The use of eye tracking technology and a digital training tool in radiographic image interpretation

by

Laura Mary McLaughlin

B.Sc., (Hons) Diagnostic Radiography and Imaging, Ulster University 2015

A THESIS

submitted in fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY (PhD)

I confirm that the word count of this thesis is less than 100,000 words.

Faculty of Life and Health Sciences
Ulster University

ULSTER UNIVERSITY
Northern Ireland, United Kingdom

2018

Dedication

To my parents for their never-ending trust and belief in my capabilities. The encouragement and

support they have given me has allowed me to come this far.

Without them, none of this would have been possible.

Thank you for everything.

The use of eye tracking technology and a digital training tool in radiographic image

interpretation

by

Laura Mary McLaughlin

B.Sc., (Hons) Diagnostic Radiography and Imaging, Ulster University, 2010

A THESIS

submitted in fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY (PhD)

Faculty of Life and Health Sciences
Ulster University

ULSTER UNIVERSITY

2018

Approved by:
Dr. Sonyia McFadden
Dr. Ciara Hughes
Dr. Raymond Bond
Dr. Jonathan McConnell

Copyright

Abstract

This thesis investigated the efficacy of eye tracking technology and a digital training tool in radiographic image interpretation. A systematic review was performed to investigate the performance of reporting radiographers completing chest image interpretation following training. The quality of evidence published in this area was high. The role of image interpretation differed between studies, ranging between: image red dot abnormality highlighting, image comment and clinical reporting. A comparison of image interpretation skills of radiographers across a range of experience was completed using eye tracking technology. Reporting radiographers trained in MSK image interpretation demonstrated statistically significant accuracy rates (p≤0.001), and confidence levels (p≤0.001) and took a mean of 2.4 s longer to clinically decide on an image compared to students. Reporting radiographers also had a statistically greater accuracy rate (p≤0.001), were more confident (p≤0.001) and took longer to clinically decide (14 s on average) on an image diagnosis (p=0.02) than radiographers. Eye tracking patterns presented within heat maps, were a good reflection of group expertise and search strategies. Eye tracking metrics were indicative of participant performance and reflected the different search strategies that each group of participants adopted during their image interpretations. A digital training tool for use in chest image interpretation was created based on evidence within the literature, using expert input and two search strategies previously used in clinical practice. Images and diagrams, aiding translation of the tool content, were incorporated where possible. Improvements were seen in interpretation performance and confidence (p<0.05). There was a decrease in FP values and increase in TN values seen in the intervention group (p<0.05). This tool therefore has the potential to be used as a training tool in chest image interpretation for reporting clinicians and healthcare professionals. This work may contribute to improving diagnosis and help reduce reporting times.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly I would like to acknowledge and thank my supervisors, Dr Sonyia McFadden, Dr Ciara Hughes, Dr Raymond Bond and Dr Jonathan McConnell, for the expert guidance and support they provided throughout my PhD journey. They have gone out of their way to provide me with prompt feedback and I'm very grateful for all of their encouragement. Thanks also to my internal examiners Dr Stephen Gallagher and Dr John Cathcart for their constructive feedback and expert advice.

I would like to give a very big thank you to Dr Nick Woznitza for supplying the images for this project, for his help in recruitment and for his contribution to the eye tracking material of the tool. A thank you also to his colleagues at Canterbury ChristChurch University for helping me to complete recruitment. I would like to thank Dr Ayman Elsayed for his help in forming the eye tracking section of the tool.

I would also like to acknowledge and thank the following people: Mrs Mary Rose Holman, subject assistant librarian for her expert guidance and patience in developing the search strategy for my systematic review (Chapter 2); Dr Ian Bradbury, for his simplistic statistical advice in Chapter 3 and Chapter 5.

This thesis would not have been possible without the participation of the radiographers, radiologists and academic staff. I would like to express my sincere thanks to all those who participated in the studies and survey. Thanks also to Canterbury Christ Church University, Reporting Radiographers Interest Group Scotland and to the Reporting Radiographers Group Northern Ireland for being so accommodating when allowing me to conduct the studies. Their help in recruitment could not have been done without.

I gratefully acknowledge the receipt of funding provided through the Department of Employment and Learning Award (DEL); this award provided me with many opportunities and experiences that enhanced my overall PhD experience.

I would like to thank my family-Daddy, Eilish, Brian, Sarah and Anna- who never let me complain about being stressed for too long and left me alone when I needed to get my head down. They have been a constant source of support and inspiration during the inevitable highs and lows of a PhD journey. A special thanks to my brother and sisters, Brian, Sarah and Anna. Sarah's wedding planning was a great distraction in the last few months. A massive thank you to their partners Breege, Dermot and Ed for making the holidays something to look forward to.

Thanks for the facetimes Anna, always nice to have as an evening break in the office or someone to de-stress with after a day's work. To my Daddy, John, whether it be my car stuck in Jordanstown car park or diesel needed to complete recruitment, you managed to sort it from Donegal. To my mum, Eilish, you are the strongest and most self-less person I know. I appreciate all of the sacrifices you made to allow me to get here and the many phone calls of a stressful Laura you had to listen to. The many trips to put a forgotten laptop charger, keys or pen-drive on the 212 was also appreciated! As you would say 'A problem isn't a problem if money can fix it', even if I was stuck in Heathrow airport at 2am or Vienna!

To all the girls in the office, thank you for the extended lunch breaks, 'threesies' and chats on all things non PhD related. I am especially grateful to all of the girls; Sarah, Adrienne, Orlagh, Alanna, Teresa, Beverly and of course Aoife for the early morning company! You have made this PhD so much more enjoyable and help me make light of the trying times. Thank you to my nextdoor neighbor Peter for the help with statistics, it also can't have been easy listening to me moan about something or other every day. To my friends from home, Stephanie, Amy, Maria and Shannon, thank you for the nights out, catch ups and gossips, all a welcome distraction. Last but definitely not least, thank you Michael. Thanks for keeping me sane with the welcome distractions and much needed breaks from studying.

Achievements

Article publications:

L McLaughlin, R Bond, C Hughes, J McConnell, S McFadden,
Computing Eye Gaze Metrics for the Automatic Assessment of Radiographer Performance during X-ray Image Interpretation, International Journal of Medical Informatics, 2017.

L McLaughlin, J McConnell, S McFadden, R Bond, C Hughes,
Methods employed for chest radiograph interpretation education for radiographers: A systematic review of the literature, Radiography 2017.

L McLaughlin, N Woznitza, A Cairns, S McFadden, R Bond, C Hughes, A Elsayed, D Finlay, J McConnell,
Digital training platform for interpreting radiographic images of the chest, Radiography 2018.

Conference presentations:

L McLaughlin, R Bond, C Hughes, J McConnell, N Woznitza, A Elsayed, A Cairns, D Finlay, S McFadden,
Development of a digital platform to aid chest radiographic image interpretation, European Congress of Radiology, 2017.

L McLaughlin, S McFadden, J McConnell, R Bond, N Woznitza, A Cairns, A Elsayed, C Hughes,
Trial of a digital training tool to support chest image interpretation in radiography, European Congress of Radiology, 2018.

L McLaughlin, C Hughes, R Bond, J McConnell, N Woznitza, A Cairns, A Elsayed, S McFadden,
Digital training platform for chest image interpretation: an RCT, 20th Internation Society of Radiographers and Radiological Technologist World Congress, 2018.

Conference posters:

L McLaughlin, S McFadden, C Hughes, J McConnell, R Bond,

The use of eye tracking technology to assess radiographer interpretation of X-ray images, UK Radiological and Radiation Oncology Congress, 2017.

L McLaughlin, S McFadden, R Bond, J McConnell, N Woznitza, A Cairns, A Elsayed, D Finlay, C Hughes,

The effect of a digital chest image interpretation training tool on interpreter performance, Clinical Translational Research and Innovation Centre 8th Annual Translational Medicine Conference, 2017.

Awards and nominations:

Awarded a bursary to attend the UK Radiological and Radiation Oncology Congress 2017.

Finalist of The Belfast HSC Trust Science driving innovation in healthcare delivery award in the Advancing Healthcare Awards Northern Ireland 2017.

Chapter 1 - Introduction

1.1 Medical image interpretation

Medical image interpretation is the process of viewing of an image to form a judgement, based on the viewer's knowledge, on the content of the image and whether pathology or trauma is present. Pathology and trauma will be considered and discussed as one from here on in. There are many levels of medical image interpretation completed by reporting clinicians, each associated with a level of responsibility and impact on patient care and diagnosis. Historically the role of radiographers has progressed through the levels of image interpretation outlined below.

Red dot system

Radiographers are tasked with annotating an image with the words 'red dot' where they believe pathology to be present within the image. Radiographers are asked or required to apply a 'red dot' to images within clinical practice according to department protocols (The Society and College of Radiographers (SCoR) 2018).

Preliminary clinical evaluation

Radiographers make a judgement based on the images they encounter whilst working in clinical practice. This role has been developed following the implementation of the red dot system as mentioned above. This process has many advantages as it enhances the practice of the radiographer, is of benefit to the referrer and can overcome ambiguities associated with the red dot system (SCoR 2018). A comment is provided in written form based on the radiographer's judgement of the image however, the image will then receive a full written report by a reporting clinician. In addition, preliminary clinical evaluation may expedite the pathway of a patient if a life threatening or high risk pathology is identified.

Clinical reporting

Clinical reporting involves the formation of a diagnosis and explanation of the trauma/pathology if present. Radiographers and other qualified reporting clinicians who have received accredited training at postgraduate level produce a diagnostic report on images. The quality of the report should meet agreed 'gold standards', irrespective of the professional background of the reporting clinician (SCoR 2018).

At present radiographers who are qualified and authorised to do so, provide written reports on an image following the completion of a postgraduate programme at Masters degree level. Within

radiography this referred to as 'advanced practice'. The postgraduate programme will provide qualifications for an individual to complete image interpretation within a specified imaging modality and anatomical area.

Clinical reporting of chest image interpretation has been implemented in England by the radiography profession. Since this implementation, strong evidence suggests that this role greatly reduces waiting times for patients and delays in patient care, however this could be further enhanced if all regions of the United Kingdom (UK) were involved (Piper et al. 2014; Woznitza et al. 2014). Therefore, this project focuses on this level of image interpretation as the author hypothesises that it has the potential to provide the greatest results and impact directly to the patient pathway.

From here on, reference to image interpretation is the clinical reporting role of medical image interpretation completed by clinicians in practice.

1.1.2 Effects of mis-diagnosis and delay in diagnosis

Medical image interpretation is a method used to detect and localise pathologies within the body. In many cases if these pathologies were not recognised or treated the consequences could be fatal or life inhibiting. Discrepancy in radiological reporting, which is confirmed by further imaging/treatment later on in patient management, ranges from 3-30%, with an estimated error rate of 3-6% per observer (Brady et al. 2012; The Royal College of Radiologists (RCR) 2014). Developments in technology have led to a variety of imaging investigations necessitating the learning of high standards in reporting skill with associated educational foundations (RCR 2012; SCoR 2010).

The pressure of health service demand can in turn negatively influence healthcare staff and image interpretation services. For example junior doctors being asked to carry out specialist radiology work has led to the failure of lung cancer cases being identified in a UK hospital (British Broadcasting Corporation (BBC) 2017). This report also acknowledged the delay in image reporting where 23,000 images acquired in the previous 12 months were not provided with a full official report by an authorised reporting clinician. Unfortunately errors such as these are not uncommon; seven serious cases including cancer were also missed by a consultant radiologist in Ireland (The Journal 2017; The Irish Times 2017). Following this, a review of

46,000 radiographs, computed tomography (CT) and Ultrasound (US) scans were completed to ensure no further errors were made. It is errors such as these which add to the delay in initiating appropriate patient treatment management.

In many cases the detection of lung pathology, often revealed on the simple projection chest radiograph prevents death and improves the chances survival. Many common errors in plain chest radiograph interpretation are frequently repeated, with discrepancy levels continuing to exist at the levels indicated above (Turkington et al. 2002; Wu et al. 2008; Nueman et al. 2012). Observer error is the most common mistake in image interpretation. This error type is particularly common in lung cancer detection whereby 90% of these cases occur on chest radiographs (World Health Organisation 2015; Del Ciello et al. 2017). It may be difficult for reporting clinicians to distinguish or identify a lung nodule or other chest pathology such as pneumonia, atelectasis or consolidation from bone, blood vessels and overlying organs. Limiting the errors made in detecting such diseases and pathologies can reduce time delays to patient care and improve patient outcomes, this can only be achieved through improving the imaging service provided (Krupinski et al. 1990; Krupinski 2000).

1.2 Reporting clinicians

Reporting clinicians of two professions are examined within this project; radiologists and radiographers.

1.2.1 Radiologists

Radiologists are qualified medical professionals who choose to specialise in the field of medical image acquisition and interpretation. Radiologists learn an in depth knowledge of all imaging modalities during their five years of training within the hospital environment (RCR 2018). Mentored by a consultant radiologist, they are normally based within the radiology department, completing and reporting on a range of specialised imaging examinations. For example, tumour biopsy, CT, MRI, NM as well as plain radiography.

1.2.2 Radiographers

After losing the reporting function to the radiologists (it was previously the medical radiographer) in the mid 1920's , the role of the radiographer has progressed since the mid 1990's to return to image interpretation in a specialised field of practice following a masters level postgraduate education. Specifically, this new role has been coined the 'reporting radiographer' (SCoR 2006). This training provides radiographers with the knowledge to interpret images within a given scope of practice and provide a written report. Stringent audits and evaluations of radiographer reporting performance occur during training, frequently with a radiologist mentor. This is followed by regular audit of performance in practice to ensure a high standard of patient care and safety is retained.

The range and number of examinations to be reported on exceeded the workload capability of the radiologists. This shortage of radiologists identified the need for role progression for the radiographer, this was then established in the mid 1990's with the approval of the professional Society of Radiographers (SCoR 2010). The reporting radiographer is deemed capable of reporting once appropriate postgraduate training is completed. This authorises them to provide a radiological report alongside the radiologist in their qualified area but only in departments where these advanced practitioner roles have been established and agreed with management (Paterson et al. 2004; Piper et al. 2005; Donovan and Manning 2006). Since then radiographer reporting accuracy and competency has been tested and proven to be of a high standard. Reporting radiographers performed better when providing a correct diagnosis (99%) than trainee radiologists (94%) when reporting on appendicular radiographs (Buskov et al. 2013). Previous authors have also highlighted that radiographers had the least overall confidence when compared with nurse practitioners and casualty officers yet the highest accuracy (Coleman and Piper 2009; Burke et al. 2013; Brealey et al. 2014). Radiographers have demonstrated an ability to work at an average speed of 47 seconds per image report which is similar to that of the radiologists who work at 43 seconds per report (Brealey et al. 2005). The advancing role of the radiographer has been welcomed by many; appendicular reporting by radiographers was featured in 57/143 radiology departments within the UK (Benger et al. 2003; Paterson et al. 2004) and also radiographers were reporting on 59% of plain radiographs within one radiology department (Woznitza et al. 2014). The high percentage of radiographs reported by radiographers within the

study of Woznitza et al. (2014) was believed to have been due to the greater number of radiographers reporting on more difficult examinations such as abdomen and chest radiographs (Woznitza et al. 2014).

Radiographer reporting of musculoskeletal images has been established throughout the UK and is a widely accepted role (Brealey et al. 2005; Piper et al. 2005). The accuracy and competency of reporting radiographers has been assessed in various studies and is consistently shown to be of a high standard (Brealey et al. 2005; Judson et al. 2009; Henderson et al. 2013; Woznitza et al. 2014).

Radiographer reporting of chest image interpretation has been established in England but there are concerns by the RCR on the role progression; RCR state that "reporting by non-medically qualified healthcare staff should involve examination types with a single organ investigation or a single suspected pathology" (RCR, 2006). Despite the concern of the RCR, a small number of published studies have evaluated the competence of reporting radiographers on chest image interpretation and have provided evidence of their ability to complete this role to a high standard subsequent to the appropriate training. Piper et al. (2014) featured the Objective Structure Examination (OSE) results of a postgraduate programme in clinical reporting of adult chest images. The 40 radiographers had a mean sensitivity and specificity of 95.4% (95% CI 94.4%-96.3%) and 95.9% (95% CI 94.9%-96.7%), respectively. Woznitza et al. (2014) completed an audit on the performance of a reporting radiographer investigating 99 chest images. The study demonstrated a very high concordance between the radiographer's reports and the radiologist interpretation (92-96%) (K> 0.8) and any discordant interpretations were noted as minor. Overall there were 8 differing reports on 7 chest images and one significant abnormality missed. This evidence supports the concept that once radiographers are provided with a high level of training they can perform the role effectively.

Postgraduate programmes in reporting radiography have exemplified the level of accuracy which students are obtaining post qualification (Piper et al. 2014; Woznitza et al. 2014). It is important that training is continuously evaluated and that high standards are met. The literature within this specified field is limited and so it is imperative to provide high quality studies to reinforce and monitor this role progression.

1.3 Chest image interpretation

The role of radiographers reporting on chest and abdomen radiographs has created controversy. Complexity generated by multiple overlying organs and the pathologies associated within the radiographic two dimensional representation are subtle, with often severe consequences of incorrect diagnosis and patient management. The reporting of chest and abdominal radiographs by radiographers has only been established within England and Wales in the UK, whilst Northern Ireland and Scotland continue to restrict the role of the reporting radiographer (Woznitza 2014). This restriction could account for 90% of reporting radiographers in the UK practicing within England, with 3 per hospital site in England as opposed to 1.8 per hospital site in the UK excluding England (Snaith et al. 2015). The advanced role of the radiographer also created scepticism and a difference of opinion amongst healthcare staff (Brealey et al. 2002a). The idea that radiographers will be forever limited in their scope of practice due to the absence of a medical degree was proposed by Donovan et al. (2006) and the caution surrounding their progression of reporting on further examinations and anatomical regions has continued. Despite the evidence of radiographers reporting to the accuracy of a radiologist, the radiologists continue to recommend that "reporting by non-medically qualified healthcare staff should involve examination types with a single organ investigation or a single suspected pathology" (RCR 2006), they also state that healthcare professionals who do not have the benefit of a medical degree should 'work in a team with ready access to a fully trained radiologist for advice'. The RCR also suggest the examination types which non-medically qualified healthcare professionals should be reporting be of a 'yes/no answer' (RCR 2006). They also go further to say radiologists and the hospital trust have a duty of care to ensure that non-radiologists undertaking reporting do not work outside their level of knowledge and expertise. The SCOR continue to support the development of the profession in this role and support the high standards maintained within it (SCOR 2010).

1.4 Image interpretation training

1.4.1 Tested methods/devices/systems to aid image interpretation training

There have been various systems and devices tested for their effectiveness in the education of healthcare and medical staff (Lison et al. 2004). Simulation-based training was investigated for its feasibility as both an education and assessment tool. Non radiology healthcare trainees

underwent training and assessment of pulmonary nodule identification at simulated radiology workstations as opposed to training via conventional methods (Aufferman et al. 2015). Participants made a significant improvement in performance following the training (P=0.015) and indicated a preference for this simulation based training over the conventional training in all five questions of the follow up evaluation questionnaire. P values of statistical significance for all questions were less than 0.01 (Aufferman et al. 2015).

Within a separate study, ten radiographers were given access to a CD ROM with various exercises and readings to complete. The CD ROM was accompanied by interactive workshops and tutorials which allowed radiographers to develop their skills in abnormality description of adult appendicular musculoskeletal trauma images. Ultimately an improvement was seen in accuracy (Freidman p=0.030/Wilcoxon p=0.021) and sensitivity (Freidman p=0.023/Wilcoxon p=0.012) by the third testing period (McConnell et al. 2012). Medical students were given access to either adaptive tutorials, online intelligent tutoring systems that deliver a personalised learning experience, or an existing peer-review web resource on the appropriate use and interpretation of common diagnostic imaging investigations (Wong et al. 2015). It was concluded that the group given the adaptive tutorials obtained significantly higher assessment scores. Better engagement was found when using the adaptive tools compared to the web resource and students rated the tutorials as a more valuable tool for learning about diagnostic imaging (Wong et al. 2016). Students also rated the tutorials as a significantly more valuable for learning (Wong et al. 2016). A web based interactive tutorial for radiology residents, which featured a checklist and a radiology consultant's guidance when viewing polytrauma CT scans, was developed and ease of use only was investigated (Schlorhaufer et al. 2012).

A chest x-ray (CXR) image reference set with tuberculosis was provided to medical professionals pre and post reviewing an image test (Waitt et al. 2013) comprising of images which contained a spectrum of Pulmonary Tuberculosis (PTB) appearances in adults. Results showed that using the reference set increased the number of correct decisions made by doctors to treat PTB, from a mean score of 60.7% to 67.1% (p=0.054). This led Waitt et al. (2013) to conclude that image reports from a reference standard or expert may help teach pathology appearances or improve image interpretation performance. McEvoy et al. (2017) developed an

alternative approach that investigated peer review as a learning and feedback tool in radiology reporting. Students were asked to complete radiology reporting and peer review at the end of a 5 week course. However, only a weak positive correlation ($r = 0.32$) was found between peer review scores received by students and the scores they obtained in a Multiple Choice Examination (MCQ). This may have been a reflection on the lack of expertise/experience for contribution to a student peer review process.

The web based decision support system developed by Wang et al. (2011) provides further evidence on the methods available to assist the learning of reporting clinicians. The web based decision support system uses features-based prediction modelling and a standardised reporting mechanism for differentiating between benign and malignant vertebral compression fractures based on Magnetic Resonance Imaging appearances. A checklist was implemented within the decision support tool and by answering and submitting the checklist a standardised image report is generated (Wang et al. 2011). The support system developed was specific to malignant vertebral compression fractures. The prediction model also supplies a figure of probability that the vertebral compression fracture was due to malignancy.

1.4.2 Search strategies implemented in image interpretation

A search strategy within image interpretation is a method employed to ensure that all aspects of the image have been checked for abnormal features (Williams 2013). Currently there is no standard systematic approach to chest image interpretation. Search strategies used by healthcare professionals and in particular within image interpretation are often based on a variety of guidelines and sources or otherwise 'self-taught' (Radiology Masterclass 2016). There are published guidelines and websites which make recommendations on how to complete an image interpretation, on how to systematically search the image, and which offer tutorials on the topic (The Society and College of Radiographers and Health Education England 2010; RCR 2011; Tamaklo 2012; Williams 2013).

Often trainee radiologists and reporting radiographers combine advice given in this guidance and a variety of recommended search techniques to form their own search strategy in image interpretation. Previous literature has stated that development of a search pattern comes only

after the development of knowledge on normal, abnormal, normal variations, characteristic features and location of individual pathologies (Kundel et al. 1972). It was also concluded that improvement of a search pattern can be achieved by concentrating on clearly defining the abnormal feature that the image interpreter is searching for and the normal tissue surrounding it (Kundel et al. 1972). In Hughes et al. (1996) at least 25/26 radiographers acknowledged tutorials incorporating a search strategy as useful, however many (n=19) claimed to have a technique for looking at chest radiographs already and therefore the true helpfulness of the search strategy implemented may be flawed. A study (Kundel et al. 1975) allowed participants to view chest images for either 0.2 second flash or either an unlimited viewing time where free search could be completed. True positives increased from 70% to 97% when viewers were given an unlimited time to interpret the image. Given the high overall accuracy when free search was undertaken, it could be questioned whether a search strategy is required. This study also identifies total search strategy as an ordered sequence of interspersed global and checking fixations (Kundel et al. 1975). It was found that although many image interpreters were taught to be systematic, directive and compare both bilateral image features, such as the right upper lobe with the left upper lobe of a CXR, that many images were read by a free search method and less than 4% of the visual activity was made up of bilateral comparisons (Carmody et al. 1984).

1.4.3 Checklists in image interpretation and healthcare

Checklists have proven to be a valuable resource within the healthcare settings. Within a systematic review, eleven of fifteen studies found a benefit of e-checklist use with regard to measured outcomes, three studies found mixed benefits on outcome measures and only one study found no benefits of an e-checklist (Kramer et al. 2016). 'Checklist fatigue' was also proposed as a possible detraction and disadvantage to the use of checklists within the healthcare system, the possible impact of this was not conclusive with only three studies focusing on the long-term impact of e-checklists. A comprehensive framework, which incorporates a safety checklist, has helped identify and mitigate hazards, standardise practice and ensure accountability (Herzer et al. 2009). A checklist has been used and implemented within a decision support tool, where each feature that was used to differentiate between benign and malignant vertebral compression fractures was supplied within a checklist and accompanied with an annotated illustration (Wang

et al. 2011). Submission of the digital checklist automatically generated a standardised image report.

1.4.4 Digital assessment methods of image interpretation training

Wright et al. (2016) tested student radiographers on their performance of musculoskeletal images pre and post university education. However, somewhat unique to their longitudinal study was the use of a specifically designed software programme 'Radbench' to record the performance measures of the student's image interpretation. The tool was used to assess student performance rather than provide any additional training to their university education (Wright et al. 2016). It was suggested that this software could be tailored to specific training needs or teamed with other key learning opportunities.

A web based resource was also formed by Subesinghe et al. (2015) to assess user performance on PET CT reporting. The resource allows users to complete an electronic form on the image under review and it supplies the user with a spreadsheet on their findings. This spreadsheet can subsequently be reviewed by the student mentor and internet based feedback template allows the mentor to supply quantitative and qualitative assessment of the trainee's reporting performance (Subesinghe et al. 2015). Both online tools strive to provide support to the user during their image interpretation learning and provide a detailed summary of the user's performance.

1.4.5 Current education of radiographers in image interpretation

Radiographers currently complete formal and informal training within chest image interpretation. Training can be completed by radiographers for many reasons; a preference to improve skills, to complete Preliminary Clinical Evaluation (PCE) or as a method of Continuing Professional Development (CPD). This training can be received through attendance of lunchtime seminars, training days off site, online tutorials etc.

A role progression was established in the mid 1990's to enable radiographers to report on medical images in order to decrease the workload of the radiologists and accommodate for the staff shortage. This role progression was achieved with the approval of the College of Radiographers (CoR 2010). Reporting radiographers may report alongside the radiologist in their

qualified area but only in departments where these advanced practitioner roles have been established and agreed with management (Paterson et al. 2004; Piper et al. 2005; Donovan et al. 2006). The reporting radiographers must have work delegated to them by a director or manager, their work will be audited and they can only work within the constraints of the Ionising Radiation (Medical Exposure) Regulations (Department of Health 2018). This role progression can be achieved through completion of postgraduate certificates within image interpretation areas of a specialised field. This role was developed following the implementation of a postgraduate programme in clinical reporting of adult chests and validation of the programme by the College of Radiographers (Canterbury Christchurch University 2002; Piper et al. 2014). Postgraduate programmes within chest radiograph reporting have a duration of 1 year and feature specific lectures within the University, formative feedback, oral/poster presentations, case studies and Objective Structured Examinations (OSE) (Canterbury Christ Church University 2017). With this role progression implemented, a recent survey recognised 27 sites within England which employed reporting radiographers within visceral (chest and abdomen) and skeletal examinations, however figures may have changed since this survey was completed (Snaith et al. 2015). Considering that the role progression of radiographers to report on musculoskeletal images was established 15 years before the survey was completed, it is surprising that a large majority of the reporting radiographers (443/494) were employed within hospitals in England. The remaining reporting radiographers who responded to the survey (n=51) were based in Scotland, Wales and Northern Ireland, where no radiographers were reporting visceral (chest and abdomen) examinations. However, it is important to note that although published in 2015, the survey was disseminated in 2011 and so figures could have risen substantially since.

1.4.6 Current education of radiologists in image interpretation

Trainee radiologists are qualified medical professionals who choose to specialise in the field of medical imaging. At undergraduate medical education, radiology is introduced and taught via modules to students. On average the largest proportion of radiology specific education across Europe is received within their fourth year of education (69%), followed by third and fifth year (53% respectively) (Korduokova 2010; European Society of Radiology ESR 2011). The European body ESR published a statement in 2011: "The learning outcomes are characterised by

an ability to detect abnormalities on chest, abdominal and skeletal radiographs and relate the findings to clinical management. Students should also display a systematic approach to comprehensive interpretation of radiographs." This statement highlights the knowledge medical students should gain during their training.

Once having chosen to specialise in radiology, trainee radiologists complete mentored training to become a radiologist. With the supervision and mentor of a consultant radiologist, the trainee radiologists learn how to undertake some medical imaging examinations and interpret all medical images on a trial and error basis. Trainee radiologists within the UK complete learning on the anatomy relevant to thoracic disease and radiological diagnosis and complete the part one examinations of the Royal College of Radiologist (FRCR) for assessment within this area (RCR 2016). It is only after their first 3 years in clinical radiology training that they are then able to complete the RCR final examinations, it is these examinations which assess their knowledge on chest radiographic interpretation and limitations (RCR 2016). Various other types of training are assessed throughout their clinical radiology training including; imaging algorithms and pulmonary disease, the role of chest radiographs, terminology, lines tubes and devices, imaging techniques and subsequent imaging appearances, clinical indication and implications, identifying and characteristic basic signs of thoracic disease on CXR (RCR 2016).

Through the use of a Likert scale on the undergraduate teaching of radiology within a UK medical school it was reported that students were most confident in their approach to interpreting chest radiographs, however they were least confident in differentiating between soft tissue, bone, fluid and air on plain radiographs as opposed to CT, MRI and US (Jacob et al. 2016). Students found CXR interpretation to be the most interesting radiology-related teaching. Over the three year undergraduate course students received three, five, and two hours of focused radiology teaching in these years and 47% of students said that they would want further radiology related teaching such as 'increased volume of teaching on image interpretation' incorporated within the curriculum (Jacob et al. 2016). Although students were most confident in their approach to chest radiograph interpretation, the results of this study show a gap of undergraduate training within image interpretation which would support these individuals if medical imaging was pursued as a speciality.

1.5 Eye tracking within image interpretation training

Eye tracking has been used to help understand the process of image interpretation and secondly to assess and provide feedback/training on the interpretation process.

Donovan et al. (2013) used eye tracking to distinguish between expertise related differences in search and decision making when viewing pulmonary nodules within medical images. Pulmonary nodules are areas of increased density present within a chest image which are roughly circular in shape and represent a benign or malignant growth within the chest cavity. The study was able to identify that naïve observers, staff and students from disciplines outside of radiography with no experience in medical images, and expert observers (radiologists and reporting radiographers) allocated less visual attention to "nodules" compared with student radiographers. This was concluded to have been due to the expert's ability to make fast and frugal decisions, by exploiting initial holistic processing and relying on their experience of viewing 'normal' images (Donovan et al. 2013). Matsumoto et al. (2011) utilised eye tracking technology to determine how neurologists deploy their visual attention when viewing brain CT images. The difference in the neurologist's dwell times over the selected regions of interest (ROI) indicated that neurologists intentionally scan clinically important areas when reading brain CT images showing cerebrovascular accidents. Participants were asked to interpret the CT images and provide a diagnosis with regard to cerebrovascular accident, this may have had an influence on the difficulty of the task presented to participants and on their search pattern (Matsumoto et al. 2011; Kok et al. 2015).

Studies by Litchfield et al. (2008) and Donovan et al. (2008) have attempted to evaluate the effect of feedback/training on lung nodule detection. The feedback based on eye tracking data from the participant (expert or novice) was shown to have an effect. Donovan et al. (2008) noted a significant improvement following feedback (p=0.021). Litchfield et al. (2008) reported an improvement in the performance of both undergraduate and postgraduate radiographers when shown a preview of eye movements before their interpretation compared to when they were instructed to 'free search' or preview the image for 20 seconds prior to their image interpretation. Perceptually based feedback was also used within a study by Kundel et al. (1990). Areas which received prolonged areas of attention (>1000ms) by the observer when interpreting the image

were shown to the observer for a second look. Use of this feedback resulted in a 16% increase in observer performance compared to showing the observer the image again with no eye tracking feedback highlighted. True positive rate increased and false positive rate decreased, indicating a true improvement in performance (Kundel et al. 1990).

1.6 Summary

To summarise, there is no standard systematic approach to chest image interpretation. Training is being provided and the impact of this assessed (Donovan et al. 2008; Piper et al. 2014; Semakula-Katende 2016). Often trainee radiologists and reporting radiographers combine advice given in guidance and a variety of recommended search techniques to form their own search strategy in image interpretation. Reporting clinicians receive formal education, however there is little evidence available on the effectiveness of training aids. Eye tracking technology has been shown to be a useful tool in providing feedback during the education of medical imaging professionals. To date no studies have been found which investigate the effect of using a digital training package based on eye tracking technology during the training of chest reporting. With the use of eye tracking technology and expert input the aim of this study is to establish and evaluate a digital training platform. The training platform is designed to help assist reporting clinicians during their image interpretation and specifically in learning a search strategy to use when interpreting chest images.

A systematic method of interpreting the entire image would limit the possibility of mis-diagnosis or failure to detect any pathology (Donovan et al. 2006; Lee et al.2013; Kok et al. 2015). Chetwood et al. (2012) suggested forming a training tool for laparoscopic surgery by projecting the eye gazes of experienced surgeons onto a trainees screen. By following their eye gaze, improved completion times and reduced errors were evident and witnessed when using the eye tracking technology with the trainee surgeons; however such training tools have yet to be formed for use in radiology. By studying the thought process of an expert radiologist's image interpretation, the proposed training tool will aim to provide a uniform standard of training which optimises learning by using the consultant radiologists' and reporting radiographer's experience, technique and verbalised thought processes.

1.6.1 Aims and objectives of the thesis

This thesis aims to investigate the image interpretation process of reporting clinicians and to develop and test a digital training platform for chest image interpretation.

Objectives:

(1) To systematically review the literature in the area of reporting radiography and image interpretation education.

(2) To investigate image interpretation performance of radiographers by computing eye gaze metrics using eye tracking technology.

(3) To develop a digital training platform for use in chest image interpretation.

(4) To investigate the effect of the digital training platform on reporting clinician performance and image interpretation learning.

(5) To propose an evidence based practice training platform which will aid the learning process of chest image interpretation

Chapter 2 - Systematic review

2.0 Abstract

As identified in Chapter 1 the role progression in image interpretation within radiography was developed and progressed into different imaging modalities and anatomical areas. The role progression can enhance waiting times, patient safety and care (Woznitza et al. 2014). However there is a further potential for patients to be misdiagnosed and treated inappropriately if such role progressions are not handled with caution and stringent audits (Paterson et al. 2004; Piper et al. 2005). In order to identify elements of image interpretation training which were lacking or not present a systematic review was completed in Chapter 2. This review helped inform the planning and development of the training platform and how the study would be organised.

2.1 Introduction

Assessments and audits carried out on the ability of radiographers to complete chest image interpretation play a crucial role in establishing the standards of performance in clinical practice. By monitoring and evaluating performance standards, attention can be brought to errors or poor performance levels (Jones et al. 2007). Training plays a large part in increasing the accuracy of clinical practice, as seen in Loughran (1994) where specificity increased from 94.4% (first 2 months) to 96.6% (final 2 months). Investigation and evaluation of the current education programmes can allow the most appropriate method of training to be applied to enhance these standards and maintain them. This was further supported by musculo-skeletal (MSK) work by Mackay et al. (2006) where a two day training programme increased sensitivity of fracture detection of radiographers from 78.9% to 88.2%. Chest image interpretation is a challenging and skillful task and therefore the most appropriate training method to be employed for it can be difficult to find. A clear knowledge of the impact of this training can allow radiographers to dedicate time within their busy schedules accordingly. Training targeted at specific aspects of image interpretation may allow users to target weaknesses. In keeping with this, the large variation in patient anatomy, the range of pathologies which can present on a chest image and the similarity in pathological presentations adds to the complications arising when undertaking this task (Woznitza et al. 2014). Initial training and education can provide a solid base to address these complications and help familiarise interpreting clinicians with them.

The greatest interpretation errors may occur when assessing radiographic chest images involving multiple organs, systems of the body and possible pathologies. A recent study by Woznitza et al. (2014) featuring the review of 99 chest radiographs (CXR) highlighted that discrepancy exists between highly qualified professionals. Of the 99 cases viewed by two radiologists and one reporting radiographer (qualified and experienced in chest radiograph interpretation) seven CXR reports were discrepant with clinical review. Mediastinal lymphadenopathy was missed by both radiologist and radiographer; linear atelectasis was reported by the radiologists but not the radiographer. Discrepancies existed in the identification of consolidation and bronchial wall thickening by the radiographer and radiologists also. Although reporting a very high concordance between radiologist and radiographer interpretation (92-96%), a discrepancy remains in the reporting of chest radiographs. Therefore monitoring of performance is an imperative to demonstrate appropriate performance and practice.

It is crucial that training and feedback on performance is available to radiographers who are eager to prosper within the role of chest image interpretation. Educational outcomes could be maximised by identifying training methods which work to their individual strengths and preferred approaches of learning (Neep et al. 2014). Nonetheless it is also important for radiographers to be able to identify a training method which complements their desired role, e.g. a qualified reporting radiographer specialising in chest image interpretation or a radiographer wishing to appropriately indicate the presence of abnormality on chest images in clinical practice.

Although the opportunities available for radiographers to interpret chest images are increasing, there remains a lack of knowledge about training methods and the effect of this on the accuracy of diagnosis. There are articles which demonstrate high levels of accuracy by reporting radiographers on musculoskeletal system (MSK) images of 91.8%-93.7%, however there is little information on the training methods which produce the highest standard of performance or whether this is affected by the training the radiographers receive (Piper et al. 2005). Postgraduate education is the accepted and preferred method of training and assessing students before a qualification is awarded (Piper et al. 2014). The effect of postgraduate training courses was

observed previously in an audit of trainee reporting radiographers whilst learning to report on images of the MSK system. In this instance the radiographer's accuracy increased from 87.8% to 100% in MSK image interpretation when compared with that of a supervising radiologist (Carter et al. 1999). The positive effect of a pilot education programme, which aimed to improve abnormality description of adult appendicular MSK trauma images, was observed when all but one radiographer demonstrated significant improvements in performances across three tests following access to the programme (McConnell et al. 2012). However, there remains uncertainty about which strategies are best utilised to prepare individuals for image interpretation of the chest radiograph or how the postgraduate programmes compare to any alternative training.

The formation of new technologies to view chest images could affect the interpretation process (Arenson et al. 1988). Within the last decade post processing abilities, the Picture Archive and Communications System (PACS) and enhanced image viewing monitors have been developed to assist the task of interpretation (Arenson et al. 1988; Arenson 1992). Furthermore, imaging equipment has developed rapidly over the past two decades from predominately film based radiography to computed radiography and then in recent years to direct digital radiography (Fajardo et al. 1989; Thaete et al. 1994). In the past films were interpreted with bright lights and magnification glasses however, digital images are now used and with the possibility of these features but sometimes at a cost to image detail due to pixel matrices. Improved speed of image acquisition and standardizsed image formats are now available with new technologies (Dougherty 2009). These equipment changes impact the format and quality of the images and this evolution of imaging allows the access and editing of images to become a much easier and more efficient to process (Dougherty 2009).

Until recently, the role of chest image interpretation has been the role of a radiologist but it is now shared with reporting radiographers. This role was developed initially following the implementation of a postgraduate programme in clinical reporting of adult chests and validation of the programme by the College of Radiographers in 2002 (Canterbury Christchurch University 2002; Piper et al. 2014). There have been several studies completed on radiologist performance, which report a mean diagnostic performance of 91.1% correct and mean 94.7% for consultant radiologist's area under the curve using Receiver Operating Characteristic (ROC) analysis (Kok

et al. 2015; Kelly et al. 2016). This high level of accuracy is supported by the mean accuracy (14.8/15, 98.66%) of consultants (n=14) and registrars (n=18) within Mehrotra et al. (2008). However, within the limited evidence available on performance by reporting radiographers, chest image interpretation mean specificity and sensitivity was reported to be 95.4% and 95.9% respectively (Piper et al. 2014). This was similar to that of the top 20 of 162 radiologists within Potchen et al. (2000) where the area under the curve of the ROC analysis was estimated to be 95% however, this result is not reflective of the performance of the entire sample size. These findings are supported in MSK image interpretation roles, where Brealey et al. (2003) noticed no difference in the Az values (ROC analysis) of clinical specialist radiographers (CSR) (0.77) and radiologists (0.85) (p=0.09). Az value being the probability of a CSR or radiologist correctly deciding whether plain radiographs are normal or abnormal. However, the evidence within this relatively new field of chest image interpretation by radiographers is limited and therefore it is vital that the standards of performance and training are monitored. A review of current education being provided to radiographers on chest image interpretation is therefore necessary to assess whether radiographers are being adequately trained and whether the methodology employed can influence the accuracy of radiographers in this area of practice.

2.2 Aims and objectives

The aim of this chapter within the thesis is to review the education and training provided to radiographers in plain radiography chest image interpretation and evaluate any assessment of the effect of this training on performance of the radiographers. A systematic literature review will also determine the quality and relevance of published material within the field of chest image interpretation and reporting by radiographers.

2.3 Methods

A systematic literature review was performed by searching the following healthcare databases: Medline (1949-present), Pubmed (1947-present), Scopus (1823-Present), Cumulative Index to Nursing and Allied Health (CINAHL) (1937-Present), the Cochrane Library Database (1974-Present) and Embase (1980-Present). The "Medical Subject Heading" (MeSH) was used to identify related keywords. The search strategy was developed using the following keywords: radiographer, radiologic technologist/technician, x-ray, image, film, radiograph, chest, thorax

and axial. (Chest OR axial OR thorax) AND (image OR radiograph OR film OR x-ray) produced 255,333 results therefore keywords were combined with each other by selecting AND in the database search.

The variation of spelling and terms used in the literature were taken into account and also searched. These included "image" OR "x-ray" OR "xray" OR "film" OR "radiograph", "chest" or "thorax" or "axial", "radiographer" OR "radiologic technician" OR "radiologic technologist". The reference list of each relevant study was searched for additional publications that involved radiographers interpreting chest radiographic images. To ensure that relevant literature was updated an alert was set up within each database from the search strategy and keywords used.

2.3.1 Inclusion/Exclusion criteria

Articles were included if they were in English, focused on chest radiograph image interpretation, involved radiographers/radiologic technicians/radiologic technologists as participants and featured a form of training in the interpretation of chest radiographic images. Articles were excluded if they featured a modality other than plain chest imaging, were articles on the imaging examination, dose, quality or technology were case specific or focused on patient safety and care/service evaluation. The candidate (LMcL) reviewed all abstracts and identified papers which met the inclusion criteria. Papers were independently screened by supervisors to ensure they met the inclusion criteria. Any ambiguous studies where discussed until consensus was reached. Data was extracted by LMcL using a predesigned form and this data was entered into the results tables.

For the purpose of the review, the quality of the studies were assessed based on an adaption of the questions provided in the critical appraisal skills programme (CASP) tools of a cohort study and diagnostic study online.
Available www.casp-uk.net/find-appraise-act/appraising-the-evidence/Oxford.CASP. Appropriate questions were identified as follows:

Q1 Did the study address a clearly focused issue?
Q2 Was the cohort recruited in an acceptable way?

Q3 Was there a comparison with an appropriate reference standard?

Q4 Was the training accurately measured to minimise bias?

Q5 Have the authors identified and taken into account confounding factors?

Q6 Were the methods for describing the test described in sufficient detail?

Q7 Do you believe the results?

Q8 Can the results be applied to the local population (local radiographers)?

Q9 Do the results of this study fit with other available evidence?

Q10 Can the results be applied to your patients/the population of interest (prevalence)?

If the answer to a question was 'yes' the article was scored 1, if the answer to a question was 'can't tell (CT)' or 'no' a score of 0 was awarded for that question. Each article in this review was independently scored by two reviewers and differences were resolved by discussion with the supervisory team.

2.4 Results

The PRISMA flow chart summarises the literature review search results (Figure 2.1). The electronic database search identified 642 studies and a further three studies were then found through hand searching and other sources. Articles were screened for duplicates, of which 287 articles were removed. Titles and abstracts of the remaining 358 articles were screened. A total 299 articles matching the exclusion criteria were removed. The remaining 59 full text articles were viewed and following the application of the inclusion criteria 45 irrelevant articles were excluded; the remaining 14 articles were reviewed by the research team. After contacting the authors for relevant clarification information and results, it was discovered that one article under review was an experiment featured within another article of the review (Litchfield et al. 2008; Litchfield et al. 2010). Litchfield et al. (2008) was excluded as it is the first experiment featured within Litchfield et al. (2010). Therefore 13 articles featured within the systematic literature review.

Figure 2-1: Summary of literature review search records using PRISMA group flow chart (2009)
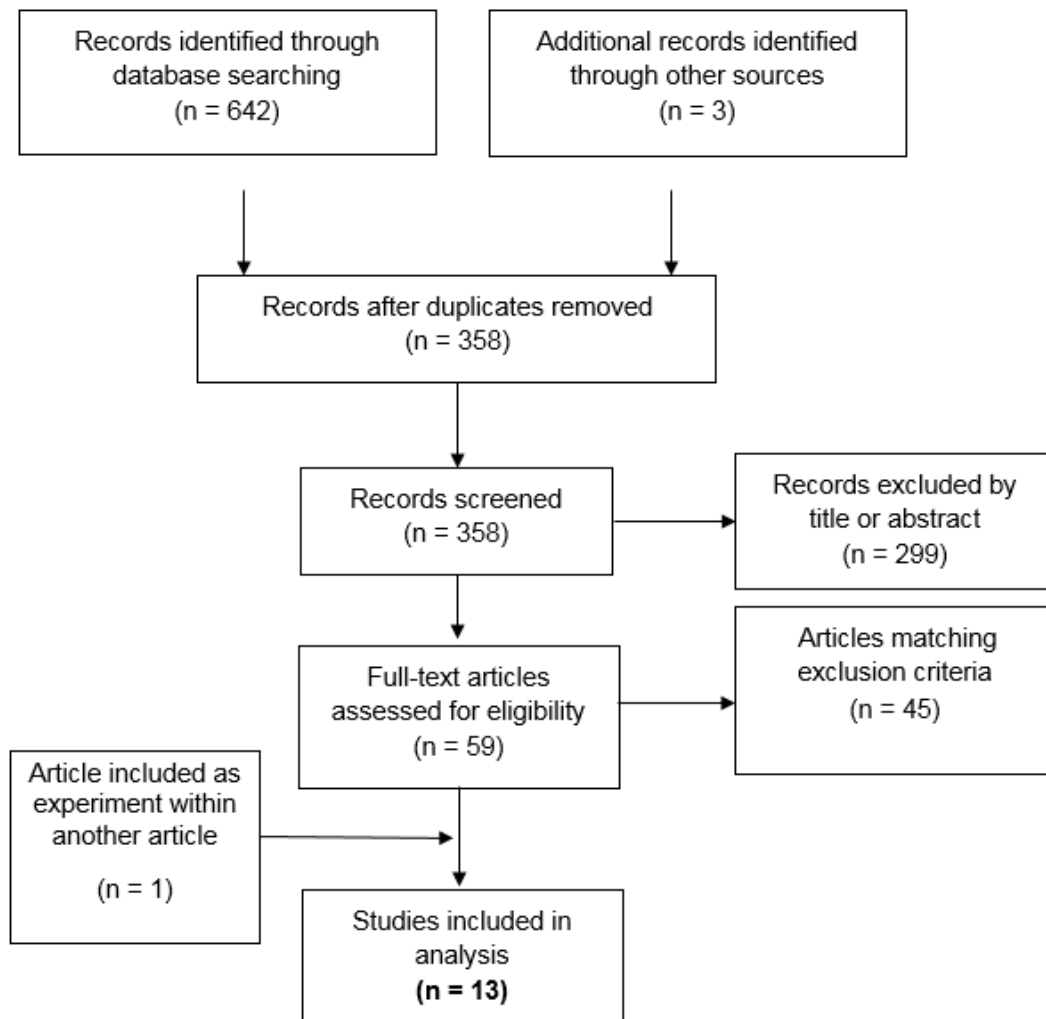
The search results included a range of studies relating to case studies, chest pathologies, imaging techniques, imaging modalities, radiation dose, image quality and patient safety/care. The characteristics of the articles relevant to the review are presented in Table 2.1. Within the 13 reviewed articles, a total population of 649 participants was assessed on their chest image/radiograph interpretation accuracy between the years 1978-2016. Of the 650 participants featured within the studies, 466 participants were students or experts within the radiography or reporting radiography profession or the equivalent. The remaining 183 participants comprised two junior radiology staff, seven consultant radiologists, 10 naïve observers, 24 physicians, 32 paediatricians, 39 clinicians, three healthcare assistants, one medical doctor and 65 radiologists. Just over 30% of the studies were completed within the last six years (2010-2016) however these studies featured approximately 69% (445/649) of overall participants and approximately 69% of radiographer (students and qualified personnel) participants (323/466) featured within the studies. The increasing frequency of publications in this area of radiography is most likely a result of the role progression of radiographer reporting and other healthcare professionals within this field in recent years. The sample size within studies varied greatly. The smallest study, completed within the UK, featured one reporting radiographer who was tasked with interpreting 100 chest radiographs (Woznitza et al. 2014). A large study, also completed within the UK, featured 148 radiography students/experienced radiographers (94 students and 54 radiographers) who interpreted 14 or 40 chest x-rays each, depending on which experiment they were recruited to participate in (Litchfield et al. 2010). Experiment one and two featured equal numbers of radiographers and students, experiment three featured 40 students as the benefit of following another's gaze primarily occurs during the early stages of learning and this is what experiment three entailed (Litchfield et al. 2010). Another large study featured within the review consisted of 256 participants but only 134 were radiographers; the remaining 122 participants were from other healthcare/medical professions (Semakula-Katende et al. 2016). The numbers of participants within the other studies ranged from 1 to 40 participants. A total of 10 studies were conducted in the United Kingdom (UK), one study was performed in Africa, one study was conducted within North America and one study was completed in the South Pacific countries. The large majority of the studies conducted within the UK (approximately 77%) is unsurprising given the role progression of chest image interpretation by reporting radiographers was established within England, UK (Woznitza 2013; Canterbury Christ Church University 2017).

The measurements recorded by each study are provided in Table 2.1. Information pertaining to participant accuracy (sensitivity, specificity, means etc.) was the most common measurement used and was recorded within all 13 articles of the review. There was a wide variation in how participant responses were requested and recorded. Studies referred to the accuracy of participants using different terminology. Observer performance, false negatives, false positives, red dot accuracy and diagnosis were used within studies when referring to the accuracy. The 'red dot' task involved participants placing the words 'red dot' on an image where they thought a pathology was present. Performance was also assessed using eye tracking technology within six studies. Time taken to interpret the images was recorded within 2 studies. A questionnaire was utilised within one study.

