain for your records. If during the course of the project you need to deviate significantly from the above-approved document please inform me since written approval will be required. Please also inform me should you decide to terminate the project prematurely.

Please contact Dr Petra Newbound, ….. Research Manager, if you have any questions.

Yours sincerely

[redacted]

………………………………………………

17 APRIL 2015

Hi Mandeep,

Thanks for your email. I've been given ethical approval to run your test and also have the approval of the Assessment Leads. I'm am waiting for approval from the Vice Dean too though, so am keeping my fingers crossed.

If approval is granted then I think Tuesday 26th May is the best date to run the test. They have Life Science teaching that day (including medical imaging) so you could make a small announcement between the sessions if that works for you? I could also give them information about this in advance, so they're aware of the test. The 1st half of the year are in teaching from 10.30-12.30 so their test could run from 1pm? The second half have teaching 1-3pm so their test could run from 3.30pm? I have tentatively booked the IT suite for the afternoon in case this is all ok. What do you think?

I've attached some comments (yellow boxes) to the participant info you set me too. Also, is the questionnaire given on paper or as part of the tool? As I'm interested to know whether you're able to group test results by gender and ethnicity? On the questionnaire I noticed you don't mention plastic models or surface/living anatomy. I was just wondering if there was a reason for this please? I hope this all helps, but let me know if you need anything changed. I'll keep you posted about the approval from the Vice dean.

Best wishes,

…………………………………………………………..

Mandeep Gill

Institute of Education

University College London

14 April 2015

Dear Mandeep,

RE: ‘Rethinking the assessment of applied anatomy knowledge of medical students: An investigation of the effect of visual resources, through contextually rich single best questions, on their performance and their views on anatomy’ –…. College London external research request permission

I am writing with regard to your recent application for permission from the ….College London Research Ethics Office to undertake the above research study, as per our external research request procedure.

I can confirm that your application for permission has been accepted and that you now have permission to undertake external research using ….College London staff or students. Your permission has been granted by the Chair of the College Research Ethics Committee. Please note that the external research request procedure does not constitute ethical review, rather it is a permission procedure put in place to ensure that only ethically acceptable studies are carried out by …. College London staff/students and premises.

Please do not hesitate to contact the Research Ethics Office should you have any queries regarding the above.

Kind regards,

Research Support Assistant

…………………………………………………

25 June 2015

Dear Mandeep

I have met with the Ethic Approval Lead and she is happy with the content, outcomes and current ethics approval from UCL.

I am able to send out you're your student research information, however before I do please could you confirm how the students log their consent.

If you would like to draft a specific email for the current year 2 students I will forward it on your behalf, as you are probably aware the student have left for their summer break but they will be checking their emails during this period.

Kind regards

……………………………………………………

Thank-you very much!