Table 2-1: Characteristics of the articles within the review

Lead author	Year	All Participants (n)	Non-radiography participants	Radiography Participants (n)	Radiography students	Radiographers	Reporting radiographers (RR)	Other qualified reporting clinicians	Country	Outcome measures	Quality score
Cowan	2007	23	3 health care assistants (1 minimal formal training, 2 no formal training) 1 doctor	19		19			South Pacific countries	Summative assessment Levels of satisfaction	6
Donovan	2008	40	10 naïve (22-40 years)	22	10 level 1(19-48 years) 10 level 2 (20-54 years)		2 chest RR (35-55 years)	8 radiologists (35-55 years)	UK	Observer performance Eye tracking	7
Flehinger	1978	2		2		2 radiologic technologists (both licensed since 1965)			USA	Observer performance	7
Hughes	1996	25/26		25/26		26 (average years qualified 10.24)			UK	Pre-tutorial questionnaire Pre-tutorial assessment Post-tutorial assessment sheet Post tutorial questionnaire	8
Litchfield	2010	148		148	24 novices 70 1st year	54 experienced radiographers (min 5 years' experience)			UK	Observer performance Eye tracking Type of preview cue Decision times	7
Manning	2003	18		12	6 novices	6 enrolled in postgraduate programme		6 radiologists	UK	Observer performance Eye tracking	7
Manning	2004	24		16	8 novices	8 before and after 6 month training programme		8 radiologists	UK	Observer performance Eye tracking	6
Manning	2006 (a)	21		13	8 novices	5 enrolled in postgraduate programme		8 radiologists	UK	Observer performance Eye tracking	7

Manning	2006 (b)	24		16	8 1st year	8 enrolled in postgraduate programme 5 re-tested after		8 radiologists	UK	Observer performance Eye tracking Time	6
Piper	2014	40		40			40		UK	Observer performance False positives, false negatives	10
Semakula-Katende	2016	256	32 pediatricians, 39 clinicians, 24 physicians including pediatricians	134		134		27 paediatric radiologists	South Africa/ UK	Pre course assessment Post course assessment	7
Sonnex	2001	23		17		17		4 consultant radiologists, 2 junior radiology staff	UK	Red dot accuracy	9
Woznitza	2014	4		1			1 trained	3 consultant radiologists	UK	Observer performance	10

2.4.1 Quality of studies

The quality of the studies varied slightly within those reviewed (Table 2.1). The approximate mean score of the studies was 7.5/10. Nine of the 13 studies scored either six or seven once critiqued with the CASP tool. Three studies were awarded high scores of 10, nine and 10 (Piper et al. 2014; Sonnex et al. 2001; Woznitza et al. 2014 respectively). The three articles scored highly as they tested the participants on a range of pathologies, included information on the reference standard to which participant's answers were compared against, were transferable to everyday clinical practice and referred to confounding factors of their study, including how this may have influenced the final scores of participants.

The published studies were low risk, they tested participants on images which were already clinically diagnosed or which would be interpreted also by a radiologist and/or reporting radiographer qualified to do so. The authors therefore probably regarded details of the recruitment process as unnecessary and inconsequential as a vulnerable population was not being impacted by the study. Generally there was limited information given on the recruitment process within the reviewed articles and so those which failed to give information on this topic were awarded a score of 0/CT for question 2 (n=10).

If only one professional (i.e. one radiologist or one reporting radiographer) was described as the reference standard or details of a reference standard were not given, this was marked as insufficient and did not receive a mark within question three of the critical appraisal (n=9). Participant's answers were marked against the reference standard. Therefore only one person acting as the reference standard does not allow an agreement to be made on the gold standard diagnosis of the image and is subject to the professional's opinion and accuracy in image interpretation (Brealey et al. 2014).

A few authors failed to list the confounding factors of their studies and so were scored poorly in question five as opposed to those who mentioned the limitations associated with their methodology or study (n=4) (Flehinger et al. 1978; Hughes et al. 1996; Manning et al. 2004; Manning et al. 2006b).

Scoring of question 10 was generally poor as there was a lack of variety in pathology used and adherence to prevalence noted within the studies. 8/13 articles focused solely on the detection of one chest pathology, which led to them scoring poorly within question 10. The most prominent example of this were the studies completed by Donovan et al. (2008), Litchfield et al. (2010) and Manning et al. which tested the participants on the presence of simulated chest nodules present within abnormal chest images. As the pathology was added to the digital image the study was not transferable to images the participants would encounter in clinical practice.

2.4.1.1 Study biases

The image inclusion and selection process has led to a series of biases presented in these studies. Spectrum bias, whereby the selection of images were unrepresentative of a population but include more difficult batches of films and a desirable variety of pathology, can be seen where test banks of images including a range of pathologies were chosen (Brealey et al. 2001). Film selection was not obvious in the studies under review. No studies detailed that the participants had a choice on which images to interpret or the choice of excluding images during the study. Population bias was seen in studies were a representative case mix was ideal for the assessment of training in particular areas such as the Accident and Emergency department (A/E) or General Practitioner (GP) referrals (Flehinger et al .1978; Sonnex et al. 2001).

The CASP tool developed and utilised for the review identified biases in the selection of the reference standard within studies. As detailed below in 2.3.13 there was a variety in the number of experts detailed as the reference standard or lack of information given in this area. Verification bias, where not all images were interpreted by the same reference standard, and work-up bias, where not all images received confirmation with the reference standard, could not be excluded within all studies given this lack of detail (Brealey et al. 2001).

2.4.2 Training received

There was a variation between the training interventions applied within the studies. A summary of the training provided is listed in Table 2.2. Six from 13 articles featured radiographers who attended postgraduate training in chest image interpretation. These postgraduate programmes were full structured courses within a University setting. Two articles supplied either feedback

through eye tracking, an image preview or an overlay of eye tracking onto an image as their intervention. Four studies provided lectures, presentations or an apprentice programme to participants. Of these four studies, two also incorporated feedback/instruction sessions into the training. The training intervention of one study was radiography clinical experience and a practical assessment of chest radiography.

Table 2-2: Training

Lead author	Training	Training impact	Performance details
Cowan 2007	1st: Lectures, guidance, tutorials, quiz, PowerPoint presentations, quiz, glossary, formative feedback. 2nd: Lectures, teaching sessions on specific abnormalities, formative film reading, tutorials, presentations on a specific case 3rd: Teaching modes used previously were repeated	1st: Scores ranged between 54-85% (mean=65%) (excluding 5 participants) 2nd: Of those who returned, two who failed the initial course lifted their performance to an acceptable standard by the end of the refresher All participants strongly agreed that they had 'gained new knowledge in this course'.	Scores ranged between 54-85% (mean=65%) (excluding 5 participants)
Donovan 2008	Cases presented, image overlaid with the scan path and fixations	Significant improvement following feedback (p= 0.021). Mean improvement 3.3% overall,8.4% mean improvement Level 1 undergraduate radiographers (p<0.05)	Mean FOM improved for all three groups following feedback. Improvement was greater for novices
Flehinger 1978	Informal apprentice program	agreement increased from 76.7% to 82.9%, agreement increased from 87.7% to 92.5%	Agreement percentage improved for both technologists Technologist A Pre 76.7%, Post 82.9%. Technologist B Pre 87.7%, Post 92.5%
Hughes 1996	Lectures introducing a pattern recognition technique, tutorials 2 (1 hour) sessions	Level of agreement pre-tutorial (K = 0.29) to post tutorial (K = 0.67).	Post study general departments/ Health centres average score: 8.45/10 A/E average score: 8.04/10
Litchfield 2010	(i)free search, image preview 20 s, eye movement preview (ii)image preview, expert search preview, unrelated preview (iii)eye movements of naïve observers no task given, naïve observers search for nodules, expert search, incongruent search	Experiment 1: Both groups performed better when shown the search behaviour of either a novice radiographer or expert radiologist. Experiment 2: Benefits in performance only when the eye movements shown were related to the search for nodules, however only the novice's performance consistently improved when shown the expert's search behaviour. Experiment 3: Novice radiographers were better at identifying nodules when shown either a naïve search behavior or an expert radiologist's search behaviour. The eye movement preview led to higher scores than the free search preview and image preview (p<0.001).	FOM improved for students and radiographers when shown an eye preview FOM decreased for students and improved for radiographers when shown an image preview Students and radiographers FOM improved when shown expert search preview compared to an image preview Unrelated preview made no difference to students FOM and slightly improved radiographers Students had the highest FOM when shown and expert search, followed by naïve search, incongruent search and naïve no task respectively
Manning 2003	Six month postgraduate programme	Radiographers reduced the number of zones they inspected quite significantly after their training. Trained participants out performed participant's performance prior to training.	AFROC area under the curve data indicated the trained participants outperformed the participants who had not received training
Manning 2004	Six month postgraduate programme	False error rates were 47% before training and 42% after training. % of total false negatives was 25% before training and 23% after training	Errors from films with multiple nodules decreased following training (25%-23%)

Manning 2006a	Six month postgraduate programme	The radiologists and radiographers significantly better than the radiographers and students (t-test p=0.02). Radiographer's performance increase from 0.70 to 0.82	RR AFROC improved following training 0.70 to 0.82
Manning 2006b	Six month postgraduate programme	The trained radiographers were quicker to find a present pathology.	
Piper 2014	Postgraduate programme	No significant differences were found. Mean sensitivity and specificity was 95.4% (95% CI 94.4%-96.3%) and 95.9% (95% CI 94.9%-96.7%) respectively.	Agreement was high 89.0 overall
Semakula-Katende 2016	30 min digital presentation (short course)	% correct diagnoses went from 47.3% pre course to 59.1% post course, statistically significant improvement in detecting tuberculosis (sensitivity) and in percentage of correct diagnoses	% CORRECT Improved within the radiographer group post training (47.3%-59.1%)
Sonnex 2001	Red dot protocol, programme of lectures, feedback sessions	The sensitivity and specificity was 90% and 99% respectively. Radiographers missed potentially important changes in 38 exams and incorrectly red dotted 100 exams.	Radiographer specificity 98.8% sensitivity 90.5%
Woznitza 2014	12 years post registration experience, consultant mentoring, 2 years postgraduate	High concordance between the reporting radiographer and radiologist (96%, 96% and 92%) (K>0.8).	Concordance between radiographer and consultant radiologists was high 96%, 96% 92%

2.4.3 Effect of chest image interpretation training/feedback

All studies within the review demonstrated positive attributions of the training under evaluation:

- A mean percentage improvement of 3.3% was presented overall and 8.4% mean percentage improvement noted within Level 1 undergraduate radiographers ($p<0.05$) (Donovan et al. 2008).

- In Flehinger et al. (1978), the initial reporting period demonstrated agreement between the radiologist and Technologist (A) of 76.7%, which increased to 82.9% in the subsequent reporting period. During the initial reporting period, agreement between the radiologist and Technologist (B) was 87.7%, this increased to 92.5% in the subsequent reporting period following training.

- Hughes et al. (1996) indicated that the level of agreement between radiographers and the reference standard of radiologists grew from pre-tutorial (kappa (K) = 0.29) to post tutorial (K = 0.67). The level of agreement increased substantially and there was a significant improvement in the predictive values overall. Cohen's kappa coefficient is a statistic which measures the inter-rater agreement for categorical items (Landis et al. 1977). A value of 0.21-0.39 indicates minimal agreement, 0.60-0.79 indicates moderate agreement and above 90 indicates almost perfect agreement (Landis et al. 1977).

- Within Litchfield et al. (2010) the Figure of Merit (FOM), used from the Jackknife Alternative Free-Response Receiver Operator Characteristic (JAFROC) analysis, showed an improvement in performance within the eye movement preview group compared to the image preview group ($p<0.001$). JAFROC, the analysis software, generated a FOM that quantification of search performance. It was defined as 'the probability that an observer will rate a lesion higher than the highest rated non-lesion on a normal image' (Donovan et al. 2008). They also showed that both sets of participants (undergraduate and postgraduate radiographers) were more likely to make the correct decision when shown either a novice radiographer or expert radiologist's eye movements compared to when they were shown the image for 20 seconds before being allowed to make a decision. Participants could either immediately make decisions regarding nodules (free search), make a decision on the presence of nodules but only after 20 seconds(s) (image preview) or were either shown eye tracking of a novice or expert for 20 seconds before making a

decision (eye movement preview). The eye movement preview led to higher scores than the free search preview and image preview (p<0.001). In conclusion, both groups benefited from viewing the eye movement preview before deciding on their diagnosis, nonetheless it was noted that there was a larger effect evident within the undergraduate group as they made the greatest performance improvement (Litchfield et al. 2010).

- Manning et al. (2003) noted that the radiologists and radiographers following training outperform the novices and radiographers prior to training. In Manning et al. (2004), after the postgraduate training radiographers' errors became more like that of the experts in image interpretation and less like that of novice radiographers. An indication of this being more search errors being made by the trained radiographers as opposed to decision errors.

- Manning et al. (2006a) mimicked the results of Manning et al. (2003) with performance measures demonstrating once again the radiologists and trained radiographers of chest image interpretation outperforming the untrained radiographers and students of radiography (t-test p=0.02). Also noted was the radiographer's performance increase from 0.70 to 0.82 (Alternative Free-Response Receiver Operating Characteristic AFROC area under the curve).

- Manning et al. (2006b) took a unique approach in investigating whether True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) related to the duration of gaze, they did this by studying the survival analysis of fixation data. Survival analysis is used in this context as a method of showing 'the porportions of decisions that are completed for each category (TP, TN, FP, FN) at increasing accumulated time intervals of visual attention' (Manning et al. 2006b). The survival analysis was studied by analysing the dwell time data for all fixations and relating this to positive and negative decisions to provide information on percentage survival of decisions over time (Manning et al. 2006b). This allowed the authors to demonstrate that trained radiographers were quicker to find visible pathologies.

- In Semakula-Katende et al. (2016), following the provision of a 30 minute course and assessment programme, radiographer's correct diagnoses increased from 47.3% (pre-course) to 59.1% (post-course). However, this improvement in performance by radiographers was the lowest of the professional groups (paediatricians, physicians, clinician/ and radiologists) and their sensitivity was significantly smaller (25.6%

[22.5%]) than the mean change in sensitivity of radiologists (41.2% [25.0%]). Comparison with other groups is not as important however due to the radiography group being much larger (n=134) than the other participant groups (n=32, n=39, n=24, n=27). It was suggested that due to their lack of prior knowledge in this area a modified training method, perhaps longer and more comprehensive, may be suitable. Nevertheless there was an increase in accuracy and a statistically significant improvement in sensitivity shown after the training by the radiographers when detecting paediatric pulmonary tuberculosis in developing countries (Semakula-Katende et al. 2016).

- Piper et al. (2014) did not test the cohort of participants' pre and post the training intervention, nonetheless, post-intervention measurements showed the radiographer's ability to report on a broad range of pathologies with high accuracy.

- Sonnex et al. (2001) also strengthened the reliability of radiographers' accuracy. They concluded that the interpretations were generally reliable with only 38/8150 examinations containing potentially important changes being missed during the introduction of the 'red dot' system.

- Woznitza et al. (2014) compared the accuracy of the reporting radiographer with that of three consultant radiologists. The concordance found was 96%, 96% and 92% (K>0.8). However, across the studies evaluated there were few negative effects found following some of the training methods. Within Litchfield et al. (2010) experiment 1, once given the image preview prior to image interpretation, there was no great impact noted within results. Students mean FOM decreased from no intervention (free search) 0.56 to 0.54 with an image preview. Similarly radiographers mean FOM remain unchanged for the free search group (0.57) and the image preview group (0.57).

2.4.4 Differences in accuracy measurements

The accuracy measurements presented in the literature are summarised in Table 2.1. All 13 articles made reference or presented results on the accuracy/performance measurements of participants within their study. However there was a large variety in how this measurement was presented and calculated within the literature.

2.4.4.1 FOM/JAFROC analysis

Donovan et al. (2008) presents a FOM/JAFROC. The analysis software used in the study, generated a FOM that allows quantification of search performance. FOM was defined as 'the probability that an observer will rate a lesion higher than the highest rated non-lesion on a normal image. Litchfield et al. (2010) also used the FOM and JAFROC analysis, however Litchfield et al. (2008) (experiment one of Litchfield et al. 2010) defined the FOM to represent 'the likelihood that a true positive will be given a higher rating than a false positive'. The majority of experiments within these two studies observed greater mean FOM for the trained groups compared to the untrained. Expert scores increased from 0.6578 to 0.6639 pre and post with no feedback and from 0.7903 to 0.8032 pre and post with feedback, overall there was a significant difference between pre and post "feedback" condition with a significant improvement overall following feedback (p=0.021) (Donovan et al. 2008). However within the expert results 10 were radiologists and only two were reporting radiographers that routinely interpreted chest radiographs. This study therefore had a majority of radiologists as 'expert participants' and did not solely focus on comparing groups of reporters. In experiment one and two completed by Litchfield et al. (2010), the majority of students demonstrated a higher mean FOM within the feedback groups; scores of 0.60, 0.66, 0.79, as opposed to the free search and image preview groups 0.56, 0.54, 0.66. Similarly radiographers in general had a higher mean FOM in feedback groups; 0.59, 0.76, 0.78, compared to free search or image preview groups; 0.57, 0.57, 0.75. In experiment 3, where various eye tracking videos were shown to participants prior to their interpretation, students performed noticeably better when shown an expert search (0.75) as opposed to an incongruent search (0.61). Both studies were difficult to compare, although both incorporated eye tracking training, the training techniques were very different.

2.4.4.2 ROC/AFROC analysis

The ROC analysis was used by Manning et al. (2002). This method quantified the inherent detectability of signals embedded within a background of noise by determining true positive (TP) and false positive (FP) rates at different criterion thresholds. A ROC curve plotted in two dimensions on axes representing probability values of TP and FP responses will have an area that is a direct measure of detectability (Manning et al. 2002). Diagnostic imaging performance can

be evaluated by ROC when four possible responses; TP, TN, FP, FN are provided. Students who had received at least 50 weeks clinical practice had a mean area under ROC curve of 0.743 whereas radiologists scored an area under ROC curve of 0.872. Radiologist's range of clinical experience was 5-22 years.

Only four studies completed and published results using AFROC methodology. A decision on lesion location and confidence level between one and four on each decision was required. False positive decisions are viewed and the highest scoring false positive decision is the only one recorded per image to avoid the possibility of infinite values in summing FP responses (Manning et al. 2004). Of the four articles, two supplied the same graph representing AFROC scores (Manning et al. 2003; Manning et al. 2006a), but only one identified specific figures for each participant group (Manning et al. 2006a). One study presented FN rates as opposed to AFROC analysis scores and will be discussed below (Manning et al. 2004) whereas another focused solely on how the four decision outcomes (TP, TN, FP, FN) related to duration of gaze (discussed below Manning et al. 2006b). Manning et al. (2006a), the only study to report figures on area under the curve for AFROC (and hence accuracy), demonstrated the following scores; radiologists (0.80) (figure taken from a graph), trained radiographers (0.82), untrained radiographers (0.70) and novices (0.63).

2.4.4.3 Mean accuracy percentages

A total of five studies provided an average accuracy or average agreement score of participants. Mean diagnostic accuracy was recorded as the lowest pre training by Semakula-Katende et al. (2016) as 47.3% and as the highest pre-training by Flehinger et al. (1978) as 87.7%. The lowest post-training accuracy was again noted by Semakula-Katende et al. (2016) as 59.1%, with the highest post-training accuracy percentages noted as 84.5% and 92.5% (Hughes et al. 1996; Flehinger et al. 1978). The most recently published articles, featuring the postgraduate training of chest image interpretation, focused on agreement percentage between the reference standard and trained reporting radiographers. These results ranged from 86.7% to 91.7% (Piper et al. 2014) and from 92% to 96% (Woznitza et al. 2014). Interestingly the oldest article within the review by Flehinger et al. (1978) also presented figures of agreement with the reference standard; however they also provided pre and post percentages as mentioned above.

2.4.4.4 Specificity and sensitivity

A total of five studies reported and published sensitivity and specificity values when measuring participant accuracy. Sensitivity values ranged from 53.5% (Semakula-Katende et al. 2016) to 87.8% in images taken in the Accident and Emergency department (A/E images) (Hughes et al. 1996) pre-training. Sensitivity values ranged from 56.7% (Manning et al. 2002) to 100% (A/E images) (Hughes et al. 1996) post-training. Specificity values ranged from 40% (A/E images) (Hughes et al. 1996) to 72.1% (Semakula-Katende et al. 2016) pre-training. Specificity values ranged from 59.3% (Manning et al. 2002) to 98.8% (Sonnex et al. 2001) post-training. Piper et al. (2014) recorded high values of sensitivity (92.8%-98.0%) and specificity (93.7%-97.4%) also. Studies asking participants to allocate images to a designated group had the lowest sensitivity and specificity values pre training (Hughes et al. 1996; Semakula-Katende et al. 2016). Manning et al. (2002) had the lowest sensitivity and specificity values post training however these were 'novice' radiography students with a minimum of 50 weeks clinical radiography experience.

2.4.4.5 Predictive power/value

Only two studies chose to present positive and negative predictive powers (Hughes et al. 1996; Sonnex et al. 2001). It has been over 15 years since both of the studies have been published. Positive predictive power ranged from 62.8% - 63.2% pre-training (Hughes et al. 1996) and negative predictive power ranged from 73.7% - 81.4% pre training (Hughes et al. 1996). Positive predictive power ranged from 67.7% (Hughes et al. 1996) to 78.4% post-training (Sonnex et al. 2001) and negative predictive power ranged from 96.3 - 100% post-training (Hughes et al. 1996).

2.4.4.6 False positives (FP), False negatives (FN), True positives (TP) and True negatives (TN)

A total of four studies recorded either FP, FN, TP and TN or a combination of the four outcome measures. These values were each presented in different formats; either pre/post training values (Hughes et al. 1996), percentages (Manning et al. 2004), on overall value (Manning et al. 2002) or single values (Piper at al. 2014), and so were difficult to compare. FP values were (50) pre and (65) post training within Hughes et al. (1996)(197 images pre and 484 images post, abnormality

rates pre study 0.47/0.54 post study 0.27/0.47) Whereas, there was a total of 83 within Piper et al. (2014) where 6 cohorts interpreting 100 images were studied (abnormal to normal 1:1). FN were 13 and eight pre and post training respectively in Hughes et al. (1996). There were 93 FN across the six cohorts within Piper et al. (2014). TP ranged from 85 within Hughes et al. (1996) to 1908 within Piper et al. (2014), with percentages of TP's decreasing from 47% to 42% following training of reporting radiographers within Manning et al. (2004) (120 images, abnormality rates 12%, 50% and 80%). TN ranged from a mean of 14.81 within (Manning et al. 2002) (120 images, abnormality rates 12%, 50% and 80%) to a total of 1917.0 across six cohorts within Piper et al. (2014).

2.4.5 Eye tracking metrics

The eye tracking metrics measured within six of the studies under review are listed in Table 2.3. Manning et al. (2003, 2004, 2006a, 2006b) did not provide figures for the eye tracking data collected however presented information in graphs. Donovan et al. (2008) assessed eye tracking metrics of participant groups pre- and post-feedback/no feedback, whereas Litchfield presented figures on the eye tracking metrics within the feedback provided to participants. Litchfield et al. (2010) used various eye tracking movements such as those from novices/experts and presented the eye tracking metrics relative to these rather than using the eye tracking to assess participant's performance as in Donovan et al. (2008). In general, time to first fixate, fixation count, average fixation length and gaze time decreased when participants were given a second look at the image (post-feedback/no feedback) compared to their first look at the image. There was a no significant effect between pre and post "no feedback" condition and a significant difference between the pre and post "feedback" condition. By studying the eye tracking metrics it was concluded that expert and naïve observers were less affected by feedback and a second look compared to the mixed results within Level 1 and Level 2 students. Therefore, perceptual feedback may be beneficial for those in early training.

2.4.5.1 Eye tracker

A total of seven studies utilised an eye tracker. Out of these, two studies mentioned the use of a standalone eye tracker (Litchfield et al. 2010; Manning et al. 2002), whilst others mentioned the use of a remote eye tracker and magnetic head tracker (Manning et al. 2003, 2006a). When allocating a task of viewing images on a screen, the standalone eye tracker can allow head movements to be more contained as opposed to the magnetic head tracker. This magnetic head tracker may influence the quality of the eye tracking data collected as participant performance could be affected due to the interference of the attached eye tracker.

Table 2-3: Eye tracking metrics

Lead author	Donovan 2008	Litchfield 2010		
Time to first fixate	Overall time to first fixate shorter for naïve, shorter for level 1, shorter for level 2, longer for the expert group with feedback			
Time to first fixate (2nd look)	Overall time to first fixate was longer for naïve, shorter for level 1, shorter for level 2, shorter for the expert group with feedback			
Fixation duration	Overall fixation duration was shorter for naïve, shorter for level 1, longer for level 2, shorter for the expert group with feedback	Novice 772.78 (596.00) Expert 859.79 (426.62)	Unrelated 305.21 (237.37) Expert 717.15 (321.99)	Naïve no task 317.53(325.35) Naïve search 1282.45(707.86) Incongruent search 230.84 (247.08) Expert 717.15 (321.99)
Fixation duration (2nd look)	Overall fixation duration was longer for naïve, shorter for level 1, longer for level 2, shorter for the expert group with feedback			
Number of fixations	Overall number of fixations was more for naïve, more for level 1, less for level 2, less for the expert group with feedback			
Number of fixations (2nd look)	Overall number of fixations was more for naïve, more for level 1, less for level 2, less for the expert group with feedback			
Dwell time	Overall dwell time was longer for naïve, longer for level 1, shorter for level 2, longer for the expert group with feedback	Novice 6.67 (5.01) Expert 7.74 (3.24)	Unrelated 1.26 (1.16) Expert 5.15 (3.27)	Naïve no task 0.74 (0.89) Naïve search5.42 (2.84) Incongruent search 1.09 (1.40) Expert search 5.15 (3.27)
Dwell time (2nd look)	Overall dwell time was longer for naïve, longer for level 1, shorter for level 2, shorter for the expert group with feedback			
Percentage of time spent looking at nodules		Novice 34.30 (26.10) Expert 43.40 (17.00)	Unrelated7.10 (6.60) Expert 27.70 (17.90)	Naïve no task 3.80 (4.50) Naïve search 27.80 (14.70) Incongruent search 6.10 (7.60) Expert search 27.70 (17.90)

2.4.6 Image viewing time

Six studies restricted the time allocated to participants; image viewing time was limited to 40 minutes (Manning et al. 2003; Manning et al. 2006a) or 1 hour (Litchfield et al. 2010; Manning et al. 2004; Manning et al. 2006b) for all images or 30 seconds per slide (Semakula-Katende et al. 2016). Some of these authors mentioned they applied this restriction on image viewing time to reduce the risk of fatigue influencing participant performance (Manning et al. 2006a; Manning et al. 2006b). No studies mentioned participants' complaints on too little time allocated for the study or the inability to return to images during the study. The approaches applied tended to control and standardise studies and disallowed these alternates from introducing differences to participant performance. One study gave an estimate of the study taking approximately one hour to complete (Litchfield et al. 2010). Two studies provided figures for the mean decision/scrutiny times per cohort (Litchfield et al. 2010; Manning et al. 2006a). Another study focused primarily on the survival analysis of the fixated data (Manning et al. 2006b). Training decreased the time taken to diagnose in Manning et al. (2006a) from a mean 33.9 seconds to 31.1seconds (p=0.02). However, feedback provided within Litchfield et al. (2010) increased and decreased decision times depending on the type of feedback provided. Likewise group decision times also varied, with students having shorter decision times than radiographers in experiment 1, and yet longer decision times than radiographers within experiment 2 following feedback (Litchfield et al. 2010).

2.4.7 Questionnaire

Only one article within the review utilised a questionnaire to gauge participant's feedback on the implemented training. A pre- and post-tutorial questionnaire was used.

A total of 25 out of 26 pre tutorial questionnaires were returned, whereas 26/26 questionnaires were completed post study. The average years qualified remained the same pre and post at 10.24. Average confidence level increased from 7.71 pre study to 8.73 post-study, however no obvious differences were noted between the various grades of radiographers. Participants who claimed to have a technique for looking at chest radiographs increased from 19 pre study to 25 post study. Four participants believe chest radiographs were adequately covered in standard teaching prior to the study whereas 26 believed the interpretation of chest radiographs were adequately covered in the tutorials they were given as the training intervention.

Following the supply of the pattern recognition technique, supplied within tutorials, at least 25/26 of the radiographers who completed the questionnaire identified the tutorials being useful to their daily work, as a valuable approach to interpretation of other radiographs and as a format which could be included similarly for chest radiographs in standard teaching.

The feedback taken from the questionnaire indicated that the intervention was perceived to have an overall positive effect as a training method.

2.4.8 Participants

As demonstrated in Table 2.1, there was a large variability of participant experience. The level of expertise varied greatly; 164 students, 260 radiographers and 43 reporting radiographers comprised the 467 radiography participants featured within the studies.

2.4.9 Report/comment/diagnosis

The instructions and guidance given to participants prior to the study varied greatly across the articles and impacted the answer and quality of the answer which the participant provided. The information given before beginning the study reflects the role the participant undertook and the expectations of the authors. Within Flehinger et al. (1978), the two radiographers asked to screen the images were based within a department which imaged patients as a part of an early cancer detection programme and so were asked to focus on the presence/absence of lung cancer. Participants within Hughes et al. (1996) were asked to use a search strategy in their interpretation and given a choice of four pathologies to diagnose. Sonnex et al. (2001) by comparison simply tasked the radiographers with placing a 'red dot' on chest images they believed to contain an abnormality. The study was featured in a cardiothoracic centre where training in management and detection of acute medical problems was a priority (Sonnex et al. 2001).

Seven articles named the pathology and asked participants to determine if images were normal/abnormal and give details on position of the pathology if they identified the image as abnormal. Beyond these articles, other authors (n=3) provided participants with choices of response such as three/four choices with which to assign to each image (Flehinger et al. 1978; Hughes et al. 1996; Semakulu-Katende et al. 2016). Where participants were given a specific

pathology to identify or a list of groups to assign an image to, it is questionable as to whether the participant has readily identified the pathology or their options for the image influenced their choice. Tasks where these options have been provided may be easier than forming a diagnosis on the image alone, the provision of specific groups/pathology can allow the participant to be cautious and mindful of these during their image viewing.

2.4.10 Imaging examination

By applying the inclusion criteria, all articles accepted for review featured plain radiography chest imaging examinations. Nonetheless, as imaging equipment is continuously being updated a difference remained in which form the x-ray examination was presented to the participant (i.e. film based radiograph or digital format image). The information given is presented in Table 2.4. 2/13 used films, 6/13 used radiographs, 2/13 used chest x-rays and 3/13 used chest images within their description of the study. With regards to chest projections used; Flehinger et al (1978) included lateral projections of the chest, 6/13 identified their use of postero-anterior (PA) chest projections and 0/13 used antero-posterior (AP) projections to test participants. Three studies identified their use of a particular age group within the test bank. Piper et al. (2014) used adult chest radiographs age>/=16, Semakula-Katende et al. (2016) used paediatric chest radiographs and Woznitza et al. (2014) used adult chest images. A total of four out of 13 articles acknowledged their use of digital display of their chest examination. The use of 30 images presented on PowerPoint presentation slides may have had an impact on image quality and degree of interaction possible for the viewer that may affect outcomes. (Semakula-Katende et al. 2016).

2.4.11 Pathology type

The pathologies featured within each article are presented in Table 2.4. Few articles tested the participants on their ability to identify more than one pathology type (n=4) (Hughes et al. 1996; Sonnex et al. 2001; Piper et al. 2014; Woznitza et al. 2014). The remaining articles (n=9) focused either on the presence of lung cancer (Flehinger et al. 1978), pulmonary nodules (Donovan et al. 2008; Litchfield et al. 2010; Manning et al. 2002-2006b) or pulmonary tuberculosis (Semakula-Katende et al. 2016). Articles by Flehinger et al. (1978), Hughes et al. (1996) and Sonnex et al. (2001) asked participants to interpret images as they encountered them in clinical practice or

were supplied with a sample of images sourced from the clinical department (Brealey et al. 2001). The participants were allocated a specific task when viewing the images in clinical practice, within an allocated time period. Elsewhere some authors chose to use images of phantoms which then had simulated digitised nodules inserted on to them to test their participants (n=7). This method of inserting the pathology on to the image allowed the appearance, occurrence and conspicuity of the pathology to be controlled. By inserting the pathology onto the image the number, size and density of lung nodules was controlled. It could be argued that clinical presentation of pathologies, with the possible distraction of pathological changes on the image, could make non simulated nodules more difficult to identify compared to inserting digitised lung nodules onto otherwise 'normal' phantom chest images. Piper et al. (2014), Sonnex et al. (2001) and Woznitza et al. (2014) tested participants' ability to identify the greatest number of pathologies.

2.4.12 Prevalence

Authors presented information on the prevalence/ratio of normal or abnormal images used within their image test banks, this information was summarised in Table 2.4. Articles by Flehinger et al (1978), Hughes et al. (1996) and Sonnex et al. (2001) focused on images which were encountered in clinical practice and the prevalence rate of abnormal images was calculated following the completion of the study by participants. However, some studies (n=7) chose to use images of phantoms which then had simulated digitised nodules inserted on to them to test their participants. This method of inserting the pathology on to the image allowed prevalence rates to be controlled. Prevalence rates of 50% (n=3) were used. Three articles contained three image test banks of prevalence rates 12%, 50% and 82% and another of prevalence rates 20%, 50% and 75%. Prevalence rates are an important consideration when testing participants on medical image interpretation. This is seen within Pusic et al. (2012) where high sensitivity was seen within a group tested using a high number of abnormal images (0.69±0.24) compared to groups trained with medium (0.63±0.21) and low (0.51±0.24) numbers of abnormal images. Also, Ryan et al. (2011) found that if image banks containing a greater number of incidental abnormalities or with more striking abnormalities were shown to participants that they were recognised more accurately, it was therefore advised that memory be taken into consideration when planning image banks.

2.4.13 Reference standard

Details of reference standards within the articles are summarised in Table 2.4. Many studies featured within the review failed to give information on how many experts formed their reference standard (n=5) or had placed the pathology on the images and a reference standard was not mentioned specifically (n=1) (Donovan et al. 2008). Articles featured a radiologist/consultant radiologist (n=3), two radiologists (n=1), three consultant radiologists (n=2) or six consultant radiologists (n=1) as their reference standard. The greater the number of experts forming the reference standard the less possibility of discrepancies in the diagnosis occurring (Hughes et al. 1996; Brealey et al. 2001; Piper et al. 2014; Woznitza et al. 2014).

Table 2-4: Imaging examination, pathologies and pathology prevalence

Lead author	Year	Imaging examination	Number of images	Pathology present	Prevalence	Reference standard
Cowan	2007	Digitised chest images	-	Common abnormal radiological appearances in the chest	-	-
Donovan	2008	Chest films	30	nodules	50%	test bank
Flehinger	1978	PA and lateral chest films	2831 Technologist A 2994 Technologist B	Lung cancer	7 cancer confirmed 5 cancer confirmed	Reading and study radiologist
Hughes	1996	Chest radiograph	197prior, 484 after	pneumothorax, effusion, collapse and pulmonary shadowing	Pre study 0.47/0.54 Post study 0.27/0.47	6 consultant radiologists
Litchfield	2010	PA chest x-rays	14, 40, 40	nodules	50% abnormal	a consultant radiologist
Manning	2003	Digitised PA chest images of adults	120	nodules	12%, 50% and 82%	confirmed reports
Manning	2004	Digitised PA chest radiographs	120	nodules	12%, 50% and 82%	confirmed reports
Manning	2006a	Digitised PA chest images	120	nodules	20%, 50% and 75%.	confirmed reports
Manning	2006b	Digitised PA chest images	120	nodules	12%, 50% and 82%	confirmed reports
Piper	2014	Adult chest radiographs (age>/ = 16 years)	100 images	Multiple pathologies	Abnormal to normal 1:1 anon a&e cases approximated 75%	3 consultant radiologists
Semakula-Katende	2016	Paediatric chest radiographs	15	pulmonary tuberculosis	33.3% normal	paediatric radiologist
Sonnex	2001	Plain chest radiographs	8614 performed	Multiple pathologies	402/8614 images abnormal,1:20 464 red dotted	a radiologist
Woznitza	2014	Digital radiography (DR) adult CXRs	149 interpretations	Multiple pathologies	59/99 normal (59.6%)	3 consultant radiologists consultant respiratory physician

2.4.14 Confidence levels

Confidence levels were recorded by Hughes et al. (1996) for pre- and post-training. There was an increase in confidence seen post training in all groups with the mean confidence increasing from 7.71/10 to 8.73/10 (median confidence value of 7 pre tutorial to 9 post tutorial). Other studies asked for confidence levels from participants in order to complete certain types of analysis (ROC, AFROC, JAFROC) however these figures were not presented directly as confidence levels. This limits the comparisons on confidence which can be made between studies.

2.5 Discussion

A broad and informative systematic review has been performed identifying 13 articles to have supplied chest image interpretation training to radiography students, radiographers or reporting radiographers and assessed their performance before and after the intervention. The review allowed the high quality and body of evidence in this area to be identified and presented. Whilst JAFROC has shown to be a popular choice of analysis within image interpretation studies, FP, TP, FN and TN are also highlighted within many of the articles. Monitoring interpretation time and setting fixed interpretation time has its benefits in avoiding fatigue (Manning et al. 2006a; Manning et al. 2006b), however it also comes with drawbacks in that participants are conscious of a time limit given to them and may cause them to rush their image interpretation at intervals throughout the study.

The role progression of chest image interpretation by reporting radiographers in recent years correlates with the number of articles recently produced on the chest image reporting as opposed to commenting by radiographers on the images. Articles published outside of the last decade have focused on participants allocating images to a given category or highlighting abnormal images (Flehinger et al. 1978; Sonnex et al. 2001), whereas in recent years most UK articles have focused on assessing performance of trained/training chest image reporting radiographers (Piper et al. 2014; Woznitza et al. 2014). Nonetheless, the most recent article by Semakula-Katende et al. (2016) has provided participants with choices of response with which to assign to each image. This type of study undertaken in South Africa is representative of the current practice of radiographers within the country and therefore is fitting to the clinical expectation of radiographers within that region. Those articles featuring chest reporting are based within the UK and arose as a result of the recently established role

of reporting radiographers within this area. Beyond the difference in interpretation task, approximately 84.6% (11/13) of articles were within the UK.

Imaging equipment has changed over recent years due to healthcare demands, technological advances and safety regulations. Flehinger et al. (1978) and Donovan et al. (2008) completed the studies in different eras however the similar terminology of 'film' was used in both. 'Film' is suggestive of the older imaging equipment techniques and viewing techniques used used and so this may have been a simple lapse in terminology used by Donovan et al. (2008) or indeed they opted to use the older imaging display format. Flehinger et al. (1978) was the only study within the review to include lateral chest radiographs and the only study undertaken within North America suggesting this was a reflection on the imaging protocols of the country of origin or the timeframe as this was the earliest study completed within the review (1978).

Although each article, in keeping with the inclusion criteria, featured reporting of projection radiography chest examinations, there was a variation in terminology used throughout the articles. The use of chest radiography, film, image or x-ray suggests information on how the chest examination was presented to participants. The type of information given (i.e. film presented over a light box or a digital image displayed on a high quality viewing monitor) could have impacted the participant's performance and ability to distinguish between absence/presence of a pathology. Viewing conditions and image presentation could impact the ability of the viewing medium to deliver sufficient spatial resolution or cause a change in contrast and density of the image (Brennan et al. 2007; Cosman et al. 1994).

The review has demonstrated that accuracy is higher in the majority of studies where specific postgraduate training is being evaluated (Piper et al. 2014; Woznitza et al. 2014). This is because a programme has been designed specifically to cover all requirements of the reporting radiographer's chest image interpretation role and it will be their sole responsibility once qualified to provide a report and diagnosis which will directly influence patient care. Other roles undertaken by radiographers in chest image interpretation such as abnormality highlighting red dotting or training which attempts to improve the radiographer's general ability to complete Preliminary Clinical Evaluation (PCE) to identify pathologies are unlike reporting radiography in that they do not carry this higher level of responsibility and liability

(College of Radiographers 2006; College of Radiographers 2013). Training within reporting radiography therefore was more thorough to ensure these high levels of accuracy are met. Nonetheless only two studies completed provided evidence within this field, one of which contained one participant, and so there remains to be a lack of knowledge on the current training and standards of this relatively new role. The 'red dot' approach to image interpretation, mentioned earlier in this chapter and in chapter 1, is becoming outdated. Perhaps the lead into Preliminary Clinical Evaluation and reporting for the profession will require app development for all professionals working clinically and not only those who have qualified from a postgraduate reporting programme.

Few studies tested the participants on their ability to identify and distinguish between a range of pathologies (n=3), with only two recent studies testing participants on a range of pathologies and their ability to provide a diagnosis (Piper et al. 2014; Woznitza et al. 2014). These two studies tested the effect of postgraduate education and so there remains a lack of variety of training tested which requires participants to provide identification details on a range of pathologies within a single examination or across several patients. Although the postgraduate programmes provide evidence of high levels of accuracy by those students, there has been no detailed investigation into training which may complement these programmes. The eye tracking feedback tested to date was used in lung nodule detection only and with little or no guidance provided with eye movements. In Hughes et al. (1996) at least 25/26 radiographers acknowledged tutorials incorporating a search strategy as useful, however many (n=19) claimed to have a have a technique for looking at chest radiographs already and therefore the true helpfulness of the search strategy implemented may be flawed. These search strategies could also have been formed for use in checking the radiographic technique of an image by radiographers rather than looking for a pathology. Tutorials, lectures and short courses proved useful also (Flehinger et al. 1978; Semakula-Katende et al. 2016). However, these were proved applicable in chest image interpretation roles other than reporting and have yet to be tested for their effect within chest image reporting by radiographers.

Postgraduate training demonstrated high levels of accuracy from all participant groups, to a standard of performance similar to radiologists. Training/feedback with little or no instructions will not be beneficial to trainees. In Litchfield et al. (2010) performance FOM

scores decreased or remained the same once participants were given time to preview the image before image interpretation. The lack of guidance on this intervention may have led it to being little or no use and rather the participants began to doubt themselves once shown the image for a longer period of time prior to interpretation rather than gain from the experience. Implementation of electronic checklists may be an option to reinforce or support training, as the majority of such checklists proved successful within healthcare during a review by Kramer et al. (2016) with only 1/15 articles identifying the electronic checklist as non-beneficial. As far as we are aware, such electronic checklists have not been tested specifically for their use in chest image interpretation by radiographers.

2.6 Future chest image interpretation training

A consensus on education, training and its effect on chest image interpretation learning can inform future practice. Publication of accuracy levels achieved within radiographer roles of chest image interpretation can help confidence grow and progression within them. Collaboration of training techniques may help maximise learning and accommodate for those with different preferred learning types, but this would need further study. Further studies could test learning techniques other than postgraduate programmes on a range of pathologies. Reporting on radiographic images is a very difficult and challenging task to learn and therefore training techniques which assist this process should be tested for their effectiveness. Given the published success of eye tracking and tutorials in chest image interpretation by radiographers, their effect on reporting of chest radiographs should be investigated.

2.7 Limitations

The variation in accuracy measurements within this literature review limits the comparisons between studies and what conclusions can be drawn. The variation in the tasks participants were given and their experience made the data difficult to compare and awkward to present. Whilst two of the most recent studies provided ample data and featured similar training of a postgraduate chest image interpretation programme, the comparisons needed to be made with caution as the experience of the reporting radiographer(s) varied greatly with participants who had just completed their training featured within Piper et al. (2014) and a reporting radiographer, having completed the CXR reporting postgraduate programme for two years, featured within Woznitza et al. (2014). Some studies used simulated nodules within phantom

chest images to test the image interpretation of participants. This task is not a true representation of the images encountered by participants in clinical practice and therefore questions the credibility of these studies. Some studies instructed participants about the pathology that was to be identified, i.e. told to spot a range of pathologies, therefore leading to a difficult comparison of these studies and studies which gave no indication of what pathologies the participants may be tested on. Those which were given no information on pathologies etc. were faced with a more complex task during their interpretation session.

A large number of the critiqued studies gave little detailed information on the accuracy performance and/or eye tracking metrics of participants. Additional information was requested from lead authors via email however not all replied with the information or they no longer had access to data. This led to few articles being included within accuracy and eye tracking comparisons. Studies within the review reported the impact of the training but did not reveal details of the accuracy of each participant/participant group.

2.8 Conclusion

Radiographers demonstrate high and improved levels of accuracy where chest image interpretation training has been undertaken. Training varied greatly in form, from relatively informal to formal, including; postgraduate programmes, eye tracking feedback, tutorials, lectures, mentoring and courses. Accuracy improved regardless of the training type however some training methods enhanced this improvement more than others. The most appropriate training method depends largely on the role of chest image interpretation the radiographer wants to learn. Although postgraduate programmes are ideal for chest image reporting, a role where patient diagnosis relies solely on the radiographer's knowledge and skill; less formal methods such as mentoring and tutorials may be useful for improving the radiographer's confidence in undertaking 'red dot' or 'image comment' roles. Eye tracking feedback proved most useful when observers were shown an expert or novice eye movements; this information was least useful when given little or no guidance. Therefore eye tracking feedback incorporating expert's guidance and instruction could be a more beneficial method of training and a tool to assist postgraduate training. It is advisable that whichever training method is chosen that it is accompanied by monitoring of student performance to ensure its worthiness, validity and success, as seen in MSK oriented studies such as by Carter et al. (1999) and Jones et al. (2007).

Chapter 3 - Experimental study 1

3.0 Introduction

There have been studies completed that investigated the techniques used by radiographers to interpret accident and emergency images which also include a combination of both the appendicular and axial skeleton (McConnell et al. 2000; Brealey et al. 2014). Whilst it is crucial to assess the radiographer's accuracy in image interpretation, it is also vital to understand their patterns and methods of image interpretation. Previous studies have used computerised eye tracking technology to assess the radiographer's ability to interpret images. These studies were carried out by Manning et al. (2006a) and Donovan et al. (2008). Studies have used participants with various levels of expertise to try and establish the differing image interpretation patterns shown by each of the groups. Eye tracking technology provides an insight into the subconscious cognitive processes of the radiographer during their image interpretation. Manning et al. (2006a) and Donovan et al. (2008) used a single or multiple simulated "nodules" or lung masses within their abnormal chest radiographic images to test the participants using the alternate free response operating characteristic (AFROC) methodology. The AFROC methodology requires the observer to decide on the presence and location of a nodule and supply a confidence level. The methodology was used to assess the interpretation of the digitally added nodules however the addition of the nodules to the images could have had an impact on the participant's ability to visualise them. Donovan et al. (2008(noted a significant difference in the group that were given personalised feedback that was based on their individual eye tracking analysis. An improvement was most evident in the performance of level 1 student radiographers, students within their first year of studying, with a percentage increase in the figure of merit (FOM) of 8.4% (p<0.05). JAFROC, the analysis software, generated a FOM that quantifies search performance. It was defined as 'the probability that an observer will rate a lesion higher than the highest rated non-lesion on a normal image' (Donovan et al. 2008). There was less of an effect noted in the performance of novice and expert participants leading to the conclusion that perceptual feedback of eye tracking may be beneficial to naïve radiographers in image interpretation. Manning et al. (2006a) utilised the eye tracking technology to monitor performance measures and noticed

the results were significantly better in the expert/trained radiographers in comparison to the rest of the studied cohort (p=0.046).

Radiographer's participation within eye tracking studies to date has focused mainly on their ability to diagnose single chest pathology and/or chest pulmonary nodules. Manning et al. (2006a) noted, by studying visual coverage of the image, that experts tended to inspect less of the area on the images compared to novices. In particular they noticed radiographers assumed this method, of inspecting fewer areas on the image, after receiving their training. Donovan et al. (2008) noted that the eye tracking data of Level 1 and Level 2 radiographers displayed a great deal of variability. Eye tracking technology used within Donovan et al. (2008) and Manning et al. (2006a) provided valuable information on how the participant groups viewed images. The use of a range of anatomical areas within differing body area radiographic images and a range of pathologies could challenge the participant and stimulate the radiographer to interpret the image using a different search strategy.

Studies focusing on the interpretation of computed tomography (CT) brain images and electrocardiograms have used a think aloud technique alongside the use of computer-based eye tracking technology. This allows generation of a comprehensive understanding of the clinician's image interpretation (Matsumoto et al. 2011; Bond et al. 2014). The 'think-aloud' technique occurs when the participant verbalises their thought processes during their interpretation. As a result, the think-aloud technique has been incorporated into the current study to elicit cognitive insight into the image interpretation carried out by this cohort.

By studying the voice recordings from the 'think-aloud' protocol, there is a potential to further understand the participant's image interpretation process which complements the eye gaze data (Bond et al. 2014). Although an anticipated higher accuracy and confidence level from the experienced and qualified reporting clinicians was to be expected, it would be interesting to guage whether the participant's level of training would be reflected through the eye gaze metrics and if particular correlations could be found within the study such as trends in the use of language of particular study groups.

Therefore, the aim of this study was to investigate the search strategies and the image interpretation techniques adopted by (1) student diagnostic radiographers, (2) diagnostic

radiographers and (3) reporting radiographers (who specialise in reporting on the musculoskeletal system) with the use of computer-based eye tracking technology using a range of anatomical areas and pathologies.

3.1 Aims and objectives

The aim of this study was to investigate the search strategies and the image interpretation techniques adopted by participant groups with the use of computer-based eye tracking technology. Also, we aimed to analyse the diagnostic accuracy amongst participants of various levels of expertise across a range of anatomical areas and pathologies. The study aimed to achieve this by identifying;

- patterns of interpretation by computing eye gaze metrics along with the duration of each interpretation for different types of pathology

- correlations between interpretation strategies and diagnostic accuracy

- inter-rater reliability amongst all participants and the common errors encountered

3.2 Methodology

3.2.1 Study approval

A research protocol and ethical approval application was completed and submitted to the Ulster University Research and Ethics Filter Committee on 11/04/2014. Following peer review the application was approved and ethical permission was granted. (See appendix 3.1)

3.2.2 Participants

Student radiographers with at least one year undergraduate education in diagnostic radiography were approached within the University by distributing an email. Radiographers were recruited to participate in the study at the Society and College of Radiographers Northern Ireland Conference in 2014 and 2015. The reporting radiographers were initially approached at a Reporting Radiographers Interest Group Scotland meeting for reporting radiographers who are qualified to reporting on images of the musculoskeletal system.

Anyone interested in participating received a participant information sheet which included a consent form and was given time to consider their participation in the study. They were then screened against the ethical protocol and if eligible a suitable time was scheduled for participation in the study. A brief overview of the study was given by the researcher and any questions or concerns the participants raised were answered honestly and in full. The participant was then asked to sign the consent form once content to complete the study. Participants were advised on the option to leave if they wished at any time but with acknowledgement that their data from any part of the process might be used prior to the decision to leave the study.

3.2.3 Images

The study included 8 images of the appendicular skeleton, axial skeleton and chest cavity; 6 musculoskeletal and 2 chest X-ray images were used. Images were formatted in the Joint Photographic Experts Group (JPEG) image file format and inserted into the eye tracking software. The same 8 images were interpreted by each participant to generate a large interpretation dataset (n=464). The image set consisted of 1 'normal' and 7 'abnormal' images. We did not want the participants to commit long periods of time (away from their study/place of work) and hence the decision to use only 8 images in this initial test was made. We chose to present a range of pathologies and one normal image to assess the reporting clinician's ability to identify and diagnose the pathology. This is similar to the studies of Brealey et al. (2014), where a range of pathologies were demonstrated on the appendicular and axial skeleton and Piper et al. (2014) where participants were assessed on image interpretation of a range of chest pathologies. This was also a study designed to investigate image interpretation strategies and accuracy and so it was decided to include a range of abnormalities and only one normal image.

Each set of 8 images were shown to the participants, who were unaware of how many images were normal/abnormal or in which order they would be presented. We aimed to include various image pathology types; one pathology, multiple pathologies, fracture, pneumothorax, lung mass etc. We chose the images at random from a test bank to represent and test participants on a range of abnormalities. The eight images chosen were shown to all participants. No clinical history or previous patient examinations were supplied with the

image; participants were asked to form a diagnosis solely on the image they were asked interpret.

3.2.4 Reference standard

The images included within the study were sourced from an online repository. The repository, an educational website supplying case studies, supplied a diagnosis with each image. For completeness, a senior reporting radiographer (member of the research team) was asked to provide a written diagnosis and the consensus agreement of image content with most likely diagnosis was agreed.

3.2.5 Prior to the study

3.2.5.1 Test environment

Each participant completed the study within a quiet, controlled and isolated environment. The researcher was present to operate the eye tracker, move from one image to the next image at the participant's desired speed and to record a written diagnosis that was verbally elicited by the participant. The researcher's presence was necessary during the image interpretation session to prompt the participant in providing a self-rated confidence level of their interpretation and diagnosis and to ensure that the equipment was operating as planned. The confidence level was asked for on a scale of 1-10 (1 being not confident and 10 being very confident in their given diagnosis).

3.2.5.2 Equipment

The Tobii Studio X60 eye tracker and the Tobii studio software© were utilised for data collection and for computing eye gaze metrics (Tobii AB 2016). The remote non-intrusive eye tracker collected the data without interference to the participant's interpretation. The eye tracker was positioned inferior to the high resolution (1440px x 900px) 24" LCD monitor that displayed the images and angled upwards (30° cranially) to align with the participant's gaze. The monitor used would be inferior to clinical reporting workstations, however the monitor was taken to data collection sites and allowed standardisation across participants for their viewing environment. This upwards angulation allowed the infrared light emitted from the eye tracker to reflect off of the participant's cornea. The angle at which the infrared light reflects off of the cornea provides information on where the eyes fixate on the image.

3.2.5.3 Participant position

When the participant was ready to begin the study they were seated directly opposite the eye tracker and the monitor. We ensured that the participant gazed at the center of the screen whilst a comfortable position was achieved for the duration of the study. Distance from the viewing monitor and chair height was altered at this point to meet the position required by the equipment to receive optimum eye tracking data.

3.2.5.4 Calibration

Prior to beginning the calibration of the eye tracking equipment, the participant was given an explanation of what to expect. They were asked to fixate their eyes on the red dot which would appear on the screen and to follow this red dot around the screen to the best of their ability. When successful calibration was achieved, the participant was instructed to maintain their position as much as possible throughout the study. The researcher's presence during the study allowed monitoring of any movement. The restricted movement of the participant aimed to limit the interruption to the eye tracking data being collected. If calibration was unsuccessful, the position of the participant was adjusted and the calibration process was repeated until a successful position and calibration was achieved.

Care was taken during the calibration process to ensure optimum eye tracking data was collected and high eye tracking quality was achieved. Eye tracking quality is defined as the "spatial and temporal deviation between the actual and measured gaze direction and the nature of this deviation, on a sample to sample basis", (Holmqvist et al. 2012). Measuring eye tracking quality allows the collection of eye tracking data from the participant's performance to be monitored. Data quality can be influenced by participants, operators, the task, recording environment, geometry or the eye tracking design (Holmqvist et al. 2012). By ensuring stringent checks were applied during the calibration process we hoped to minimise the effect of features listed above on the eye tracking data quality obtained.

3.2.5.5 Instructions given prior to the presentation of the first image of the study

Since the 'think aloud' method was used during the study, all participants were reminded to verbalise their thought processes as much as they could. The participant was asked to indicate when they were ready for the first image to appear on the display and for the study to begin. Participants were asked not to talk about the images with fellow students/colleagues which were due to complete the study also.

3.3 Participant groups

A total of 21 undergraduate radiography students were recruited, each of whom had progressed to either the second or third year of a three year undergraduate diagnostic radiography and imaging degree within Ulster University. A further 19 experienced radiographers of various specialities and years of experience were recruited through their attendance of the UK Society and College of Radiographers conference within Northern Ireland and 18 reporting radiographers experienced in reporting on images of the musculoskeletal skeleton were recruited within University settings.

3.4 Outcome measures

Participant accuracy (proportion of correct interpretations) was measured within the study. Images were marked as correct (1) if the reference diagnosis was stated or similar to what the participant described. Interpretations that were inconclusive, stated that they could not provide a diagnosis or provided multiple incorrect answers were awarded a score of 0 for the image. The researcher calculated the participant's accuracy in image interpretation.

Participant confidence was measured. Once the diagnosis was provided, the researcher requested and recorded a confidence level on the given diagnosis. This confidence level was given on a scale of 1 to 10.

The following eye gaze metrics were also computed:

- Fixation duration: Measure of the sum of the duration for all fixations within a defined area of interest (AOI).

- Fixation count: Measure of the number of times the participant fixated on an AOI.

- Time to first fixation: Measure of how long it took before a test participant fixated on an AOI

- Visit duration: Measure of the duration of all visits within an AOI.

- Visit count: Measure of the number of visits within an AOI.

- Fixation frequency (fixation duration/ fixation count)

Each eye gaze metric was analysed for all three groups of participants for the selected area(s) of pathology (AOP) within each abnormal image and for each entire image (when appropriate to do so).

Decision time was measured for the time spent interpreting the image and providing a diagnosis. A questionnaire was given to the participant following each session. The questionnaire contained a range of questions on the experience and training received by the participant and their thoughts on eye tracking technology.

3.5 Data analysis

Descriptive statistics (mean, standard deviation, minimum, maximum and inter-quartile ranges) and Kolmogorov-Smirnov normality tests were also completed before deciding on which hypothesis (inferential and statistical) tests to use. Spearmans *rho* correlation coefficient was used to investigate the correlation between accuracy, confidence, and decision time and fixation frequency. A one way analysis of variance (ANOVA) was completed to investigate statistical significant differences between the three groups for each of the normally distributed outcome measures. When using non-parametric data (data that was not normally distributed), a Kruskal-Wallis test was completed to investigate whether there was an overall significant difference within the three groups. To strengthen the validity of results several Mann-Whitney U tests were completed using two groups to obtain statistically significant values for each significant difference noted within the Kruskal-Wallis tests. An alpha value of 0.05 was used to determine statistical significance.

3.6 Results

3.6.1 Descriptive statistics

The greatest eye tracking sampling quality was collected from the reporting radiographers (82.5%), followed then by data collected from the radiographers (80%) and subsequently then

by the students (74%). The eye tracking sampling quality is lower in the student cohort than reporting radiography cohort (p=0.02). It was noticed that students tended to look away from the monitor following and in between their image interpretations, therefore this may have contributed to the lower sampling quality obtained.

Table 3-1: Participant group demographics

	All	**Students**	**Radiographers**	**Reporting radiographers**
Age (years)	34.6 ± 14.0	21.4 ± 2.5	44.1 ± 12.9	40.0 ± 11.3
Experience interpreting images (years)	9.7 ± 11.3	1.6 ± 0.9	17.9 ± 13.6	10.5 ± 8.5
Experience reporting images (years)	1.3 ± 3.0	0	0	4.3 ± 4.1
Gender	Female 87.9% Male 12.1%	Female 81% Male 19%	Female 89.5% Male 10.5%	Female 94.4% Male 5.6%

3.6.2 Confidence and accuracy

Participants were asked to rate their confidence in the given diagnosis on a scale of 1-10, with 1 being not confident and 10 being very confident in their given diagnosis. Reporting radiographers were more confident in their given diagnosis than radiographers ($p<0.001$) and students ($p<0.001$) (Table 3.2). Reporting radiographers had a greater median confidence level of 2.2 compared to students and also a greater median confidence of 0.8 than radiographers. In addition, radiographers had a 1.4 greater median confidence than students ($p \leq 0.001$). Reporting radiographers were more accurate than radiographers ($p<0.001$) and students ($p<0.001$). Radiographers were more accurate than students ($p=0.03$) (Figure 3.1).

Table 3-2: Total confidence in diagnosis of each participant group

	Students (n=21)	Radiographers (n=19)	Reporting Radiographers (n=18)
Confidence	5.9 (4.8 - 6.8)	7.3 (6.4 - 7.8) *	8.1 (7.8 - 8.6) * #

*Total confidence levels collected from students, radiographers and reporting radiographers. All values are medians (inter-quartile ranges). Total confidence was calculated over the total confidence given for 8 images on a scale of 1-10 (1 being not confident and 10 being very confident in their given diagnosis). *indicates significantly different to students (P<0.05) # indicates significantly different to radiographers (P<0.05)*
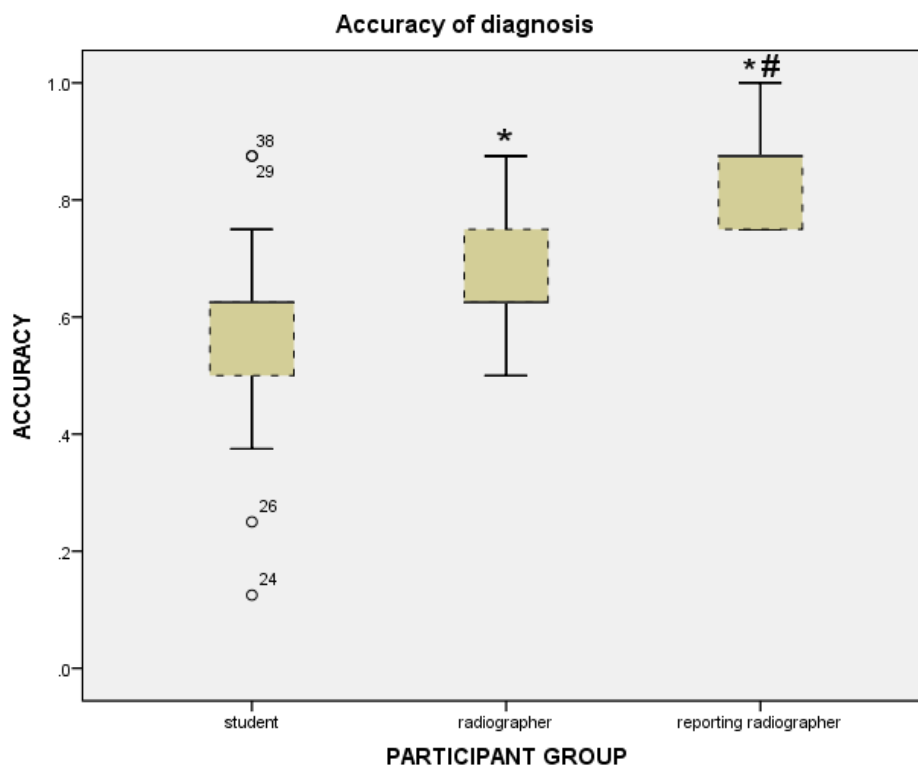
= Different compared to students (P<0.05) # = Different compared to radiographers (P<0.05) o = Outlier

Figure 3-1: Accuracy of diagnosis within the student, radiographer and reporting radiographer cohort

3.6.3 Eye tracking

The time to first fixation decreased with experience in that the most experienced group, the reporting radiographers, had taken the shortest time (4.3s) before fixating on the pathology (Table 3.1); radiographers took 5.2s to first fixate on the pathology and students took the longest time to first fixate on the pathology (5.5s) (Figure 3.2). However, there was no significant difference between the total times to first fixate (Table 3.3).

When compared to students, reporting radiographers had a greater mean fixation duration (p=0.01), mean fixation count (p=0.04) and mean visit count (p=0.04) on the areas of pathology (Table 3.3; Figure 3.3). There were no statistically significant differences noted between the radiographers and reporting radiographers when the eye gaze metrics within the area/s of pathology were compared (Table 3.3). However, we can see a trend in the results between the groups for mean fixation duration, mean fixation count and mean visit count for the areas of pathology. These eye gaze metrics tended to increase as the level of expertise/capability increased (Table 3.3; Figure 3.3).

Table 3-3: Eye tracking data for each participant group which was collected from the area of pathology within each image

	Students (n=21)	Radiographers (n=19)	Reporting radiographers (n=18)
Mean time to first fixation (secs)	5.5 (3.4 - 8.4)	5.2 (2.7 - 7.0)	4.3 (2.1 - 6.5)
Mean fixation duration (secs)	6.0 ± 5.2	8.0 ± 3.4	11.3 ± 6.6 *
Mean fixation count (n)	20.4 ± 15.8	27.2 ± 11.1	32.7 ± 17.8 *
Mean visit count (n)	9.9 ± 6.1	10.6 ± 3.5	14.6 ± 7.4 *

*Eye gaze metrics collected from students, radiographers and reporting radiographers for each area of pathology within each abnormal image. Time to first fixate is presented in median (inter-quartile range). Remaining data is presented in mean ± standard deviation. *indicates significantly different to students (P<0.05) # is significantly different to radiographers (P<0.05)*
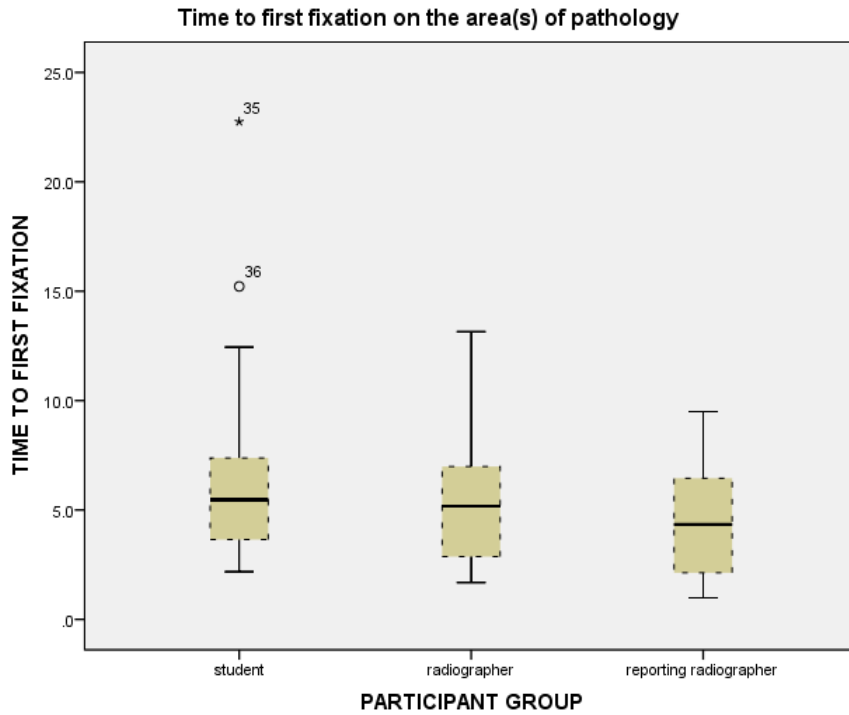
Figure 3-2: Time to first fixate on the areas of pathology within the student, radiographer and reporting radiographer cohort

o = Outlier
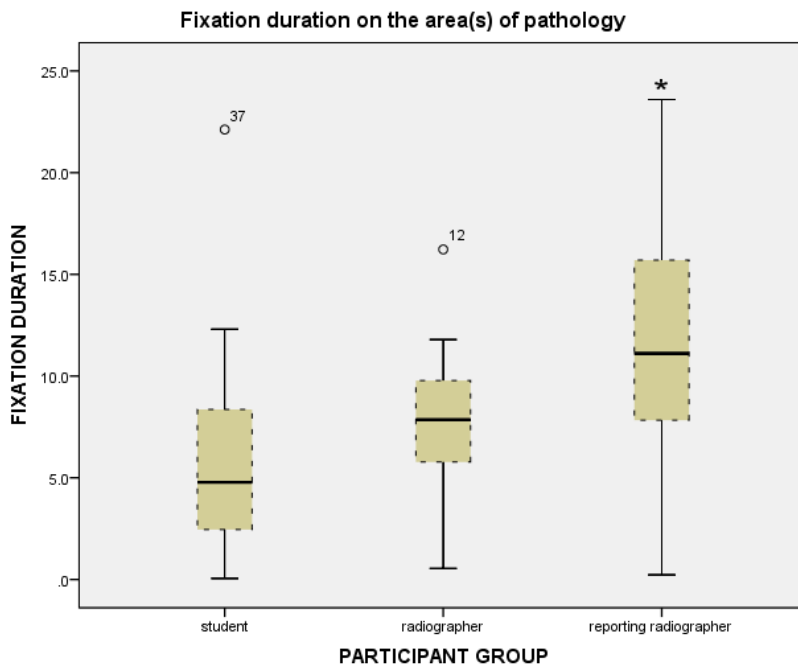
Figure 3-3: Mean fixation duration on the areas of pathology within the student, radiographer and reporting radiographer cohort

* = Different compared to students (P<0.05) o = Outlier

Reporting radiographers had the longest mean fixation duration over the entire image. Their mean fixation duration (44.1s) was 16.0s longer than radiographers (28.1s) (p=0.05). In addition, reporting radiographers also demonstrated the largest number of fixation counts for the entire image (143.1) of the three groups (Table 3.4).

Radiographers spent less time viewing the images before coming to a decision (49.4s) than students (p=0.04) and reporting radiographers (p=0.02). Students and reporting radiographers spent longer viewing the image, 63.4s and 65.8s respectively (Table 3.4), before coming to a decision on the diagnosis. Yet students had the lowest median accuracy (%) and reporting radiographers had the highest accuracy (%) (Figure 3.1).

Table 3-4: Eye tracking data for each participant group which was collected from the entire image

	Students (n=21)	Radiographers (n=19)	Reporting radiographers (n=18)
Mean fixation duration	32.9 ± 19.4	28.1 ± 12.1	44.1 ± 26.7 #
Mean fixation count	124.1 ± 66.3	110.0 ± 45.9	143.1 ± 68.5
Mean decision time	63.4 ±18.5	49.4 ± 14.0 *	65.8 ± 19.0 #

*Data is presented in mean ± standard deviation. *indicates significantly different to students (P<0.05) # indicates significantly different to radiographers (P<0.05)*

3.6.4 Correlations

There was a weak negative correlation between accuracy and decision time of the reporting radiographers (r=-0.20, P<0.001). If reporting radiographers spent longer interpreting the image then they were more likely to be inaccurate in their diagnosis, however because of their overall high accuracy rate of 87.5%, it was rare that they were wrong in their diagnosis. Within this study, reporting radiographers demonstrated 100% accuracy in the interpretation of the musculoskeletal system images (Figure 3.1).

A weak negative correlation existed between confidence and mean decision time (r=-0.22, P<0.001). When studied further, a moderate negative correlation was found between these two variables within the radiographer (r=-0.68, P<0.001) and reporting radiographer (r=-0.45, P<0.001) groups but not within the student group (r=-0.06, P<0.001). This would imply that with expertise, the more time spent interpreting an image, the less likely the participant was to be confident with the diagnosis they give. There was no correlation noted between the confidence and mean decision time of students, indicating that the mean decision time taken by the student to interpret an image is unlikely to indicate a level of confidence in their diagnosis.

3.6.5 Fixation frequency

The fixation frequency is the number of fixations per second (hertz or Hz). A high fixation frequency could indicate that the participant rapidly gazed over a large area of the screen and was more sporadic in their image interpretation. A low fixation frequency indicates that the participant had steady eye movements during their interpretation; in addition these individuals were more likely to be controlled in where they fixated within the image. Students had a higher fixation frequency than reporting radiographers (p=0.03) for the area of pathologies within each image. Inexperienced participants were more erratic during the process of image interpretation, compared with the experts in image interpretation who were trained to interpret the image systematically. Radiographers were more accurate than students (p=0.03). The significance seen between these two groups was not as great as that between students and reporting radiographers (p≤0.001) or reporting radiographers and radiographers (p≤0.001). Mimicking these results, there was a small difference in the fixation frequency of students (3.7Hz) and radiographers (4.0Hz) on the entire images also.

There was a positive correlation between total confidence and total fixation frequency for students (r=0.21, P<0.001). The more sporadic the students were in their interpretation, the more confident they were in their diagnosis. However, there was a negative correlation between total confidence and total fixation frequency for radiographers (r=-0.62, P<0.001) and for reporting radiographers (r=-0.20, P<0.001). The more sporadic they were in their interpretation, the less confident they were in their given diagnosis. As the participants with greater experience became more sporadic their confidence levels decreased (Table 3.2).

3.6.6 Heat map results

Due to the vast number of images and heat maps which can be generated by the eye tracking technology, it was chosen to include those which supplied a good visual representation of each cohort's performance and search strategies.

Heat maps taken from the interpretations of each group of participants for the first 10 seconds demonstrate the number of fixation areas observed for reporting radiographers and radiographers are similar (where red areas represent areas of high numbers of fixation counts and green areas represent lower numbers of fixation counts). Less variability is shown by the reporting radiographers in their fixation areas, as they began to "zone in" on the areas of concern. However, the students continue to demonstrate a large variability in their gaze.
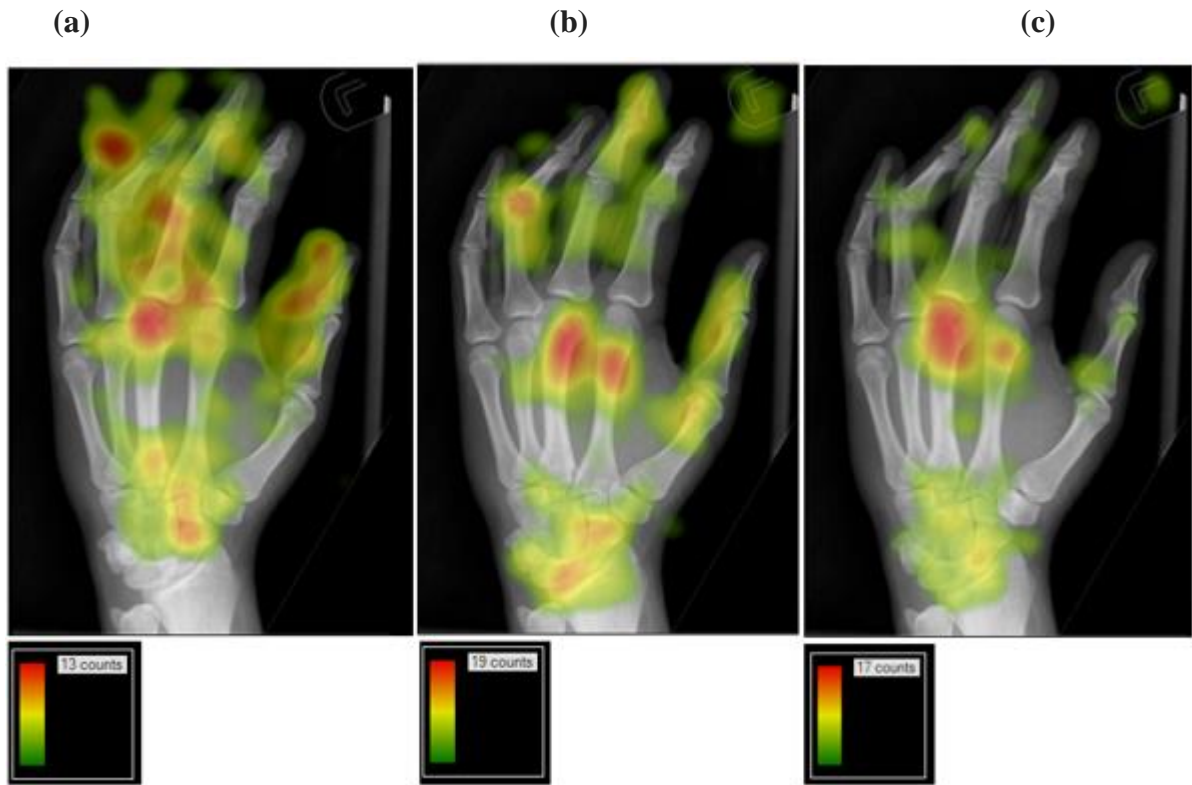
Figure 3-4: Heat maps for the first 10 seconds of image 3.

Heat maps containing the fixation counts of (a) students, (b) radiographers and (c) reporting radiographers during the first 10 seconds of their interpretation of a hand image.

Radiographers and students demonstrated fewer fixations on the second pathology of the fractured ulna styloid within image 4 than the reporting radiographers. 13/19 radiographers and 11/21 students failed to report the secondary pathology of the fractured ulna within image 4, whereas all of the reporting radiographers identified the second pathology.
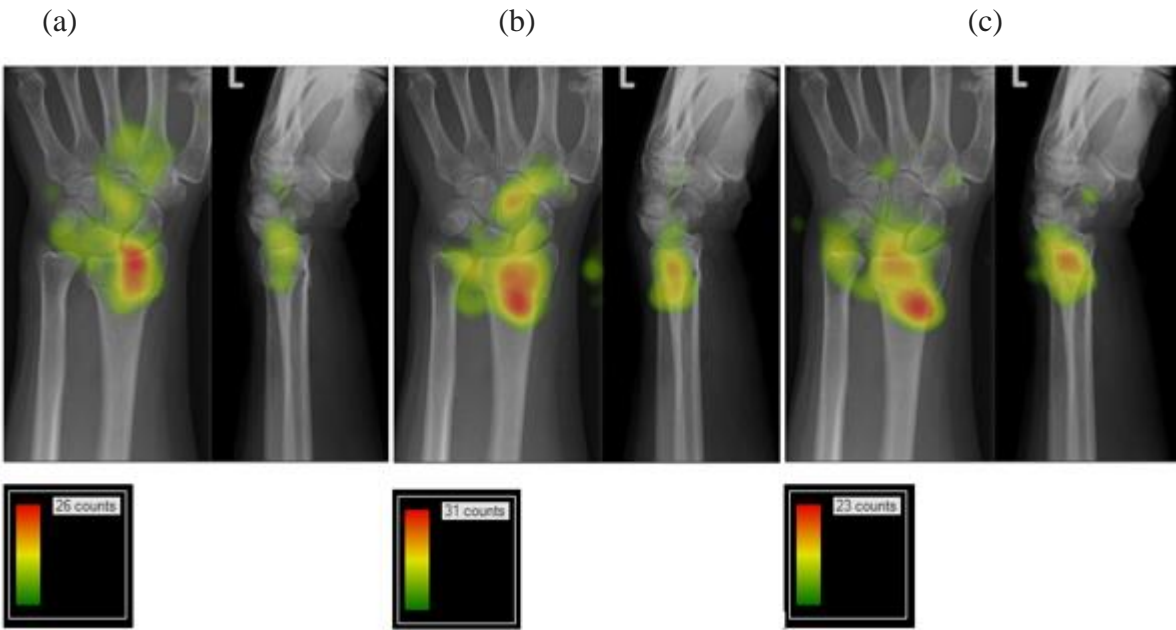
Figure 3-5: Heat maps for 3 images of the first 10 seconds of image interpretation. The heat maps contain the fixation counts of (a) students, (b) radiographers and (c) reporting radiographers during the first 10 seconds of their interpretation of a wrist image.

The reporting radiographers had a greater number of fixations at each of the areas of pathology during the first 10 seconds of their interpretation but importantly it is the only group to have a high fixation count on the area of pathology in each image (including chest images in which they have not received training).

The reporting radiographers demonstrated greater variation in their gaze patterns when viewing the chest images than the appendicular images. This was most likely due to their training in appendicular images only.
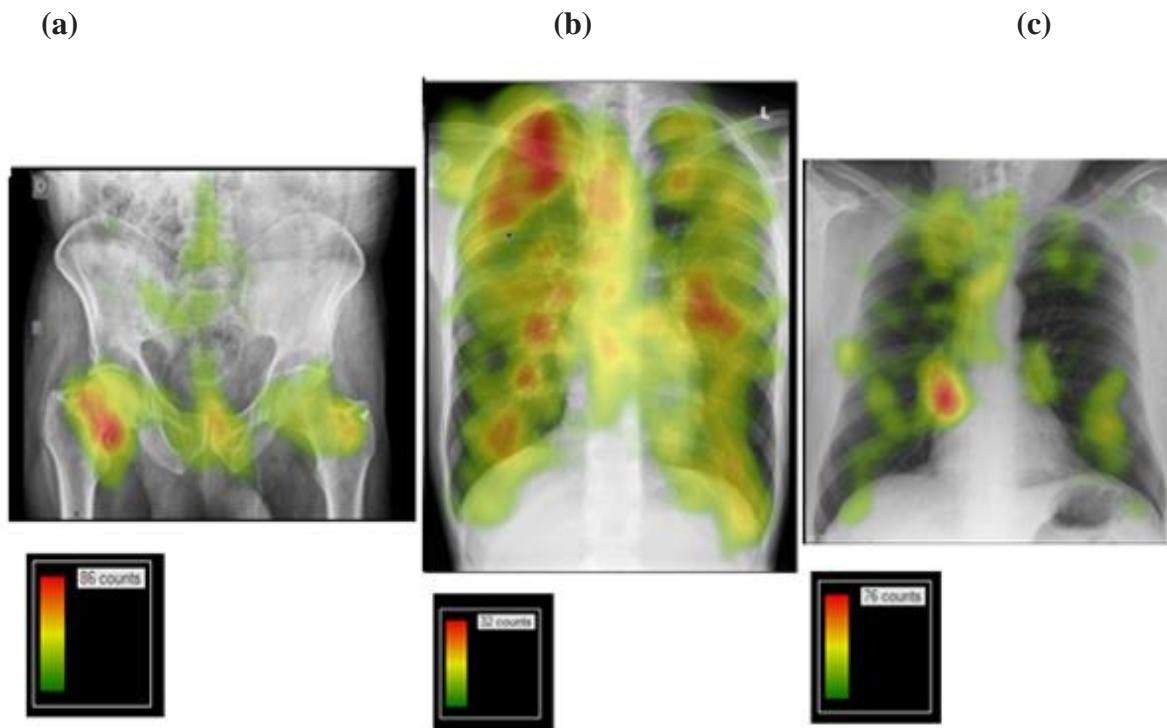
Figure 3-6: Heat maps for 3 images of the entire image interpretation duration of reporting radiographers.

Heat maps contain the fixation counts of reporting radiographers during their

interpretation of (a) pelvis image, (b) chest image and (c) chest image.

When viewed on a heat map, the radiographers had fewer and smaller areas of fixations than the students and reporting radiographers during their first 5 seconds of the image interpretation. Nonetheless they were also the group which did not have any 'high fixation areas' over the area of pathology within the first 5 seconds of interpretation. The reporting radiographers demonstrated greater variation in their gaze patterns when viewing the chest images than the appendicular images.
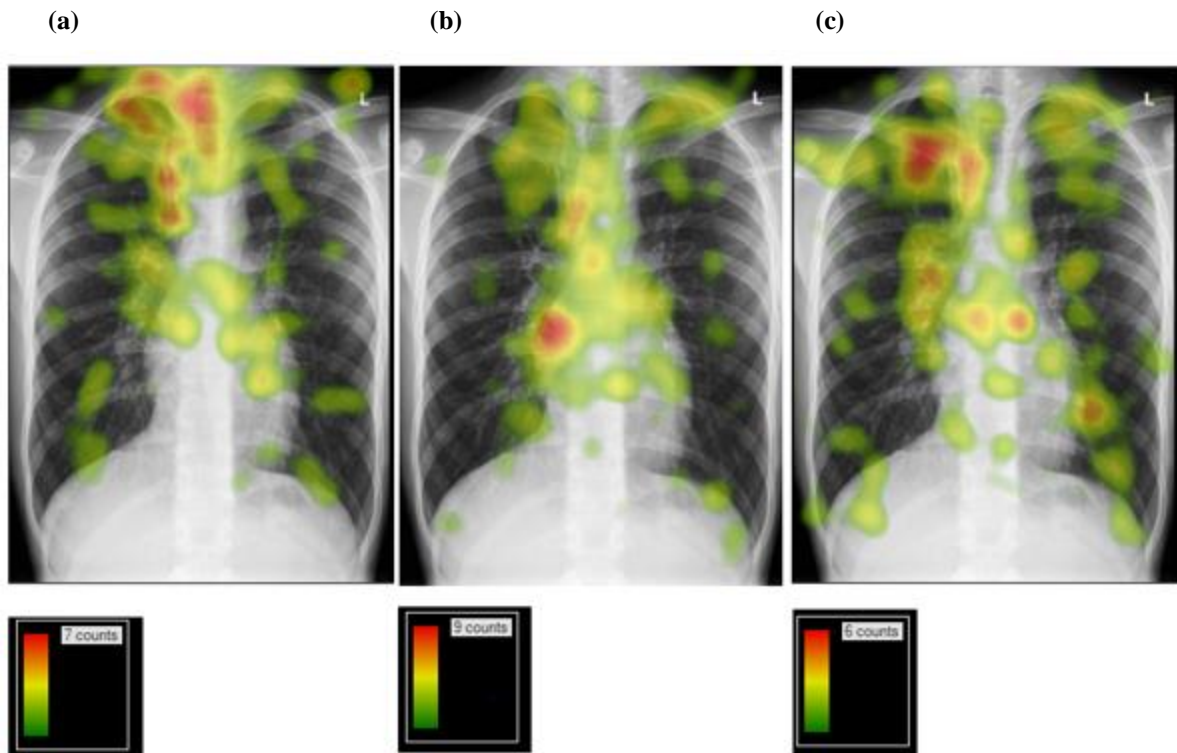
Figure 3-7: Heat maps for the first 5 seconds of image interpretation.
Heat maps contain the fixation counts of (a) students, (b) radiographers and (c)
reporting radiographers during their interpretation of a chest image.

3.7 Discussion

There was a greater level of accuracy in diagnosis demonstrated by the reporting radiographers. This was expected given the training reporting radiographers receive. This supports previous evidence which claims that appropriately trained professionals within this field can complete their work to a high level of accuracy (Piper et al. 2005; Piper et al. 2014; Woznitza et al. 2014). Accuracy of axial and appendicular reporting by radiographers was demonstrated to be between 91.8%-93.7% post training (Piper et al. 2005) and reporting radiographers had a mean sensitivity and specificity of 95.4% (95% CI 94.4%-96.3%) and 95.9% (95% CI 94.9%-96.7%, respectively when reporting on clinical chest radiographs (Piper et al. 2014).The high median accuracy of reporting radiographers (median score of 87.5% of the 6 musculoskeletal and 2 chest images within the study) may reflect the higher fixation count and visit count from this group. Reporting radiographers often gave a more detailed explanation of the pathology, and so could have fixated on and visited the areas of pathology more to assist their explanation of the diagnosis. Reporting radiographers had 100% accuracy in their reporting of the musculoskeletal images and this group therefore only incorrectly diagnosed the chest radiographic images in which they had no specific training. Reporting radiographers spent more time concentrating on the images which they were less familiar with in interpreting (chest radiographic images) and hence led to a negative correlation within this group between decision time and accuracy. Due to the reporting radiographers having received no training and their lack of experience in reporting radiographic chest images, many verbalised their uncertainty in interpreting chest images and were more likely to take longer in forming a diagnosis on the image or to form a conclusion on the pathology present. There may be a fear of missing a pathology also present due to the doubt presented by the RCR in this role progression of image interpretation within radiography (RCR 2012). As decision time increased for this particular group their accuracy decreased. These results are supported by evidence of Manning et al. 2006b whereby incorrect negative decisions were characterised by longer dwell times. However, Manning et al. (2006b) noted the longer fixation times to be more obvious in novice participants whereas within our results there were no correlations seen between accuracy and decision time of radiographers or students.

High accuracy was accompanied by high confidence levels. Radiographers had a slightly lower mean confidence rating (7.3) in comparison to reporting radiographers (8.1). The

training which reporting radiographers have received can provide them with confidence in their professional role. Coleman et al. (2009) identified radiographers to have the lowest confidence and yet the highest accuracy when testing the interpretation of the appendicular skeleton radiographic images by different healthcare professionals. We expected that those who practise image interpretation on a daily basis in clinical practice and those who had received the appropriate training would be more confident in their given diagnosis. Students, as expected, were less confident and often expressed their uncertainty in the given diagnosis or provided the diagnoses with doubt. There was a moderate positive correlation found between the radiographers perceived image interpretation abilities and their achieved score. We also found a small positive correlation ($r=0.36$) between radiographer's accuracy and confidence, indicating that they may be reliable in predicting their performance and ability to provide the correct diagnosis. Reporting radiographers often gave a more detailed explanation of the pathology. This was more than likely due to their experience, training, their duty to provide a full written report in clinical practice, their role to advise on patient care and their experience on the impact a full report can have on the patient's management as imaging or further treatment. These participants fixated on and visited the areas of pathology more often. The voice recordings and eye tracking videos allowed an observation to be made that this group looked at the pathology more to assist their explanation of the diagnosis and provide a full report on the image. Also a reflection on their experience and training, the reporting radiographers were generally first to fixate on the pathology. Time to first fixation was a mean of 1.2 seconds faster than the students and 0.9 seconds faster than the radiographers, however this was not statistically significant.

Experienced reporting radiographers took a longer time to reach a decision in comparison to the students and the radiographers. Again this could have been due to the completeness of the reports provided by the experienced reporting radiographers. This evidence is not what one would expect and contradicts previous evidence that the more experienced observers spent less time viewing images in comparison to novice radiographers (Manning et al. 2006a). Manning et al. (2006a) asked participants to 'decide on a nodule's presence and its location' whereas within this study, participants were asked to interpret the image and provide a diagnosis. This difference in instruction could account for the shorter decision time by experts seen in Manning et al. (2006a) as they are only asked to identify and locate. Furthermore, in agreement with previous evidence, within our study the time taken by

experienced radiographers to reach a decision was faster in comparison to the less experienced students. Participants within our study were not restricted to time, however Manning et al. (2006a) permitted a maximum observation time of 40 minutes. This methodological feature may again have affected how each cohort of participants approached each study. It is also possible there is a satisfaction of search aspect i.e. the abnormality is spotted and then the participant moves on.

There was an increase in variability and widespread fixations observed on the heat maps produced by the eye gaze patterns of the students and reporting radiographers. Variability was expected within the student group due to their lack of experience. However, the variability shown by the reporting radiographers (the most experienced group) was unexpected. The increased variability demonstrated by the reporting radiographers was possibly due to their search strategies or adopting the principle of satisfaction of search. Satisfaction of search suggests that once any pathology has been identified, the image interpreter applies further diligence to continue in searching the image for more than one abnormality that could be pathology related (Berbaum et al. 2010; Krupinski et al. 2010). Failure to continue searching the image once an initial pathology has been identified could lead to a clinically significant abnormalities being missed and as a result these experts were perhaps taught to interpret an image fully. Their search patterns and need for a satisfaction of search could have led them to demonstrate an analysis of the entire image rather than a series of fixations on a few areas within the image (Berbaum et al. 2000; Piper et al. 2014). The increased variability shown by reporting radiographers because of their training was supported by the high confidence and accuracy shown (8.1/10 ±0.8, 87.5% ± 0.1 respectively). This high accuracy is a confirmation of the level of performance as seen in Piper et al. (2005). Kok et al. (2015) supports the increased variability shown by experts having noticed that experts were significantly more systematic than students. They noted a correlation between systematic viewing and coverage which may explain the increased coverage/variability shown by the reporting radiographers within our study, assuming the reporting radiographers viewed the images systematically. Supplying further evidence to support this, the voice recordings of reporting radiographers demonstrated that many of the reporting radiographers immediately stated their recognition of the pathology but adapted a full assessment process to interpret the image before focusing on the pathology once this was completed. The findings of Donovan et al. (2008) where Level 1 and Level 2 groups

demonstrated a great deal of variability within their eye tracking data is similar to the large variability shown within the student group of our study. However they also suggested radiographers are more regimented in how they scan films, whereas our study showed they had in general the least variability within the heat maps. Manning et al. (2006a) noticed fewer areas on the image were inspected by radiographers following training, this is similar to the reduced variability within musculoskeletal images interpreted by reporting radiographers. The training delivered within Manning et al. (2006a) was six months chest image interpretation training and therefore although the reporting radiographers within this study demonstrated increased variability within chest images, their lower variability within musculoskeletal imaging is similar to the findings by Manning et al. (2006a). Lower variability, of the areas fixated on, was seen both in this study and previous studies within participant groups which were trained to interpret images relevant to their training. However, the reporting radiographers demonstrated greater variation in their gaze patterns when viewing the chest images than the appendicular images. This increased variability could have been due to a number of reasons; the added challenge of many chest pathologies, less information given to the radiographers regarding the pathology before the study began or their lack of formal training within this role.

In contrast, the variability demonstrated on the heat maps produced by student eye gazes may have been due to their lack of experience and confidence (experience ranging from 1-3 years interpreting images and mean confidence 5.9/10 ± 2.0). The radiographer's less erratic eye gazes suggest that although they do not possess the uncertainty of a student radiographer, they have not yet established a method of systematically searching the image but rather focus on 'key' areas.

As expected, in general the reporting radiographers had a greater number of fixations at each of the areas of pathology during the first 10 seconds of their interpretation but importantly it is the only group to focus on the area of pathology in every image (including chest images in which they have not received training). Their training to identify abnormalities within the musculoskeletal system could have allowed them to transfer their skills in finding areas of abnormality/pathologies within the chest images. Their ability to systematically search an image could have allowed them to interpret the chest images fully and find pathology(ies)

even when they have not been specifically trained within this area. Their experience of working as a radiographer and looking at chest images in practice, along with their experience in reporting, is opined to have contributed to how they approached the task of interpreting the chest images, i.e. knowing about key areas of pathology such as costal angles for pleural effusion, lung apices for pneumothorax and lung fields for black or white densities.

The reporting radiographers have developed methods of systematically searching the appendicular images beyond initial education for reporting, however given their lack of experience and training in interpreting chest images, it was expected that there would be a lack of certainty when viewing these images compared to the appendicular images. However, the heat maps suggest that the reporting radiographers adopted a systematic approach within the chest images also and the greater variability of their eye gazes mimic their aim to achieve satisfaction of search. Although reporting radiographers tackled the unfamiliar task of interpreting chest images by employing a systematic approach, some participants mentioned the increased difficulty performing this task compared to the interpretation of the musculoskeletal images.

Radiographers and students demonstrated fewer fixations on the second pathology of the fractured ulnar styloid in image 4 than the reporting radiographers. Reporting radiographers are trained to interpret the image fully and rule out more than one area of pathology, they have further knowledge about aspects such as biomechanics, modes of injury or common sites for pathology to be found. Therefore, the increased fixations of the reporting radiographers on the second discrete pathology within image 4 could have been due to the reporting radiographer's need to achieve the 'satisfaction of search' which they have been trained to complete, and this is combined with their knowledge of common mechanisms of injury and their patterns of abnormality. The radiographers lack of 'satisfaction of search' when identifying the second pathology of the fractured ulna styloid process reflects the 67% of radiographers within the Coleman et al. (2009) study who failed to notice further fractures once having identified one fracture within an image.

In general, the heat maps provided information on each of the groups' approach to image interpretation. Heat maps were generated for each participant group during the first 5

seconds, the first 10 seconds and through the entire duration of the image interpretation. As there was a large variability among participants, the heat maps alone were a poor indication of whether the participant would identify the pathology successfully. However, for the more complicated images, such as image 6 (an image of the chest), the students had only provided a small number of fixations on the chest pathology within the first 5 seconds. Moreover, they had not fixated on the chest pathology at 10 seconds and they had the lowest number of fixation counts within the chest pathology area during the entire duration of their interpretation of this image. Only eight out of 21 students correctly identified the chest pathology, therefore the heat maps could be a good indicator of whether the participant group will diagnose accurately in extreme cases, where the variation in participants is subtle.

3.8 Limitations

The researcher's presence within the study may have posed a distraction and unease to the participant. Unfortunately, this was necessary to maximise the data collection from the eye tracking software and ensure the participant remained focused on their task of image interpretation. Completing the study within a test environment, rather than a clinical environment, was a limitation of this study, perhaps causing participants to err on the side of caution in their interpretation and contributing to their stress during the study.

The monitor used within the study is not of the quality which would be used within a reporting room in clinical practice; however, students and radiographers would be familiar with viewing images on such monitors within the radiology department on a daily basis. The inability to change size and window width/level during viewing was a limitation of the monitor and study approach. Reporting conditions were replicated as much as possible. Dimmed lighting and a comfortable environment was provided to participants to enhance the reporting experience.

Prevalence of normal images (12.5%) was a poor representation of the prevalence of normal images that the reporting clinicians would encounter in daily clinical practice. A consideration to the prevalence of pathologies and normal images could have allowed the study to be more realistic to the daily practice of the reporting clinician (Flehinger et al. 1978; Sonnex et al. 2001).

The eye tracking sampling quality collected from the participants varied. This was not ideal however it was thought best to include all of the participants rather than excluding them due

to the eye tracking quality received, which cannot be completely controlled. The quality of data can be influenced by participants, operators, the task, eye strength, recording environment, geometry or the eye tracking design (Holmqvist et al. 2012). Any of these could have hindered the eye tracking data collection. The eye tracking sampling quality was significantly different between the students and reporting radiographers (p=0.02) and so this was a limitation of the study.

3.9 Conclusion

Reporting radiographers were more confident in their interpretation and given diagnosis than radiographers (p<0.001) and students (p<0.001). Radiographers were more confident than students (p<0.001). Reporting radiographers were more accurate than radiographers (p<0.001) and students (p<0.001). Radiographers were more accurate than students (p=0.03). The time to first fixation decreased with experience in that the most experienced group, the reporting radiographers, fixated on the pathology first, followed by radiographers. Students took the longest time to fixate on the pathology. Reporting radiographers had a greater mean fixation duration (p=0.01), mean fixation count (p=0.04) and mean visit count (p=0.04) than students on the areas of pathology. There was also a trend noted within these eye gaze metrics across groups, in that they tended to increase as the level of expertise increased. This could suggest experts recheck areas and cross reference more in interpreting images. Reporting radiographers spent longer fixating on the entire image than radiographers (p=0.05). Radiographers were quicker at identifying the major abnormality within the images than students (p=0.04) and reporting radiographers (p=0.02).

The less experienced participant, when able to identify an abnormality, often gave little detail or description of the pathology and its consequence to the patient. Radiographers tended to supply detailed information on the technical adequacy of the images, seen also in Manning et al. (2006a). Reporting radiographers, as expected, were more thorough in their explanation, detail and description of the pathology identified. Surprisingly within the first 5-10s of viewing the images, students and reporting radiographers demonstrated similar variable patterns in their interpretation as demonstrated by the eye tracking data. However, on further inspection of the voice recordings and confidence levels it became clear that the variability could be reflected on to the search patterns employed by the reporting radiographers and lack of search patterns or strategy employed by the student cohort.

Reporting of musculoskeletal skeleton images by reporting radiographers is an established role progression within the radiographic profession and is supported by evidence of accuracy provided within previous studies (Brealey et al. 2014). This study reinforces evidence for the ability of radiographers to complete a role successfully which they have been appropriately trained to complete. Reporting radiographers had a 100% accuracy level on their diagnosis of musculoskeletal images and their training allowed them to complete the image interpretation systematically to assess all areas of the image.

This is the first study to utilise eye tracking technology to test image interpretation skills between these various groups of individuals within the radiography field on a combination of images of the musculoskeletal system, chest cavity and a variety of pathologies. The eye tracking technology supplied a valuable insight into the interpretation process and its use should be incorporated within further research of this area. The computed eye gaze metrics in this study show that eye tracking could be used to automatically assess a radiographer or to identify different levels of competencies, however further work is needed to provide additional evidence. This study is a baseline evaluation of a more involved investigation for chest image interpretation and aimed to establish breadth of interpretive differences of different anatomical examinations and cohorts. Further study is undertaken on the effect of training on the image interpretation of participants and detailed in Chapter 5.

Chapter 4 - Formation of the digital training tool

4.0 Introduction

Chest image interpretation training and education is accessed in many different forms. This is readily apparent from the systematic review completed in the earlier chapter 2. The systematic review identified that although there were several methods available for training in chest image interpretation, the techniques varied and there were limited studies and evidence available on each of these methods. The systematic review identified the weaknesses of the evidence supporting this topic namely; participants were tested on one or several different chest pathologies, the expected answer given by participants was of different standards and the variation in training methods made comparison of results and outcome measures difficult. The most popular methods of formal training identified were the postgraduate programme in chest image interpretation for reporting radiographers or the mentoring/education that trainee radiologists receive having chosen to specialise in the field of imaging (Piper et al 2014; Woznitza et al. 2014). Beyond these methods there are several devices and systems available to aid chest image interpretation training (as described in the next section). These devices may be used either during formal training as mentioned above or simply by a professional aiming to broaden their knowledge within this area and complete work for their continuing professional development (CPD).

The systematic review ultimately identified that although there were various education methods available, the effect of accessory training aids could not be identified. The impact of eye tracking feedback was positive, furthermore the evidence obtained within the review highlights the possibility of including this technology within an effective training aid (Donovan et al. 2008; Litchfield et al. 2010). The evidence from the systematic review has therefore indicated a lack of verification for the effectiveness of specific training devices/tools and so provided the rationale for the development of this training aid and its content.

There have been various systems and devices tested for their effectiveness in the education of healthcare and medical staff (Piper et al. 2014; Woznitza et al. 2014). A search strategy within image interpretation is a method employed to ensure that all aspects of the image have been checked for abnormal features (Williams et al. 2013). Search strategies used by

healthcare professionals and in particular within image interpretation are often based on a variety of guidelines and sources or otherwise 'self-taught' (Health Education England 2010; RCR 2011; Williams et al. 2013). Checklists have also proven to be a valuable resource within the healthcare settings (Hughes et al. 1996; Sonnex et al.2001; Wang et al. 2011; Kramer et al. 2016).

Eye tracking has been used to help understand the process of image interpretation and secondly to assess and provide feedback/training on the interpretation process. The feedback based on eye tracking data from the participant (expert or novice) was shown to have an effect, with a mean percentage improvement of 3.3% was presented overall and 8.4% mean percentage improvement noted within Level 1 undergraduate radiographers ($p<0.05$) ($p=0.021$) (Donovan et al. 2008). Litchfield et al. (2008) reported an improvement in the performance of both undergraduate and postgraduate radiographers when shown a preview of eye movements before their interpretation compared to when they were instructed to 'free search' or preview the image for 20 seconds prior to their image interpretation. The eye movement preview led to higher scores than the free search preview and image preview ($p<0.001$). Use of eye tracking feedback resulted in a 16% increase in observer performance compared to showing the observer the image again with no eye tracking feedback highlighted (Kundel et al. 1990). True positive rate increased and false positive rate decreased, indicating a true improvement in performance.

To date no studies have been found which investigate the effect of using a digital training package based on eye tracking technology during the training of chest reporting. With the use of eye tracking technology and expert input we aim to establish and evaluate a digital training tool.

4.1 Aims and objectives

The aim of this study was to devise a novel training tool to enhance and aid chest radiographic image interpretation. The training tool was formed to supply healthcare professionals with educational eye tracking videos and a search strategy training tool when practicing and learning about image interpretation of the chest cavity.

4.2 Structure of the training tool:

The evidence and literature above informed the choices made regarding content and design of the digital training tool. A tool was developed to include: A) a search strategy training tool to assist reporters during their interpretation of images, and B) an educational tool to communicate the search strategies to trainees using eye tracking technology.

A). Search strategy training tool
4.2.1 Formation of the search strategy training tool:

The first section of the training tool is a search strategy training tool for use in chest image interpretation. Members of the research team have collaborated to develop a robust search strategy which is suitable for use in chest image interpretation.

A consultant reporting radiographer with an education background in image interpretation, generated a paper based checklist for use in chest abnormality searching formulated from the work of Hughes et al. (1997). A clinical academic reporting radiographer within chest image interpretation, supplied the research team with a PowerPoint presentation on the search strategy which he devised and uses within clinical practice and for teaching purposes. The research team combined both approaches to chest image interpretation. The research team, which consisted of skilled reporting clinicians and academics and a reporting radiographer received the search strategy training tool and were asked to comment and provide iterative feedback. After considering all feedback, amendments were made to the search strategy training tool. With the addition of further content and scrutiny a comprehensive search strategy was developed and finalised following evaluation of the pilot packages.

4.2.2 Use of the search strategy training tool:

The search strategy training tool is to be used when practicing chest image interpretation. The checklist should be used when viewing an image so that both can be viewed simultaneously. It is envisaged that over time, when using the online checklist that the search strategy becomes second nature to the image interpreter. Therefore it is predicted that the user can avoid employing the online search strategy and instead simply follow the method of image interpretation they have adopted and adapted through using the tool when initially practicing and developing a chest image interpretation methodology suitable to the individual.

4.2.3 Layout of the search strategy training tool:

The search strategy comprises of a series of questions and prompts to guide the user to exclude pathologies, systematically search the image and form a diagnosis (finalised word document of the search strategy training tool can be seen in appendix 4.1). The search strategy begins by allowing reporting clinician's to focus on the 'general considerations' of the image presentation. By encouraging participants to firstly acknowledge the chest image projection and additional image details (i.e. technical factors such as anatomical markers, post processing labels) they have been presented with, the image interpretation process and expectations of the image presentation may be influenced by this content. Following this initial image analysis prompt, the search strategy leads the user through the image and encourages them to consider different parts of the image individually. The search strategy comprises six sections which focus on different anatomy, pathologies and artefacts which may be present within the image. The sections are:

(1) General image considerations
(2) Tubes/lines/devices
(3) Bony thorax, soft tissues
(4) Diaphragm/heart/mediastinum
(5) Lung zones
(6) Lung shadows

Each section provides specific questions to the area in focus and encourages the user to think about pathologies and abnormal presentations within the image. A combination of open and closed questions, diagrams and guidelines point out areas on the image that the observer should interpret to complete the checklists within the pro-forma. Following the completion of each section of the search strategy training tool, users are asked to complete a preliminary diagnosis. After completion of the search strategy training tool, viewers asked to form a final and complete diagnosis on the image; as part of the process users are then presented with the preliminary diagnosis' which they supplied at each section. This supplies the user with their thoughts throughout each section of the image interpretation and will help them provide a cohesive and clear image report following their interpretation.

B). Educational programme

4.2.4 Layout of the educational programme

The educational tool consists of videos comprised of expert eye gazes and scan paths recorded during chest image interpretation and collected whilst the expert used the search strategy training tool. Expert input was from qualified reporting clinicians who specialise in chest image interpretation. The expert's eye gaze behaviours were recorded as well as their verbalisation of their thought processes during their interpretation which provides a clear description of their search strategy. The training tool, once finalised, was transformed into an online digital format for participant's ease of use.

4.2.5 Expert eye tracking data collection

The Tobii Studio X60 eye tracker and the Tobii studio software© were utilised for data collection and for computing eye gaze metrics (Tobii AB 2016). The remote non-intrusive eye tracker collected the data without interference to the participant's interpretation. The eye tracker was positioned inferior to the high resolution (1440px x 900px) 24" LCD monitor that displayed the images, and angled upwards (30° cranially) to align with the participant's gaze. The eye tracking data collection was completed during the interpretation of 20 chest images (see section 4.2.8) by a reporting radiographer trained to interpret chest images which were included within the training tool supplied to the reporting radiographers. The same images were interpreted by a consultant radiologist and included within the training tool supplied to trainee radiologists. The two reporting clinicians from different backgrounds were chosen to complete the image interpretation for the educational multimedia tool. This was firstly to encourage acceptance of the tool across disciplines and secondly to allow a comparison between backgrounds to be drawn, however this is a somewhat limited comparison given that some training reporting radiographers are mentored by radiologists through their training period.

We asked the two expert reporting clinicians to speak aloud during the image interpretation session, both to verbalise the search strategy and help translate the search strategy to users watching the eye tracking videos.

The two reporting clinicians were asked to achieve a position in front of the eye tracker, similar to how they would sit before beginning a chest image reporting session in clinical practice. The height of the chair and distance from the monitor could be adjusted until the reporting clinicians were comfortable and at a desired position for the image interpretation session. The eye tracking technology was then calibrated or adjusted in position until successful calibration was achieved. The image interpretation session was completed in a room with dimmed/ambient lighting of the expert's choice.

The eye tracking data was collected in a bank of four images at a time, i.e. the reporting clinicians were asked to complete a calibration test of the equipment before interpreting four chest images whilst thinking aloud and implementing the search strategy. Following this, the clinicians were asked if they were happy to move on to the next set of four images, where calibration was again completed before their image interpretation. Interpretation of four images within a data collection group allowed calibration of the equipment every four images and therefore maximised the eye tracking sampling quality obtained. The highest achievable eye tracking sampling quality would ensure that maximum eye gazes were presented to the training tool users when translating the search strategy through the eye tracking videos.

Eye tracking sampling quality for each of the data collections was above 73% for both the reporting radiographer and consultant radiologist on all occasions. Of the 18 recordings taken during the eye tracking collection, 16 had an eye tracking sampling quality between 90%-100%, one of 87% and another of 73%. Due attention was given to the eye tracking sampling quality to ensure it was as high as reasonably achievable. The importance of eye tracking quality is referred to in Chapter 3.

4.2.6 Expert eye tracking presentation

Following data collection of the expert eye tracking, the image interpretations were divided into separate videos, each video contained the image interpretation of one chest radiographic image. This was for ease of use; the user could then access one image at a time and did not have to scroll through the video of twenty images to find where they had left off previously when using the tool. The eye tracking of the expert was superimposed onto the chest image being interpreted. No clinical history was given with the images, the videos were targeted at translating the eye tracking and search strategy, clinical history may have distracted from this

and also may have contributed to bias. These biases are listed further in Chapter 5 and the methods employed to reduce the effect of bias are included also.

Eye tracking data such as scan paths were displayed over the image content. Fixations, where the participant concentrated on a specific area of the image, were demonstrated as coloured circles on the image with the area of the circle increasing as more time was spent fixating on an area. The fixations were commonly connected with saccades or a line joining the two areas of fixations. A saccade represents a quick movement of both eyes between areas of fixations. Combination of the saccades and fixations allowed the formation of scan paths to be developed. By following these, users can observe where the expert views on the image, which areas they gave greater attention to and how they viewed the image or the search strategy implemented (McLaughlin et al. 2017).

Voice recordings of the expert were also presented with the eye tracking data. The voice recordings were collected during the eye tracking data collection and translate the expert's thoughts and methods of interpreting the image. They coincide with the eye tracking data and allow the expert to explain how he is systematically searching the image and why he is looking at specific areas of the image (Bond et al. 2014; Matsumoto et al. 2011). The voice recording concludes once the image interpretation has been completed and a full diagnosis has been given on the image.

4.2.7 Study conditions

Both reporting clinicians completed their interpretation sessions at a personally selected distance from the monitor and an optimum position for successful calibration to be achieved by the eye tracking technology. The data collection was completed in dimmed lighting in both interpretation settings; the reporting radiographer completed their interpretation in an office and the consultant radiologist completed his interpretation in a University education room used previously for reporting images. Both interpretation sessions were completed in quiet, isolated environments without disturbance.

Both reporting clinicians were given similar instructions prior to their image interpretation session, such as speaking aloud as much as possible to ensure the search strategy they used

during their interpretation was clear. All efforts were made to ensure the image interpretation environments were standardized as much as possible.

The high resolution (1440px x 900px) 24" LCD monitor was used for both data collections. The monitor displayed the same 20 images to both reporting clinicians, images were displayed in the same sequence and same format, and both reporting clinicians were asked to interpret four images before re-calibrating the eye tracking equipment (Venjakob et al. 2016). We aimed to maximise the image quality used within the training package. Images were presented on the eye tracking software in Tagged Image File Format (TIFF). Images could not be presented in Digital Imaging and Communications in Medicine (DICOM) format on a calibrated monitor used in clinical practice, due to limited access to equipment. Therefore, we chose to convert and present the images on the display monitor via a lossless compression method of TIFF.

4.2.8 Image Selection

Images were chosen from a test bank previously used in research with a confirmed reported diagnosis agreed by three radiologists (Woznitza et al. 2014). The reference standard was therefore a strong source and images were already anonymised for patient protection. Permission to use these images was achieved from the test bank source. Of the twenty chest images interpreted by the reporting clinicians; 14 images were completed with the patient positioned postero-anterior (PA), six were completed with the patient positioned antero-posterior (AP), four chest examinations were completed outside the permanent radiology rooms, two were completed within the resuscitation department, two were completed using mobile x-ray equipment, 11 images were normal in appearance with no clinically significant pathologies evident and nine images were abnormal in appearance with at least one significant pathology presented in the image. An example of the images incorporated within the tool is given in appendix 4.2. A normal image included the presence of pleural plaques, this image was regarded as normal as the appearances were not significant to patient pathway or care. Abnormal images were chosen to include a range of pathologies, for example lung nodules, atelectasis, consolidation or pneumothorax, this was to demonstrate the use of the search strategy training tool in a wide range of image appearances.

4.3 Digitising the tool

Following many changes and much discussion by the research team a final version of the search strategy training tool was structured in paper format. This was then submitted to a fellow PhD student, of the School of Computing and Mathematics, who developed the training tool into a web-based interactive tool, which could be easily accessed online. The PhD student supplied the paragraph below on digitisation of the tool:

The digital training tool was developed using web technologies enabling interpreter engagement across various tools and devices. Hypertext Mark-up Language version 5 (HTML5) was employed to structure and display webpages across numerous web browsers, whilst an engaging user experience was created through implementing Cascading Style Sheets (CSS3). The web-scripting language, JavaScript, in combination with JQuery (a Document Object Model (DOM) manipulation library), was implemented to facilitate interactive participation from interpreters. Toggling buttons and text inputs were used to collect interpreter annotations. Reactive animations were employed via JavaScript as a form of interpreter feedback when collecting interpreter annotations. The Hypertext Pre-processing language (PHP) was used to save data to a MySQL database. An example of the layout and display of the tool, post digitization, is demonstrated within appendix 4.3.

4.3.1 Changes made following first draft of digitisation:

Two diagrams (appendix 4.4) were initially drafted for inclusion within the digitised training tool. However once visualised within the web based tool, the diagrams were confusing and difficult to interpret as each contained 4 overlapping devices. The diagrams were of little use in explaining the correct position of a device and so it was decided that separate diagrams (i.e. one for each tube, line or device), would present this information in a clearer format. This alteration allows the user to firstly decide if the line/tube/device is present and if so they are presented with a unique diagram demonstrating the preferred and correct position for patient safety and care (see appendix 4.5).

The user is asked to provide a preliminary diagnosis following each completed section of the search strategy training tool. This prompts the user to summarise their thoughts during their image interpretation and provide information on a particular aspect of the image, for example

bony structures, mediastinum or hilar regions. The inclusion of this preliminary diagnosis allows the user to comment on specific pathologies/abnormalities as they are detected and therefore minimises the possibility of the abnormality being excluded within the final diagnosis. Originally this was included within the search strategy training tool, however on further consideration, it was decided this feature would be most beneficial if presented to the user again once forming their final diagnosis. Therefore, preliminary diagnoses made during image interpretation are presented as a list before and during the formation of their completed final diagnosis. This was also amended within the web-based interactive tool to ensure its visibility prior to the final diagnosis being typed into the conclusive diagnosis box in the tool.

4.4 Accessibility of the training tool

The digital package is easily accessed on any device with an internet connection and therefore can be readily transferred from one department to another. Instructions on use of the training tool are inserted where required. The layout and display is simple and easy to follow, diagrams are supplied with labels to guide the user and provide them with examples of pathologies.

4.5 Monitoring use of training tool

A login feature monitors the use of the eye gaze enhanced videos and the search strategy training tool. Participant's performance is expected to enhance greatly over their formal chest image interpretation training, and so investigation on the use of the tool, participant performance and eye tracking data when analysed together should supply a clear insight on the impact of the training tool. Users must enter a user identification code before being granted access to the tool. User ID's are unique to the individual, which will allow investigation of both the use of the tool and users performance. Users are asked to select 'submit' once they have finished their session using the tool, allowing the collection of timeframes of use. Once submitted, time of use, the answers selected, the preliminary diagnosis sections and the full diagnosis section are automatically stored within an excel spreadsheet.

4.6 Discussion

A digital training tool for use in chest image interpretation was created based on evidence within the literature, expert input and two search strategies previously used in clinical practice. Images and diagrams, aiding translation of the tool content, were incorporated where possible. The images and diagrams feature colour and labels to ensure information is supplied clearly (Blake et al. 2014). The tool is structured to allow the chest image interpretation process to be clear, concise and methodical. A search strategy was incorporated within the tool to ensure users would devote attention to all aspects of a chest image. Search strategies and checklists have been shown to be useful in the healthcare setting prior to this (Hughes et al. 1996; Herzer et al. 2009; Wang et al. 2011). Given the lack of evidence based practice on use of a particular search strategy within image interpretation, it was important to investigate the use of a search strategy, with the possibility that it could be recommended as an evidence based approach for use by reporting clinicians.

The digital tool accommodates and can be used by all learner types. The checklist featuring a list of questions and images optimises the use of the tool for visual and reading/writing learners. The educational videos, comprising the voice over of experts' thoughts, are ideal for visual and auditory learners. Whereas the overall use of the tool to be used when practicing image interpretation can appeal to those which benefit most from kinesthetic learning (Illeris 2009; Learning Rx 2017).

The digitisation of the training tool was completed for ease of use. As the tool is presented and used in digital format, it can be reused, used easily within learning/clinical environments and avoids the use of CD/downloads etc. to watch the eye tracking videos.

A checklist, to allow ease of use of the training tool, was included within the search strategy training tool. Although checklist fatigue was suggested as a possible problem by Kramer et al. (2016), the implementation of this checklist within practicing and learning about chest image interpretation allows the load this checklist imposes on possible 'checklist fatigue' to be minimised. The participants will be likely cease to use this online checklist once they have completed their initial image interpretation training, it is simply to supply them with an initial checklist and search strategy when developing their image interpretation skills.

The tool is aimed for use in practicing chest image interpretation. Its use is targeted to support the formal training received by reporting radiographers and radiologists. It is a device which can provide a search strategy to the user and also offers them an opportunity on how the search strategy can be used in practice, by viewing the eye tracking videos, and provides an opportunity to practice implementing the search strategy, by applying the ideals expressed in the search strategy training tool. However, this tool could also be provided to undergraduate students and other healthcare professionals who wish to adapt a search strategy in their chest image interpretation, become familiar with the appearance and position of common chest pathologies, or who wish to improve their knowledge of the correct positioning of tubes, lines and devices on a chest image. Healthcare professionals may be asked to interpret images, the use of this tool could help them build skills to recognise pathology/abnormaltiy and bring it to the attention of others. Roles such as these may lead to quicker patient treatment and diagnosis where implemented.

Within McEvoy et al. (2017) only a weak positive correlation was found between peer review scores received by students and scores obtained by the students in a Multiple Choice Question (MCQ) examination. This provides evidence to support that student/peer feedback may not be a reliable or optimal source of information to be implemented within image interpretation learning programmes. However, Litchfield et al. (2010) found little difference in participant performance when providing radiographers (undergraduate and postgraduate) with eye tracking feedback from a novice compared to eye tracking feedback generated by an expert's image interpretation. What must be considered is that the eye tracking feedback provided by Litchfield et al. (2010) was unaccompanied by instructions or guidance on the image interpretation and could have possibly hindered the transfer of learning from the feedback, leading to little difference seen when comparing the source of the eye tracking feedback and the effect of each. Given this evidence, it was decided to implement expert eye tracking and voice recordings to explain the search strategy which is being used and to help explain how it is used during the interpretation process.

Online tools were developed which provided support to the user during their image interpretation learning and a detailed summary of the user's performance (Wright 2014; Subesinghe et al. 2015; Wright et al. 2017). Similar to these studies, the training tool has the potential to be used for user's self-assessment and image interpretation learning support. The

'preliminary diagnosis' sections and the 'final diagnosis' section are recorded and stored within an excel database. This feature of the tool can be used retrospectively to pinpoint common user errors in their image interpretations or to record their performance as an assessment method. The presentation of the 'preliminary diagnosis' sections to the user when making their decision, prompts them to consider a differential diagnosis for each image. This element is similar to the image report generated by the decision support tool of Wang et al. (2011). This is an important element as it allows the user to provide a full and thorough diagnosis, reminding them of all abnormal elements they found on the entire image.

4.7 Conclusion

A training tool for use in chest image interpretation learning has been designed, created and digitised. The original documents used to develop the training tool (the PowerPoint search strategy and the written checklist) were used in practice by reporting clinicians and complemented each other well to form a search strategy training tool. The PowerPoint presentation supplied a visual method of explaining the search strategy and the written checklist provided a structured set of questions to guide the user through the image.

The eye tracking videos help translate the search strategy and allow the use of a new and innovative technology to be employed in an image interpretation learning tool. Many users will not be familiar with viewing eye gazes when learning about image interpretation, in which case this could pose an obstacle for their learning or alternatively represents an exciting new method of learning. Eye tracking videos were aided by the voice recordings of the expert, which has the potential to reduce any issues users have when viewing and comprehending the eye tracking videos.

Once an internet connection is opened the display of the tool is structured, with instructions of its use provided where necessary. A user login feature allows monitoring of the use of the tool and enables investigation into whether there is a relationship between its use and users image interpretation. This tool was then investigated for any potential benefits of its use during training in chest image interpretation. The results of these studies are presented in subsequent chapters of this thesis.

Chapter 5 - Eye tracking study

5.0 Introduction

Preliminary work in this thesis has identified the lack of resources available to aid the learning of chest image interpretation by reporting clinicians (see Chapter 2). Use of eye tracking technology and search strategies were two of the least tested methods of interpretation learning however, when they were tested results showed a positive effect on participant performance (Donovan et al. 2008; Litchfield et al. 2010). Hence, it was decided to form a digital training tool to implement and measure these two components. Previous research has investigated the use of eye tracking technology as a feedback tool (Donovan et al. 2008; Litchfield et al. 2010) and also as an outcome measure (Donovan et al. 2008); both studies were completed successfully. However, these studies focused solely on pulmonary nodule identification and were completed as short term studies, whereby participants were tested immediately after they received the eye tracking feedback. These applied methodologies are weakened by these limiting factors.

Recent research has investigated image interpretation techniques in terms of both "formal" and "informal" methods, as discussed in the systematic review (chapter 2). Studies have been carried out to test the effect of postgraduate programmes in chest image interpretation (Piper et al. 2014; Wozntiza et al. 2014). Wide variation in testing methods of chest image interpretation education was identified within the systematic review. Studies failed to test participants before training and so although the accuracy level was relatively low, there was no initial accuracy level to compare this with (Cowan 2007; Semekula-Katende 2016). Some methods were less formal and effective in identifying the true significance of the education method. Only one study tested the effect of the training pre and post implementation (Hughes et al. 1996), however even then the study involved participants allocating images to one of the four pathology options given. Deciding whether a chest image includes one of four types of given pathologies, does not mimic the task which chest image interpretation poses in clinical practice.

This chapter describes the study carried out to test the effect of the digital training tool on participant performance. A variety of outcome measures were used, participants were tested

pre and post implementation of the tool, a range of chest abnormalities were included and the study was executed over a nine month period. These factors were employed to attempt to correct the limitations of previous studies, which were identified within the systematic review, in testing the effect of an education method in this field.

5.1 Aims and objectives

(i) The purpose of this study was to determine the impact and effectiveness of the digital training tool (described in Chapter 4) on participant performance.

The objectives were to determine;

- performance levels pre and post implementation of the tool and between intervention and control groups;
- confidence levels for each group of participants pre and post study;
- any variation in eye tracking metrics between the groups and pre and post intervention.

5.2 Methodology

5.2.1 Study approval

Ethical approval was obtained from the Research and Ethics Filter Committee at Ulster University. Confirmation of the approval was received on 22/02/2016 (see Appendix 5.1).

An amendment was requested on 16/08/2016 to recruit a further cohort of reporting radiographers trained in interpreting musculoskeletal images. This was approved on 23/08/2016. Another amendment was requested on 22/02/2017 for the inclusion of a survey within the study. This was governed by chairs action and approved on 15/03/2017. The filter committee ethics application and approval can be seen in Appendices 5.1 and 5.2.

5.2.2 Study design

A quasi-experimental study was carried out over a nine month period with the reporting radiographers who began training in chest image interpretation. A randomised trial was carried out over a nine month period with the reporting radiographers who were trained in musculoskeletal (MSK) image interpretation. Participants completed an initial assessment at recruitment and were asked to re-attend nine months later for a follow-up assessment. The

intervention group were given unlimited access to the training tool during the nine month period. The control group had no access to the training tool during this time.

During the assessments, each participant's diagnosis was recorded. A confidence level between one and ten was provided on each given diagnosis. In addition, a questionnaire on radiographers' clinical experience, participant thoughts on the eye tracking technology and their experience of image interpretation was completed. Participants completed the assessment of 20 images using the eye tracking technology enabling eye gaze metrics to be collected during the image interpretation session. Verbalised thoughts of the participant were also collected by voice recordings taken during the image viewing session.

5.2.3 Informed consent

Written informed consent was obtained from participants who were willing to complete the study. Following this, participants were screened against the inclusion/exclusion criteria before participation in the study (Appendix 5.3).

5.2.4 Inclusion and exclusion criteria

Data was collected with the aid of the following participants:

Inclusion criteria:

 (i)   Trainee reporting radiographers undertaking postgraduate education in chest image interpretation.

 (ii)   Reporting radiographers trained to report on the musculoskeletal system but with no experience in chest image interpretation.

 (iii)   Those willing to dedicate their time to the study and those who supplied written informed consent.

Exclusion criteria:

 (i)   Those with complete loss of vision in one eye, those with astigmatism.

 (ii)   Those who withdrew consent or participation in the study.

 (iii)   Participants taking any drugs that may affect vision.

5.2.5 Participant groups

Four groups of participants were required for this study.

Quasi-experimental trial:

Trainee reporting radiographers were recruited through the chest image interpretation postgraduate programme at Canterbury Christchurch University, England. A convenience sample was used for radiographers who were registered on the postgraduate programme to train in reporting chest images (referred to as CXR reporting radiographers from here on). Participants who enrolled on the postgraduate programme in March 2016 were the control group (group 1 – no access to the training tool) and participants who enrolled on the postgraduate programme in October 2016 were the intervention group (group 2 – access to the training tool).

Randomised trial:

Reporting radiographers trained to report on the musculoskeletal system but not currently trained to report on chest images were recruited through their attendance at Continuing Professional Development (CPD) events within University premises.

The group of radiographers trained to report on images of the musculoskeletal system were randomly allocated to a control group (group 3 – no access to the training tool) or intervention group (group 4 - access to the training tool) (Figure 5.1 and Figure 5.2).

Figure 5-1: Groups of reporting radiographers beginning chest image interpretation training

Reporting radiographers beginning training in CXR image interpretation
(n=12)

Control group
(n=7)

Intervention group
(n=5)
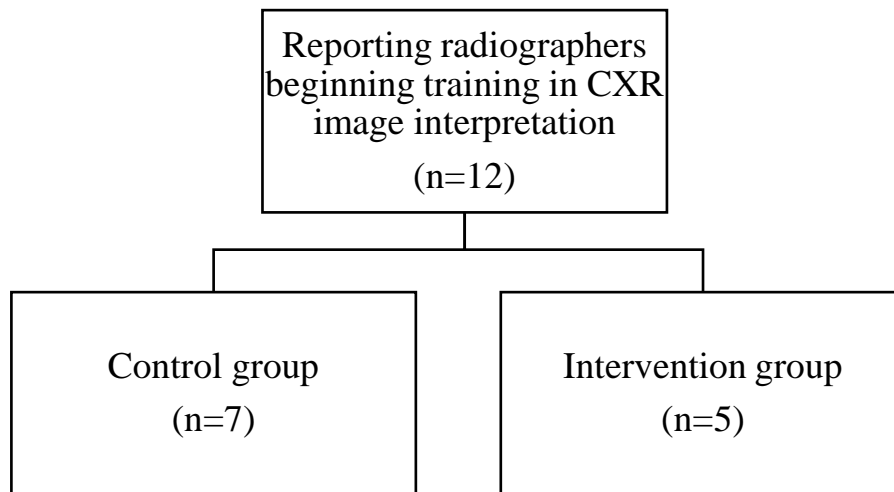
Figure 5-2: Groups of reporting radiographers trained in musculoskeletal image interpretation

Reporting radiographers trained in MSK image interpretation
(n=23)

Control group
(n=13)

Intervention group
(n=10)
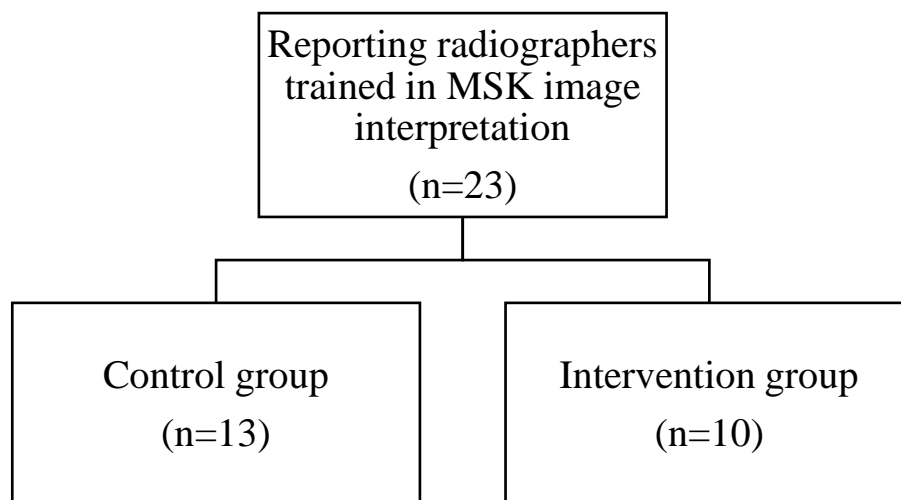
5.2.6 Data collection

Eye tracking data collections were carried out by the PhD student pre and post implementation of the tool to the intervention groups. Once signed consent was obtained, participants were seated in front of the Tobii XPro eye tracker. Participants were shown twenty chest images in each eye tracking data collection session. A data collection sheet was utilised by the researcher to record the diagnosis and confidence level provided by the participant during the study.

The participants were asked to state whether the image was normal or abnormal in appearance and to describe and state the pathology if identified. They were then asked for a confidence level between one and ten in each diagnosis they provided (one being least confident and ten being most confident in the given diagnosis). Following this, participants completed a short questionnaire on radiographer experience, image interpretation experience and their thoughts on the eye tracking technology.

Study protocol

Initial assessment: Participants were asked to interpret 20 chest plain radiographic images whilst thinking aloud and using eye tracking technology. Eye tracking data, a completed questionnaire and a diagnosis of the images were collected along with confidence levels on their given diagnosis. The initial assessments were carried out on the trainee reporting radiographers following enrolment on the postgraduate programme. The initial assessments of the reporting radiographers who were trained to report on the musculoskeletal system was carried out at CPD events within University premises following recruitment into the study.

Intervention and control group: Participants within the intervention group had access to the training tool for 9 months following their initial assessment. Control group participants had no access to the training tool during this 9 month period, however they were given access to the training tool following completion of their participation in the study. Student reporting radiographers from both groups were attending the University at different block training sessions which ensured limited interaction or possibility for the training tool to be shared with the control group. Both intervention groups were asked to refrain from sharing the training tool with the control group. Login credentials were required to use the training tool, which also reduced the likelihood of participant groups sharing the training tool.

Follow up assessment: Image interpretation sessions using the eye tracking equipment were carried out at approximately 9 months following the initial assessment (Figure 5.3). Delays of two to three weeks occurred when data collections could not be arranged for specific dates and depending on the dates the participants were attending the postgraduate programme within the University. Participants were asked to interpret 20 different chest images to those seen in the initial batch (pre-intervention) so that memory effects were avoided. This was completed whilst thinking aloud and using eye tracking technology (Ryan et al. 2011). The outcome measures were repeated following the use of the chest image training tool by the intervention group. All participants were requested to provide feedback on features of the training tool and its utility/ease of application by completing an online survey once they had access to it. Feedback on the training tool and its usefulness were welcomed to gain a consensus on its effectiveness.

Figure 5-3: Study plan

Initial eye tracking data collection

Intervention group given access to the training tool (9 months)

Follow up eye tracking data collection
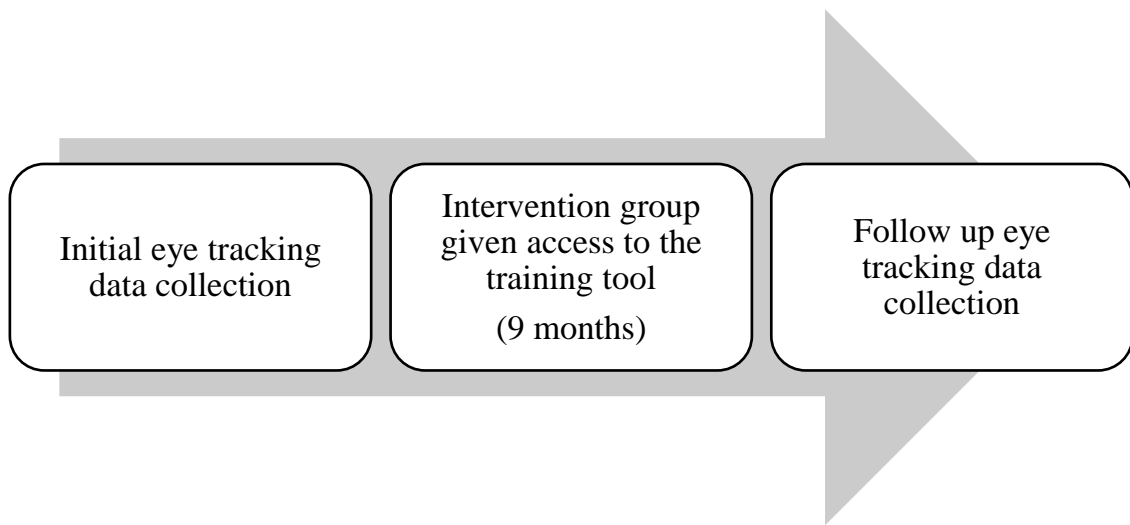
5.2.7 Images

Images were selected from a previously compiled test bank used in previous studies (Woznitza et al. 2014). 100 chest images were utilised for this study. A total of 20 images were used to form the educational programme. An additional set of 20 images were interpreted by each participant during each session of image interpretation as part of the outcome measures for the study (pre and post training). Images shown to a participant changed pre and post training to avoid the possibility of memory influencing the results (Ryan et al. 2011). The image banks selected contained both normal and abnormal images (Nocum et al. 2013). There was a prevalence of abnormal images of 45% in the initial assessment, 45% in the training tool and 50% in the follow up assessment (Table 5.1). The participants were unaware of this diagnosis to avoid any bias in their results, (Brealey et al. 2002c; Brealey et al. 2014). Given the lack of studies which tested participants on a variety of chest pathologies (as identified by the systematic review in Chapter 2) we aimed to implement a range of chest pathologies. Normal images were incorporated to account for the variety of images encountered in clinical practice by reporting clinicians. The tool encouraged users to recognise and diagnose a range of abnormalities. The content of the tool also influenced the selection of chest images, in that the pathologies taught in the tool were included within test images given to participants. The prevalence rate of normal images (50-55%) was similar to that of studies featured within the systematic review (Donovan et al. 2008; Hughes et al. 1996; Litchfield et al. 2010; Manning et al. 2003; Manning et al. 2004; Manning et al. 2006a; Manning et al. 2006b; Woznitza et al. 2014). The images chosen for the initial and follow up assessment were counter balanced to maintain the level of difficulty pre and post the study. Similar pathologies were included in the initial and follow up assessment to ensure both interpretation tasks were a similar level of difficulty but without the possibility of memory influencing the interpretations.

Table 5-1: Pathologies present within abnormal images included within the study

Abnormal image content pre intervention	Abnormal image content post intervention
Consolidation, effusion	Multiple rib fractures
Cancer, lung collapse, deviated trachea	Nodule, effusion
Atelectasis, opacity	Consolidation, naso-gastric tube
Cancer	Metastatic spread
Lung nodule	Nodule, haematoma, pnuemothorax
Consolidation	Cancer
Atelectasis, lines	Cancer
Consolidation, scoliosis	Bilateral effusion
Pnuemothorax, chest drain, possible emphysema	Consolidation
	Tuberculosis

5.2.8 Reference standard

Each of the images within the test bank were interpreted by three consultant radiologists and a consultant reporting radiographer. All professionals agreed on a diagnosis that would be the gold standard in this study. Individuals from both professions were approached to ensure both would be satisfied that the knowledge included in the tool was relevant to their professional background. Four individuals interpreted the images to ensure there was consensual agreement about image content.

5.2.9 Equipment

The Tobii Studio X60 eye tracker and the Tobii studio software©, used within Phase 1 (detailed in Chapter 3 Section 3.2.5.2) were used within this study. The remote non-intrusive eye tracker collected data without interference to the participant's interpretation. All participants were reminded to verbalise their thought processes as much as they could. The eye tracker was positioned inferior to the monitor that displayed the images and angled upwards to align with the participant's gaze. The angle at which the infrared light reflects off of the cornea provides information on where the eyes fixate on the image.

5.2.10 Test Environment

A test environment similar to that used in the first study was utilised. Information on this is detailed in Chapter 3 Section 3.2.5.1. (See Figure 5.4 and 5.5).

Figure 5-4: Diagram of the test environment
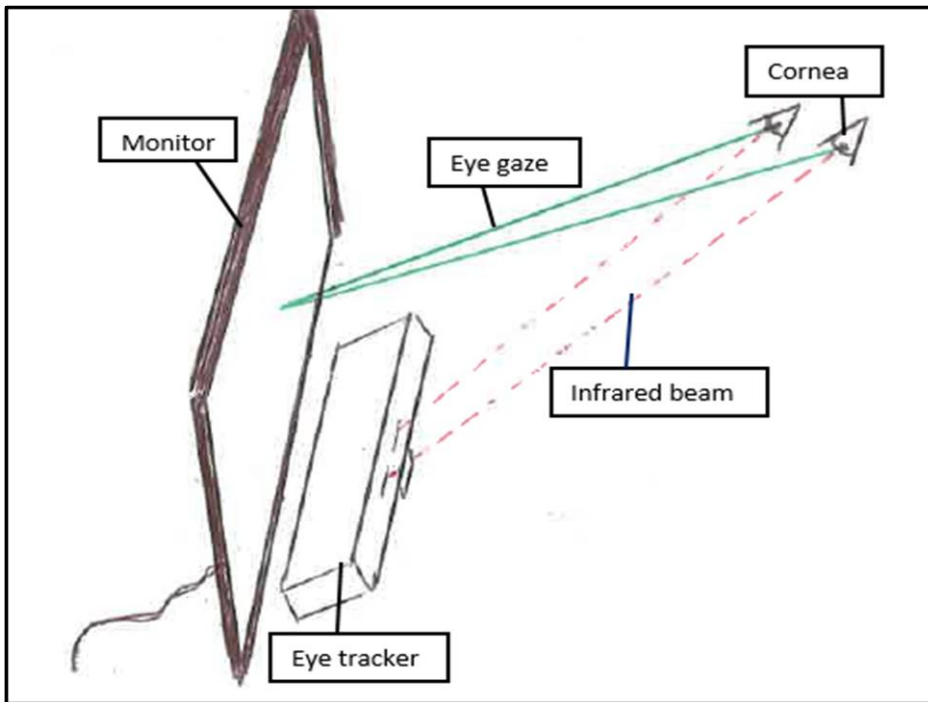
Figure 5-5: Photograph of the test environment
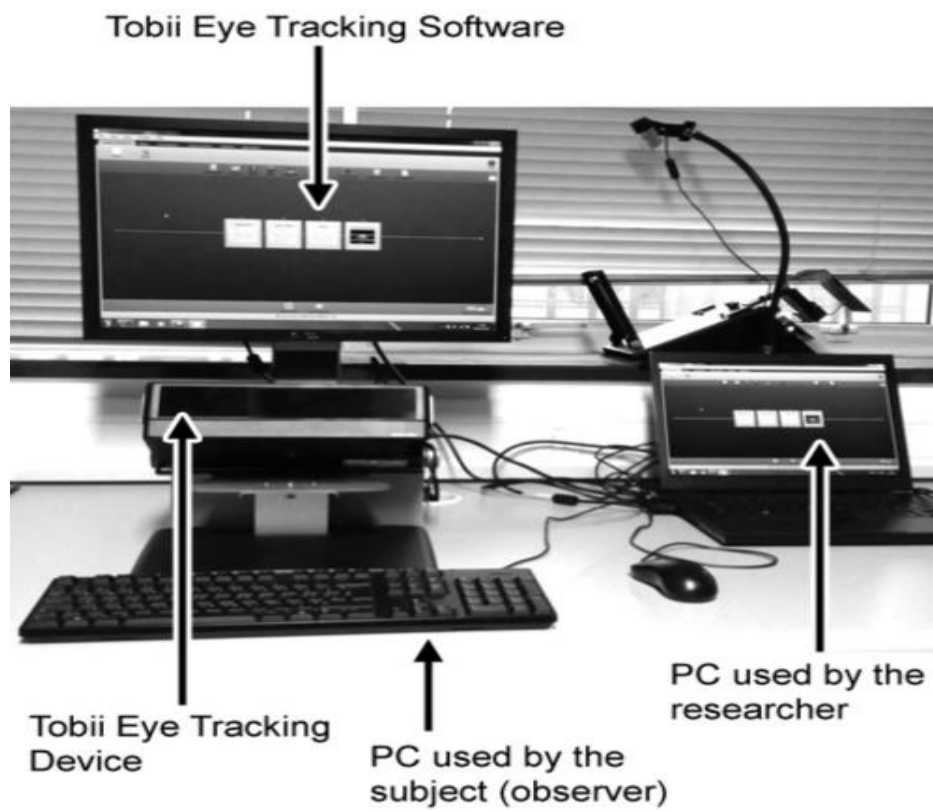
5.2.11 Bias

Bias was taken in to consideration in the planning of the study and its effect was limited where possible (Table 5.2).

Table 5-2: List of biases and how the study navigates to overcome each

Type of bias	Method employed to eradicate/lesson impact of the bias
Cognitive biases:	The initial test bank of images was entirely different to the follow up test bank of images to avoid memory influencing the results (Ryan et al. 2011).
Satisfaction of search	The tool aims to promote satisfaction of search by encouraging the user to search all areas of the image, this element aims to lesson the effect of this bias.
Availability bias	As participants were professionals working in clinical practice the effect of mental shortcut and immediate examples couldn't be minimized easily. Participants had no access to additional information such as patient history, examination referral or previous images. This was decided to minimise the effect of outside influences on image interpretation and to standardise the eye tracking study as much as possible.
Knowledge of clinical history	No clinical history was provided to participants during the study to avoid influencing their decision process.
Premature closure	The systematic method of image interpretation implemented in the tool, which aims to promote satisfaction of search, aims to reduce the possibility of participants latching on to a diagnosis early in the interpretation.
Anchoring bias	Similar to premature bias, the systematic method of image interpretation promoted in the training tool aims to reduce anchoring bias also. The tool aims at encouraging participants to search the entire image rather than deciding on a diagnosis in the early moments of interpreting an image and dismissing the possibility of an alternative diagnosis.
Confirmation bias	The training tool mentions specific pathologies and image appearances throughout to avoid participants trying to confirm what pathology they originally suspect is present on the image at the beginning of their image interpretation.

5.3 Outcome measures

Eye tracking data: The screen of the monitor was video recorded to view eye fixations and saccades of the participant overlaid on each chest X-ray image pre and post use of the training tool (previously defined in Chapter 3). Total fixation duration, fixation counts, time to first fixation, visit count and total visit duration were drawn from each participant's interpretation and used within data analysis.

Diagnosis and confidence levels: The participants were asked to think aloud to gain an insight into their cognitive processes and to establish when they first acknowledge the pathology during their image interpretation (Gegenfurtner et al. 2013). The participants were asked to provide a diagnosis on the images and self-rated confidence levels between 1 and 10 were requested on their given answers. The participants' comments were voice recorded to enable further analysis. By listening to the voice recordings a deeper understanding of the participant's method of interpretation and rationale for their diagnosis was gained. Greater detail of the participant's response was recorded for analysis following the data collection period and more information on the participant's thoughts were achieved. Without the voice recordings such detail would not have been manually recorded by the researcher within the short time frame of the data collection. The participant's given diagnosis and confidence level was written by the investigator on a data collection sheet.

Questionnaire: A questionnaire containing open and closed questions was completed. This helped gain an insight into the individual's experience/ field of practice, their opinion on the expected level of accuracy they should be achieving, the level of accuracy they believe they had achieved and the participant's perception of the use of the eye tracking for the improvement of image interpretation (see Appendix 5.4).

Survey: A survey seeking user feedback on the online training tool was given to all participants following their participation within the study (see section 5.8)(see Appendices 5.5 and 5.6). The participants were asked to include detailed feedback on the training tool and the study they had participated in. Participant inputs were coded allowing the survey response to be related to the data gathered during the eye tracking study. The Qualtrics software was used to present the survey. Participants were asked to complete the survey online via email.

5.4 Data analysis

This investigation is a quasi-experimental study to establish the usefulness of a training tool in image interpretation of plain chest radiographs with normal and a range of abnormalities evident in the images. Eye tracking data, accuracy of diagnosis, confidence levels, and participant's feedback were analysed and discussed below. Descriptive statistics included the mean, median and standard deviation. Box plots were generated with SPSS software version 24. Analysis of Variance (ANOVA) and Analysis of Covariance (ANOCOVA) as parametric testing for statistical significance were used following the completion of normality testing.

Additional quality scores

Quality scores were awarded to participants for mentioning additional information regarding the quality of each image. Additional information such as positioning errors (for example a lordotic appearance), areas of abnormality, artefacts, lines/tubes/devices, normal variants or image features (for example raised diaphragm, enlarged heart etc.) which may look abnormal but were not the pathological abnormality associated for the chest image. A series of possible quality scores were determined following discussion by the research team and participants were then scored using these. Participants were awarded a score of 1 for each criterion if they mentioned the issue/aspect of the image verbally during their interpretation process. Radiographers are trained in their undergraduate degree to perform diagnostic imaging examinations with emphasis on radiographic technique, artefact presence and patient positioning. The nature of this education may sway opinion of radiographers about image content, more so perhaps than a radiologist would. Therefore the quality scores were formed to take in to account additional correct information provided by participants on the image quality/content however, which were not directly applicable to the TP, TN. FP and FN scoring criteria.

5.5.1 Eye tracking sampling quality

Eye tracking sampling quality, as explained in Chapter 3, Section 3.6.3, was extracted and analysed for the participants in this study. These figures were extracted for all groups of participants and compared. Quality decreased overall in the post data collection, however this was not significant. Attempts were made in data collection to maximise eye tracking data quality. Please see Tables 5.3, 5.4, 5.5, 5.6.

114

Table 5-3: Eye tracking sampling quality for all participants pre and post study

	All pre (n=35)	All post (n=35)
Eye tracking sampling quality	70.91 ± 28.82	61.09 ± 31.09

Table 5-4: Eye tracking sampling quality for control and intervention participants pre and post study

	Control pre (n=20)	Control post (n=20)	Intervention pre (n=15)	Intervention post (n=15)
Eye tracking sampling quality	73.10 ± 28.72	61.10 ± 31.52	68.00 ± 29.68	61.07 ± 31.62

Table 5-5: Eye tracking sampling quality for control and intervention CXR RR participants pre and post study

	Control pre (n=7)	Control post (n=7)	Intervention (n=5)	Intervention post (n=5)
Eye tracking sampling quality	83.57 ± 13.56	56.86 ± 29.81*	72.40 ± 29.41	44.60 ± 33.51*

Table 5-6: Eye tracking sampling quality for control and intervention MSK RR participants pre and post study

	Control pre (n=13)	Control post (n=13)	Intervention (n=10)	Intervention post (n=10)
Eye tracking sampling quality	67.46 ± 33.40	63.38 ± 33.35	65.80 ± 31.14	69.30 ± 28.81

*Data is presented in mean ± standard deviation. # significant difference between groups at baseline * significant difference within group ~significant difference between groups post intervention (P<0.05)*

5.5 Intervention and Control Results

5.5.2 Demographics

The control group was qualified for longer and had a greater experience interpreting images on average (Table 5.7). However, both the control and intervention group had similar experience reporting images of approximately 6-7 years. There was a high majority of females in the intervention group (86.7%) (Table 5.7).

Table 5-7: Descriptive statistics of reporting radiographers in the control and intervention group

	Control (n=20)	Intervention (n=15)
Qualified (years)	22.18 ± 10.28	14.97 ± 7.22
Experience interpreting images (years)	21.68 ± 9.65	13.90 ± 6.48
Experience reporting images (years)	6.70 ± 5.77	6.03 ± 5.41
Sex	Female 55% Male 45%	Female 86.7% Male 13.3%

Mean ± standard deviation

5.5.3 Scores

FP scores decreased for both the control and intervention group (p<0.05), however there were fewer FP decisions in the intervention group compared to the control group. TN scores increased for the intervention group also (p<0.05). TP scores increased for both groups. Quality scores decreased for the control and the intervention group at the follow up testing (p<0.05) (Table 5.8). Confidence increased for both groups however, this was only significant in the intervention group (Table 5.8). No significant differences were observed between groups post intervention.

5.5.4 Eye tracking data

Fixation count and visit duration on the entire image decreased for the control group following the intervention period. Interpretation time for this group decreased significantly also following this period (Table 5.9). No significant changes were observed for the intervention group for these parameters. There was no significant change of the eye tracking metrics within the areas of pathology (AOP) (Table 5.10).

5.5.5 Fixation frequency

Intervention group participants had a greater fixation frequency than the control group (*P<0.05*). This increased by a mean of 0.68 Hertz (Hz) in the follow up testing period (Table 5.11).

Table 5-8: Scores and confidence for intervention and control groups

	Control pre (n=20)	Control post (n=20)	Intervention pre (n=15)	Intervention post (n=15)
FP	4.20 ± 1.70# (38.18%)	3.20 ± 1.96 (32.0%)	5.87 ± 1.96# (53.36%)	3.27 ± 1.62* (32.7%)
FN	5.95 ± 1.73 (66.11%)	5.60 ± 1.47 (56.0%)	6.53 ± 1.25 (72.56%)	6.00 ± 1.25 (60.0%)
TP	3.05 ± 1.73 (33.89%)	4.40 ± 1.47* (44.0%)	2.47 ± 1.25 (27.44%)	4.40 ± 1.47* (44.0%)
TN	6.80 ± 1.70# (61.82%)	4.40 ± 1.47 (44.0%)	5.13 ± 1.96# (46.64%)	6.73 ± 1.62* (67.30%)
Quality scores	13.20 ± 2.80 (44.0%)	10.30 ± 2.89* (46.82%)	13.20 ± 2.96 (44.0%)	9.67 ± 2.29* (43.95%)
Confidence	6.17 ± 1.23 (61.7%)	6.43 ± 1.20 (64.3%)	5.32 ± 1.70 (53.2%)	6.31 ± 1.07* (63.1%)

*Data is presented as mean ± standard deviation #significant difference between groups at baseline *significant difference within group ~significant difference between groups post intervention (P<0.05)*

Table 5-9: Eye tracking data for the entire image interpretation of control and intervention group participants pre and post implementation of the training tool

	Control pre (n=20)	Control post (n=20)	Intervention pre (n=15)	Intervention post (n=15)
Mean fixation duration (secs)	44.50 ± 22.02	41.53 ± 28.13	40.87 ± 28.33	49.08 ± 30.00
Mean time to first fixate (secs)	2.31 ± 5.71	2.74 ± 5.86	4.43 ± 12.45	4.08 ± 8.99
Mean fixation count (n)	156.91 ± 72.52	104.31 ± 59.64*	152.22 ± 95.57	130.32 ± 86.23
Mean visit duration (secs)	60.90 ± 21.70	44.37 ± 23.56*	61.44 ± 25.47	59.68 ± 25.27
Mean visit count (secs)	2.73 ± 2.16	2.98 ± 2.80	4.02 ± 5.05	2.91 ± 2.73
Mean decision time (secs)	66.19 ± 17.00	41.75 ± 27.65*	68.55 ± 22.33	54.51 ± 40.02

*Data is presented as mean ± standard deviation. All values are means ± SD. *significant difference within group (P<0.05)*

Table 5-10: Eye tracking data for the control and intervention group which was collected from the Areas of Pathology (AOP)

	Control pre (n=20)	Control post (n=20)	Intervention pre (n=15)	Intervention post (n=15)
Mean fixation duration (secs)	8.88 ± 5.95	7.93 ± 4.70	7.02 ± 4.65	7.95 ± 5.62
Mean time to first fixate (secs)	7.86 ± 5.03	8.94 ± 8.07	13.34 ± 15.50	11.50 ± 15.91
Mean fixation count (n)	27.82 ± 15.97	27.40 ± 14.49	23.30 ± 13.36	31.76 ± 20.29
Mean visit duration (secs)	10.10 ± 6.21	9.81 ± 5.00	8.49 ± 4.41	10.72 ± 6.13
Mean visit count (n)	11.78 ± 5.91	9.09 ± 4.85	11.04 ± 6.36	11.44 ± 7.14

Data is presented in mean ± standard deviation. All values are means ± SD. # significant difference between groups at baseline ~significant difference between groups post intervention (P<0.05)

Table 5-11: Fixation frequency for reporting radiographers in the control and intervention group

	Control pre (n=20)	Control post (n=20)	Intervention pre (n=15)	Intervention post (n=15)
Fixation frequency (Hertz)	3.82 ± 0.84	3.92 ± 1.54#	4.34 ± 1.37	5.02 ± 1.44

*Data is presented as mean ± standard deviation. # significant difference between groups at baseline *significant difference within group ~significant difference between groups post intervention (P<0.05)*

Data was then analysed further to delve deeper into the participant groups. Data collected from reporting radiographers beginning their training in chest image interpretation training (CXR reporting radiographers) and reporting radiographers trained in MSK image interpretation (MSK reporting radiographers) was analysed to investigate whether the changes observed above were evident in specific participant groups.

5.6 CXR Results

5.6.1 Demographics

Control group participants were qualified for longer, had a greater experience interpreting images and a greater experience reporting images on average. A high majority of males were in the control group (71.4%) whereas a high majority of females were in the intervention group (80%) (Table 5.12).

5.6.2 Scores

The intervention group had fewer FP values following the intervention period ($p<0.05$). The number of FP values decreased from approximately 5.20 (47.3%) to 2.8 (28.0%) following the intervention (Table 5.13). However, this was not significantly different to the control group. The intervention group were better at identifying a normal image after the study. The intervention group had a greater number of TPs following the intervention period ($p<0.05$). TPs increased approximately from 2.60 (28.9%) to 3.8 (38.0%) following the intervention period (Table 5.13). They were better at identifying an abnormal image and pathology (ies) at the end of the study also. However, after adjusting for pre intervention scores using the Analysis of Covariance (ANCOVA), there was a significant difference between the two groups on post-intervention scores of TP and FN ($p=0.049$; $p=0.049$) (Figure 5.6 and 5.7). The control group had fewer FN's 4.71 (47.1%) compared to the intervention group 6.20 (62.0%) post intervention period ($p=0.049$) (Table 5.13). There was a significant difference found between these two groups for the initial scoring period for both FN and TP scores also. Both groups increased their quality score at the follow up test period (Table 5.13).

The control group were more confident than the intervention group at the initial testing period ($p<0.05$) and at the follow up testing period ($p<0.05$). Both groups were more confident in the diagnoses they provided at the end of the study, however this increase in confidence was more apparent in the intervention group, with an increase of 16.3%, compared to the 3.4% rise seen within the control group (Table 5.13).

Table 5-12: Descriptive statistics for reporting radiographers training in chest image interpretation

	CXR reporting radiographers control pre (n=7)	CXR reporting radiographers intervention pre (n=5)
Mean qualified (years)	22.5 ± 8.7	16.6 ± 4.5
Mean experience interpreting images (years)	22.4 ± 9.0	14.6 ± 5.9
Mean experience reporting images (years)	10.9 ± 4.8	7.5 ± 5.7
Sex	Female 28.6% Male 71.4%	Female 80% Male 20%

Mean ± standard deviation

Table 5-13: False Positives (FP) for reporting radiographers' training in chest image interpretation pre and post study

	Control pre (n=7)	Control post (n=7)	Intervention pre (n=5)	Intervention post (n=5)
FP	3.29 ± 1.11 (29.91%)	3.57 ± 3.05 (35.7%)	5.20 ± 0.84 (47.27%)	2.80 ± 1.30* (28.0%)
FN	5.29 ± 1.50# (58.78%)	4.71 ± 1.38~ (47.1%)	6.40 ± 0.89# (71.11%)	6.20 ± 0.48~ (62.0%)
TP	3.71 ±1.50# (41.22%)	5.29 ± 1.38~ (52.9%)	2.60 ± 0.89# (28.89%)	3.80 ± 0.48*~ (38.0%)
TN	7.71 ± 1.11# (70.09%)	6.43 ± 3.05 (64.3%)	5.80 ± 0.84# (52.73%)	7.20 ± 1.30 (72.0%)
Quality scores	14.86 ± 3.02 (49.53%)	12.14 ± 2.41* (55.18%)	15.4 ± 3.13 (51.33%)	11.80 ± 1.92 (53.64%)
Confidence	7.18 ± 0.95 # (71.8%)	7.52 ± 0.42 # (75.2%)	4.78 ± 1.91 # (47.8%)	6.41 ± 1.05 # (64.1%)

*Data is presented as mean ± standard deviation *significant difference within group (P<0.05) # significant difference between groups at baseline ~significant difference between groups post intervention (P<0.05)*

Figure 5-6: True positive values for reporting radiographers training in chest image interpretation pre and post study

Figure: True positive values pre and post study
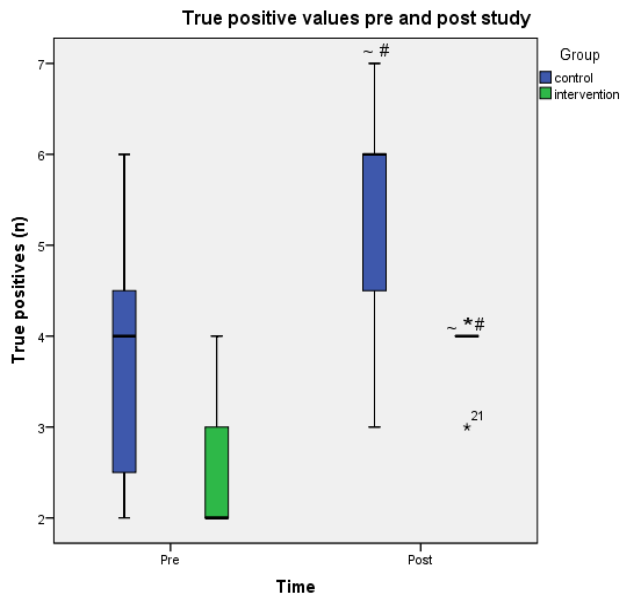
 # significant difference between groups at baseline *significant difference within group

~significant difference between groups post intervention (P<0.05)

Figure 5-7: False negative values for reporting radiographers training in chest image interpretation pre and post study

Figure: False negative scores pre and post study

 # significant difference between groups at baseline ~significant difference between groups

post intervention (P<0.05)

126
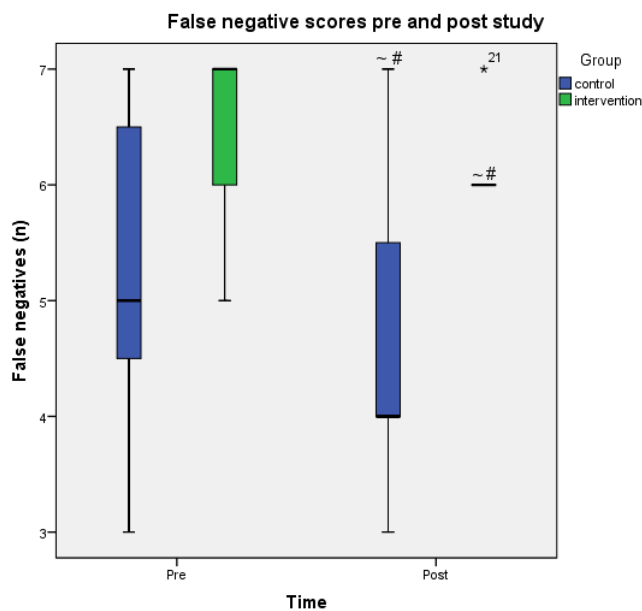
5.6.4 Eye tracking data

Mean fixation duration and mean fixation count significantly decreased for both groups following the intervention period ($p<0.05$) (Table 5.14). Mean visit duration decreased for the control group.

Decision time amongst the CXR reporting radiographers increased from approximately 67.7 secs per image at the initial testing period to 76.7 secs per image at the follow up testing period for the intervention group participants ($p<0.05$) (Table 5.15).

5.6.3 Correlations of reporting radiographers training in chest image interpretation

There was a large positive correlation found between reporting experience and confidence of control participants ($r=0.788$, $n=7$, $p<0.05$). The greater experience the RR's had in reporting images the more likely they were to be confident in the diagnosis they gave (Figure 5.8). There was also a medium positive correlation found in interpretation time and the number of TP's ($r=0.385$, $n=7$) (Table 5.16, Figure 5.9). Control participants who spent more time interpreting the images were more likely to identify abnormal images.

A large negative correlation was noted in reporting experience and the number of TP's ($r= -0.837$, $n=5$), whereby the more experienced the intervention group were the less likely they were to identify an abnormal image. There was also a large negative correlation identified between the interpretation time and experience of the intervention group ($r= -0.548$, $n=5$), more experienced participants tended to spend less time interpreting the chest images. A large positive correlation existed between interpretation time and the number of TP's ($r=0.577$, $n=5$), the more time they spent looking at the image the more likely they were to successfully identify an abnormal image (Table 5.16).

5.6.4 Fixation frequency

Fixation frequency increased for both groups in the follow up testing period ($p<0.05$). This was most obvious in the intervention group with an increase of 2.62 Hz in the follow up testing period (see Table 5.17).

Table 5-14: Eye tracking data for the entire image interpretation of reporting radiographers training in chest image interpretation

	Control pre (n=7)	Control post (n=7)	Intervention pre (n=5)	Intervention post (n=5)
Mean fixation duration (secs)	49.50 ± 21.32	26.01 ± 10.93*	47.01 ± 23.62	20.66 ± 28.06*
Mean time to first fixate (secs)	0.15 ± 0.15	3.65 ± 7.86	1.57 ± 2.87	5.02 ± 7.30
Mean fixation count (n)	169.11 ± 68.37	97.19 ± 37.41*	169.29 ± 89.39	99.39 ± 109.43*
Mean visit duration (secs)	63.12 ± 26.37	46.46 ± 14.11*	67.72 ± 19.76	53.11 ± 31.10
Mean visit count (secs)	2.51 ± 2.11	3.99 ± 3.90	8.31 ± 7.31	3.57 ± 4.31

*Data is presented as mean ± standard deviation. All values are means ± SD. *significant difference within group*

Table 5-15: Eye tracking data for the Areas of Pathology (AOP) of reporting radiographers training in chest image interpretation

	Control pre (n=7)	Control post (n=7)	Intervention pre (n=5)	Intervention post (n=5)
Mean fixation duration (secs)	11.69 ± 8.12	7.33 ± 3.36	7.93 ± 4.24	5.16 ± 6.70
Mean time to first fixate (secs)	8.32 ± 5.61	11.15 ± 10.17	13.28 ± 6.62	11.44 ± 9.14
Mean fixation count (n)	35.38 ± 20.13	25.08 ± 8.85	24.74 ± 13.02	23.95 ± 22.91
Mean visit duration (secs)	13.06 ± 8.53	9.29 ± 3.11	9.09 ± 4.01	9.30 ± 7.17
Mean visit count (n)	13.89 ± 6.77	9.70 ± 4.03	12.17 ± 6.32	9.12 ± 7.95
Mean decision time (secs)	58.80 ± 16.77	65.81 ± 28.09	67.72 ± 17.12	76.71 ± 18.86*

*Data is presented as mean ± standard deviation. All values are means ± SD. *significant difference within group*

Table 5-16: Correlations post intervention

	All (n=12)	Control (n=7)	Intervention (n=5)
TP and confidence	0.426	0.125	0.167
TN and confidence	-0.114	-0.227	0.267
TN and experience	-0.021	0.116	-.0249
TP and experience	0.149	0.149	-0.837
Confidence and experience	0.548	0.788	0.387
Interpretation time and TP	0.177	0.385	0.577
Interpretation time and experience	-0.266	-0.62	-0.548

Figures are presented as r values

Figure 5-8: Reporting experience and confidence levels of reporting radiographers at the follow up testing period (r=0.788, n=7, p<0.05) (r=0.387, n=5)

129
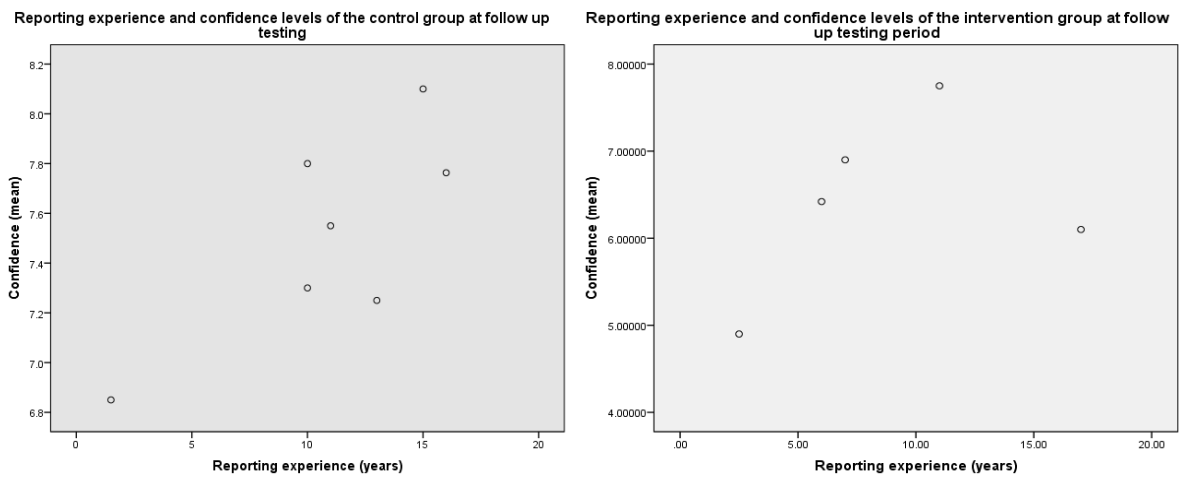
Figure 5-9: True positive values and interpretation times of reporting radiographers at the initial testing period (r=0.962, n=5, p<0.001).
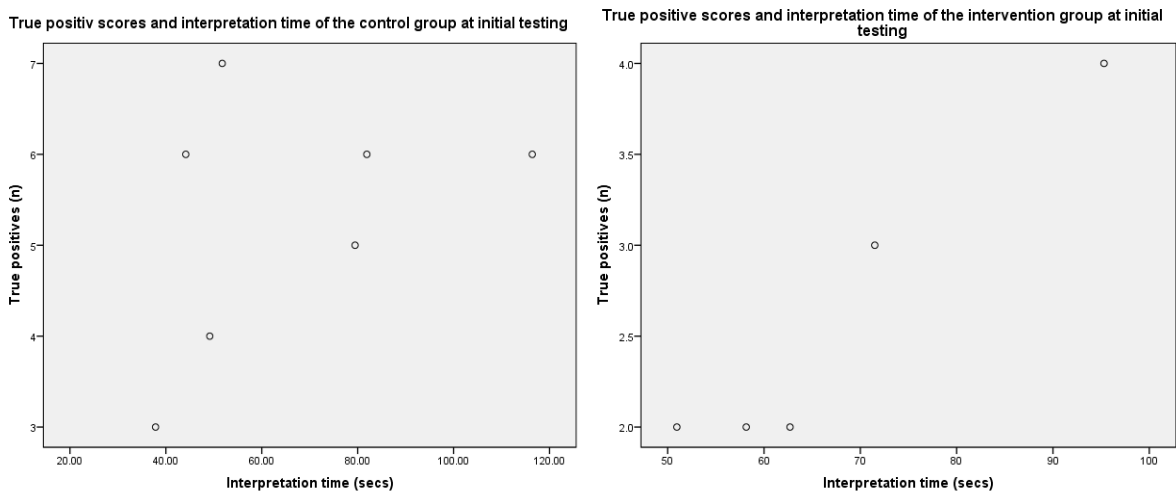
Table 5-17: Fixation frequency for reporting radiographers training in chest image interpretation

	Control pre (n=7)	Control post (n=7)	Intervention (n=5)	Intervention post (n=5)
Fixation frequency (Hertz)	3.43 ± 0.26#	3.84 ± 0.40*~	3.66 ± 0.50#	6.28 ± 1.54*~

*Data is presented in mean ± standard deviation. # significant difference between groups at baseline *significant difference within group ~significant difference between groups post intervention (P<0.05)*

5.7 MSK Results

5.7.1 Demographics

Control group participants were qualified for longer and had a greater experience interpreting images. However, intervention group participants had a greater mean experience of 1.07 years in reporting MSK images (Table 5.18).

5.7.2 Scores

TP, TN scores increased and FP scores decreased for the intervention group. As TP and TN scores increased, it was expected that FP scores would have decreased, this decrease was significant. For both groups there were fewer errors in identifying a normal image following the intervention period. This was a mean difference of 11.14% for the control group and a mean difference of 21.36% for the intervention group (Table 5.19).

There were minimal decreases in the quality scores of the control and intervention group.

The intervention group's confidence increased following the implementation of the training tool. A slight increase of 2.1% was seen in the control group, compared to an increase of 6.7% in the intervention group (Figure 5.10).

5.7.3 Eye tracking data

In the groups, none of the eye tracking metrics were significantly different. However, the intervention group spent longer fixating, took longer to fixate, fixated a greater number of times and visited on the area of pathology for longer or more times (Table 5.20).

Table 5-18: Demographics for reporting radiographers trained in MSK image interpretation

	Reporting radiographers trained in MSK image interpretation control pre (n=13)	Reporting radiographers trained in MSK image interpretation pre intervention (n=10)
Qualified (years)	22.0 ± 11.37	14.15 ± 8.36
Experience interpreting images (years)	21.31 ± 10.31	13.55 ± 7.05
Experience reporting MSK images (years)	4.23 ± 4.89	5.3 ± 5.42
Sex	78.3% Female (n=18) 21.7% Male (n=5)	

Mean ± standard deviation

Table 5-19: Scores and confidence of reporting radiographers trained in MSK image interpretation

	Control pre (n=13)	Control post (n=13)	Intervention pre (n=10)	Intervention post (n=10)
FP	4.69 ± 1.80 (42.64%)	3.00 ± 1.16* (30.0%)	6.20 ± 2.30 (56.36%)	3.50 ± 1.79* (35.0%)
FN	6.31 ± 1.80 (70.11%)	6.08 ± 1.32 (60.8%)	6.60 ± 1.43 (73.33%)	5.90 ± 1.52 (59.0%)
TP	2.69 ± 1.80 (29.89%)	3.92 ± 1.32 (39.2%)	2.40 ± 1.43 (26.67%)	4.10 ± 1.52* (41.0%)
TN	6.31 ± 1.80 (57.36%)	6.85 ± 1.35 (68.5%)	4.80 ± 2.30 (43.64%)	6.50 ± 1.78* (65.0%)
Quality scores	12.31 ± 2.32 (41.03%)	9.31 ± 2.69* (42.32%)	12.10 ± 2.28 (40.33%)	8.60 ± 1.65* (39.09%)
Confidence	5.63 ± 1.02 (56.3%)	5.84 ± 1.06 (58.4%)	5.60 ± 1.61 (56.0%)	6.27 ± 1.13* (62.7%)

*Data is presented as mean ± standard deviation *significant difference within group*

Figure 5-10: Confidence levels of reporting radiographers trained in MSK image interpretation

Confidence levels pre and post study
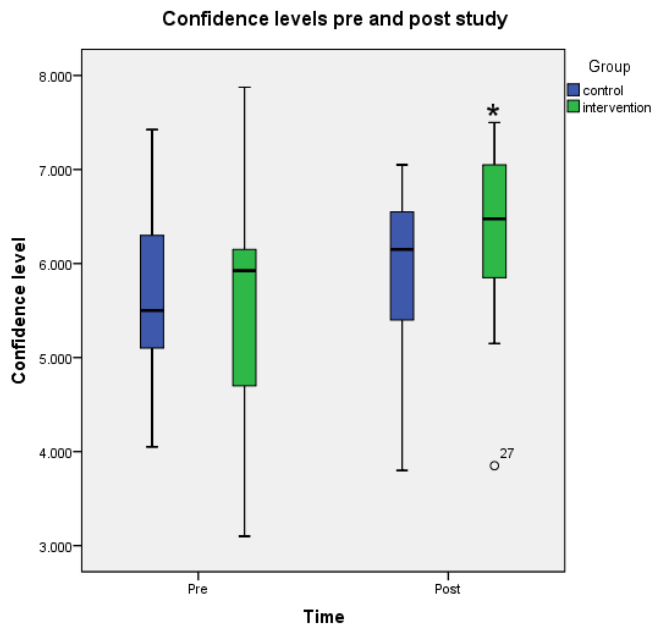
Table 5-20: Eye tracking data for the entire image interpretation of reporting radiographers trained in MSK image interpretation

	Control pre (n=13)	Control post (n=13)	Intervention pre (n=10)	Intervention post (n=10)
Mean fixation duration (secs)	41.81 ± 22.77	28.46 ± 18.17*	37.80 ± 31.14	35.26 ± 18.42
Mean time to first fixate (secs)	3.47 ± 6.89	2.25 ± 4.77	5.86 ± 15.18	3.61 ± 10.06
Mean fixation count (n)	150.33 ± 76.53	108.14 ± 69.90	143.68 ± 102.05	145.79 ± 73.80
Mean visit duration (secs)	59.71 ± 19.83	43.25 ± 27.85	58.29 ± 28.38	62.97 ± 22.96
Mean visit count (secs)	2.85 ± 2.27	2.44 ± 1.98	1.88 ± 0.82	2.59 ± 1.74
Mean decision time (secs)	70.17 ± 16.37	55.87 ± 24.04	68.97 ± 25.39	72.63 ± 37.16

*Data is presented in mean ± standard deviation. *significant difference within group*

133

Table 5-21: Eye tracking data for the Areas of Pathology (AOP) of reporting radiographers trained in MSK image interpretation

	Control pre (n=13)	Control post (n=13)	Intervention pre (n=10)	Intervention post (n=10)
Mean fixation duration (secs)	7.37 ± 4.00	8.25 ± 5.39	6.57 ± 5.00	9.34 ± 4.76
Mean time to first fixate (secs)	7.62 ± 4.91	7.75 ± 6.86	13.36 ± 18.82	11.53 ± 18.88
Mean fixation count (n)	23.76 ± 12.25	28.65 ±19.66	22.59 ± 14.17	35.66 ± 18.87
Mean visit duration (secs)	8.51 ± 4.11	10.08 ± 5.83	8.19 ± 4.78	11.44 ± 5.82
Mean visit count (secs)	10.64 ± 5.33	8.76 ± 5.37	10.48 ± 6.65	12.60 ± 6.83

*Data is presented as mean ± standard deviation. *significant difference within group*

5.7.4 Correlations of reporting radiographers trained in MSK image interpretation

There was a weak positive correlation seen in the TN scores and experience of control participants. The more experienced they were the more likely they were to diagnose a normal image correctly (Table 5.22).

There was a weak negative correlation in the TP scores and experience of participants in the intervention group, this was a moderate weak correlation in the control group. The more experienced the participants were the less likely they were to diagnose an abnormal image correctly.

Control participants were more confident with greater experience in image interpretation, given the moderate positive correlation seen (Table 5.22).

5.7.5 Fixation frequency

Fixation frequency slightly decreased for both groups but this was not significant (Table 5.23).

Table 5-22: Correlations of reporting radiographers trained in MSK image interpretation post intervention

	All (n=23)	Control (n=13)	Intervention (n=10)
TN and confidence	-0.104	-0.241	0.065
TP and confidence	0.120	0.182	0.031
TN and experience	0.163	0.269	0.109
TP and experience	-.263	-.330	-0.230
Confidence and experience	0.191	0.376	-0.040
Interpretation time and TP	-.015	0.170	-0.187
Interpretation time and experience	-0.190	-0.505	-0.066

Table 5-23: Fixation frequency for reporting radiographers trained in MSK image interpretation. Data presented are r values

	Control pre (n=7)	Control post (n=7)	Intervention pre (n=5)	Intervention post (n=5)
Fixation frequency (Hertz)	4.02 ± 0.97	3.96 ± 1.91	4.68 ± 1.56	4.40 ± 0.93

*Data is presented as mean ± standard deviation Mean fixation frequencies of reporting radiographers training in chest image interpretation. # significant difference between groups at baseline *significant difference within group ~significant difference between groups post intervention (P<0.05)*

5.8 Survey

5.8.1 Design

A survey was designed to gain feedback on the digital training tool (appendix 5.6). Firstly the survey was piloted among academic staff who were also radiographers to check for ambiguity of questions. All staff were satisfied with the survey layout and phrasing, therefore no modifications were made to the survey before distribution.

The survey sought to determine opinions on the tool and in particular its:

- convenience
- layout
- usability
- display
- content

Participants were also given the opportunity to offer recommendations or changes which would benefit the tool and its purpose. The tool was given to the intervention and control group to maximise feedback on the tool, participants from both groups responded to the survey.

The survey was transformed into an electronic format using the Qualtrics Software. This University has a license for this software and allows the data collected from the survey to be stored securely and extracted efficiently for analysis.

5.8.2 Recruitment

The survey was distributed via e-mail (through Qualtrics software) to the intervention group participants following their completion of their follow up assessment. Control participants were provided with access to the training tool following the study. The survey was then e-mailed to this group of participants through the software and all feedback was welcomed on the tool.

5.8.3 Results

Nine reporting radiographers completed the survey online.

<u>Access to the digital training tool</u>

The digital tool was accessed on work computer/laptop by four participants. It was accessed on computer/laptop by four participants and one participant accessed it on their tablet and smartphone.

Frequency of use

One participant answered that they used the tool once a week. Five participants said they rarely used the tool. Three participants answered 'never used' for this question however all three were allocated to the control group previously for the study and so would only have access to the tool following the study.

Three participants mentioned time as an issue when explaining their frequency of use of the tool. One participant mentioned the difficulty accessing the tool during working hours and image quality being the cause of not using the tool often. One participant mentioned the tool was too lengthy and another admitted that they had forgotten to use the tool.

Search strategy used to interpret chest images

Five participants answered 'yes' to using a search strategy during chest image interpretation, four answered 'no', three of which were control group participants. Four said they used a search strategy devised by themselves whilst two said they used the search strategy supplied within the training tool.

Improved skills of chest image interpretation

Participants were asked if they thought the training tool improved their skills in image interpretation (e.g. speed, accuracy, confidence). Four answered 'yes' and three answered 'no'. Two of the three participants who answered 'no' were control participants. The four participants who answered 'yes' mentioned this self-perceived improvement was due to the search strategy included within the tool. Three of these were reporting radiographers trained in MSK image interpretation, one was a reporting radiographer training in CXR image interpretation and all four were in the intervention group. This included the benefit of having a systematic approach to the image interpretation and ensuring all areas on the image were considered. One participant said they used the tool as a refresher in the beginning which helped to increase their confidence.

Most useful features

Four participants detailed the features they found most useful in the training tool. These included; using it to check all areas on an image, the narrated videos, the consistent approach to all images and the structure when interpreting images.

Least useful features

The only 'least useful feature' noted was 'lack of accessibility at work'.

<u>Suggestions to improve the digital training tool</u>

A participant suggested the videos could be prefaced with a brief intro, explaining what the videos are suggesting and one noted confusion in wondering where the image was for interpreting in the search strategy training tool section.

Table 5-24: Mean scores of rated features of the tool

	Layout	Accessibility	Visualisations	Content	Eye tracking videos	Search strategy training tool
Mean score	8.6	7.2	5.2	7.6	7.6	8.0

Table 5-25: The System Usability Scale (SUS)/Likert scale of survey responses

Statement / Likert Question	Participant individual responses within a Likert scale				
I think I would like to use the training tool frequently	2	4	4	3	5
I found the training tool unnecessarily complex	2	1	2	3	1
I thought the training tool was easy to use	4	5	4	4	5
I think I would need the support of a technical person to use this system	1	1	2	1	4
I found the various functions within the training tool were well integrated	3	5	4	4	5
I thought there was too much inconsistency in this training tool	2	1	2	1	1
I would imagine that most people would learn to use this training tool very quickly	4	5	4	5	5
I found the training tool very cumbersome to use	2	2	2	3	1
I felt very confident using the training tool	4	5	4	3	5
I needed to learn a lot of things before I could get going with this training tool	2	4	2	3	1
SUS score (added together and multiplied by 2.5)	77	82.5	75	75	82.5

1=strongly disagree 2= disagree 3=neutral 4=agree 5=strongly agree

Continued use

Three participants, a reporting radiographer training in chest image interpretation and two reporting radiographers trained in MSK image interpretation, will continue to use the training tool, three participants will not continue to use the training tool.

Further comments/additional feedback

One participant mentioned their inability to view a chest image when filling in the training tool as a limitation to the tool. Four participants mentioned their participation in the study as a participant allocated to the control group or their lack of interpreting chest images currently were reasons why they have not engaged with the tool.

5.9 Training tool use

Use of the training tool varied throughout the study timeline. A summary of the eye tracking videos and search strategy training tool is detailed in Appendix 5.7

5.10 Heat maps

Heat maps were extracted using the eye tracking technology software. All images below are overlaid with a heat map. Green areas represent areas of low fixation counts and red areas represent areas of high fixation counts.

<u>Initial testing period</u>
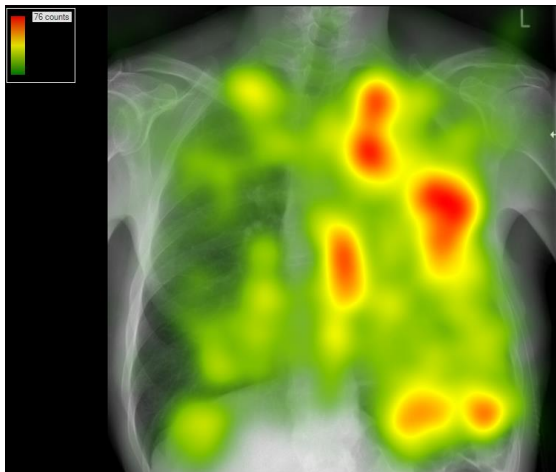
Image A   Image B
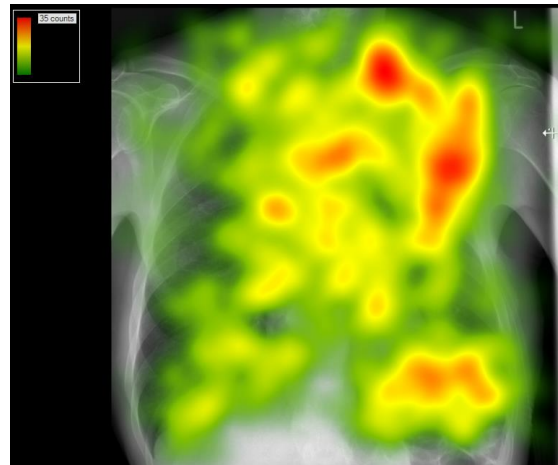
Image A and Image B are representative examples of the image interpretation by reporting radiographers trained in MSK image interpretation (Image A) and from reporting radiographers beginning training in CXR image interpretation (Image B) at the initial testing (pre). The pathology was a pneumothorax in the left apex. Both images are the same.
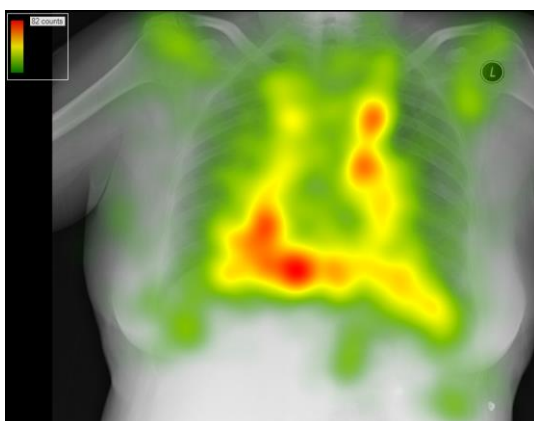
Image C   Image D
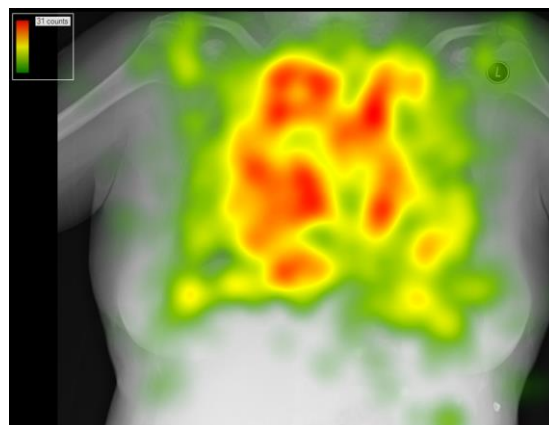
Image C and Image D are a normal image interpreted by reporting radiographers trained in MSK image interpretation (Image C) and reporting radiographers beginning CXR image interpretation at the initial testing (pre) (Image D).

Reporting radiographers beginning CXR image interpretation training tended to be more widespread in their image interpretation, with both abnormal and normal image heat maps showing higher numbers and more diffusion of high fixation count areas.

Follow up testing period-MSK
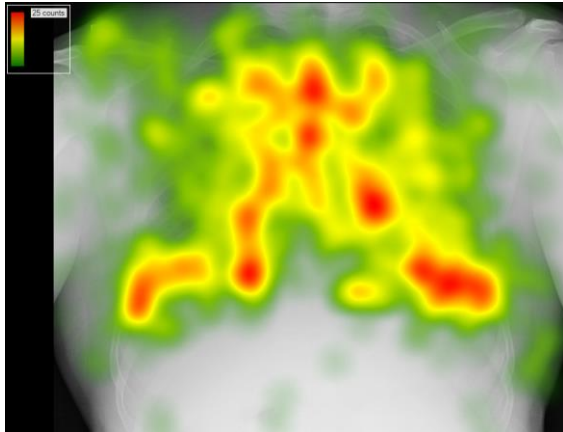
Image E   Image F
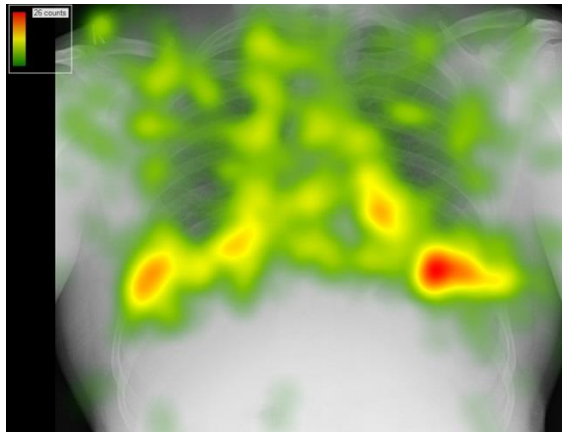
Image E and Image F are representative examples of image interpretation by reporting radiographers trained in MSK image interpretation at the follow up testing (post intervention period). Image E is an example of a participant from the intervention group and Image F is an example of an image interpreted by a participant from the control group. Both are normal images.
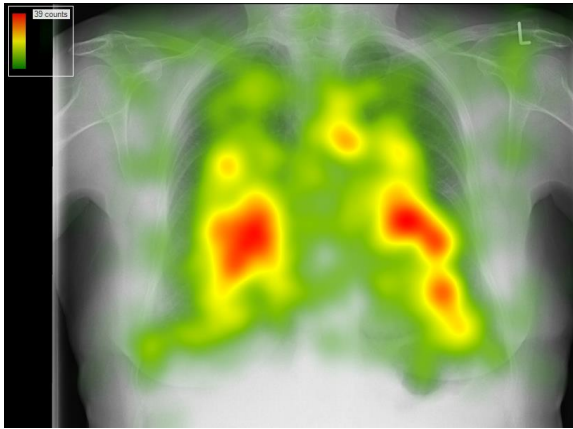
Image G   Image H
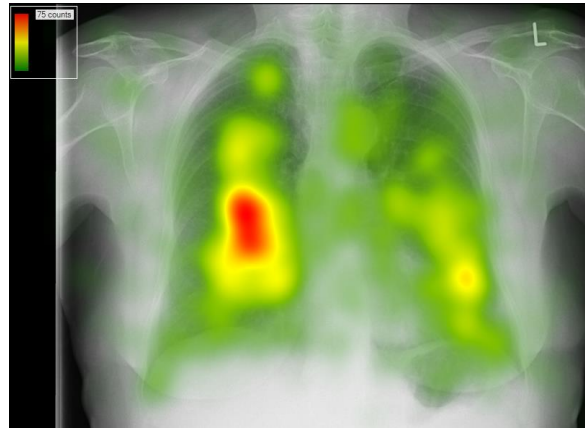
Image G and Image H are representative examples of the image interpretation by reporting radiographers trained in MSK image interpretation at the follow up testing (post intervention period). Image G is an example of interpretation by a participant who was in the intervention group and Image H is an example of interpretation by a control participant. Both are abnormal images, the pathology present was a lung nodule in the right middle zone.

Areas of high fixation counts are greater in the intervention group compared to the control group in the follow up testing period.

Follow up testing period-CXR
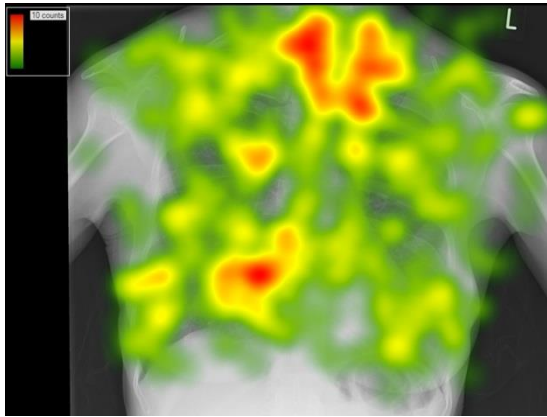
Image I   Image J
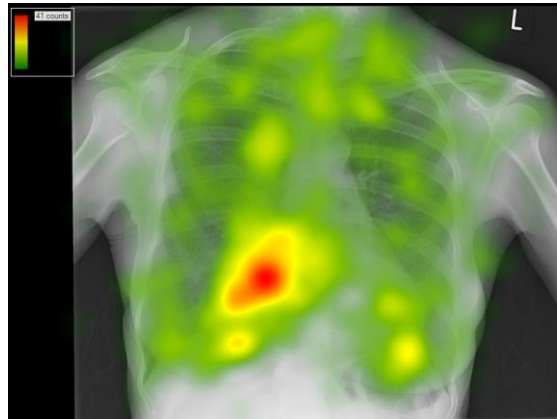
Image I and Image J are the same normal image interpreted by reporting radiographers beginning CXR image interpretation training at the follow up testing (post intervention period). Image I is an example of interpretation by a participant from the intervention group and Image J was interpreted by a control participant.
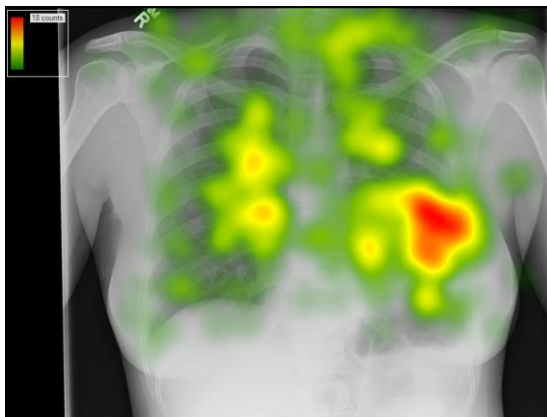
Image K   Image L

Image K and Image L are the same abnormal image interpreted by reporting radiographers beginning CXR image interpretation training at the follow up testing (post intervention period). Image K is an example of interpretation by a participant from the intervention group and Image L is an example of interpretation by a control participant. The pathology present was consolidation in the left lower zone.

In the normal image, the intervention group have a high number of fixation counts throughout the image. The control group have fewer high fixation count areas.

Both groups fixate greatly over the pathology in the left lower zone, however the intervention group have slightly more fixation counts throughout the image compared to the control group.

<u>Control and intervention groups-follow up testing period</u>
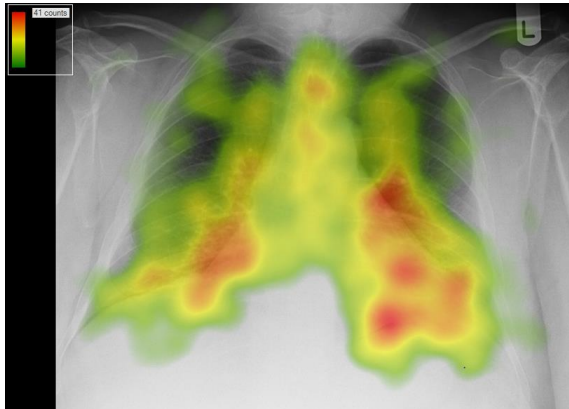
Image M   Image N
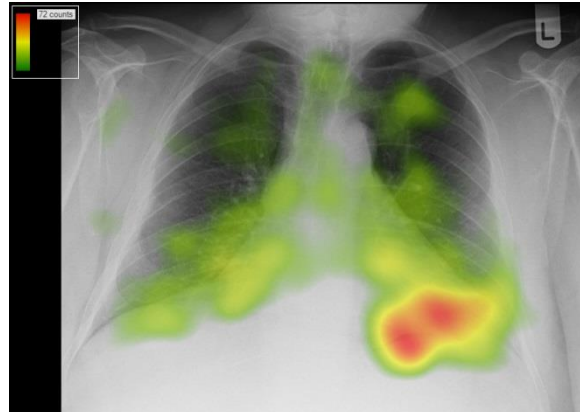
Image M and Image N are a normal image as interpreted by a representative participant from the intervention group (Image M) and control group (Image N) (post intervention period).
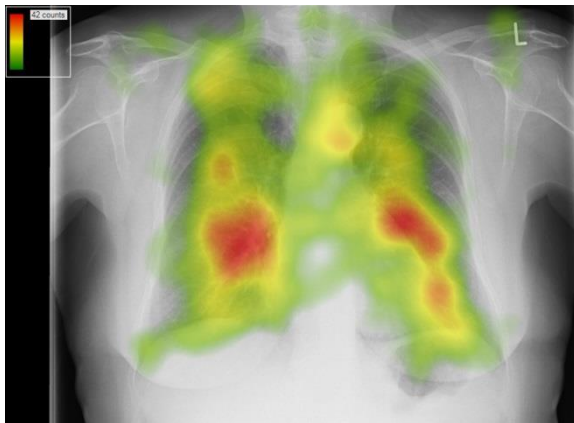
Image O   Image P

Image O and Image P are the same   abnormal image as interpreted by a representative participant from the   intervention group (Image O) and control group (Image P) (post intervention period) The pathology was a lung nodule in the right middle zone. Both the control and intervention group participants demonstrate areas of high fixation counts in the right middle zone, where the pathology is present. The intervention group had more variation in their eye gazes and this perhaps is an indication of a systematic approach being applied which the tool encourages.

5.10 Discussion

Control and intervention group differences

Decrease in FP values and increase in TN values were seen in the intervention group. FP scores decreased for the control group also. These improvements in performance were similar to those identified in Chapter 2 (Sonnex et al. 2001; Piper et al. 2014; Woznitza et al. 2014). The improvements seen in this study would suggest the tool has potential to improve and greatly aid chest image interpretation for reporting clinicians. As this study included only reporting radiographers and chest images there is further work to be investigated in other professions and anatomical areas (McLaughlin et al. 2017a).

Control participants increased their comments given on radiographic technique and additional image appearances beyond the pathology in the follow up testing period, whereas intervention participants tended to mention these slightly less in the follow up. This may have been due to their familiarity with the task allocated which was primarily to recognise whether the image was normal/not normal and to describe/state a pathology if identified. Differences in instruction given to participants were identified in McLaughlin et al. (2017b) as a possible influence to the interpretation task undertaken.

The improvements in performance and confidence were most evident in the intervention group of both MSK and CXR radiographers, indicating that the tool potentially had a positive influence on performance and confidence.

The tool may have led to the longer interpretation times and increased number of eye gaze metrics seen in the intervention group as they tried to follow the search strategy. Longer interpretation times were seen also in McLaughlin et al. (2017a) by the experienced reporting radiographer group which searched the entire image for pathology. The time implications of implementing the tool should be taken into consideration when deciding how to use it and in what environment it is to be used in when training. Although the intervention group took longer than the control group to interpret the images, the intervention group and control group were both quicker in the follow up compared to the initial testing period. Increased familiarity with the task at the follow up period may have contributed to this.

CXR and MSK reporting radiographers

The improvements in performance seen in the intervention groups following access to the training tool may provide a less expensive and quicker means of educating reporting radiographers in chest image interpretation than paying for training or attending courses. Reporting radiographers are a cost-effective alternative to radiologist reporting in lung cancer diagnosis (AuntMinnieEurope 2017; Bajre et al. 2017), an educational tool such as this has the potential to assist the training of staff in image interpretation. In the boxplots presented, there appear to be some big outliers that are probably a result of the participant sample being small. It is interesting to note that the CXR RR post intervention plots have a very small variation even if the mean is lower for TP values suggesting an effect throughout the whole group. Both CXR RR groups improve but all of the intervention group have improved given the small variation, whereas there is still considerable variation in the control group. These changes may have been due to the postgraduate course which the participants were attending for nine months however these improvements are not observed within the control group who are attending the same course. It is therefore possible to suggest that the effect was a result of the use of the training tool.

Significant improvement in TP rate of the intervention group is seen post receipt of and working with the tools. Intervention group participants were better at identifying an abnormal image and pathology once the tool had been implemented ($p<0.05$). Improved performance in the intervention groups' ability to identify a normal image following the 9 month training period ($p<0.05$) was identified. For both CXR groups there were fewer errors in identifying a normal image following the intervention period. This was a mean difference of 11.14% for the control group and a mean difference of 21.36% for the intervention group. The improvement in performance mimics the improvements seen in the review focusing on chest image interpretation training (McLaughlin et al. 2017b).

TP, TN scores increased and FP scores decreased for the MSK RR intervention group. As the TP and TN scores are directly linked to FP and FN scores, changes in FP scores were expected following the increase in TN scores. However, significant decrease was only noted in FP scores for the MSK RR control group. This increased TP and TN scores and increased confidence ($p<0.05$) seen within the MSK intervention group may be a reflection of the value of using the tool given that this change was not seen in the control group (McLaughlin et al. 2018). Increases in TP and TN were also seen in Hughes et al. (1996) post intervention. The

148

implementation of a pattern recognition technique in chest image interpretation had similar positives to the search strategy and tool implemented in this study (Hughes et al. 1996; McLaughlin et al. 2018). FN scores decreased for both groups, similar to Manning et al. (2004) where FN in their study decreased following the implementation of a 6 month training programme. The training programme in Manning et al. (2004) was a postgraduate programme, differing greatly from the less formal training provided by the training tool. Both of these training types help decrease FN scores, therefore both have merit in improving the education of chest image interpretation.

TN rates increased for the intervention group as a whole (both MSK and CXR radiographers) when analysed as one group (p<0.05). When analysed further, this significant difference was not seen for the TN scores of CXR radiographers, however an increase was observed in the MSK intervention group. This could therefore indicate that the postgraduate programme attended by the CXR radiographers masked this effect, where a difference wasn't between the control and intervention group, whereas those with only access to the tool, the MSK radiographers, demonstrated an effect on TN scores.

There was a majority of male participants in the control CXR group and a majority of female in the CXR intervention group. Participants were in small cohorts and attended the same University, therefore to minimise the effect of the intervention group participants sharing the tool with those in the control group a convenience sample was used.

The quality score percentage was increased for both CXR groups at the follow up test period. Quality scores were awarded for the participant acknowledging error in radiographic technique, normal variants or image appearances, which could be indicative of potential pathology. CXR RR's tended to mention these more in the follow up testing period. There were fewer quality scores available in the follow up batch of images so although the figures decreased in the tables (Table 5.6.2), the overall percentage increased. The cause for this is uncertain but may be linked to the systematic approach of interpreting the chest images present in the tool or as a result of participant progression through the formal training over the 9 month period. The postgraduate programme participants were attending and the tool could have emphasised the impact radiographic appearances have on abnormality detection (Canterbury Christ Church University 2018).

Quality scores varied for the initial test bank and follow up test bank images for the MSK group. There were minimal decreases in both the control and intervention group. This is in contrast to the CXR RR groups where larger increases were seen. Although the tool and postgraduate training may have influenced the CXR RR group to think more on radiographic technique and varied image appearances, the MSK RR group may have focused more on identifying the pathology in the follow up phase and referred less to the quality score criteria (McLaughlin et al. 2018). The task of chest image interpretation is very different to their daily tasks of MSK interpretation (Woznitza et al. 2014), possibly influencing how they approached their interpretation and the image appearances they mentioned. Radiographers concern themselves primarily with image quality and radiographic technique, to ensure the best possible image provided for diagnosis. As radiographers complete image reporting courses they become more concerned with the consequences of missing a pathology. There is less overall educational grounding for this with the MSK reporting radiographers compared to CXR reporting radiographers given the greater difficulty and risk which comes with interpreting chest images (RCR 2006). Therefore the decrease in quality scores may have been a reflection of MSK reporting radiographer learning to behave similar to those qualified to report chest images, where an incorrect diagnosis could have detrimental consequences.

All groups were more confident in the diagnoses they provided at the end of the study. This enhancement in confidence is similar to the increase of 10.2% seen in Hughes et al. (1996). The increase in confidence observed within this study could be a reflection of the use of the tool and in the participant being more confident that all areas of the image have been checked for the presence of a pathology.

Decision time increased for both CXR groups, however within the intervention group this was statistically significant. The increase in decision time may have occurred due the implementation of a search strategy and the additional time required to incorporate this element to the interpretation process. This is similar to Litchfield et al. (2010) where the type of training provided determined whether the decision time increased or decreased. Utility function feature, whereby participants may take longer to decide on a diagnosis when there is a consequence of an incorrect diagnosis, does not influence participants in this study, but should be considered when considering participant decision time (McConnell et al. 2005; Kahneman et al. 2013). Images were already reported on and the patient had received their

diagnosis, therefore patient treatment or care was not influenced by the participant's decisions. The CXR RR group have more responsibility to get the diagnosis right compared to the MSK RR group, given their career progression into this field, this may account for the slightly longer interpretation time of the reporting radiographers in the CXR group compared to the MSK group. The training tool has the potential to improve interpretation performance however the time implications of its use must be taken into consideration also.

The additional time and eye gaze metrics seen in some of the intervention groups may have been due to the additional questions and search strategy posed within the tool. Although these outcome measures increased, the overall increases were minimal (10-20secs) and outweigh the potential of missing an abnormality. Overtime, as participants were to become accustomed to using the search strategy and tool in practice this slight delay may reduce and become obsolete.

Few significant differences in the eye tracking metrics were seen between groups.

The MSK control group spent longer fixating on the image at the follow up testing period whereas the MSK intervention group spent slightly less time fixating on the image following the intervention. Intervention group participants spent longer fixating and visiting the areas of pathology.

Fixation duration significantly reduced for the MSK control group but only slightly decreased for the intervention group. Fixation count decreased for the control group however, increased slightly for the intervention group, this is similar to the CXR RR results. These prolonged fixation durations and fixation count for the intervention group may be a result of introducing the systematic search strategy via the training tool to the intervention group. This is likely a result of viewing the expert's eye tracking videos and attempting to mimic their systematic and thorough eye gaze on the images within the education programme video.

The MSK intervention group saw improved performance and increased confidence however, increase in fixation duration, fixation count, time to first fixation and visit count on the area of pathologies were also seen.

Although no other significant differences were noted in the AOP eye gaze metrics, in general the eye gaze metrics tended to decrease at the follow up testing period compared to the initial testing period. The number and duration of fixations significantly reduced in both the control

and intervention CXR group, perhaps indicating the tool or postgraduate training encouraged the participants to scan over the image rather than fixate on images for longer. One increase was recognized in the control group where participants took longer to fixate on the pathology (approximately 11.2 secs) at the follow up testing period. This is similar to the time to first fixate taking longer for the less experienced groups in McLaughlin et al. (2017a). The control group, with no access to the training tool, performed similarly by taking longer to fixate on the AOP in the follow up testing therefore indicating that this may be an effect of the postgraduate training which both groups of participants were attending.

Both MSK groups spent longer fixating on the areas of pathology in the follow up data collection. This was more obvious in the intervention group compared to the control group. The intervention group were quicker to fixate on the AOPs following the intervention period. The amount of times both groups looked at the AOPs increased in the follow up test period. This was greater in the intervention group where by the number of fixations increased by 13.07 compared to 4.89 in the control group. The tool was given to intervention participants and participants were then self-driven, in that they were asked to use it at their choice. MSK control participants most likely completed little to no training in chest image interpretation over the nine month period other than those encountered through CPD activities. These changes therefore may have been a result of the satisfaction of search element embedded in the search strategy training tool. The number of times the participant visited the AOPs decreased for the control group following study period but increased for the intervention group. However, these were not significantly different. The intervention group spent longer fixating, took longer to fixate, fixated a greater number of times and visited on the area of pathology for longer or more times. It is opined the questions asked in the search strategy and the detail required for the questions contributed to the extra time and attention given to the areas of pathology. In comparison, the changes within the control groups were minimal or decreased instead of increasing, as they did not have access to the tool they would have had little to no changes to their method of interpreting images.

The fixation frequency is the number of fixations per second (hertz). A high fixation frequency indicates that the participant rapidly gazed over a large area of the screen and was more sporadic in their image interpretation (McLaughlin et al. 2017a). Both CXR group's fixation frequency increased at the follow up testing period. This increase in fixation

frequency could have been due to the participants now implementing a systematic method of interpreting chest images. The fixation frequency within the CXR intervention group increased more than that of the control group at the follow up testing period ($p<0.05$). The training tool incorporated a search strategy which aimed to help teach the participants how to interpret an entire image systematically. This is believed to be due to the fact that the search strategy gave a routine to follow an associated increase in fixation frequency for this particular group is likely. When following a search strategy, participants would be more likely to fixate on a greater number of areas over a short period, hence demonstrating an increase in fixation frequency. Alternatively, they may have developed a scanning technique that helped them concentrate and look at key areas where pathologies are more likely to be hidden, aiding them to become more efficient with their interpretation process. Participants attending the postgraduate programme were mentored and told areas in which pathologies are normally found; instructions such as these would influence eye gaze metrics also. Fixation frequency decreased for the MSK groups but these were not significant.

The more experienced the radiographers were in reporting MSK images, the more confident they tended to be in interpreting chest radiographic images. This was evident in the control group with medium positive correlation of 0.376. This is believed to be due to their confidence in their role as a reporting radiographer and interpreting images compared to those with no formal postgraduate training in image reporting. A positive correlation existed between interpretation time and the number of TP's of both MSK and CXR group. The formal training supplied to participants potentially increased their ability to identify an abnormal image. The more experienced the CXR intervention group were, the more likely they were to miss an abnormal image. This was an unexpected finding, it is not clear why this may be.

Survey

Feedback from the survey was mostly positive with the benefits of the tool highlighted. Feedback given in the survey helped identify the specific issues that users had with the tool. Inability to view a chest image when filling in the training tool was highlighted as a limitation. This point will be considered carefully when continuing the research in use of the tool.

Four participants mentioned their participation in the study as a participant allocated to the control group or their lack of interpreting chest images currently were reasons why they have not engaged with the tool.

The SUS score was above 75 for all participants who completed the survey, the score was 82.5 for two participants. Given any scores above 68 are regarded as above average the tool scored highly. This is a positive reflection on the tool, its content and usability (Brooke 1996).

Five participants responded to the survey from both the control and intervention group. The small sample size is a limitation of the survey.

The tool has potential to make a positive impact on participant performance. Positives were highlighted in the survey feedback. The tool could be a cost effective measure for the health system. By training radiographers or other healthcare professionals with the tool there is a possibility of saving costs in paying radiologists to report on the images, either by reducing numbers employed or using external providers to address reporting turn around times. Furthermore, the tool as an accessory to formal training could provide a means of support and further education for reporting clinicians.

Heat maps

In general, the high fixation count areas and fixation count areas tended to be more widespread in the intervention group participants in the follow up testing period. This could potentially be an effect of the training tool which aims to encourage users to search the entire image to exclude pathologies. Areas of pathology tended to have high fixation counts by both control and intervention group participants. This may indicate that although both groups were able to identify an area of abnormality, the method of interpretation may have changed in the intervention groups due to the systematic viewing approach encouraged by the tool.

5.11 Limitations

Two reporting radiographers trained in MSK image interpretation completed the initial data collections and were randomly allocated to the intervention group however, they were unable to complete the follow up data collections. This was due to restricted schedules and not as a result of the study contents. The participant data was excluded from analysis.

Lack of use of the training tool could be a concern. However, if the training tool was used only once the search strategy incorporated in it could be learnt and implemented in clinical practice routinely.

The emails including the survey were sent via the Qualtrics software; it was later discovered that some NHS hospitals blocked emails from this server and not all participants received the email successfully. Emails were then sent again from the university portal rather than the Qualtrics software to try to overcome this limitation.

The reporting radiographers trained in MSK image interpretation may have also had some chest image interpretation training during their attendance of the MSK reporting postgraduate programme. They therefore may not be a complete control group with no access to CXR image interpretation.

The small sample size acquired in this study is a limitation. Measures were implemented to maximise recruitment to the study such as the inclusion of MSK and CXR reporting radiographers. However, due to the relatively small numbers of qualified reporting radiographers across the UK and limited resources, the sample size obtained was small.

5.12 Conclusion

Within the current study, the implementation of the training tool provided benefits and improvements in participant performance. Drawbacks in the lack of time participants had to use the tool and the practicalities of using it should be taken into consideration when planning where and when it could be implemented. The introduction of the tool may cause longer decision times but could lead to an overall cost effective and clinically safe service in reporting radiography. Decision times can be reduced when using the tool over time especially when the tool becomes familiar to staff and the tool itself is iteratively refined and optimised. Research can be expanded in this area; the use of the tool can be investigated in other professions and the tool can be developed to be used in the interpretation of images of different anatomical areas. Recommendations made by users in the survey should be considered and changes made to improve the quality of the tool. The tool has the potential to ensure satisfaction of search is achieved and to aid chest image interpretation training.

Chapter 6-Summary and Conclusions

6.0 Introduction

This chapter combines and discusses the main findings from the research studies carried out in this thesis. It will consider the rationale, discussion points from each chapter, summarise results and draw conclusions from each. The overall impact of the thesis will be highlighted and future research will be proposed.

6.1 Rationale for the research

The systematic literature review has identified the volume of published data discussing radiographer performance and chest image interpretation learning. Despite an evolution in the role of radiographers performing chest image interpretation over the past two decades, there is still a lack of evidence for training methods available other than that of the certified postgraduate programmes. Two of the studies which scored highest in the systematic review included participants who had completed postgraduate programmes in chest image interpretation (Piper et al. 2014; Woznitza et al. 2014). These studies tested participants on a range of pathologies, included information on the reference standard to which participants' answers were compared with, identifying if the findings were transferable to everyday clinical practice and identified confounding factors of their study. The study by Sonnex et al. (2001), which scored highly in the systematic review, included these attributes, however this study focused on assisting radiographers to 'red dot' (i.e. highlight abnormality on the image after identification) rather than provide a diagnosis on the image following interpretation (Sonnex et al. 2001). The literature review highlighted the limitations of studies completed and elements required for high quality investigations to be conducted. These elements have only been considered when the investigations focused on chest image interpretation reporting when the influence of postgraduate programmes were being assessed, or when used to measure performance during the task of 'red dotting'. Studies which included eye tracking technology in the intervention or assessment of participants were shown to have a positive effect on performance (Donovan et al. 2008; Litchfield et al. 2010). However, Donovan and Litchfields' studies were conducted over a short time period, as the effects of an intervention on participant training were assessed directly using eye tracking. The eye tracking technology included little to no guidance with it, in that expert eye tracking was shown to participants with no explanation or instruction on how the expert was viewing the images. Hence, for this thesis the decision was made to include the "think out loud" voice over of the reporting

clinician in the eye tracking videos of the training tool. This allowed the reporting radiographer to explain what they were looking at on the image, why and how they search the image for pathology. The implementation of guided eye tracking, tested over a longer time period and with the inclusion of some or all of the above mentioned high quality elements, has provided a robust study and is discussed in the chapter content below. The combination of these factors had the potential to further increase the improvements in competency, efficiency and sensitivity noted in the systematic review. The research also identified that there was no published evidence based search strategy available to reporting clinicians when interpreting or reporting images (Chapter 1). Consequently, the series of studies within this thesis were devised and executed with subsequent analysis of the results obtained.

6.2 Key findings

Chapter 2

The systematic review identified 13 relevant articles focusing on chest image interpretation learning by radiographers. The quality of evidence published in this area was high. The devised CASP tool was used to score studies in the systematic review. There was a mean of 7.5/10 scored. Studies scored six or seven out of ten when they failed to test the participants on a range of pathologies, include information on the reference standard to which participant's answers were compared against, were not transferable to everyday clinical practice. The reduced score was also applied when a study failed to refer to confounding factors of their investigation, including how this may have influenced the final scores of participants. There were limitations on the comparisons which could be made between studies because the accuracy measurements and training/education methods varied greatly. The review highlighted that performance increased following the implementation of all training methods. Accuracy was found to be highest in those studies featuring postgraduate programmes (Piper et al. 2014; Woznitza et al. 2014). The role of image interpretation differed between studies, ranging between: image red dot abnormality highlighting, image comment and clinical reporting. The red dot system describes when radiographers are tasked with annotating an image with the words 'red dot' where they believe a pathology to be present within the image. Preliminary clinical evaluation or 'image commenting' occurs when radiographers make a judgement based on the image appearance they encounter in radiographer clinical practice. An image comment is made in written form on the radiographer's judgement of the image however, the image will receive a full written report

by a reporting clinician following this. Clinical reporting involves the formation of a diagnosis and explanation of pathology.

Implications and recommendations

The amount of evidence available in this specific area is relatively small with only 13 articles identified in the review. Difficulties were posed in comparing the accuracy levels obtained by participants and the training implemented within the studies. With various analysis methods used for accuracy levels then comparisons between results must be made with caution. The evidence is limited in this specific area and so the training tested in the studies varied greatly. As radiographer chest image interpretation is expected to expand greatly over the next decade, a systematic review completed when more evidence becomes available would be beneficial. More comparisons of similar training types could be made and larger participant sample sizes will be available as the population of reporting radiographers in this area increases.

Chapter 3

A comparison of image interpretation skills of radiographers across a range of experience was completed using eye tracking technology. Reporting radiographers trained in MSK image interpretation demonstrated statistically significant accuracy rates ($p \leq 0.001$), and confidence levels ($p \leq 0.001$) and took a mean of 2.4 s longer to clinically decide on an image compared to students (as written in Chapter 3, Section 3.6.3, Table 3.4). Reporting radiographers also had a statistically greater accuracy rate ($p \leq 0.001$), were more confident ($p \leq 0.001$) and took longer to clinically decide (14 s on average) on an image diagnosis ($p = 0.02$) than radiographers. Reporting radiographers had a greater mean fixation duration ($p = 0.01$), mean fixation count ($p = 0.04$) and mean visit count ($p = 0.04$) within the areas of pathology compared to students. These figures were greater in the radiographer group compared to students and greater in the reporting radiographer group compared to radiographers, however these were not statistically significant. Eye tracking patterns presented within heat maps (as illustrated in Chapter 3, Section 3.6.6), were a good reflection of group expertise and search strategies. Eye gaze metrics such as time to first fixate, fixation count, fixation duration and visit count within the areas of pathology were indicative of the radiographers' and reporting radiographers' accuracy. The accuracy and confidence of students, radiographers and reporting radiographers could be reflected in the variability of their eye tracking heat maps. Participants' thoughts and decisions were quantified using the

eye tracking data. Eye tracking metrics also reflected the different search strategies that each group of participants adopted during their image interpretations.

Implications and recommendations

The computed eye gaze metrics in this study show that eye tracking could be used to automatically assess a radiographer's accuracy or to identify different levels of competencies, however further work is needed to provide additional evidence. A similar study with a larger sample size of participants, a larger number of images and a clinically relevant representation of normal and abnormal images would be recommended. The inclusion of these features and the eye tracking technology would make this a unique and substantial piece of work.

Chapter 4

A digital training tool for use in chest image interpretation was created based on evidence within the literature, using expert input and two search strategies previously used in clinical practice. Images and diagrams, aiding translation of the tool content, were incorporated where possible. The tool is structured to allow the chest image interpretation process to be clear, concise and methodical. A search strategy was incorporated within the tool to investigate its use. Eye tracking, a checklist and voice recordings were combined to form a multi-dimensional learning tool, which to date has not previously been used in chest image interpretation learning. The training tool for use in chest image interpretation learning has been designed, created and digitised.

Implications and recommendations

The first digital training tool incorporating a search strategy and eye tracking technology to interpret chest images has been developed. This tool has the potential to improve image interpretation and enhance the learning experience of image interpretation by reporting clinicians. The survey feedback highlighted the limitations and changes which could be made to the tool to improve its usability (Chapter 5). These included the need for chest images to be supplied with the tool to practice, graphics such as images, diagrams and videos was the feature marked lowest scoring 5.2/10 on the likert scale and a brief introduction to the tool was highlighted as a potential positive feature to be added to the tool.

Chapter 5

Improvements were seen in interpretation performance and confidence ($p<0.05$). There was a decrease in FP values and increase in TN values seen in the intervention group ($p<0.05$). The

tool may have led to the longer interpretation times and increased number of eye gaze metrics seen in the intervention group as they tried to follow the search strategy, whereas control group participants had a shorter interpretation time in the follow up study ($p<0.05$). Eye gaze metrics and decision time increased for some of the intervention groups, whereas these figures decreased for the control groups. This may imply, by following a standardised routine for every image, the tool and its content require the viewer to take additional time to apply in image interpretation until the viewer develops a routine for themselves.

Implications and recommendations

A novel training tool has been tested for its effect on user performance. Positive improvements in performance measures and confidence were found. This tool therefore has the potential to be used as a training tool in chest image interpretation for reporting clinicians and healthcare professionals. Recommendations made by users in the survey should be considered and changes made to improve the quality of the tool. Research can be expanded in this area by implementing the tool in other healthcare professionals (for example physiotherapists and nurses) who aim to improve their image interpretation skills. Currently such healthcare professionals may be asked to interpret patient images and yet have little formal training in this area at undergraduate level. This tool may provide a simple guide to users to ensure all areas of the image are being considered for pathology and may provide reassurance in completing a search of the entire image for an abnormality. Use of the tool and maintenance effect should be considered and investigated also. Use of the tool was relatively low and so attention and further research should be given to the continuous usability of the tool.

6.3 Achievement of aims and goals

The overall aim of this thesis was to investigate the image interpretation process of reporting clinicians and to develop and test a digital training package for chest image interpretation. In order to determine the effect of the digital training package on chest image interpretation, accuracy, confidence, eye tracking metrics and participant feedback were assessed.

The objectives of this research were:

(6) *To systematically review the literature in the area of reporting radiography and image interpretation education.*

This aim was achieved by completing a systematic review of the literature regarding chest image interpretation education and radiographers. Six large databases were searched, search keywords were formed and relevant articles were extracted. Outcome measures of accuracy, confidence, eye tracking, interpretation times and participant feedback were summarized. A trend in education methods and effects on performance were collated. A summary of the literature available on this topic was presented in Chapter 2.

(7) To investigate image interpretation performance of radiographers by computing eye gaze metrics using eye tracking technology.

An eye tracking study was set up to investigate trends in interpretation of radiographers. To compare trends of radiographers who have completed different training levels, radiography students, radiographers and reporting radiographers trained in completing MSK image interpretation were recruited to the study. Participants were asked to interpret both MSK and CXR images to identify whether the anatomical area in the image influenced how the participant completed the image interpretation. Eye gaze metrics were studied and used to identify similarities in interpretations of different groups. Accuracy as an outcome measure allowed performance levels to be demonstrated. It was identified that accuracy and confidence levels increased with experience and different eye tracking trends were seen in each participant group (Chapter 3).

(8) To develop a digital training package for use in chest image interpretation.

A digital training package was created for use in chest image interpretation training. Information collected from the systematic review, the results of the eye tracking study detailed in Chapter 3 and expert opinion were used to inform the development and formation of the training package. A search strategy was developed to ensure satisfaction of search on an image was achieved during interpretation, this was integrated into and became a central component of the training package. The digital training package that was produced is detailed in Chapter 4.

(9) To investigate the effect of the digital training package on reporting clinician performance and image interpretation learning.

The digital training package was implemented in the training and reporting practice of reporting radiographers. The package was shown to have an overall positive effect on reporting clinician performance. Improvements were observed in performance and confidence of the intervention group, improvements were seen across a range of parameters in the intervention group compared to the control group. Reporting time issues were highlighted as a possible inconvenience of the tool as the control group took a shorter time to interpret images in the follow up test ($p<0.05$) whereas intervention group participants took an average of 12.76 secs longer than this. Participant feedback identified benefits and flaws of the package (see Chapter 5 Section 5.8.1). The data analysis, results and discussion were presented to demonstrate the true influence of the package on performance in Chapter 5.

(10)   To propose an evidence based practice training package which will aid the learning process of chest image interpretation

The data collected in the trial of the digital training package and the survey encouraging participant feedback on the use of the training package, provided a comprehensive insight to the effect, benefits and weaknesses of the intervention. Benefits were observed in participant performance. Changes recommended by participants will be incorporated into the training package and so an evidence based practice training package has been developed to aid the learning process of chest image interpretation. The data analysis and survey results demonstrating the effect of the training package were demonstrated in Chapter 5.

6.4 Limitations of the research

6.4.1 Recruitment

Due to the nature of the research and its implementation in an overall small sample size of reporting radiographers, recruitment was difficult and relatively small numbers of

participants were achieved. Travel required for the recruitment was expensive and so the number of sites attended for recruitment was limited.

6.4.2 Time commitments of participants

The participants were either attending a postgraduate programme and/or working full time in busy clinical practice. Time allowing for the use of the intervention was limited and as a result many struggled to use the tool at all or to use it repetitively.

6.4.3 Image quality

The Tobii studio software did not permit the use of DICOM images. Images needed to be converted into an appropriate format to be used with the eye tracking technology and to be used by participants when present within the digital training tool. Images were converted into TIFF format, the highest quality radiology general use format of an image which could be used with the Tobii software to minimise the loss in image quality.

6.4.4 Image manipulation tools

Post processing or image manipulation tools were not available for use with the eye tracking technology. A few participants remarked on this directly following the eye tracking data collection and in the training tool evaluation questionnaire, suggesting the image viewing did not closely match the features available on a viewing or reporting work station. For the eye tracking technology to record data accurately the image could not be manipulated (pan, magnification or image rotation functions) during the image interpretation process, unfortunately this was one of the disadvantages of using the current software.

6.4.5 Eye tracking sampling quality

The eye tracker was calibrated successfully before each participant completed the study. Unfortunately, due to participant movement, glasses or contact lens use by individuals, there was a decrease in some of the eye tracking sampling qualities. The eye tracking sampling quality was controlled as much as possible by applying a standardised calibration strategy and by instructing participants to limit their movement during the eye tracking study as much as possible. More stringent inclusion criteria could limit these factors influencing the eye tracking data, such as excluding those with glasses/contacts or poor vision and should be considered in further studies. This was not ideal in this study as the sample size was small

and participants often required their glasses when spending periods of time looking at a computer screen.

6.5 Recommendations

1. Based on the evidence presented throughout Chapter 5, the proposed digital training tool will be recommended to those aiming to improve chest image interpretation skills as evidence based practice. The digital training tool can be used for Continuing Professional Development (CPD), for individuals to improve interpretation skills when the task forms part of their daily workload, to aid a formal training programme and to instill confidence in their ability to complete this task.

2. The time implications which occurred when applying this tool to practice could be minimal compared to an inaccurate diagnosis. The tool should be implemented in practice and the additional time allowed to construct a complete and thorough diagnosis. It is also anticipated the time implications are to be expected at the outset when beginning training, and that these time delays should decrease with experience of using the training tool. The time delays are minimal (10-20secs) and therefore outweigh the possibility and detriments of a pathology being missed. The tool might be most useful during initial training, until a routine is developed by the individual that has been informed by the content of the instrument. Perhaps further insight could then be gained to identify whether experience changes the approach to image interpretation, notwithstanding the fact that the viewing approach is non-standard and may affect users nonetheless.

3. The results of the training tool evaluation questionnaire have identified the need for changes to be made to the layout, design and content. These changes should be made and measured so that further feedback to maximise the potential successful future use of the tool is gained.

4. The developed education tool should enable a search strategy to be implemented in chest image interpretation to ensure the entire image is interrogated for pathology or abnormal appearances.

5. The training tool evaluation questionnaire results highlighted that most participants believed eye tracking could aid or improve image interpretation learning. This technology should be investigated further in this field of study to establish how this system of work may be applicable in general image reporting scenarios.

6.6 Future research

The data collected within both Chapter 3 and Chapter 5 can be analysed further and used to extract supplementary relational information to establish if other links are present or could be developed. The database has been prepared for analysis of all participant data from the studies presented in this thesis. A computer algorithm will be used with this data to investigate the possibility of predicting participant performance based on their eye tracking data. This analysis could be used to predict whether a participant is likely to be correct or incorrect in their diagnosis of an image interpretation following computer analysis of their eye gaze metrics and confidence levels. This data is currently being prepared to test the data using several computer algorithms to complete the analysis described above.

Due to the improvements seen in performance and confidence of participants receiving the training tool as an intervention, the content could be altered to focus on a range of anatomical parts other than for chest image interpretation. The tools formed could be tested and investigation of its use on areas other than chest image interpretation evaluated The tool could be used to aid the education of students who are not completing a reporting course, such as undergraduate diagnostic radiographers, and its effect investigated. As an alternative series of investigations, research should be completed on healthcare professionals other than radiographers to investigate whether the tool is transferable across a range of professions.

This thesis has identified that a search strategy training tool using eye tracking technology can make a positive difference to reporting clinician performance. Further work can be completed in this specific area. The use of eye tracking combined with real time computer learning could enhance the use of this technology and provide instant feedback to the user on their performance (Choy et al. 2018; Yates et al. 2018).

6.7 Dissemination of results

The findings from this work have been presented at conferences and scientific meetings locally, nationally and internationally (see Appendice 6.a). These events have included the UK Radiological and Radiation Oncology Congress, the European Congress of Radiology, the International Society of Radiographers and Radiological Technologists Congress, the Human and Computer Interaction conference, the Reporting Radiographers Interest Group Scotland study day and the Clinical Translational Research and Innovation Centre conference. A further presentation of the work is due to be completed at the Leading the Way: Radiographer Advanced Practice conference later this year in October. These presentations have been recorded in the Ulster University Institutional Repository. The systematic review of this thesis (Chapter 2) was published by the Radiography Journal in 2017 (McLaughlin et al. 2017b). The experimental eye tracking study (Chapter 3) was published in the International Journal of Medical Informatics in 2017 (McLaughlin et al. 2017a). The development of the training tool (Chapter 4) was also published in the Radiography Journal in 2017 (McLaughlin et al. 2018).

A bursary was awarded for the attendance to the UK Radiological and Radiation Oncology Congress in June 2017 where a presentation was given on the work described in Chapter 3. In October 2017 the work contained in the thesis was shortlisted as a finalist of the Belfast Health and Social Care Trust Science driving innovation in healthcare delivery award of the Advancing Healthcare Northern Ireland Awards 2017. The results of this work, in particular those from Chapter 5, will be presented and published further following the submission of this thesis.

6.8 Impact of research

The research was highly relevant and timely given the role progression within chest image reporting by radiographers. Recent publications have highlighted the reluctance of other professions to support this role development and this role is not yet widely accepted as a reporting radiographer scope of practice within Scotland and Northern Ireland (RCR 2012; Howard 2013; Milner et al. 2016; AuntMinnieEurope.com 2018). Limited knowledge exists regarding the most appropriate or useful training aids to assist this role of image interpretation. As this role continues to evolve and become more widespread it is important

that the standards of practice are monitored and the most efficient training aids are maximised in use.

The results of this study are highly important not only to develop training methods but may also impact directly or indirectly on other areas of radiographer reporting and the potential for role progression into different anatomical areas and imaging examinations/modalities. This research could impact on patient care, decrease waiting times for reports, decrease the time taken to diagnose a patient and hence appropriate treatment could be delivered sooner. This research offers a support and educational tool to users. Modifications to chest reporting education have been recommended which can be further applied to other anatomical areas. There is a low level of information available on chest reporting by radiographers given the recent role progression into this particular area of anatomy. For the first time, the present study has collated information in this area and the training methods employed to improve knowledge levels. This information and the proposed training method of an eye tracking and search strategy tool will add to the body of knowledge.

6.9 Conclusion

The systematic review has identified significant amounts of data relevant to the field of chest image interpretation learning by radiographers. The body of evidence was of an overall high quality with a mean score of 7.5/10 identified using the CASP tool. However, despite the strengths of the high rating studies, elements which improved the quality of the studies failed to be used in investigations other than those testing the effect of postgraduate education or in those seeking a full diagnosis from the participants. The most recent study (Semakula-Katende et al. 2016) featured 134 radiographers but only provided training and testing on the identification of tuberculosis. A recent study completed, which scored highly using the CASP tool, featured only one reporting radiographer (Woznitza et al. 2014). This review identifies the lack of robust, high quality evidence, outside that applied to the measurement of training in the form of a postgraduate education in this area.

The variability in interpretation processes of different groups of radiographers were demonstrated through eye gaze metrics, accuracy, confidence levels and heat maps. This knowledge provides a further understanding of how radiographers of a particular group of experience react when interpreting different images. This data combined with computer

learning could be used to predict performance levels, aid training and indicate when further education is required by a radiographer or other reporting clinician.

Significant improvements were observed in the intervention groups following implementation of the tool, TP and TN scores increased. Significant increases in confidence were also seen when compared to the control group. The tool has the potential to be a useful aid in chest image interpretation and may assist formal training in image interpretation also. The tool has the potential to instill confidence in users and reassurance that the entire image has been searched to exclude pathology.

An increase in interpretation time and eye gaze metrics were noted in intervention groups compared to control groups following access to the digital training tool compared with those who did not. The systematic search strategy may take longer to implement in image interpretation given these figures. Over time as experience is gained, it is envisaged participants will become accustomed to using this search strategy and there is an expectation that interpretation times will return to previous figures seen in the initial testing period. An overall delay of approximately 5-20 seconds and the correct diagnosis being formulated, greatly outweighs the risk of an area on the image not being interpreted or the incorrect diagnosis being provided.

The digital training tool is currently available via an internet connection. It can be easily accessed and is transferable across educational/clinical sites. Recommendations provided by participant feedback can greatly improve the quality and features of the tool to allow it to become more intuitive for the user. Features on the tool will be adjusted following user feedback. The images, videos and diagrams will be improved, chest images will be supplied with the tool to practice using the search strategy tool with, and a brief introduction will be supplied with the tool explaining how to use it as a stand alone system. Consequently the main areas highlighted in the feedback which could be improved will be addressed. A further study or survey to gain feedback on the new version of the tool would be beneficial.

The work has provided a digital training tool which can be used to enhance chest image interpretation skills. The results of this thesis have contributed to the field of knowledge

within this area and will influence the learning methods employed in chest image interpretation.

References

Anderson, B. and Shyu, C.R. (2011) Studying visual behaviors from multiple eye tracking features across levels of information representation. *AMIA Annual Symposium proceedings* 2011 72-79.

Andronikou S. and SemakulaKatende, N. (2014.) Crash course for improving non-radiologists interpretation of chest radiographs for suspected TB in children: A 'task-shifting' exercise. *Pediatric radiology,* 44 S315.

Arenson, R.L., Seshadri, S.B., Kundel, H.L., DeSimone, D., Van der Voorde, F., Gefter, W.B., Epstein, D.M., Miller, W.T., Aronchick, J.M. and Simson, M.B. (1988) Clinical evaluation of a medical image management system for chest images. *AJR.American Journal of Roentgenology,* 150 (1), 55-59.

Auffermann, W.F., Henry, T.S., Little, B.P., Tigges, S. and Tridandapani, S. (2015) Simulation for Teaching and Assessment of Nodule Perception on Chest Radiography in Nonradiology Health Care Trainees. *Journal of the American College of Radiology,* 12 (11), 1215-1222.

Bajre, M.K., Pennington, M., Woznitza, N., Beardmore, C., Radhakrishnan, M., Harris, R. and McCrone, P. (2017) Expanding the role of radiographers in reporting suspected lung cancer: A cost-effectiveness analysis using a decision tree model. *Radiography,* 23.

Berbaum, K., Franklin Jr, E., Caldwell, R. and Schartz, K. (2010) Satisfaction of search in traditional radiographic imaging. *The Handbook of Medical Image Perceptio and Techniques,* 107-138.

Berbaum, K.S., Franken, E.A., Dorfman, D.D., Caldwell, R.T. and Krupinski, E.A. (2000) Role of faulty decision making in the satisfaction of search effect in chest radiography. *Academic Radiology,* 7(12), 1098-1106. Available at: http://www.sciencedirect.com/science/article/pii/S107663320080063X

Bertram, R., Helle, L., Kaakinen, J.K. and Svedstrom, E. (2013) The effect of expertise on eye movement behaviour in medical image perception. *PloS one,* 8 (6), e66169.

Blake, A., Stapleton, G., Rodgers, P. and Howse, J. (2014) How should we use colour in euler diagrams? *In:* How should we use colour in euler diagrams? *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction.* ACM, 149.

Board of the Faculty of Clinical Radiology, (2015) *Standards for the Reporting and Interpretation of Imaging Investigations.* London: The Royal College of Radiologists. Available from: https://www.rcr.ac.uk/publication/standards-reporting-and-interpretation-imaging-investigations [Accessed 13/11/2015].

Bond, R.R., Finlay, D.D., Breen, C., Boyd, K., Nugent, C.D., Black, N.D., Macfarlane, P.W. and Guldenring, D., (2012) Eye tracking in the assessment of electrocardiogram interpretation techniques. *In: Computing in Cardiology (CinC), 2012.* IEEE, 581-584.

Bond, R., Zhu, T., Finlay, D., Drew, B., Kligfield, P., Guldenring, D., Breen, C., Gallagher, A., Daly, M. and Clifford, G. (2014) Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach. *Journal of Electrocardiology,* 47 (6), 895-906.

Bond, N. (1999) Optimization of image quality and patient exposure in chest radiography. *Radiography,* 5 (1), 29-31.

Bond, R.R., Kligfield, P.D., Zhu, T., Finlay, D.D., Drew, B., Guldenring, D., Breen, C., Clifford, G.D. and Wagner, G.S. (2015) Novel approach to documenting expert ECG interpretation using eye tracking technology: a historical and biographical representation of the late Dr Rory Childers in action. *Journal of Electrocardiology,* 48 (1), 43-44.

Brady, A., Laoide, R.Ó., McCarthy, P. and McDermott, R. (2012) Discrepancy and error in radiology: concepts, causes and consequences. *The Ulster Medical Journal,* 81 (1), 3.

Brealey S., Scally A., Hahn S., Thomas N., Godfrey C. and Crane, S. (2006) Accuracy of radiographers red dot or triage of accident and emergency radiographs in clinical practice: a systematic review. *Clinical Radiology,* 61 (7), 604-615.

Brealey, S., King, D. and Warnock, N. (2002) An assessment of different healthcare professionals' attitudes towards radiographers' reporting A&E films. *Radiography,* 8 (1), 27-34.

Brealey, S., King, D., Crowe, M., Crawshaw, I., Ford, L., Warnock, N., Mannion, R. and Ethell, S. (2014) Accident and emergency and general practitioner plain radiograph reporting by radiographers and radiologists: a quasi-randomized controlled trial. *The British Journal of Radiology,*

Brealey, S., King, D., Hahn, S., Godfrey, C., Crowe, M., Bloor, K., Crane, S. and Longsworth, D. (2014) The costs and effects of introducing selectively trained radiographers to an A&E reporting service: a retrospective controlled before and after study. *The British Journal of Radiology,*

Brealey, S. and Scally, A.J. (2008) Methodological approaches to evaluating the practice of radiographers' interpretation of images: A review. *Radiography,* 14 e46-e54.

Brealey, S. and Scally, A.J. (2001) Bias in plain film reading performance studies. *The British Journal of Radiology,* 74 (880), 307-316.

Brealey, S., Scally, A.J., Hahn, S. and Godfrey, C. (2014) Evidence of reference standard related bias in studies of plain radiograph reading performance: a meta-regression. *The British Journal of Radiology,*

Brealey, S. and Scuffham, P.A. (2014) The effect of introducing radiographer reporting on the availability of reports for Accident and Emergency and General Practitioner examinations: a time-series analysis. *The British Journal of Radiology,*

Brealey, S. (2001) Measuring the effects of image interpretation: an evaluative framework. *Clinical Radiology,* 56 (5), 341-347.

Brealey, S., Scally, A., Hahn, S., Thomas, N., Godfrey, C. and Crane, S. (2006) Accuracy of radiographers red dot or triage of accident and emergency radiographs in clinical practice: a systematic review. *Clinical Radiology,* 61 (7), 604-615.

Brealey, S., Scally, A.J. and Thomas, N.B. (2002) Methodological standards in radiographer plain film reading performance studies. *British Journal of Radiology,* 75 (890), 107-113.

Brealey, S., Scally, A.J. and Thomas, N.B. (2002) Presence of bias in radiographer plain film reading performance studies. *Radiography,* 8 (4), 203-210.

Breen, C.J., Bond, R. and Finlay, D. (2014) An evaluation of eye tracking technology in the assessment of 12 lead electrocardiography interpretation. *Journal of electrocardiology,* 47 (6), 922-929.

Brennan, P.C., McEntee, M., Evanoff, M., Phillips, P., O'Connor, W.T. and Manning, D.J. (2007) Ambient lighting: effect of illumination on soft-copy viewing of radiographs of the wrist. *American Journal of Roentgenology,* 188 (2), W177-W180.

British Broadcasting Corporation BBC (2017). Portsmouth hospital missed lung cancer cases [online]. Available at: http://www.bbc.co.uk/news/uk-england-hampshire-42178069 [Last accessed 25th January 2018].

Broeders, M.J., Ten Voorde, M., Veldkamp, W.J., van Engen, R.E., van Landsveld-Verhoeven, C., 't Jong-Gunneman, M.N., de Win, J., Greve, K.D., Paap, E. and den Heeten, G.J. (2015) Comparison of a flexible versus a rigid breast compression paddle: pain experience, projected breast area, radiation dose and technical image quality. *European radiology,* 25 (3), 821-829.

Brooke, J., 1996. SUS-A quick and dirty usability scale.*Usability evaluation in industry*, *189*(194), pp.4-7.

Brown S.E., MacAnovic M. and Williams, M.P. (2012) Oxygen reservoir bags simulating chest pathology: A case series. *Journal of Emergency Medicine,* 43 (6), 1045-1048.

Burke, S., Patel, K., Amin, S., Grayson, K. and Woznitza, N. (2013) Chest X-Ray Interpretation: Agreement Between Consultant Radiologists And A Reporting Radiographer In Clinical Practice In The United Kingdom. *Am J Respir Crit Care Med,* 187 A2229.

Burton, K.R., Perlis, N., Aviv, R.I., Moody, A.R., Kapral, M.K., Krahn, M.D. and Laupacis, A. (2014) Systematic review, critical appraisal, and analysis of the quality of economic evaluations in stroke imaging. *Stroke,* 45(3), 807-814

Buskov, L., Abild, A., Christensen, A., Holm, O., Hansen, C. and Christensen, H. (2013) Radiographers and trainee radiologists reporting accident radiographs: A comparative plain film-reading performance study. *Clinical Radiology,* 68 (1), 55-58.

Caffery, L.J., Armfield, N.R. and Smith, A.C. (2015) Radiological interpretation of images displayed on tablet computers: A systematic review. *British Journal of Radiology,* 88 (1050),

Cameron H.D. and Stafrace, S. (2012.) Assessment of neonatal chest radiograph quality; a comparison between three health professional groups. *Pediatric Radiology,* 42 S473.

Canterbury Christ Church University (2002), *PgC clinical reporting (adult chest) validation document.*

Canterbury Christ Church University (2017), *MSc Postgraduate Diploma, Postgraduate Certificate Clinical Reporting,* England: Canterbury Christ Church University. Available at: https://www.canterbury.ac.uk/study-here/courses/postgraduate/clinical-reporting.aspx [Accessed 3rd March 2017].

Carmody, D.P., Kundel, H.L. and Toto, L.C. (1984) Comparison Scans While Reading Chest Images Taught, But Not Practiced. *Investigative Radiology,* 19(5), 462-466.

Carter, S. and Manning, D. (1999) Performance monitoring during postgraduate radiography training in reporting—a case study. *Radiography,* 5 (2), 71-78.

Cassidy, S., Eyres, R., Thomas, N., Whitehouse, R. and Williams, P. (1999) A comparative study of the reporting accuracy of clinical radiographers and radiology registrars. *The British Journal of Radiology,* 72 19.

Chen, Y., James, J.J., Turnbull, A.E. and Gale, A.G. (2015) The use of lower resolution viewing devices for mammographic interpretation: implications for education and training. *European Radiology,* 25 (10), 3003-3008.

Chetwood, A.S., Kwok, K., Sun, L., Mylonas, G.P., Clark, J., Darzi, A. and Yang, G. (2012) Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical Endoscopy,* 26 (7), 2003-2009.

Clark, T.J., Flood, T.F., Maximin, S.T. and Sachs, P.B. (2015) Lung CT Screening Reporting and Data System Speed and Accuracy Are Increased With the Use of a Semiautomated Computer Application. *Journal of the American College of Radiology,* 12(12), 1301-1306.

Coleman, L. and Piper, K. (2009) Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers. *Radiography,* 15 (3), 196-202.

Cowan, I., Smith, T., Nakabea, P., Ajibulu, O., Hennessy, O. and WHO/RANZCR/ISRRT (2007) Developing the image interpretation skills of South Pacific radiographers: a joint WHO/RANZCR/ISRRT project. *Australasian Radiology,* 51 (6), 527-531.

Cowling, C. (2008) A global overview of the changing roles of radiographers. *Radiography,* 14, Supplement 1 e28-e32.

Crichton, M.C. (1987) Aberdeen radiographer and a Mediterranean emergency. *Radiography,* 53 (608), 71-73.

Cullinan, J.E. (1981) A 'perfect' chest radiograph--or a compromise? the John B. Cahoon Memorial Lecture. *Radiologic technology,* 53 (2), 121-137 17p.

Cullinan, J.E.,Jr (1960) The role of the x-ray technician in the cardiopulmonary laboratory. *The X-ray technician,* 31 623-7 passim.

Czuczman, G.J., Pomerantz, S.R., Alkasab, T.K. and Huang, A.J. (2013) Using a Web-Based Image Quality Assurance Reporting System to Improve Image Quality. *American Journal of Roentgenology,* 201 (2), 361-368.

Del Ciello, A., Franchi, P., Contegiacomo, A., Cicchetti, G., Bonomo, L., & Larici, A. R. (2017). Missed lung cancer: when, where, and why? *Diagnostic and Interventional Radiology*, *23*(2), 118–126. http://doi.org/10.5152/dir.2016.16187

Department of Health. (2017) *Ionising Radiation (Medical Exposure) Regulations 2000.* London: The Stationery Office. Available at: https://www.gov.uk/government/publications/the-ionising-radiation-medical-exposure-regulations-2000 [Accessed 8th March 2017].

Diaz, I., Schmidt, S., Verdun, F.R. and Bochud, F.O. (2015) Eye-tracking of nodule detection in lung CT volumetric data. *Medical physics,* 42 (6), 2925-2932.

Donovan, T., Manning, D.J. and Crawford, T., (2008) Performance changes in lung nodule detection following perceptual feedback of eye movements. *In: Medical Imaging.* International Society for Optics and Photonics, 691703-691703-9.

Donovan, T., Manning, D.J., Phillips, P.W., Higham, S. and Crawford, T., (2005) The effect of feedback on performance in a fracture detection task. *In: Medical Imaging.* International Society for Optics and Photonics, 79-85.

Donovan, T. and Litchfield, D. (2013) Looking for Cancer: Expertise Related Differences in Searching and Decision Making. *Applied Cognitive Psychology,* 27 (1), 43-49.

Donovan, T. and Manning, D.J. (2006) Successful reporting by non-medical practitioners such as radiographers, will always be task-specific and limited in scope. *Radiography,* 12 (1), 7-12.

Ekpo E.U., Egbe N.O. and Akpan, B.E. (2015) Radiographers' performance in chest X-ray interpretation: The Nigerian experience. *British Journal of Radiology,* 88 (1051).

European Society of Radiology (ESR (2011) Undergraduate education in radiology. A white paper by the European Society of Radiology. *Insights into Imaging,* 2 (4), 363-374.

Flehinger, B.J., Melamed, M.R., Heelan, R.T., McGinnis, C.M., Zaman, M.B. and Martini, N. (1978) Accuracy of chest film screening by technologists in the New York early lung cancer detection program. *AJR.American Journal of Roentgenology,* 131 (4), 593-597.

Gant, D. (2007) Image conscious: establishing a scheme for assessing chest x-rays. *Synergy: Imaging & Therapy Practice,* 10-12 3p.

Gegenfurtner, A. and Seppänen, M. (2013) Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education,* 63 393-403.

Gutmark, R., Halsted, M.J., Perry, L. and Gold, G. (2007) Use of computer databases to reduce radiograph reading errors. *Journal of the American College of Radiology,* 4(1), 65-68.

Hardy, M., Flintham, K., Snaith, B. and Lewis, E.F. (2016) The impact of image test bank construction on radiographic interpretation outcomes: A comparison study. *Radiography,* 22 (2), 166-170.

Hardy, M., Hutton, J. and Snaith, B. (2013) Is a radiographer led immediate reporting service for emergency department referrals a cost effective initiative? *Radiography,* 19 (1), 23-27.

Hardy, M. and Snaith, B. (2011) The impact of radiographer immediate reporting on patient outcomes and service delivery within the emergency department: Designing a randomised controlled trial. *Radiography,* 17 (4), 275-279.

Hardy, M., Snaith, B. and Smith, T. (2008) Radiographer reporting of trauma images: United Kingdom experience and the implications for evolving international practice. *RADIOGRAPHER-EAST MELBOURNE-,* 55 (1), 16.

Hardy, M. and Snaith, B. (2011) The impact of radiographer immediate reporting on patient outcomes and service delivery within the emergency department: Designing a randomised controlled trial. *Radiography,* 17 (4), 275-279.

Health Education England (The Society and College of Radiographers) (2010) Interpretation of Radiological Images [eLearning for Healthcare], [online]. Available at: http://www.e-lfh.org.uk/programmes/image-interpretation/ [Accessed 27th July 2016].

Herzer, K.R., Rodriguez-Paz, J.M., Doyle, P.A., Flint, P.W., Feller-Kopman, D.J., Herman, J., Bristow, R.E., Cover, R., Pronovost, P.J. and Mark, L.J. (2009) A practical framework for patient care teams to prospectively identify and mitigate clinical hazards. *The Joint Commission Journal on Quality and Patient Safety,* 35(2), 72-81.

Hlongwane, S.T. and Pitcher, R.D. (2013) Accuracy of after-hour 'red dot' trauma radiograph triage by radiographers in a South African regional hospital. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde,* 103 (9), 638-640.

Holmqvist, K., Nyström, M. and Mulvey, F., (2012) Eye tracker data quality: what it is and how to measure it. *In: Proceedings of the symposium on eye tracking research and applications.* ACM, 45-52.

Howard, M.L. (2013) An exploratory study of radiographer's perceptions of radiographer commenting on musculo-skeletal trauma images in rural community based hospitals. *Radiography,* 19 (2), 137-141.

Hughes H., Hughes K. and Hamill, R. (1996) A study to evaluate the introduction of a pattern recognition technique for chest radiographs by radiographers. *Radiography,* 2 (4), 263-288.

Illeris, K. (2009) *Contemporary theories of learning: learning theorists... in their own words.* Routledge.

Jacob, J., Paul, L., Hedges, W., Hutchison, P., Cameron, E., Matthews, D., Whiten, S. and Driscoll, P. (2016) Undergraduate radiology teaching in a UK medical school: a systematic evaluation of current practice. *Clinical radiology,* 71 (5), 476-483.

Jones T., Booker M., Hobbins S. and Dawkins, D. (2009) Documentation of radiographic findings by non-radiologists - An audit. *European Journal of Radiography,* 1 (2), 48-51.

Jones, H. and Manning, D. (2008) A survey to assess audit mechanisms practised by skeletal reporting radiographers. *Radiography,* 14 (3), 201-205.

Jones, H. and Robinson, L. (2008) A reflection on the role of an emergency care Consultant Radiographer in achieving appropriate imaging conditions for nasogastric tube positioning. *Radiography,* 14 e11-e14.

Judson, E. and Nightingale, J. (2009) An evaluation of radiographer performed and interpreted barium swallows and meals. *Clinical radiology,* 64 (8), 807-814.

Kelly, B.S., Rainford, L.A., Gray, J. and McEntee, M.F., 2012. Collaboration between radiological technologists (radiographers) and junior doctors during image interpretation improves the accuracy of diagnostic decisions. *Radiography*, *18*(2), pp.90-95.

Kok, E.M., de Bruin, A.B., Leppink, J., van Merriënboer, J.J. and Robben, S.G. (2015) Case Comparisons: An Efficient Way of Learning Radiology. *Academic Radiology,* 22 (10), 1226-1235.

Kok, E.M., Jarodzka, H., de Bruin, A.B.H., BinAmir, H.A.N., Robben, S.G.F. and van Merriënboer, J.J.G. (2015) Systematic viewing in radiology: seeing more, missing less? *Advances in Health Sciences Education,* 189-205.

Kramer, H.S. and Drews, F.A. (2016) Checking the lists: A systematic review of electronic checklist use in health care. *Journal of Biomedical Informatics,* Available from: http://www.sciencedirect.com/science/article/pii/S1532046416301186

Krupinski, E.A. (2010) Current perspectives in medical image perception. *Attention, Perception, & Psychophysics,* 72 (5), 1205-1217.

Krupinski, E.A. (2000) The importance of perception research in medical imaging. *Radiation Medicine,* 18 (6), 329-334.

Krupinski, E.A., Tillack, A.A., Richter, L., Henderson, J.T., Bhattacharyya, A.K., Scott, K.M., Graham, A.R., Descour, M.R., Davis, J.R. and Weinstein, R.S. (2006) Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human pathology,* 37 (12), 1543-1556.

Krupinski, E. and Borah, J. (2006) Eye tracking helps improve accuracy in radiology. *Biophotonics International,* 13 (6), 44-49.

Krupinski, E.A. (2009) Medical grade vs off-the-shelf color displays: Influence on observer performance and visual search. *Journal of Digital Imaging,* 22 (4), 363-368.

Kumar, S., Pandey, A.K., Sharma, P., Shamim, S.A., Malhotra, A. and Kumar, R. (2012) Instantaneous exposure to nuclear medicine staff involved in PET-CT imaging in developing countries: experience from a tertiary care centre in India. *Japanese journal of radiology,* 30 (4), 291-295.

Kundel, H.L., (2004) Reader error, object recognition, and visual search. *In: Medical Imaging 2004.* International Society for Optics and Photonics, 1-11.

Kundel, H.L. and La Follette Jr, Paul S (1972) Visual search patterns and experience with radiological images 1. *Radiology,* 103 (3), 523-528.

Kundel, H.L. and Nodine, C.F. (1975) Interpreting Chest Radiographs without Visual Search 1. *Radiology,* 116 (3), 527-532.

Kundel, H.L., Nodine, C.F. and Krupinski, E.A. (1990) Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. *Investigative radiology,* 25 (8), 890-896.

Kundel, H.L., Nodine, C.F., Krupinski, E.A. and Mello-Thoms, C. (2008) Using Gaze-tracking Data and Mixture Distribution Analysis to Support a Holistic Model for the Detection of Cancers on Mammograms. *Academic Radiology,* 15 (7), 881-886.

Lancaster, A. and Hardy, M. (2012) An investigation into the opportunities and barriers to participation in a radiographer comment scheme, in a multi-centre NHS trust. *Radiography,* 18 (2), 105-108.

Landis, J., & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Langen H.J., KohlhauserVollmuth C., Muras S., Stenzel M. and Beer, M. (2009) Training program for radiologic technologists for performing chest X-rays at inspiration in uncooperative children. *RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin,* 181 (3), 237-241.

Learning Rx. (2017) *Types of Learning Styles.* Available at: http://www.learningrx.com/types-of-learning-styles-faq.htm [Accessed 9th March 2017].
Lin, H., Yang, X. and Wang, W. (2008) A Content-Boosted Collaborative Filtering Algorithm for Personalized Training in Interpretation of Radiological Imaging. *Journal of Digital Imaging,* 27(4), 449-456.

Lison, T., Günther, S., Ogurol, Y., Pretschner, D.P. and Wischnesky, M. (2004) VISION2003: Virtual learning units for medical training and education. *International journal of medical informatics,* 73 (2), 165-172.

Litchfield D., Ball L.J., Donovan T., Manning D.J. and Crawford, T. (2010) Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied,* 16 (3), 251-262.

Litchfield, D., Ball, L.J., Donovan, T., Manning, D.J. and Crawford, T., (2008) Learning from others: Effects of viewing another person's eye movements while searching for chest nodules. *In: Medical imaging.* International Society for Optics and Photonics, 691715-691715-9.

Loughran, C. (1994) Reporting of fracture radiographs by radiographers: the impact of a training programme. *The British journal of radiology,* 67 (802), 945-950.

Mackay, S. (2014) The impact of a short course of study on the performance of radiographers when highlighting fractures on trauma radiographs:"The Red Dot System". *The British Journal of Radiology,*

Mallett, S., Phillips, P., Fanshawe, T.R., Helbren, E., Boone, D., Gale, A., Taylor, S.A., Manning, D., Altman, D.G. and Halligan, S. (2014) Tracking Eye Gaze during Interpretation of Endoluminal Three-dimensional CT Colonography: Visual Perception of Experienced and Inexperienced Readers. *Radiology,* 273 (3), 783-792.

Manning D., BarkerMill S.C., Donovan T. and Crawford, T. (2006) Time-dependent observer errors in pulmonary nodule detection. *British Journal of Radiology,* 79 (940), 342-346.

Manning, D., Ethell, S., Donovan, T. and Crawford, T. (2006) How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography,* 12 (2), 134-142.

Manning, D., Ethell, S. and Crawford, T., (2003) An eye-tracking AFROC study of the influence of experience and training on chest X-ray interpretation. *In: Proceedings of SPIE - The International Society for Optical Engineering.* 257-266. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0042358956&doi=10.1117%2f12.479985&partnerID=40&md5=38d189ed45fe39aa049c50bba4124d0d [Accessed 20 March 2017].

Manning, D., Ethell, S. and Donovan, T., (2004) Categories of observer error from eye-tracking and AFROC data. *In: Progress in Biomedical Optics and Imaging - Proceedings of SPIE.* 90-99. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-12144274406&doi=10.1117%2f12.533489&partnerID=40&md5=97d58df0d185c7b934b52ecb6c70b8b9 [Accessed 20 March 2017].

Manning, D., Leach, J. and Bunting, S. (2000) A comparison of expert and novice performance in the detection of simulated pulmonary nodules. *Radiography,* 6 (2), 111-116.

Manning, D.J., Bunting, S. and Leach, J. (1999) An ROC evaluation of six systems for chest radiography. *Radiography,* 5 (4), 201-209 9p.

Manning, D.J. and Leach, J. (2002) Perceptual and signal detection factors in radiography. *Ergonomics,* 45 (15), 1103-1116 14p.

Matsumoto, H., Terao, Y., Yugeta, A., Fukuda, H., Emoto, M., Furubayashi, T., Okano, T., Hanajima, R. and Ugawa, Y. (2011) Where do neurologists look when viewing brain CT images? an eye-tracking study involving stroke cases. *PLoS one,* 6 (12),

Mc Entee, M.F. and Dunnion, S. (2009) A FROC analysis of radiographers performance in identification of distal radial fractures. *European Journal of Radiography,* 1 (3), 90-94.

McEvoy, F.J., Shen, N.W., Nielsen, D.H., Buelund, L.E. and Holm, P. (2016) Online Radiology Reporting with Peer Review as a Learning and Feedback Tool in Radiology; Implementation, Validity, and Student Impressions. *Journal of Digital Imaging*, 1-8.

McConnell, J., Devaney, C. and Gordon, M. (2013) Queensland radiographer clinical descriptions of adult appendicular musculo-skeletal trauma following a condensed education programme. *Radiography,* 19 (1), 48-55.

McConnell, J. (2009.) The diaphragm: Some simple rules of how to recognise the features and conditions that affect and interact with it. *Journal of Medical Imaging and Radiation Oncology,* 53 A188.

McConnell, J.R. and Webster, A.J. (2000) Improving radiographer highlighting of trauma films in the accident and emergency department with a short course of study--an evaluation. *The British journal of radiology,* 73 (870), 608-612.

McConnell, J., Devaney, C., Gordon, M., Goodwin, M., Strahan, R. and Baird, M. (2012) The impact of a pilot education programme on Queensland radiographer abnormality description of adult appendicular musculo-skeletal trauma. *Radiography,* 18 (3), 184-190.

McEntee M., Frawley H. and Brennan, P.C. (2007) A comparison of low contrast performance for amorphous Silicon/caesium iodide direct radiography with a computed radiography: A contrast detail phantom study. *Radiography,* 13 (2), 89-94.

McEntee M.F., Brennan P.C. and Connor, G.O. (2004) The effect of X-ray tube potential on the image quality of PA chest radiographs when using digital image acquisition devices. *Radiography,* 10 (4), 287-292.

McEntee M.F., Houssein N. and Alazawi, D. (2014) Use of incentive spirometry in portable chest radiography. *Radiologic technology,* 85 (3), 261-269.

McEntee, M.F. and Bergin, N., (2010) The effect of image interpretation training on the fracture recognition performance of radiographers. *In: SPIE Medical Imaging.* International Society for Optics and Photonics, 762712-762712-11.

McEntee, M.F., Ryan, J., Evanoff, M.G., Keeling, A., Chakraborty, D., Manning, D. and Brennan, P.C., (2007) Ambient lighting: setting international standards for the viewing of softcopy chest images. *In: Medical Imaging.* International Society for Optics and Photonics, 65150M-65150M-10.

McEvoy, F.J., Shen, N.W., Nielsen, D.H., Buelund, L.E. and Holm, P. (2016) Online Radiology Reporting with Peer Review as a Learning and Feedback Tool in Radiology; Implementation, Validity, and Student Impressions. *Journal of Digital Imaging,* 1-8.

McLaughlin, L., Bond, R., Hughes, C., McConnell, J. and McFadden, S. (2017a) Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation. *The International Journal of Medical Informatics,* 11-21. Available from: http://www.sciencedirect.com/science/article/pii/S1386505617300540

McLaughlin, L., McConnell, J., McFadden, S., Bond, R. and Hughes, C. (2017b) Methods employed for chest radiograph interpretation education for radiographers: A systematic review of the literature. *Radiograph,* 350-357. Available from: http://www.sciencedirect.com/science/article/pii/S1078817417301281

McLaughlin, L., Woznitza, N., Cairns, A., McFadden, S.L., Bond, R., Hughes, C.M., Elsayed, A., Finlay, D. & McConnell, J. 2018, Digital training platform for interpreting radiographic images of the chest. *Radiography*, 159-164.

McNamara, K. (2006) Observer variation in chest radiograph interpretation. *Synergy: Imaging & Therapy Practice,* 8-11 4p.

Mehrotra, P., Bosemani, V. and Cox, J. (2009) Do radiologists still need to report chest x rays? *Postgraduate medical journal,* 85 (1005), 339-341.

Milner R.C., Culpan G. and Snaith, B. (2016) Radiographer reporting in the UK: Is the current scope of practice limiting plain-film reporting capacity?. *British Journal of Radiology,* 89 (1065)

Neep, M.J., Steffens, T., Owen, R. and McPhail, S.M. (2014) A survey of radiographers' confidence and self-perceived accuracy in frontline image interpretation and their continuing educational preferences. *Journal of Medical Radiation Sciences,* 61 (2), 69-77.

Neuman MI, Lee EY, Bixby S, Diperna S, Hellinger J, Markowitz R, Servaes S, Monuteaux MC, Shah SS. Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. Journal of Hospital Medicine. 2012 Apr;7(4):294-8.

Nocum D.J., Brennan P.C., Huang R.T. and Reed, W.M. (2013) The effect of abnormality-prevalence expectation on naive observer performance and visual search. *Radiography,* 19 (3), 196-199.

Nunn, H. and Nunn, D.L. (2011) Determination of difficult concepts in the interpretation of musculoskeletal radiographs using a web-based learning/teaching tool. *Radiography,* 17 (4), 311-318.

Ohkado A., Luna P., Querri A., Mercader M., Yoshimatsu S., Coprada L., Banares R., Garfin A.M. and Date, T. (2015) Impact of a training course on the quality of chest radiography to diagnose pulmonary tuberculosis. *Public Health Action,* 5 (1), 83-88.

Paterson, A. (2010) *Medical Image Interpretation by Radiographers: Definitive Guidance. The Society and College of Radiographers.* Available from: http://www.sor.org/learning/document-library/medical-image-interpretation-radiographers-definitive-guidance

Paterson, A.M., Price, R.C., Thomas, A. and Nuttall, L. (2004) Reporting by radiographers: a policy and practice guide. *Radiography,* 10 (3), 205-212.

Piper K., Cox S., Paterson A., Thomas A., Thomas N., Jeyagopal N. and Woznitza, N. (2014) Chest reporting by radiographers: Findings of an accredited postgraduate programme. *Radiography,* 20 (2), 94-99.

Piper, K., Paterson, A. and Godfrey, R. (2005) Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases. *Radiography,* 11 (1), 27-34.

Pook, J. (2015) The role of radiography in the diagnosis of pulmonary tuberculosis in the developing world. *Journal of Medical Radiation Sciences.Conference: 2015 NZIMRTAIR-Scientific Meeting "The Cloud: Shaping our Future".New Zealand.Conference Start: 20150724.Conference End: 20150726,* 62 74.

Potchen, E.J., Cooper, T.G., Sierra, A.E., Aben, G.R., Potchen, M.J., Potter, M.G. and Siebert, J.E. (2000) Measuring Performance in Chest Radiography 1. *Radiology,* 217 (2), 456-459.

Price R.C., Miller L.R. and Mellor, F. (2002) Longitudinal changes in extended roles in radiography. *Radiography,* 8 (4), 223-234.

Price, R. and Le Masurier, S. (2007) Longitudinal changes in extended roles in radiography: a new perspective. *Radiography,* 13 (1), 18-29.

Price, R.C. (2001) Radiographer reporting: origins, demise and revival of plain film reporting. *Radiography,* 7 (2), 105-117.

Pusic, M.V., Andrews, J.S., Kessler, D.O., Teng, D.C., Pecaric, M.R., Ruzal-Shapiro, C. and Boutis, K. (2012) Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation. *Medical education,* 46 (3), 289-298.

Radiology Masterclass. (2017) *Chest X-ray Systematic Approach.* Available at: http://www.radiologymasterclass.co.uk/tutorials/chest/chest_system/chest_system_start [Accessed 9th March 2017].

Robinson, P.J., Wilson, D., Coral, A., Murphy, A. and Verow, P. (1999) Variation between experienced observers in the interpretation of accident and emergency radiographs. *The British journal of radiology,* 72 (856), 323-330.

Rudd, P. (2003) The development of radiographer reporting 1965–1999. *Radiography,* 9 (1), 7-12.

Ryan, J.T., Haygood, T.M., Yamal, J., Evanoff, M., O'Sullivan, P., McEntee, M. and Brennan, P.C. (2011) The "memory effect" for repeated radiologic observations. *American Journal of Roentgenology,* 197 (6), W985-W991..

Schlorhaufer, C., Behrends, M., Diekhaus, G., Keberle, M. and Weidemann, J. (2012) Implementation of a web-based, interactive polytrauma tutorial in computed tomography for radiology residents: How we do it. *European Journal of Radiology,* 81(12), 3942-3946.

Semakula Katende N.S., Andronikou S. and Lucas, S. (2016) Digital platform for improving non-radiologists' and radiologists' interpretation of chest radiographs for suspected tuberculosis - a method for supporting task-shifting in developing countries. *Pediatric radiology,* 46 (10), 1384-1391.

Smith S. and Reeves, P. (2009) The extension of the role of the diagnostic radiographer in the UK National Health Service over the period 1995-2009. *European Journal of Radiography,* 1 (4), 108-114.

Smith T. and Fisher, K. (2011) Self-reported competency and continuing education needs of limited licence remote X-ray operators in New South Wales, Australia. *Rural and remote health,* 11 1560.

Snaith, B., Hardy, M. and Lewis, E.F. (2015) Radiographer reporting in the UK: a longitudinal analysis. *Radiography,* 21 (2), 119-123.

Snaith B. and Flintham, K. (2015) Radiology responsibilities post NPSA guidelines for nasogastric tube insertion: A single centre review. *Radiography,* 21(1), 11-15.

Snaith, B. and Hardy, M. (2008) Radiographer abnormality detection schemes in the trauma environment—An assessment of current practice. *Radiography,* 14(4), 277-281.

Society of Radiographers (2007) *Radiographer reporting of trauma images: evaluation of practice in the United Kingdom.* London, UK: Society of Radiographers. Available from: http://www.sor.org/system/files/article/201109/radiographer_reporting020707.pdf [Accessed 06/01/2016].

Society of Radiographers (2004) *Annual Report: Campaigning together growing together serving together.* London, UK: Society of Radiographers.(1). Available from: http://www.sor.org/system/files/article/201109/SoR-REPORT-04.pdf [Accessed 06/01/2016].

Sonnex, E., Tasker, A. and Coulden, R. (2001) The role of preliminary interpretation of chest radiographs by radiographers in the management of acute medical problems within a cardiothoracic centre. *The British Journal of Radiology,* 74(879), 230-233.

Soultanis, K.C., Tsiavos, K., Grivas, T.B., Stavropoulos, N.A., Sakellariou, V.I., Mavrogenis, A.F. and Papagelopoulos, P.J. (2015) Reliability study for the Rib Index in chest radiographs of a control group. *Scoliosis,* 10 (Suppl 2), S9-7161-10-S2-S9. eCollection 2015.

Stephenson, P., Hannah, A., Jones, H., Edwards, R., Harrington, K., Baker, S., Fitzgerald, N. and Belfield, J. (2012) An evidence based protocol for peer review of radiographer musculoskeletal plain film reporting. *Radiography,* 18 (3), 172-178.

Subesinghe, M., Goldstone, A.R., Patel, C.N., Chowdhury, F.U. and Scarsbrook, A.F. (2015) Design and implementation of a web-based PET-CT reporting assessment and e-portfolio tool. *Clinical Radiology,* 70(2), 123-127.

Tamaklo, E., (2012) Can e-learning improve medical student's ability to interpret chest x-rays in comparison with electronic text? Poster session presented at: ASME Annual Scientific Meeting 18-20.

The Irish Times (2017). Woman paid damages for delay in breast cancer diagnosis in Kerry [online]. Available at: https://www.irishtimes.com/news/health/woman-paid-damages-for-delay-in-breast-cancer-diagnosis-in-kerry-1.3337120 [Last accessed 25th January 2018].

The Society and College of Radiographers. (2010). Medical image interpretation of radiographers definitive guidance, [online]. Available at: https://www.sor.org/system/files/document-library/public/sor_Definitive_Guidance_May_2010.pdf [Accessed 25th January 2018].

The Journal 2017. Doctor who misdiagnosed seven serious cases at Kerry hospital was not a permanent staff member http://www.thejournal.ie/kerry-hospital-misdiagnosis-3758691-Dec2017/ [online]. Available at: [Last accessed 25th January 2018].

The Royal College of Radiographers, (2018). Policy, Public and Media [online]. https://www.rcr.ac.uk/public-and-media/common-medical-terms/radiology-faqs Available at: [Last accessed 25th January 2018].

The Society and College of Radiographers (2018). Society responds to radiologist who says 'Radiographers must not read chest x-rays' [online]. Available at: https://www.sor.org/news/society-responds-radiologist-who-says-radiographers-must-not-read-chest-x-rays

The Society and College of Radiographers, Sean Kelly (2018). Reporting [online]. Available at: https://www.sor.org/practice/reporting [Last accessed: 25th January 2018].

The Royal College of Radiologists, (2011). Standards for the Reporting and Interpretation of Imaging Investigations [Clinical Radiology], [online]. Available at: https://www.rcr.ac.uk/publication/standards-reporting-and-interpretation-imaging-investigations [Accessed 14th April 2016].

The Royal College of Radiologists. (2016) *Speciality training curriculum for clinical radiology.* London: Clinical Radiology The Royal College of Radiologists. Available at: https://www.rcr.ac.uk/clinical-radiology/specialty-training/radiology-curriculum [Accessed 8th March 2017].

Tobii Pro, (2016) Envision human behaviour, [online]. Available: http://www.tobiipro.com/ [19/09/2016].

Tosh N. and Mander, G. (2014.) Radiographer commenting in emergency with no formal training: Results of a blind trial in an Australian regional teaching hospital. *Journal of Medical Imaging and Radiology Oncology,* 58 152.

Turkington PM, Kennan N, Greenstone MA (2002) Misinterpretation of the chest x ray as a factor in the delayed diagnosis of lung cancer, Postgraduate Medical Journal;**78:**158-160. Weyant, E. (2014) Chest X-Ray Resources. *Journal of Electronic Resources in Medical Libraries,* 11(3), 159-164.

Waitt, C.J., Joekes, E.C., Jesudason, N., Waitt, P.I., Goodson, P., Likumbo, G., Kampondeni, S., Faragher, E.B. and Squire, S.B. (2013) The effect of a tuberculosis chest X-ray image reference set on non-expert reader performance. *European radiology,* 23 (9), 2459-2468.

Wang, K.C., Jeanmenne, A., Weber, G.M., Thawait, S. and Carrino, J.A. (2011) An online evidence-based decision support system for distinguishing benign from malignant vertebral compression fractures by magnetic resonance imaging feature analysis. *Journal of Digital Imaging,* 24 (3), 507-515.

Wang, X.H., Durick, J.E., Lu, A., Herbert, D.L., Golla, S.K., Foley, K., Piracha, C.S., Shinde, D.D., Shindel, B.E. and Fuhrman, C.R. (2008) Characterization of radiologists' search strategies for lung nodule detection: slice-based versus volumetric displays. *Journal of Digital Imaging,* 21 (1), 39-49.

Williams, I.J. (2013) Appendicular skeleton: ABCs Image Interpretation Search strategy. *The South African Radiographer,* 51(2), 9-14.
Wong, V., Smith, A.J., Hawkins, N.J., Kumar, R.K., Young, N., Kyaw, M. and Velan, G.M. (2015) Adaptive Tutorials Versus Web-Based Resources in Radiology: A Mixed Methods Comparison of Efficacy and Student Engagement. *Academic Radiology,* 22(10), 1299-1307.

Wood, G., Knapp, K.M., Rock, B., Cousens, C., Roobottom, C. and Wilson, M.R. (2013) Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal radiology,* 42 (2), 165-172.

Woznitza N., Burke S., Patel K., Amin S. and Grayson, K. (2013.) Chest X-ray interpretation: Agreement between consultant radiologists and a reporting radiographer in clinical practice in the United Kingdom. *American Journal of Respiratory and Critical Care Medicine,* 187

Woznitza N., Piper K., Burke S., Patel K., Amin S., Grayson K. and Bothamley, G. (2014) Adult chest radiograph reporting by radiographers: Preliminary data from an in-house audit programme. *Radiography,* 20 (3), 223-229.

Woznitza N., Piper K., Rowe S. and West, C. (2014) Optimizing patient care in radiology through team-working: A case study from the United Kingdom. *Radiography,* 20 (3), 258-263.

Woznitza N. and Piper, K. (2015) Radiographer errors in adult chest x-ray reporting. *Journal of Medical Radiation Sciences.Conference: 2015 NZIMRTAIR-Scientific Meeting "The Cloud: Shaping our Future".New Zealand.Conference Start: 20150724.Conference End: 20150726,* 62 31.

Woznitza, N. (2014) Radiographer reporting. *Journal of Medical Radiation Sciences,* 61 (2), 66-68.

Woznitza, N., Piper, K., Burke, S. and Bothamley, G. (2016) Chest x-ray interpretation by radiographers: diagnostic accuracy and influence on diagnostic reasoning.

Woznitza, N. (2017) Chest x-ray interpretation by radiographers. *Journal of Medical Radiation Sciences.Conference: 12th Annual Scientific Meeting of Medical Imaging and Radiation Therapy.Australia.Conference Start: 20170324.Conference End: 20170326,* 64 35.

Wright, C. (2014) RadBench: benchmarking image interpretation skills on a global scale.

Wright, C. and Reeves, P. (2017) Image interpretation performance: A longitudinal study from novice to professional. *Radiography,* 23 (1), e1-e7.

Venjakob, A.C., Marnitz, T., Phillips, P. and Mello-Thoms, C.R. (2016) Image Size Influences Visual Search and Perception of Hemorrhages When Reading Cranial CT: An Eye-Tracking Study. *Human Factors,* 58(3), 441-451.

World Health Organization. Cancer. Fact Sheet Number 297. [Accessed July 10, 2015]. Available at:www.who.int/mediacentre/factsheets/fs297/en/Last updated July 2015.

Wu MH, Gotway MB, Lee TJ, Chern MS, Cheng HC, Ko JC, Sheu MH, Chang CY. Features of non-small cell lung carcinomas overlooked at digital chest radiography. Clinical Radiology. 2008 May 1;63(5), 518-28.

Yates, E.J., Yates, L.C. and Harvey, H., 2018. Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. *Clinical Radiology*.

.

Appendix A - Chapter 2

Appendix 2.1 CASP tool

CASP tool   **Article**

no.____

Are the results of the study valid?

<u>Screening questions</u>

1. Did the study address a clearly focused issue?

Yes	☐
No	☐
Can't tell	☐

Hint: A question can be 'focused' in terms of

- The population studied

  YES ☐   NO ☐   CAN'T TELL ☐
  _____

- The risk factors studied

  YES ☐   NO ☐   CAN'T TELL ☐
  _____

- The outcomes considered

  YES ☐   NO ☐   CAN'T TELL ☐
  _____

- Is it clear whether the study tried to detect a beneficial or harmful effect?

  YES ☐   NO ☐   CAN'T TELL ☐
  _____

2. Was the cohort recruited in an acceptable way?

Yes ☐

No ☐

Can't tell ☐

Hint: Look for selection of bias which might compromise the generisability of the findings:

- Was the cohort representative of a defined population?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Was there something special about the cohort?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Was everybody included who should have been included?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

3. Was there a comparison with an appropriate reference standard?

Yes ☐

No ☐

Can't tell ☐

- Hint: Is this reference the best available indicator in the circumstances?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

Is it worth continuing?

4. Was the training accurately measured to minimise bias?

Yes ☐

No ☐

Can't tell ☐

Hint: Look for measurement of classification bias:

- Did they use subjective or objective measurements?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Do the measurements truly reflect what you want them to (have they been validated)?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Were all subjects classified into exposure groups using the same procedure?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Were the measurement methods similar in the different groups?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Were the subjects and/ or the outcome assessor blinded to the training (does this matter)?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

5. Have the authors identified and taken into account all important confounding factors?

Yes ☐

No ☐

Can't tell ☐

List the ones you think might be important, that the author missed.

_____

—

Hint: Look for restriction in design and techniques e.g. modelling, stratified-, regression-, or sensitivity analysis to correct, control or adjust for confounding factors

6. Were the methods for performing the test described in sufficient detail?

Yes ☐

No ☐

Can't tell ☐

(B) What are the results?

7. Do you believe the results?

Yes ☐

No ☐

Can't tell ☐

Hint: Consider

- Big effect is hard to ignore!
- Can it be due to bias, chance or confounding?

  YES ☐   NO ☐   CAN'T TELL ☐

  _____

- Are the design and methods of this study sufficiently flawed to make the results unreliable?

YES ☐　　　NO ☐　　　CAN'T TELL ☐

_____

(C) Will the results help locally?

8. Can the results be applied to the local population?

Yes ☐

No ☐

Can't tell ☐

Hint: Consider whether

- A cohort study was the appropriate method to answer this question
- The subjects covered in this study could be sufficiently different from you population to cause concern
- Your local setting is likely to differ much from that of the study
- You can quantify the local benefits and harms

9. Do the results of this study fit with other available evidence?

Yes ☐

No ☐

Can't tell ☐

10. Can the results be applied to your patients/ the population of interest?
Hint: Prevalence of pathologies

Yes ☐

No ☐

Can't tell ☐

Appendix B - Chapter 3

Appendix 3.1 Ethics approval

UNIVERSITY OF ULSTER   RESEARCH GOVERNANCE

RG1a APPLICATION TO UNDERTAKE RESEARCH ON HUMAN SUBJECTS

PLEASE REFER TO THE NOTES OF GUIDANCE BEFORE COMPLETING THIS FORM.
(Available from the Research Governance website at
http://www.ulster.ac.uk/research/rg/)

All sections of this form must be completed (use minimum font size 11). If the form is altered in any way it will be returned unconsidered by the Committee.

This form should be used for research in categories A, B and D

Do not use this form for research being conducted in collaboration with the NHS/HPSS (category C).

SECTION A

Chief Investigator	Dr Raymond Bond
Title of Project	*An assessment of clinical interpretation of medical images and diagnostic accuracy with the aid of eye tracking technology*
Student and course (if applicable)	
Additional	Dr Sonyia McFadden School of Health Sciences

Investigators

Declaration - Chief Investigator:

I confirm that

A

- this project meets the definition for research in category* *(please insert)*
- this project is viable and is of research or educational merit;
- all risks and ethical and procedural implications have been considered;
- the project will be conducted at all times in compliance with the research description/protocol and in accordance with the University's requirements on recording and reporting;
- this application has not been submitted to and rejected by another committee; and
- Permission has been granted to use all copyright materials including questionnaires and similar instruments

Signed: _(signature)_   **Date: 11/04/14**

---

*In addition, you should complete form RG1d for all category D research and form RG1e for both category B and D research

---

Peer Review

- *Those conducting peer review should complete form RG2 and attach it to this form (RG1). RG1, RG2 and all associated materials should then be returned to the Chief Investigator.*

- *Depending upon the outcome of peer review, the Chief Investigator should arrange to submit to the Filter Committee, resubmit the application for further review or consider a new or substantially changed project. The application must not be submitted to the Filter Committee until the peer review process has been completed (except as permitted below)*

- ***Please note that peer review can be conducted by the Filter Committee if time and***

Filter Committee

- *The application must be considered by the Filter Committee in accordance with the requirements of the University*

- *The Filter Committee should complete form RG3 and write to the Chief Investigator indicating the outcome of its review*

- *Depending upon the outcome of the Filter Committee review, the Chief Investigator should arrange to proceed with the research OR submit to the University's Research Ethics Committee OR resubmit the application for further review OR consider a new or substantially changed project*

SECTION B

1. Where will the research be undertaken?

Within the University as part of student practical lessons. Delegates will be recruited as part of a conference to be held on Saturday 11th October 2014 at the La Mon Hotel and Country Club.

2. a. What prior approval/funding has been sought or obtained to conduct this. research? Please also provide the UU cost centre number if known

N/A

b. Please indicate any commercial interest in/sponsorship of the study

N/A

3. Duration of the Project

Start: 25/09/14	End: 25/12/15	Duration: 15 months

4. Background to and reason(s) for the Project

Please provide a brief summary in language comprehensible to a lay person or non-expert. Full details must be provided in the description/protocol submitted with this application (see Notes of Guidance)

Approximately 100,000 patients die every year as a result of a medical error and one million patients are injured as a result of human errors made by clinical staff. Furthermore, clinicians often misdiagnose patients, which can result in a lack of appropriate treatment or indeed unnecessary treatment. This research looks to gain insight into how both student and expert radiographers and radiologists interpret important radiographic images. We will do this through the use of an unobtrusive eye tracker, which will be used to record data that can objectively and quantifiably detail the area where each subject studies the least and the most. We will also ask subjects to think-aloud and give a verbal diagnosis. This will allow the investigators to gain insight into the subject's cognitive processes. Both the audio and the eye tracking data will be synchronously recorded.

5. Aims of the Project

Please provide a brief summary in language comprehensible to a lay person or non-expert. Full details must be provided in the description/protocol submitted with this application (see Notes of Guidance)

- To gain insight into how student and expert radiographers and radiologists interpret important radiographic images.
  To identify:
  - o patterns of interpretation e.g. where subjects look the least and the most on radiographic images
  - o duration of each interpretation for different types of pathology
  - o correlations between interpretation methods, diagnostic accuracy and confidence levels
  - o diagnostic accuracy amongst different groups, i.e. students and experts
  - o inter-rater reliability amongst all participants
  - o common interpretation errors and pitfalls

6. Procedures to be used

a. Methods

Please provide a brief summary in language comprehensible to a lay person or non-expert. Full details must be provided in the description/protocol submitted with this application (see Notes of Guidance)

1. Each subject recruited will read an information sheet and give consent.
2. Each subject will then interpret a series of medical images whilst thinking aloud. The think-aloud protocol is a well-known method often used for the elicitation of cognitive processes.
3. Whilst thinking-aloud, the Tobii Eye Tracking device will be used to non-invasively and unobtrusively track their eye movement patterns.
4. After each interpretation, the subject will be asked to give a verbal diagnosis, their recommended treatment and their level of confidence.
5. After the study each subject will complete a questionnaire, which will be used to collect demographics and attitudes towards interpretation of medical images.

b. Statistical techniques

Please provide details of the statistical techniques to be used within the   project description/protocol (see Notes of Guidance)

Several of the questions will contain 'semantic differential scales', thus assisting to categorise the answers given by the subjects. These categories will then form the basis for the assessment of statistically significant differences between the different age groups and expert groups using Chi-Square and/or *t*-tests. Pearson's product moment correlation coefficient will also be used to determine correlations. Fleiss' Generalized Kappa will also be used to determine inter-rater reliability.

7. Subjects:

a. How many subjects will be recruited to the study (by group if   appropriate)?

Students	Min.15 Max. 20
Conference delegates/experts	Min.15 Max. 20

b. Will any of the subjects be from the following vulnerable groups -

	YES	NO
Children under 18		√
Adults with learning or other disabilities		√
Very elderly people		√
Healthy volunteers who have a dependent or subordinate relationship to investigators		√
Other vulnerable groups		√

If YES to any of the above, please specify and justify their inclusion

c. **Inclusion and exclusion criteria**

Please indicate, with reasons, the inclusion criteria for the project

Inclusion criteria: Involved as a student, novice or expert in recording or interpreting radiographic images. Subjects will be determined as a student, novice or 'experienced' based on their occupation and years of experience. For example, if the subject is studying an undergraduate programme in radiography then they are obviously classified as a student. If the subject has less than 3 years clinical experience then they are classified as a novice and if the subject has more than 3 years clinical experience then they are classified as 'experienced'.

N/A

d. **Will any inducements be offered? If 'Yes', please describe**

N/A

e. Please describe how and where recruitment will take place

Students, staff members and conference delegates will be recruited in university classrooms and/or academic research conferences. The information sheet will be emailed to students and another information sheet will be disseminated via

8. Ethical implications of the research

Please provide an assessment of the ethical implications of the project

Typically ethical issues in research may arise from recruitment, consent, confidentiality and data protection. In this study we have identified the risk of embarrassment during the recordings however we have put in place precautions as discussed in section 11. There is also no medical assessment involved in this study, therefore there is no risk in detecting unknown conditions, nor any risk of experiencing pain or physical discomfort. Subjects are also not required to wear additional devices, as the eye tracking device is non-invasive and unobtrusive. Also there are no video recordings of the subject's face or body. Only the screen is recorded using screen-casting software and the subject's verbalisation (from thinking-aloud) when interpreting images is recorded. However, all data will be stored on a protected hard drive. And any survey results and data collected will be anonymised (i.e. no names will be stored).

9. Could the research identify or indicate the existence of any undetected healthcare concern?

Yes ☐ **No** ☑

If **Yes**, please indicate what might be detected and explain what action will be taken (e.g. inform subject's GP)

10. Risk Assessment **

Please indicate any risks to subjects or investigators associated with the project

The risk assessment did not identify any risks associated with this study.

If you wish, you can use form RG1c – Risk Assessment Record (available from the Research Governance website) to help you assess any risks involved

11. Precautions

Please describe precautions to be taken to address the above

> The risk of embarrassment during the 'Think Aloud' phase of the test is limited given the area where the tests are performed will be screened (i.e. private) so that the subject can only be viewed by the researcher. Although subjects will be encouraged to be actively involved in the test, it will be explained to them that their degree of involvement is a personal decision and that it will be respected at all times. If the researcher detects distress of a participant, or the participant discusses his/her distress with the researcher,

12. Consent form

It is assumed that as this study is being conducted on human subjects, an information sheet and associated consent form will be provided. A copy of the information sheet and form must be attached to this application. See Notes of Guidance.

If a consent form is not to be used, please provide a justification:

> Consent form and information sheet is included in Appendix 1.

13. Care of personal information

Please describe the measures that will be taken to ensure that subjects' personal data/information will be stored appropriately and made available only to those named as investigators associated with the project.

> All electronic data will be stored on a password protected University of Ulster hard drive in accordance with the Data Protection Act (1998). Only those persons directly involved in the project will have access to the data. Each subject will be issued with a number to protect their identity; this number will then be used in any presentation, or published work of the results.
>
> Screen casts and audio recordings will be stored immediately to the above-mentioned external hard drive. No copy is retained on the computer used to make the recordings. All other data and documentation will be destroyed at the end of the project, in October 2013, through confidential waste management and data storage media reformatting.

14. Copyright

Has permission been granted to use all copyright materials including questionnaires and similar instruments?

Yes ☑ **No** ☐

If **No**, please provide the reason

```



```

Once you have completed this form you should also complete form RG1d for all category D research and form RG1e for both category B and D research

University of
ULSTER

Title of Study: An assessment of clinical interpretation of medical images and diagnostic accuracy with the aid of eye tracking technology

Investigators   Dr Raymond Bond, Dr Sonyia McFadden

Information Sheet and Consent Form for Participants

You are invited to participate in a University of Ulster research project. Before you decide whether or not to take part, it is important that you understand what the research is for and what you will be asked to do. Please read the following information and do not hesitate to ask any questions about anything that might not be clear to you. Make sure that you are happy before you participate. Thank you for considering this invitation!

Correct interpretation of medical images is vitally important to patient diagnosis and subsequent therapy. However very little research has been done to investigate how novices and experts interpret these images through the use of an eye tracker and the think-aloud protocol. The think-aloud protocol is where each subject is asked to think-aloud whilst interpreting a medical image. This allows the researcher to gain insight into how the subject actually reads and processes the medical data.

Project Aim

To gain insight into how students, radiographers and radiologists interpret important medical images.

We aim to identify:

- o   patterns of interpretation e.g. where subjects look the least and the most on medical images
- o   duration of each interpretation for different types of diseases
- o   correlations between interpretation methods and diagnostic accuracy
- o   diagnostic accuracy amongst students and experts
- o   inter-rater reliability amongst all participants
- o   common interpretation errors and pitfalls

The session will last no longer than 20 minutes. You will be asked to look at a sample of medical images on a computer monitor and you are asked to provide a verbal diagnosis. Your eye gaze path on the computer screen will be recorded using a non-obtrusive Tobii eye tracking technology.

Your participation in this study is voluntary and you are able to withdraw at any point without giving a reason. All data will be stored securely and will be made available only to persons conducting the study. No reference will be made in any oral or written reports that could link you to the study. Your data collected will be anonymous, i.e. we will not record your name. The results of the study will be used to further research.

This study has been reviewed by the Faculty of Computing and Engineering Research Ethics Filter Committee and is in accordance with the University of Ulster research governance guidance. If you have questions at any time about the study or the procedures, you may contact **Dr Raymond Bond** via email or phone:

Email: rb.bond@ulster.ac.uk   **Tel**:   028 90 368156

Address: Room 16G06, University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB

Information Sheet and Consent Form for Participants (1.0) (continued)

Title of Project: An assessment of clinical interpretation of medical images and diagnostic accuracy with the aid of eye tracking technology

Name of Investigators:

Dr Raymond Bond, Dr Sonyia McFadden

Please initial each point:

• I confirm that I have been given and have read and understood the information sheet for the above study and have asked and received answers to any questions raised

[   ]

• I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without my rights being affected in any way

[   ]

• I understand that the researchers will hold all information and data collected securely and in confidence and that all efforts will be made to ensure that I cannot be identified as a participant in the study (except as might be required by law) and I give permission for the researchers to hold relevant anonymised personal data

[   ]

• I agree to take part in the above study   **[   ]**

• I agree to photography of the experiment   **[   ]**

_____   _____   _____

Name of Subject   *Signature*   *Date*

_____   _____   _____

Name of researcher   *Signature*   *Date*

You may contact **Dr Raymond Bond** via email or phone:

Email: rb.bond@ulster.ac.uk   **Tel**:   028 90 368156

Address: Room 16G06, University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB

Appendix 3.3 Data collection sheets

Image 1

Correct Diagnosis: Pneumothorax, bilateral. (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 2

Correct Diagnosis: Femoral neck fracture on right (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 3

Correct Diagnosis: 3rd metacarpal fracture. (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 4

Correct Diagnosis: Colles' fracture Colles' fracture of the left wrist with associated ulnar

styloid fracture. (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 5

Correct Diagnosis: Radial head fracture (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 6

Correct Diagnosis: CXR shows is a round opacity lesion locate on right lower lobe,

projecting through the right hilum (hilum overlay sign). (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 7

Correct Diagnosis: Lateral tibial plateau fracture (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Image 8

Correct Diagnosis: Normal (hidden from subject)

Diagnosis given by Subject:_____

_____

Confidence level (circle a number):

Low											High
	1	2	3	4	5	6	7	8	9	10	

Appendix 3.4 Questionnaire

User_ID: _____

Questionnaire

Please answer the questions below.

_____

Age: _____ **years**

Gender: Male | Female   *(Circle one)*

I am a radiologist /radiographer /student radiographer (delete as appropriate) specialised in the field of:

_____

Are you a reporting radiographer?

☐ Yes ☐ No

How many years have you been viewing and interpreting radiographic images:

_____ years

How many years have you been officially reporting radiographic images?

_____ years

On a scale of 1 to 10 (1 being poor and 10 being excellent), rate your level of expertise in medical image interpretation?

(Circle a number)

Poor											Excellent
	1	2	3	4	5	6	7	8	9	10	

In your opinion do you believe that we need to gain insight into how clinicians interpret medical images?

☐ Yes   ☐ No

Do you believe clinicians should be regularly assessed for their competency?

☐ Yes   ☐ No

Would you support the development of best practice guidelines on the process of interpreting different radiographic images?

☐ Yes   ☐ No

Rate how useful you believe eye tracking would be in assessing clinical competency in interpreting radiographic images?

(Circle a number)

Poor											Excellent
	1	2	3	4	5	6	7	8	9	10	

To prove competency, how many radiographic images would a clinician need to correctly diagnose as part of an assessment?

_____ out of _____ radiographic images.

How many radiographic images do you believe you diagnosed correctly?

_____ *(out of 8)*

10. Do you have any other questions or comments?

Thank you for filling out the questionnaire – it is appreciated as part of our ongoing research.

Appendix C - Chapter 4

Appendix 4.1 Written search strategy training tool

CHEST IMAGE INTERPRETATION

GENERAL CONSIDERATIONS

1. **Look across the top of the image from LEFT-RIGHT to check details (usually in the DICOM header on a PACS station)**
   a. Name
   b. Date of examination
   c. Age of patient
   d. Side marker   YES ☐   NO ☐

   e. Lung apices   NORMAL ☐

                                      ABNORMAL ☐

   Preliminary diagnosis
   _____

2. **Look down the middle of the chest image from TOP-BOTTOM**

   a. Projection   PA ☐

                                      AP ☐

   b. Was the image taken   ERECT ☐
                              SUPINE ☐
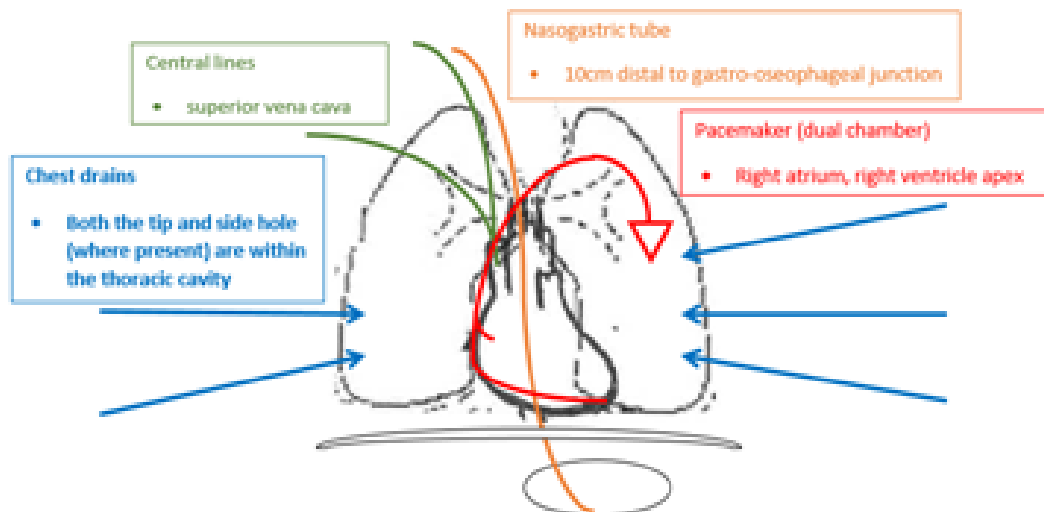   OTHER ☐

   c. Are the lungs   OVER INFLATED ☐
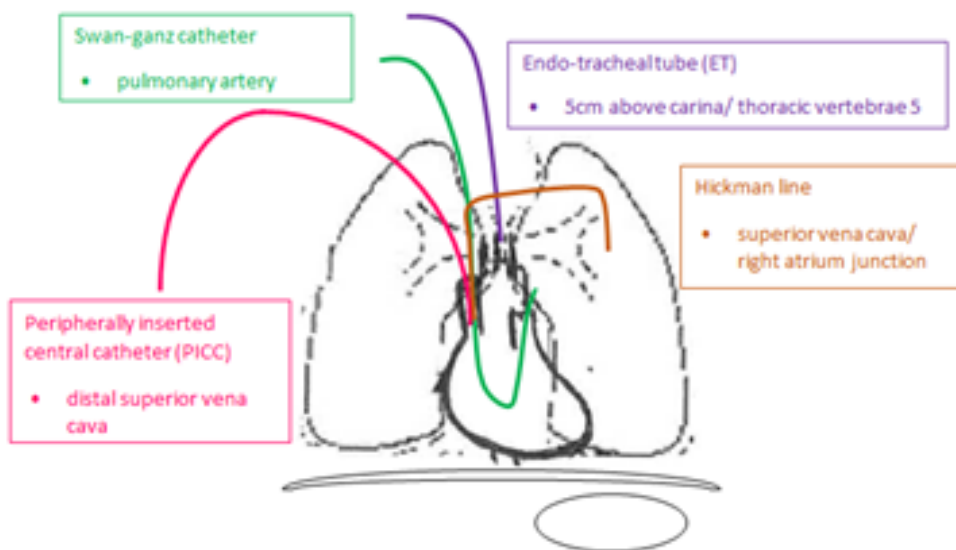   ADEQUATELY INFLATED ☐
                              UNDER INFLATED ☐

   d. Is image penetration   OVER ☐
                              ADEQUATE ☐
                                 UNDER ☐

e. Rotation
Are clavicles equidistant from spine?   YES ☐   NO ☐

Is the trachea central?   YES ☐   NO ☐

Are the lung fields of equal density and size?   YES ☐   NO ☐

f. Is there evidence of kyphosis or lordosis?   YES ☐   NO ☐

Medial clavicle is over posterior end of 4th rib   YES ☐   NO ☐

g. Does the chest appear tilted?   YES ☐   NO ☐

3. Look across the bottom of the chest image from LEFT-RIGHT

a. Is all the chest included?   YES ☐   NO ☐

POSITION OF COMMON TUBE, LINES AND DEVICES

Central lines
• superior vena cava

Nasogastric tube
• 10cm distal to gastro-oseophageal junction

Pacemaker (dual chamber)
• Right atrium, right ventricle apex

Chest drains
• Both the tip and side hole (where present) are within the thoracic cavity

Swan-ganz catheter
- pulmonary artery

Endo-trachealtube (ET)
- 5cm above carina/ thoracic vertebrae 5

Hickman line
- superior vena cava/ right atrium junction

Peripherally inserted central catheter (PICC)
- distal superior vena cava

LINES, TUBES AND DEVICES

Is there a line, tube and device present	YES ☐	NO ☐
Nasogastric tube visible	YES ☐	NO ☐
10cm below gastro-oseophageal junction	YES ☐	NO ☐
Central line visible	YES ☐	NO ☐
Tip present within the superior vena cava	YES ☐	NO ☐
Chest drain inserted	YES ☐	NO ☐
Tip and side hole (where present) within the thoracic cavity	YES ☐	NO ☐
Pacemaker visible	YES ☐	NO ☐
Leads on the right atrium and right ventricle apex	YES ☐	NO ☐
Swan-ganz catheter visible	YES ☐	NO ☐
Tip within pulmonary artery	YES ☐	NO ☐
Endo-trachael tube visible	YES ☐	NO ☐
Tip 5cm above the carina	YES ☐	NO ☐
Hickman line visible	YES ☐	NO ☐
Tip within superior vena cava/ right atrium junction	YES ☐	NO ☐

Peripherally inserted central catheter visible	YES ☐	NO ☐
Distal superior vena cava	YES ☐	NO ☐
Other line, tube, device seen	YES ☐	NO ☐
Correctly positioned	YES ☐	NO ☐

BONY THORAX

Look at the image from TOP-BOTTOM and ALONG the ribs. Look for any shape or density changes, callus formation or fractures.

a. Is the rib curvature   NORMAL ☐

                                 ABNORMAL ☐

   Preliminary diagnosis

_____

b. Are the clavicles   NORMAL ☐

                                 ABNORMAL ☐

   Preliminary diagnosis

_____

c. Are the following	NORMAL	ABNORMAL
Ribs along their length	☐	☐

Right scapula ☐ ☐

Left scapula ☐ ☐

Right humeral head ☐ ☐

Left humeral head ☐ ☐

Preliminary diagnosis

_____

THE NECK AND SOFT TISSUES

Assess the breasts / pectoral region
Evidence of mastectomy / breast implants
or other abnormality

START | FINISH
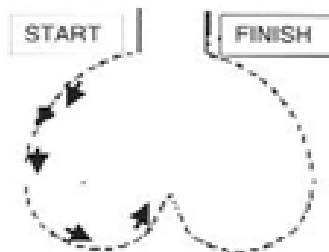
YES ☐
NO ☐

Any evidence of surgery or other
intervention
YES ☐
NO ☐

THE DIAPHRAGM, HEART AND MEDIASTINUM

Any free air   YES ☐

                        NO ☐

Is the cardiothoracic ratio   >50% ☐

                        <50% ☐

Is the aortic knuckle   NORMAL ☐

                        ☐

	NO	☐
If yes;	UNILATERAL	☐
	BILATERAL	☐
If bilateral;	SYMETRIC	☐
	ASYMETRIC	☐

Preliminary diagnosis

_____

	NORMAL	☐
Are the hilar vessels		
	ABNORMAL	☐

Preliminary diagnosis

_____

Look around the heart border

Right atrial enlargement	YES ☐	NO ☐
Left atrial enlargement	YES ☐	NO ☐
Ventricular enlargement	YES ☐	NO ☐

Look around the mediastinal contours
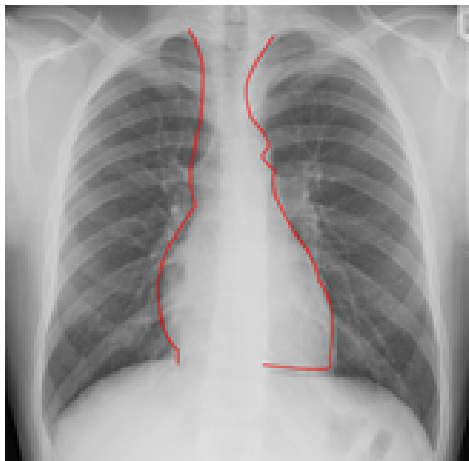
Look around the edges of both lungs
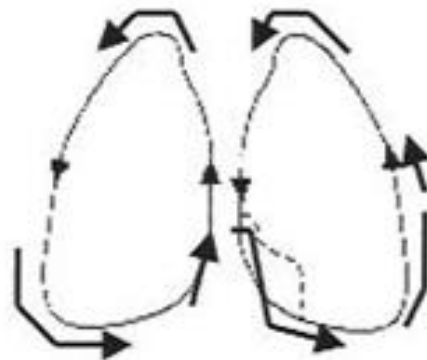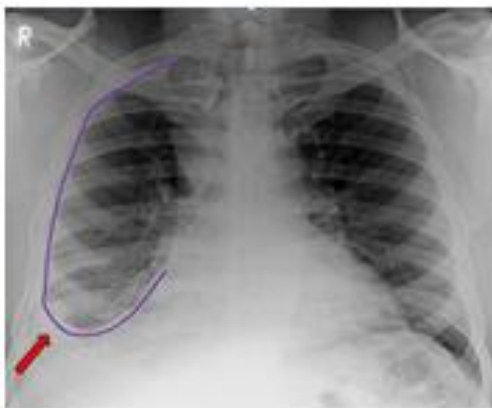
Are the costophrenic and cardiophrenic angles

NORMAL ☐

ABNORMAL ☐

If abnormal can you see evidence of

PNEUMOTHORACES ☐

EFFUSIONS ☐

PLEURAL THICKENING / TAGGING ☐

Look at and compare both lung fields from top to bottom STARTING MEDIALLY
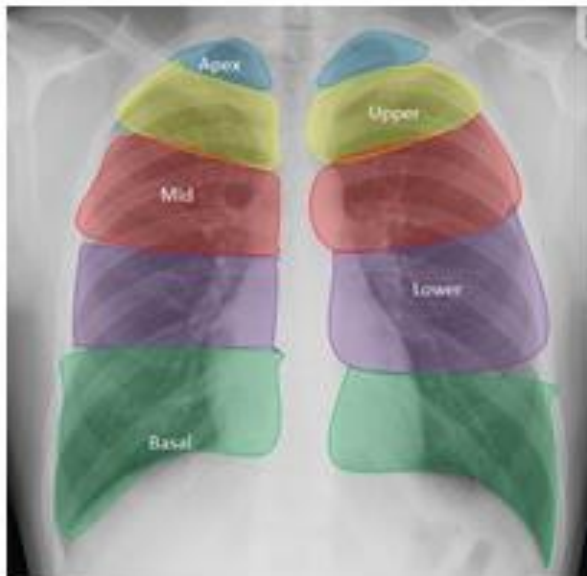
COMPARE

Right with left upper zone

Right with left middle zone

Right with left lower zone

Are there any abnormalities present   YES ☐   NO ☐

If YES are they   BLACK ☐

                                                           WHITE ☐
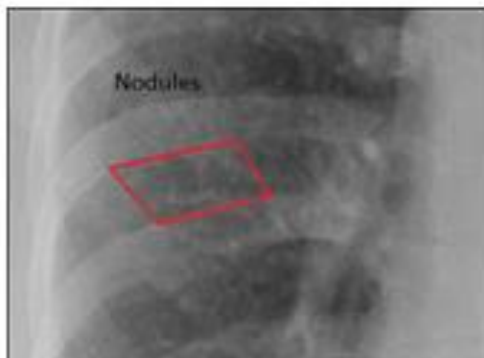
SHADOWS

AREAS OF LUCENCY

a. No lung markings present and extends inwards from periphery

                                                           YES ☐   NO ☐

b. Pneumothorax?
   Preliminary diagnosis _____

c.

Is the anatomy rotated?   YES ☐   NO ☐

Evidence of hyperinflation?   YES ☐   NO ☐
Preliminary diagnosis _____

Hyperinflation is seen with a well-defined area of lucency?

     YES ☐   NO ☐

Does this represent a   BULLA   ☐

                         ABSCESS   ☐

                         AIR TRAPPING eg COAD   ☐

                         CAVITY   ☐

AREAS OF OPACIFICATION

EFFUSION
Is there a basal meniscus at the costophrenic angle(s)   YES ☐   NO ☐

Is there bilateral or contralateral mediastinal shift?   YES ☐   NO ☐

COLLAPSE
Have fissures and diaphragm moved?   YES ☐   NO ☐

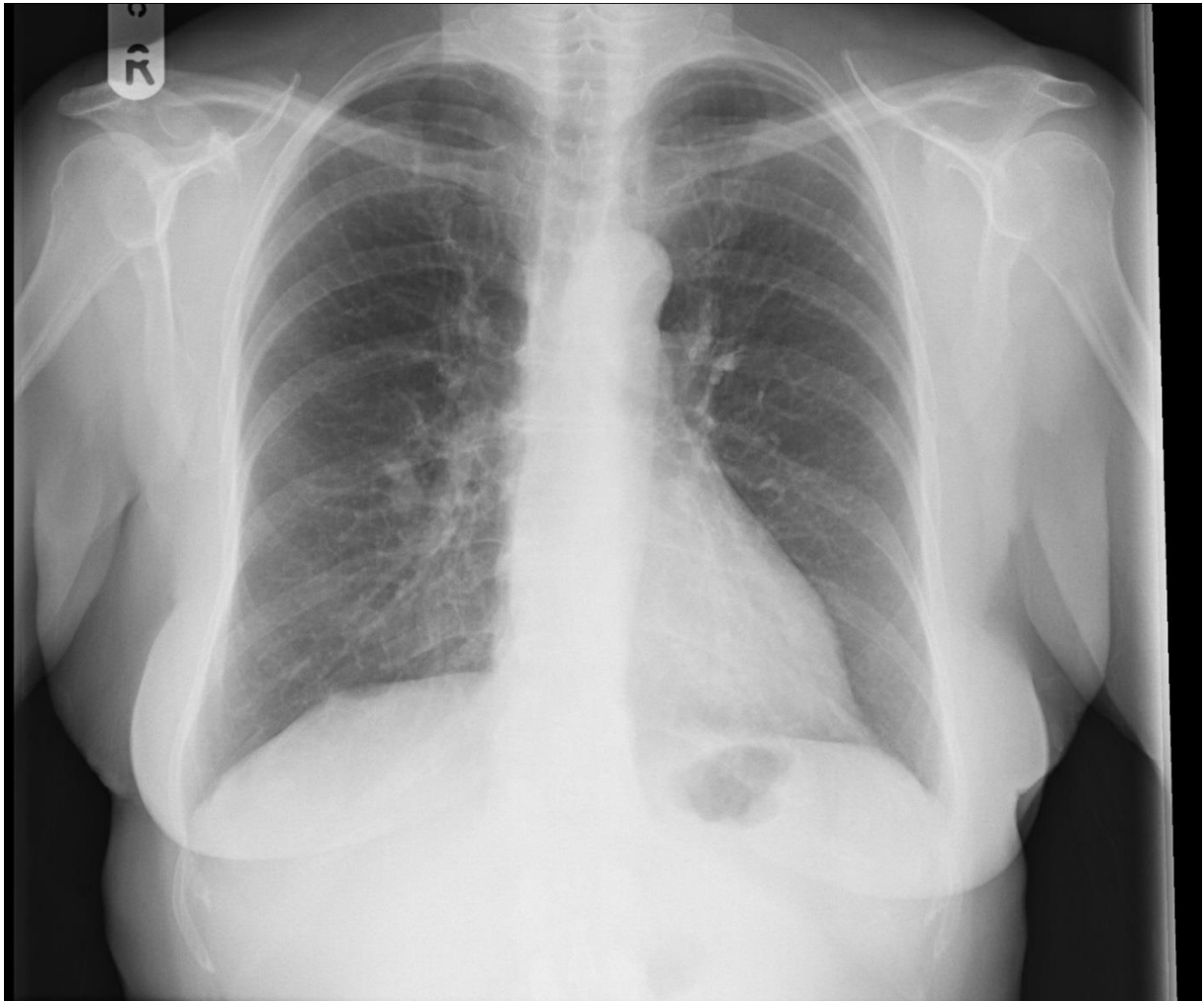Has size and direction of vascular markings changed?   YES ☐   NO ☐

Are ribs and mediastinum pulled to the shadow?   YES ☐   NO ☐

Is there rib crowding?   YES ☐   NO ☐

NORMAL VARIANTS_____

DIAGNOSIS AND FULL REPORT_____

Appendix 4.2: Example of a chest image used within the educational programme

Appendix 4.3: Training tool following digitisation

General considerations

1. Look across the top of the image from LEFT-RIGHT to check details (usually in the DICOM header on a PACS station)?

    a. Name

    b. Date of examination

    c. Age of patient

    d. Sidemarker   NO ⬤ YES

    e. Lungapices   NORMAL ⬤ ABNORMAL

    Preliminary diagnosis:

    [   ]

---

2. Look down the middle of the chest image from TOP-BOTTOM

    a. Projection   PA ⬤ AP

    b. Was the image taken | ERECT | SUPINE | OTHER |

    c. Are the lungs | OVER INFLATED | ADEQUATELY INFLATED | UNDER INFLATED |

    d. Is image penetration | OVER | ADEQUATE | UNDER |

    e. Rotation

        i. Are clavicles equidistant from spine?   NO ⬤ YES

The diaphragm, heart and mediastinum

6. Assess the diaphragm, heart and mediastinum

a. Any free air?

NO ⬤ YES

b. Is the cardiothoracic ratio?

> 50% ⬤ < 50%

c. Is the aortic knuckle

NORMAL ⬤ ABNORMAL
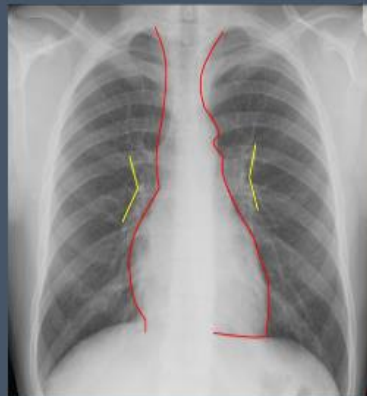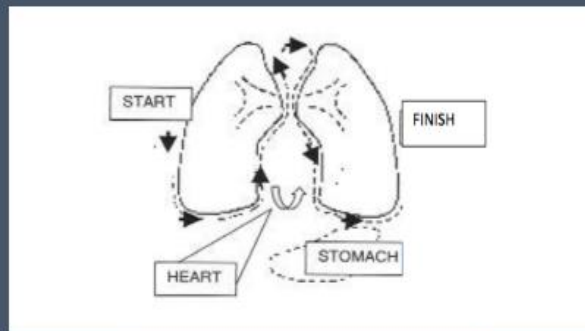
Bony Thorax

4. Look at the image from TOP-BOTTOM and ALONG the ribs. Look for any shape or density changes, callus formation or fractures?
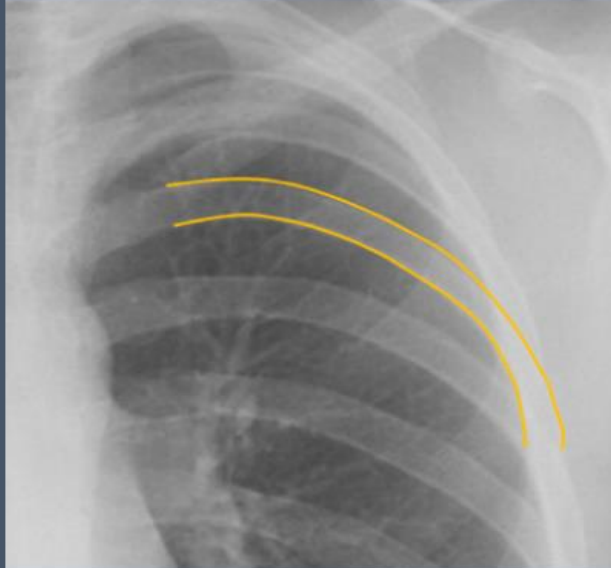
a. Is the rib curvature   NORMAL ⬤ ABNORMAL

Preliminary diagnosis:

b. Are the clavicles   NORMAL ⬤ ABNORMAL

Preliminary diagnosis:

Educational programme

Please enter your subject ID below:

General considerations

View each video and then either log out or continue to training tool

1. **Video 1**
2. **Video 2**
3. **Video 3**
4. **Video 4**
5. **Video 5**
6. **Video 6**
7. **Video 7**
8. **Video 8**
9. **Video 9**
10. **Video 10**

0:00 / 2:49

238
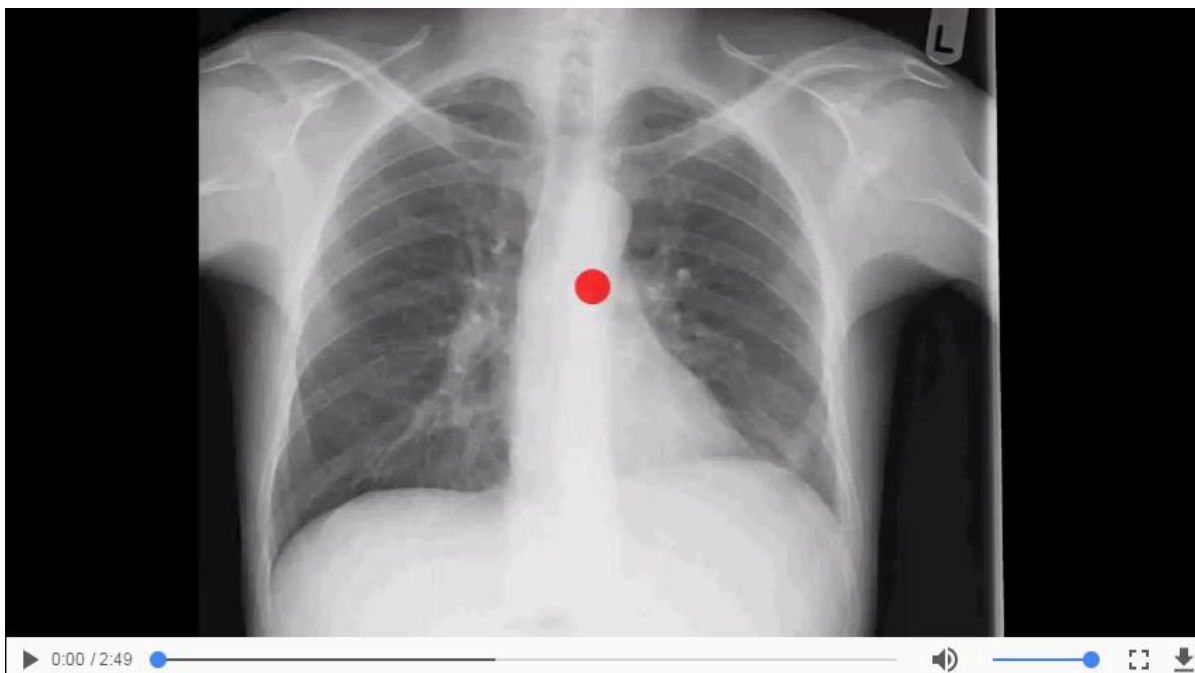
Appendix 4.4: Original diagrams used within search strategy training tool

Nasogastric tube
- 10cm distal to gastro-oseophageal junction

Central lines
- superior vena cava

Pacemaker (dual chamber)
- Right atrium, right ventricle apex

Chest drains
- Both the tip and side hole (where present) are within the thoracic cavity

Swan-ganz catheter
- pulmonary artery

Endo-tracheal tube (ET)
- 5cm above carina/ thoracic vertebrae 5

Hickman line
- superior vena cava/ right atrium junction

Peripherally inserted central catheter (PICC)
- distal superior vena cava

Appendix 4.5: Diagrams following amendments within the digitised training package
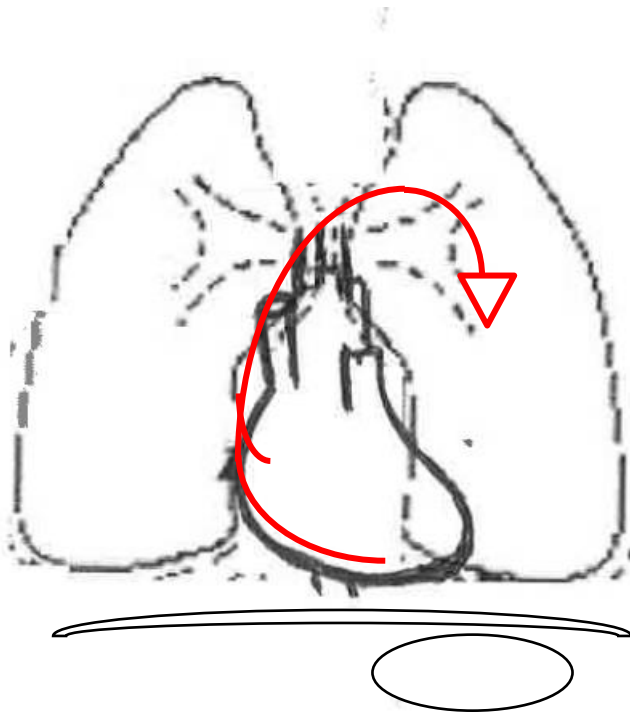
Pacemaker: Right atrium, right ventricle apex

Nasogastric tube: 10cm distal to gastro-oseophageal junction
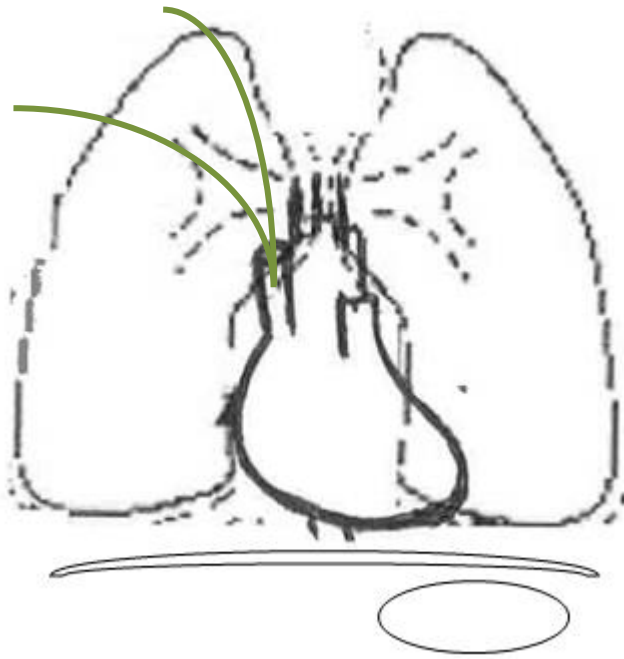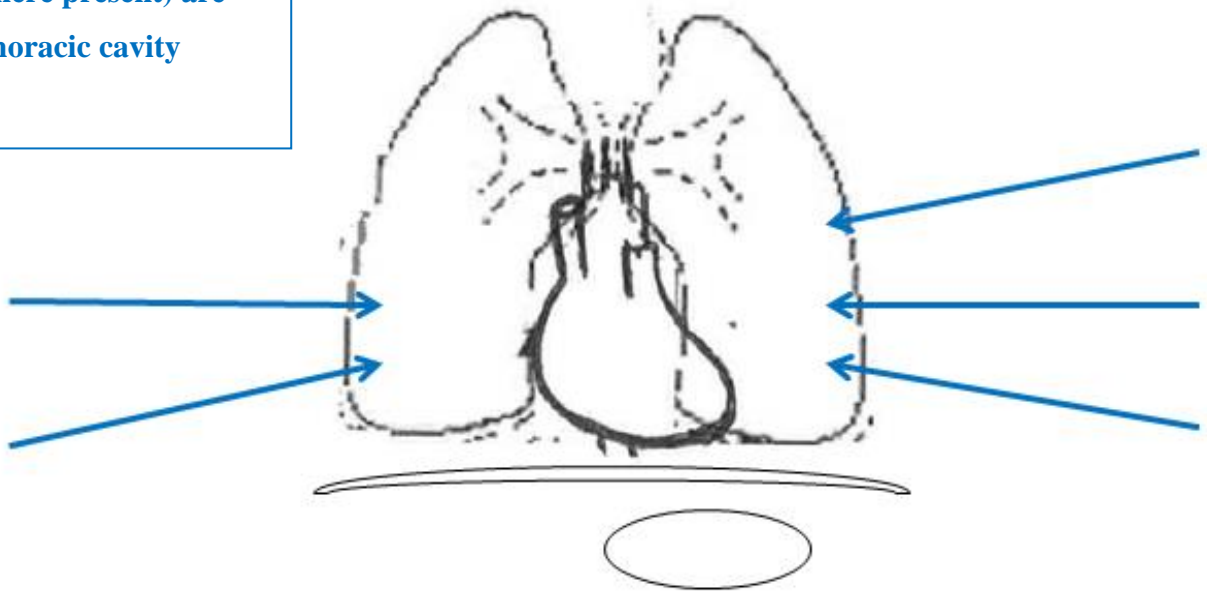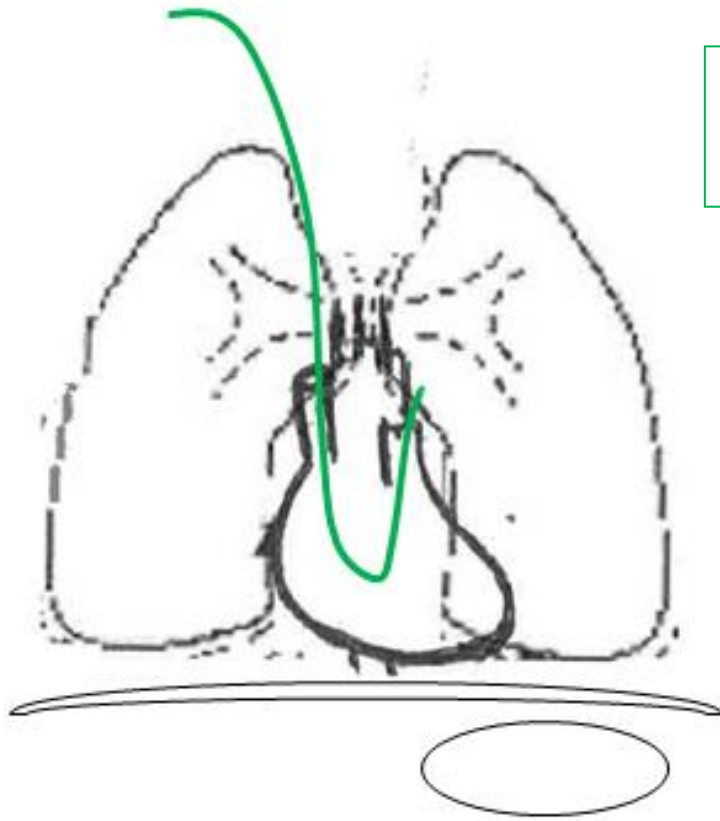
Swan-ganz catheter: pulmonary artery
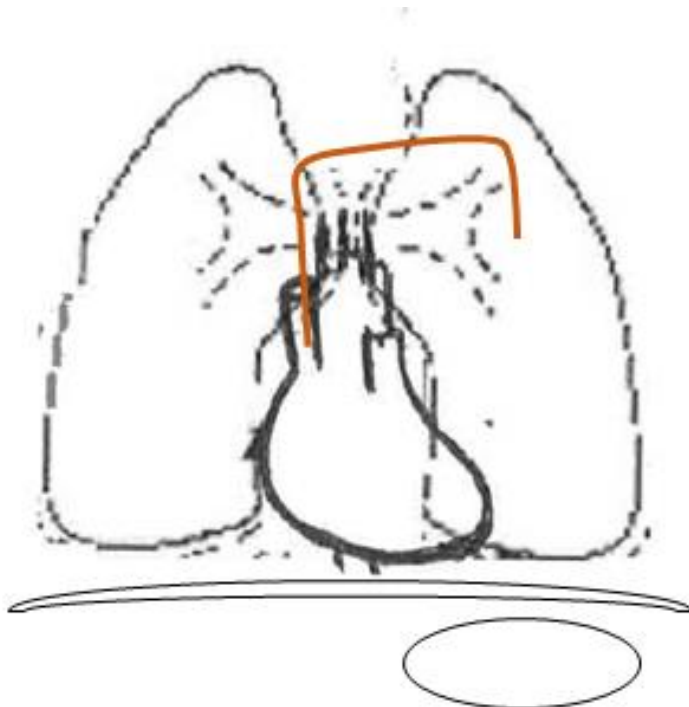
Hickman line: superior vena cava/ right atrium junction
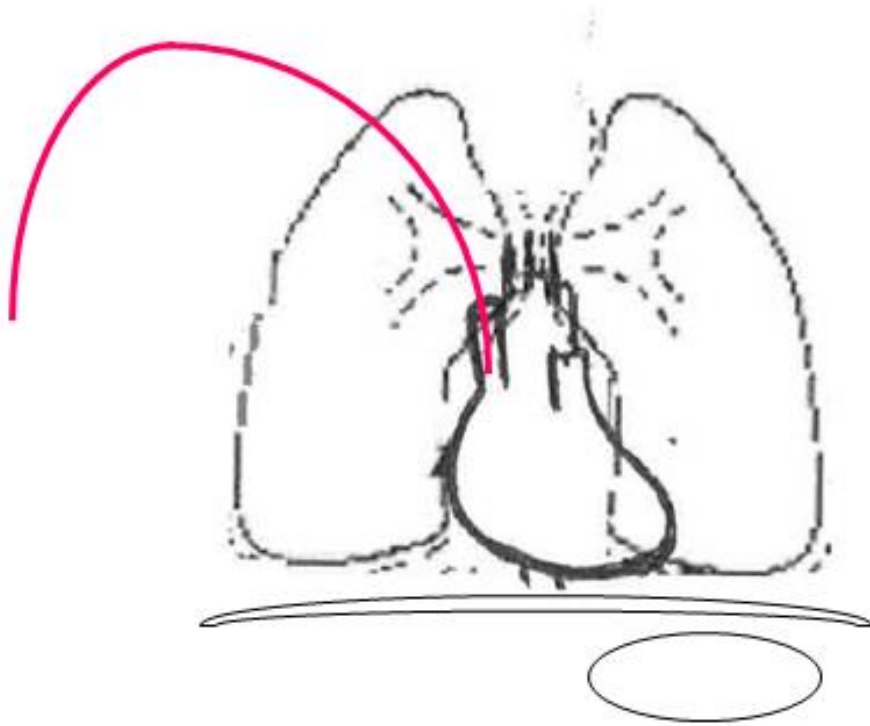
Peripherally inserted central catheter (PICC): distal superior vena cava
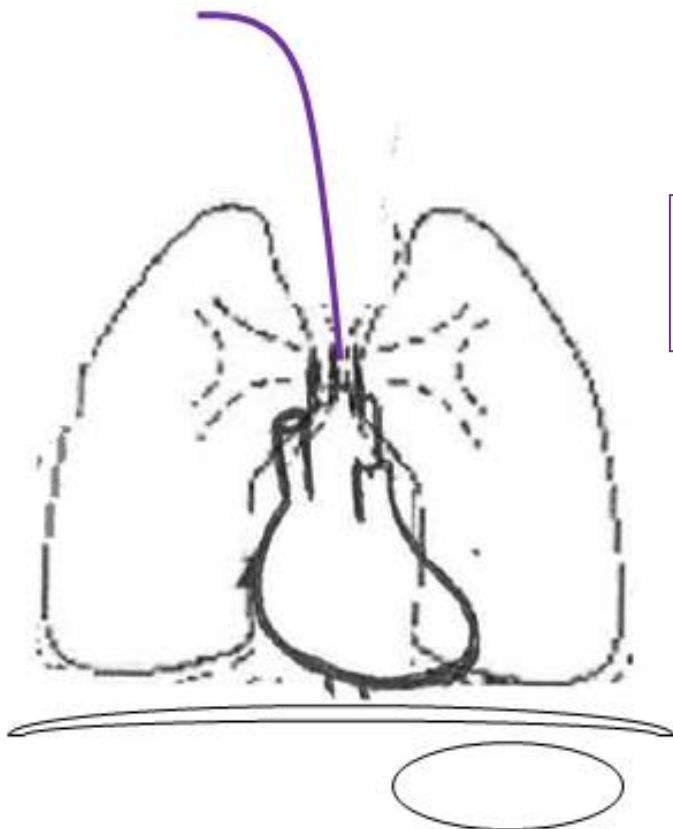
Endo-tracheal tube (ET): 5cm above carina/ thoracic vertebrae 5

Appendix D - Chapter 5

Appendix 5.1 Ethics approval

UNIVERSITY OF ULSTER RESEARCH GOVERNANCE	UU Ref No:

Form RG6	Notification of a proposed substantial amendment

Chief Investigator:
Dr. Ciara Hughes

Approved Study Title:

An evaluation of a training package in chest image interpretation with the aid of eye tracking technology

New/Amended Title (if appropriate):

n/a

Type of Amendment (please indicate any that apply):

- Amendment to application form   [X]

- Amendment to description/protocol   [X]

- Amendment to the information sheet/consent or other   [   ]
  supporting information

Please submit the appropriate amended documentation in each case, ensuring that new text is highlighted to enable comparison with the previous version to be made.

Summary of Changes:
Another participant group has been added to the recruitment process.
This additional group has been added in order to investigate the impact of the training tool on a group of professionals in isolation ie who are not also currently attending a training course.

Reporting radiographers who have previously been trained to report on images of the musculoskeletal system will be recruited through their attendance at Continuing Professional Development (CPD) events within University premises.
The initial outcome measure assessment for this group of participants will be carried out following recruitment into the study.
These participants will then be randomly allocated to a control group or intervention group.
The intervention group they will have access to the training package for 9 months. The control group will not have access to the training package.
Post-intervention assessment will be completed after 9 months on both groups when participants are attending CPD events within University premises.

Additional ethical considerations:
No vulnerable participants are going to be recruited within the new cohort. This study remains to be low risk. No additional ethical considerations exist following the inclusion of the new cohort.

List of enclosed documents:

RG1a
Study protocol
Information sheet
Consent form
Questionnaire
Risk assessment
RG2
Lead reviewers feedback
Rebuttal
Data flowchart
Short CV for all research team members

Declaration:

I confirm that the information in this form is accurate and that implementation of the proposed amendment will benefit the study appropriately.

Signed .... Date 18/8/16
(Chief Investigator)

Filter Committee Decision

This amendment:

is appropriate to the needs of the study, is in category A and should be implemented   [✓]
is appropriate to the needs of the study, is in category B and should be considered by the University REC   [ ]
is NOT appropriate and should be reconsidered or withdrawn   [ ]

Signed .... George Kernshon .... Date 23/08/16
(Chair of Filter Committee)

Appendix 5.2 Ethics protocol

UNIVERSITY OF ULSTER   RESEARCH GOVERNANCE

RG1a APPLICATION TO UNDERTAKE RESEARCH ON HUMAN SUBJECTS

PLEASE REFER TO THE NOTES OF GUIDANCE BEFORE COMPLETING THIS FORM.

(Available from the Research Governance website at

http://www.ulster.ac.uk/research/rg/)

All sections of this form must be completed (use minimum font size 11). If the form is altered in any way it will be returned unconsidered by the Committee.

This form should be used for research in categories A, B and D

Do not use this form for research being conducted in collaboration with the NHS/HPSS (category C).

SECTION A

Chief Investigator	Dr. Ciara Hughes

Title of Project	An evaluation of a training package in chest image interpretation with the aid of eye tracking technology

Student and course (if applicable)	Laura McLaughlin PhD

Additional Investigators	Dr. Sonyia McFadden	
	Dr. Raymond Bond	
	Dr. Jonathan McConnell	

Declaration - Chief Investigator:

I confirm that

$\boxed{A}$

- this project meets the definition for research in category* *(please insert)*
- this project is viable and is of research or educational merit;
- all risks and ethical and procedural implications have been considered;
- the project will be conducted at all times in compliance with the research description/protocol and in accordance with the University's requirements on recording and reporting;
- this application has not been submitted to and rejected by another committee; and
- Permission has been granted to use all copyright materials including questionnaires and similar instruments

Signed: **Date:**

Once complete, this application and all associated materials must be submitted for peer review

Peer Review

- *Those conducting peer review should complete form RG2 and attach it to this form (RG1). RG1, RG2 and all associated materials should then be returned to the Chief Investigator.*

Filter Committee

- *The application must be considered by the Filter Committee in accordance with the requirements of the University*

- *The Filter Committee should complete form RG3 and write to the Chief Investigator indicating the outcome of its review*

- *Depending upon the outcome of the Filter Committee review, the Chief Investigator should arrange to proceed with the research OR submit to the University's Research Ethics Committee OR resubmit the application for further review OR consider a new or substantially changed project*

SECTION B

1. **Where will the research be undertaken?**

    Canterbury Christchurch University in England and other university sites where appropriate

2. **a. What prior approval/funding has been sought or obtained to conduct this. research? Please also provide the UU cost centre number if known**

    Department of Education and Learning Award

    b. Please indicate any commercial interest in/sponsorship of the study

    N/A

3. **Duration of the Project**

    | Start: 01/03/2016 | End: 01/03/2018 | Duration: 2 years |

4. **Background to and reason(s) for the Project**

Please provide a brief summary in language comprehensible to a lay person or non-expert. Full details must be provided in the description/protocol submitted with this application (see Notes of Guidance)

Reporting of chest images usually lies in the domain of the radiologist but changes in the last decade has led to role development of radiographers to report chest images. Currently training involves the completion of a clinical log of practice and 'shadow' reporting the radiologist. There are no defined teaching tools and often a variety of techniques are used in different clinical departments. To date there has been no research undertaken to test these training tools or validate their use in chest image interpretation. With the use of eye tracking technology and expert consensus this research aims to establish a training package which aids the chest image interpretation of reporting radiographers. A uniform standard of training may then be established.

5. Aims of the Project

Please provide a brief summary in language comprehensible to a lay person or non-expert.

Full details must be provided in the description/protocol submitted with this application (see Notes of Guidance)

-To use eye tracking technology to investigate if an expert informed training package can improve the accuracy of diagnosis of reporting radiographers.

- To develop a training package incorporating eye tracking data and expert consensus to aid chest interpretation for reporting clinicians

- To test the effect of the training package on accuracy of diagnosis. Adherence of radiographers currently training to report on chest images and reporting radiographers previously trained to report on images of the musculoskeletal system.

- To compare eye gazes of these participants to identify: patterns of interpretation e.g where subjects look the least and the most on radiographic images, duration of each interpretation, correlations between interpretation methods, adherence to the training package, diagnostic accuracy and confidence levels, inter-rater reliability amongst all participants, common interpretation errors and pitfalls

6. Procedures to be used

a. Methods

Please provide a brief summary in language comprehensible to a lay person or non-expert.

Full details must be provided in the description/protocol submitted with this application (see

1. This is a quasi-experimental study. A training package will be formed consisting of a search strategy training tool and an educational programme. The educational programme will be videos of expert's eye gazes and a voice over of their search strategy during chest image interpretation.
2. Participants will be approached in the university setting as postgraduate students
3. Each participant recruited will read an information sheet and give consent
4. Four groups of participants will be required for this study. A convenience sample will be used for radiographers who are currently registered on a postgraduate programme training to report on chest images. Participants who enrol on the postgraduate programme in March 2016 will be the control group (group 1 – no access to the training package) and participants who enrol on the postgraduate programme in October 2016 will be the intervention group (group 2 – access to the training package). An additional group of radiographers trained to report on images of the musculoskeletal system but who are not trained in chest reporting will be randomly allocated to a control group (group 3 – no access to the training package) or intervention group (access to the training package).
5. Each participant will then interpret a series of medical images whilst thinking aloud. Participants will be shown images from a training repository used within previous studies (Woznitza et al. 2014). The think aloud protocol is a well-known method often used for the elicitation of cognitive processes
6. Whilst thinking aloud, the Tobii Eye Tracking device will be used to non-invasively and unobtrusively track their eye movement patterns
7. After each interpretation, the subject will be asked to give a verbal diagnosis and indicate their level of confidence. After the study each subject will complete a questionnaire and survey (see Appendix 2.0).
8. The intervention group will be given access to the training package and asked to re-attend and steps 3-7 repeated

b. Statistical techniques

Please provide details of the statistical techniques to be used within the   project description/protocol (see Notes of Guidance)

7. Subjects:

a. How many subjects will be recruited to the study (by group if   appropriate)?

Radiographers currently training in chest image interpretation	Max. 40
Reporting radiographers trained to report on the musculoskeletal system but not currently trained to report on chest images.	Max. 30

b. Will any of the subjects be from the following vulnerable groups -

                                                                    YES NO

	X

	X
Children under 18	

	X
Adults with learning or other disabilities	

	X
Very elderly people	

	X
Healthy volunteers who have a dependent or subordinate relationship to investigators	

Other vulnerable groups

If YES to any of the above, please specify and justify their inclusion

c. Inclusion and exclusion criteria

Please indicate, with reasons, the inclusion criteria for the project

Inclusion criteria: Trainee reporting radiographer undertaking postgraduate education in chest image interpretation, reporting radiographer trained to report on the musculoskeletal system and those willing to dedicate their time to the

Please indicate, with reasons, any exclusion criteria for the project

Exclusion criteria: Those with complete loss of vision in one eye, those with astigmatism, those which withdraw consent or participation in the study and participants currently taking experimental or cytotoxic drugs affecting vision

d. Will any inducements be offered? If 'Yes', please describe

N/A

e. Please describe how and where recruitment will take place

Trainee reporting radiographers on a postgraduate chest image interpretation programme will be recruited through university tutors. Reporting radiographers trained in musculoskeletal system reporting will be recruited through their attendance at Continuing Professional Development (CPD) events within

8. Ethical implications of the research

Please provide an assessment of the ethical implications of the project

There is a risk of embarrassment to the participants and a fear of error as they are outside the clinical environment and within a somewhat more test like environment. This will be addressed by reminding the participant that all of their eye tracking data and accuracy levels will be completely anonymised and stored in a secure manner at all times. Each participant's data will be allocated a number to ensure it is non-identifiable following data collection (i.e no names will be stored).

All patient's identifiable information present on images such as the patient's name, date of birth and health and care number will be removed from images before the study. Images included will not contain any rare abnormalities or pathologies which could readily identify an individual. The medical images accessed will have been previously reported on by a qualified professional and the patient management will already be underway, there is therefore no risk of impact on patient pathway or new medical revelations being made from the chest image interpretations. Participants are not required to wear additional devices, as the eye tracking device is non-invasive and unobtrusive, so there is no risk of experiencing pain or physical discomfort. Also there are no video recordings of the participant's face or body. Only the screen is recorded using screen-casting software and the participant's verbalisation (from thinking aloud) when interpreting the images is recorded.

9. Could the research identify or indicate the existence of any undetected healthcare concern?

Yes ☐ No ☒

If **Yes**, please indicate what might be detected and explain what action will be taken (e.g. inform subject's GP)

```



```

10. Risk Assessment **

Please indicate any risks to subjects or investigators associated with the project

```
The risk assessment did not identify any risks associated with this study.

```

**If you wish, you can use form RG1c – Risk Assessment Record (available from the Research Governance website) to help you assess any risks involved

11. Precautions

Please describe precautions to be taken to address the above

```
The risk of embarrassment and fear of error during the image interpretation will be
limited by the location of the study. The image interpretation will be performed in a
private area so that the participant can only be viewed by the researcher.
Although participants will be encouraged to be actively involved in the study, it will
be explained to them that their degree of involvement is a personal decision and
that it will be respected at all times. If the researcher detects distress of a
participant, or the participant discusses his/ her distress with the researcher, the
participant will be offered to terminate the study.
```

12. Consent form

It is assumed that as this study is being conducted on human subjects, an information sheet and associated consent form will be provided. A copy of the information sheet and form must be attached to this application. See Notes of Guidance.

If a consent form is not to be used, please provide a justification:

Consent form and information sheet is included in appendix 1.1

14. Care of personal information

Please describe the measures that will be taken to ensure that subjects' personal data/information will be stored appropriately and made available only to those named as investigators associated with the project.

All electronic data will be stored on a password protected Ulster University computer in accordance with the Data Protection Act 1998. Only those persons within the research team will have access to the data. Each participant will be issued a number to protect their identity and this number will then be used in any

14. Copyright

Has permission been granted to use all copyright materials including questionnaires and similar instruments?

Yes ⊠ **No** ☐

If **No**, please provide the reason

Once you have completed this form you should also complete form RG1d for all category D research and form RG1e for both category B and D research

Appendix 5.3 Informed consent

<u>**Appendix 1.0**</u>

Title of the study:
An evaluation of a training package in chest image interpretation with the aid of eye tracking technology

Investigators: Dr. Ciara Hughes, Dr. Sonyia McFadden, Dr. Raymond Bond, Dr. Jonathan McConnell

Doctoral Student: Laura McLaughlin

<u>Information Sheet and Consent Form for Participants</u>

You are invited to participate in an Ulster University research project. Before deciding whether or not you wish to be involved in the study it is important you understand what the research is for and what it will require you to do. Please read the following information and do not hesitate to ask questions about anything that may not be clear to you. Make sure you are happy before you decide to participate. Thank you for taking the time to read this invitation.

Correct interpretation of medical images is vitally important to patient diagnosis and subsequent treatment. However, relatively little research has been completed to evaluate the search strategies implemented by observers during image interpretation. Eye tracking has been used to demonstrate the thought processes of observers. It allows the researcher to gain an insight into how the subject actually reads and processes the medical data. The research seeks to develop and implement a training package which consists of a search strategy formed by expert opinion and a set of educational videos based on expert's eye gazes during chest image interpretation. We will then observe the effects of introducing the training package and its usefulness in image interpretation.

Project Aim:
To use eye tracking technology to investigate if a research informed training package can improve the accuracy of diagnosis of reporting radiographers.

We aim to identify:

- patterns of interpretation e.g. where subjects look the least and the most on medical images
- duration of each interpretation of different types of pathologies
- correlations between interpretation methods and diagnostic accuracy
- inter-rater reliability amongst all participants
- common interpretation errors and pitfalls
- adherence to the training package and its effect on accuracy of diagnosis

Each image interpretation session will last no longer than 30 minutes. You will be asked to look at a sample of chest images on a computer monitor and you will be asked to think aloud and provide a verbal diagnosis, which will be audio recorded. Your eye gaze path on the computer screen will be recorded using the non-obtrusive Tobii eye tracking technology. This will give us a greater insight into your thought processes.

Following your initial image interpretation we will then randomly allocate you to a control group or an intervention group, where you will not or will have access to the training package for practicing your image interpretation. Following this you will be asked to complete another image interpretation session using the eye tracking technology.

Do I have to take part?

Your participation in this study is voluntary and you are able to withdraw at any point without any given reason.

Will my taking part in this study be kept confidential?

All data will be stored securely and will be made available only to persons directly involved in conducting the study. No reference will be made in oral or written reports that could link you to the study. Your data collected will be anonymous i.e. we will not record your name. The results of the study will be used to further research. All information generated from this study will be kept in accordance with the Ulster University regulations. This will involve all participant data being stored within a data protection office for a minimum of 10 years following the study.

What if something goes wrong?

As this study has been carefully planned and approved by the Ulster University Ethics Committee, it is extremely unlikely that something will go wrong during this study. However, you should know that the university has procedures in place for reporting, investigating, recording and handling adverse events and complaints from study volunteers. In addition the university routinely insures for its staff to carry out research involving people. Further information on the complaints procedure can be found at the University's ''Research Ethics and Governance'' webpage (Internet address: http://research.ulster.ac.uk/rg/0208ResearchVolunteerComplaintsProcedure.pdf). Any complaint or concerns should be made, in the first instance, to the Chief Investigator identified for this particular study (contact details are below); complaints will be treated seriously and reported to the appropriate authority. The Chief Investigator will try their best to resolve this concern or complaint, however should this attempt fail the Research Ethics and Governance should be contacted (contact details below).

Who is organising the funding for this research?
This study is being funded by the Northern Ireland Department for Employment and Learning and will form part of a PhD study being undertaken at the Ulster University.

How do I go about participating?
If you have any questions at any time about the study or the procedures, you may contact Laura McLaughlin via email or phone:

Email: McLaughlin-L16@email.ulster.ac.uk **Tel:** 02890366191
Address: Room 01F125, Ulster University, Jordanstown campus, Shore Road, Newtonabbey, Co. Antrim, BT370QB

Further information:
Thank you for reading this information sheet, if you would like further information about the research study please contact:

Chief Investigator
Dr. Ciara Hughes
Address: Room 01B118, School of Health Sciences
University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB

Email: cm.hughes@ulster.ac.uk

Tel: 02890366227

Investigators:

Dr. Sonyia McFadden   s.mcfadden@ulster.ac.uk

Dr. Raymond Bond   rb.bond@ulster.ac.uk

Dr. Jonathan McConnell   jonathan.mcconnell@ggc.scot.nhs.uk

Research Ethics and Governance:

Mr. Nick Curry

Address: Room 26A17, Research & Innovation, University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB

Email: n.curry@ulster.ac.uk

Tel: 028903666229

Information Sheet and Consent Form for Participants (1.1) (continued)

Title of project:

An evaluation of a training package in chest image interpretation with the aid of eye tracking technology

Name of Investigators: Dr. Ciara Hughes, Dr. Sonyia McFadden,

Dr. Raymond Bond, Dr. Jonathan McConnell

Doctoral student: Laura McLaughlin

Please tick each box:

- I confirm that I have been given and have read and understood the information sheet for the above study and have asked and received answers to any questions raised

  [ ]

- I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without my rights being affected in any way

  [ ]

- I understand that researchers will hold all information and data collected, including audio recordings, securely and in confidence and that all efforts will be made to ensure that I cannot be identified as a participant in the study (except as might be required by law) and I give permission for the researchers to hold relevant anonymised personal data

  [ ]

- I agree to take part in the above study   **[ ]**

- I agree to photography of the experiment   **[ ]**

_____   _____   _____
Name of participant   *Signature*   *Date*

_____   _____   _____
Name of researcher   *Signature*   *Date*

Chief Investigator

Dr. Ciara Hughes

Address: Room 01B118, School of Health Sciences
University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB
Email: cm.hughes@ulster.ac.uk
Tel: 02890366227

Research Ethics and Governance:

Mr. Nick Curry

Address: Room 26A17, Research & Innovation, University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB

Email: n.curry@ulster.ac.uk

Tel: 028903666229

You may contact Laura McLaughlin via email or phone:

Email: McLaughlin-L16@email.ulster.ac.uk   **Tel:** 02890366191
Address: Room 01F125, Ulster University, Jordanstown campus, Shore Road, Newtonabbey, Co. Antrim, BT37 0QB

Appendix 5.4 Questionnaire

User_ID: _____

Questionnaire

Please answer the questions below.

_____

1. Number of years qualified as a radiographer: _____ **years**

2. Gender: Male | Female   *(Circle one)*

3. How many years have you been working clinically as a radiographer routinely producing and viewing radiographic images:

_____ years

4. Have you specialised in a field of radiography before enrolling in the chest image interpretation postgraduate programme?

☐ Yes   ☐ No

Please specify which field:

_____

5. Are you a reporting radiographer currently?

☐ Yes ☐ No

If you answered yes, in which speciality are you a reporting radiographer in?

_____

If you answered yes, how many years have you been officially reporting radiographic images?

_____ years

6. On a scale of 1 to 10 (1 being poor and 10 being excellent), rate your level of expertise in medical image interpretation?

(Circle a number)

Poor											Excellent
	1	2	3	4	5	6	7	8	9	10	

7. In your opinion do you believe that we need to gain insight into how clinicians interpret medical images?

☐ Yes ☐ No

8. Do you believe clinicians should be regularly assessed for their competency in reporting medical images?

☐ Yes ☐ No

9. Would you support the development of best practice guidelines on the process of interpreting different radiographic images?

☐ Yes   ☐ No

10. Rate how useful you believe monitoring eye tracking would be in assessing clinical competency in interpreting radiographic images?

(Circle a number)

Poor											Excellent
	1	2	3	4	5	6	7	8	9	10	

11. To prove competency, how many radiographic images would a clinician need to correctly diagnose as part of an assessment?

_____ out of _____ radiographic images.

12. How many radiographic images do you believe you diagnosed correctly?

_____ *(out of 20)*

13. Do you believe a training package including a log of images and the associated expert eye gazes could be beneficial to image interpretation training?

☐ Yes   ☐ No

14. Do you think the eye tracking system influenced your image interpretation?

☐   ☐

Yes   No

15. Did you find the eye tracking system easy to use?

☐ Yes   ☐ No

16. Do you have any other questions or comments?

| |
| |

Thank you for filling out the questionnaire – it is appreciated as part of our ongoing research.

Appendix 5.5 Survey amendment

UNIVERSITY OF ULSTER

UU Ref
No:

RESEARCH GOVERNANCE

Form RG6 **Notification of a proposed substantial amendment**

Chief Investigator:

Dr. Ciara Hughes

Approved Study Title:

An evaluation of a training package in chest image interpretation with the aid of eye tracking technology

New/Amended Title (if appropriate):

n/a

Type of Amendment (please indicate any that apply):

- Amendment to application form   [X]

- Amendment to description/protocol   [X]

- Amendment to the information sheet/consent or other   [X]
  supporting information

Summary of Changes:

A survey has been added to the end of the study.

This additional survey has been added to further investigate user feedback of the training package. It will be supplied with the questionnaire to participants.

The Qualitrix software will be used to present the survey. Participants will be asked to complete the survey online via email, the email will also include a link with which to access the survey. All participants will be asked to complete the survey following their participation

Additional ethical considerations:

This study remains low risk. No additional ethical considerations exist following the inclusion of the survey on the training package.

List of enclosed documents:

RG1a

Study protocol

Information sheet

Consent form

Questionnaire

Participant information letter-survey

Survey

Declaration:

I confirm that the information in this form is accurate and that implementation of the proposed amendment will benefit the study appropriately.

Filter Committee Decision

This amendment:

is appropriate to the needs of the study, is in category A and should be implemented   [   ]

is appropriate to the needs of the study, is in category B and should be considered by the University REC   [   ]

is NOT appropriate and should be reconsidered or withdrawn   [   ]

Appendix 5.6 Survey information

Appendix 2.0

Participant recruitment letter

Survey of the digital chest image interpretation training package

First of all, may I thank you for taking the time to read this information. We would like to invite you to complete an online survey on the chest image interpretation training package. Before you decide to do so, it is important to understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with others if you wish. Please feel free to ask us if there is anything that is not clear or if you would like more information.

Volunteering for the study

This survey aims to provide an opportunity for you to provide feedback on the chest image interpretation training package. We would sincerely appreciate your thoughts on the use of the training package, aspects of it which could be improved or features which you found particularly helpful. We also wish to explore general willingness of reporting clinicians to adopt currently available technologies aimed at supporting image interpretation. This information will allow us to consider changes to the training package. We can also combine the outcome measures of the eye tracking study with the thoughts and opinions of participants to provide us with a clear insight into the use of the training package in chest image interpretation performance.

Your role in the study

The survey is electronic, anonymous and confidential. All information provided is only available to the research team. The questions ask for information about your experience of using the training package and its use in chest image interpretation.

Most of the questions require a tick and, depending on your opinion, a small number may require a few words from you.

The survey will take approximately 10-15 minutes to complete.

To allow for the research group to associate each reply with the control or intervention group, we do ask that you input your User ID before completing the survey. This will not be used to identify you, but will only be used to associate your response to the survey with the relevant study group.

How your information will be treated?

The data achieved from your completed survey will be anonymised and stored on a locked computer in Ulster University. No one can access it apart from the members of the research team.

Useful contacts
If at any time you have any questions about this research project, please contact:
Ms Laura McLaughlin 02890366191
Email: McLaughlin-L16@email.ulster.ac.uk
Dr Ciara Hughes 02890366227 (Chief Investigator)
Email: cm.hughes@ulster.ac.uk
Dr. Sonyia McFadden 02890366224
Email: s.mcfadden@ulster.ac.uk
Dr. Raymond Bond 02890368156
Email: rb.bond@ulster.ac.uk
Dr. Jonathan McConnell +44 141 452 3629 (ext 83629)
Email: jonathan.mcconnell@ggc.scot.nhs.uk

Should you wish to make a complaint about any aspect of the conduct of this study please contact:
Ulster University Research Ethics and Governance:
Mr. Nick Curry
Address: Room 26A17, Research & Innovation, University of Ulster, Jordanstown campus, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB
Email: n.curry@ulster.ac.uk
Tel: 028903666229

Consent
Your consent to complete the survey is voluntary and confirmed by clicking the **Happy to Proceed** button at the bottom of this information sheet. This means that you have read and understand all the information provided above, and have no further questions about the study. If you have any further questions, please contact any member of the research team before completing the survey. Contact details are provided above.
Once you have clicked the **Happy to Proceed** button you will automatically be taken to the start of the survey. When you have completed the survey, please click on the **submit** button and the survey will be automatically loaded into a separate file for analysis. You do not have to complete the survey and you can stop at any time and close the programme without giving any reason. This will have no effect on you. Only after you have clicked the final submit button will it be assumed that you have given consent for the data to be used for the study.

Survey

If you wish to take part in the survey, please enter the address below into your computers web browser:
www...... <mark>(To be confirmed)</mark>

User ID: _____

<u>Survey</u>
Please could you complete the below survey to help us to understand your impressions of the online training package any problems you may identify within it can be addressed.
1. Please state your current job position:

Please type answer here – box extends as you type

2. How did you access the online training package? please circle the appropriate option(s)

(a) computer/laptop
(b) work computer/laptop
(c) tablet
(d) smartphone

3. Please estimate the frequency you used the training package:
(a) once or twice a day
(b) 2-3 times a week
(c) once a week
(d) rarely used
(e) never used

If you answered (d) rarely used or (e) never used in Q3. Why did you rarely/never use the training package?

Please type answer here – box extends as you type

4. Do you use a search strategy to interpret chest images?
(a) Yes
(b) No

5. If you answered yes in Q4. What search strategy do you prefer to use?
(a) the search strategy presented within the training package

(b) a search strategy devised by yourself

(c) a search strategy from another source

6. Please score the following features of the training package: (1= poor, 10= excellent)

Overall Layout

1   2   3   4   5   6   7   8   9   10

Accessibility

1   2   3   4   5   6   7   8   9   10

Visualisations (videos/images)

1   2   3   4   5   6   7   8   9   10

Content

1   2   3   4   5   6   7   8   9   10

Educational eye tracking videos

1   2   3   4   5   6   7   8   9   10

Search strategy training package

1   2   3   4   5   6   7   8   9   10

7. Do you feel that the training package improved your skills in interpretation (e.g. speed, accuracy, confidence)?

(a) Yes

(b) No

If you answered yes in Q7 please give details:

Please type answer here – box extends as you type

8. What features did you find most useful in the online training package?

Please type answer here – box extends as you type

9. What features did you find least useful in the online training package?

Please type answer here – box extends as you type

10. What suggestions (if any) would you make to improve the training package?

Please type answer here – box extends as you type

Please rate the below statements, based on how you feel, by circling a value you wish to select on the scale:

11. I think that I would like to use this training package frequently:

Strongly Disagree	1	2	3	4	5	Strongly Agree

12. I found the training package unnecessarily complex:

Strongly Disagree	1	2	3	4	5	Strongly Agree

13. I thought the training package was easy to use:

Strongly Disagree	1	2	3	4	5	Strongly Agree

14. I think I would need the support of a technical person to be able to use this system:

Strongly Disagree	1	2	3	4	5	Strongly Agree

15. I found the various functions within the training package were well integrated:

Strongly Disagree	1	2	3	4	5	Strongly Agree

16. I thought there was too much inconsistency in this training package:

Strongly Disagree	1	2	3	4	5	Strongly Agree

17. I would imagine that most people would learn to use this training package very quickly:

Strongly Disagree	1	2	3	4	5	Strongly Agree

18. I found the training package very cumbersome to use:

Strongly Disagree	1	2	3	4	5	Strongly Agree

19. I felt very confident using the training package:

Strongly Disagree	1	2	3	4	5	Strongly Agree

20. I needed to learn a lot of things before I could get going with this training package

Strongly Disagree	1	2	3	4	5	Strongly Agree

21. Will you continue to use the online training package following this study?
(a) Yes
(b) No

22. Please supply any further comments/additional feedback on the training package:

Please type answer here – box extends as you type

Thank you for taking the time to complete this survey on the training package and for your participation within the study, your help and input in the study is really appreciated.

Please select 'save' before exiting.

Appendix 5.7 Training tool use

Educational programme use

Participant	Duration	Date
IA4	45.95	Oct 24th 2016
IA5	28.072	Nov 10th
IA5	3.973	
IC5		
IC2	220.301	Nov 29th 2016
IC5		
IC5		
IC1	3.727	
IC1	47.826	Dec 22nd 2016
IB1		Dec 28th 2016
A1		Jan 3rd 2017
A1	6.942	Jan 3rd 2017
IC2	1070.394	Jan 31st 2017
IA3	322.169	Feb 2nd 2017
IC2	5.943	
IC5		Feb 4th 2017
IB1		Feb 14th 2017
IC2	503.444	Mar 6th 2017
IC2	2.209	
IC2	2.01	
IC2	1.917	
IC2	230.858	Mar 29th 2017
IC1		Apr 7th 2017
IC2	235.054	Apr 10th 2017
IB1	5.607	
IB1	6.499	
IC2	281.628	May 8th 2017

IC7		May 9th 2017
IC7		May 9th 2017
Control	14.32	May 23rd 2017
Control	10.297	May 23rd 2017
Control	5.164	May 23rd 2017
IC15	277.741	May 25th 2017
IC14	381.952	May 25th 2017
IC16	65.858	May 25th 2017
IC11	5.627	
IC11	9.653	
IC1	6.172	
IC2	3.169	
IC5	2874.352	Jul 19th 2017
IC5	7.499	Jul 19th 2017
IC5	361.755	Jul 19th 2017
Control	34.234	Jul 25th 2017
Control	492.081	Jul 25th 2017
IC14	473.931	Aug 14th 2017
Control		Aug 18th 2017
Control		
Control		
Control		
Control		
Control		
Control		
Control		
Control		
IC11	157.561	Aug 31st 2017
IC11	160.628	Aug 31st 2017
IA1	3.764	
IC15	953.149	

IC15	2.827	
Control		
Control		
IC11	4.764	
IC14	6.808	
IC14	5.627	
ID14	980.446	
C17	87.501	Feb 16th 2018
C17		Feb 16th 2018
C18		Feb 16th 2018
C17		Feb 16th 2018
C12	16.131	
C12		Feb 21st 2018
C15	4.031	

Search strategy training tool use

Participant	Date
IC2	494.786
IA3	634.438
IC2	786.069
IC2	514.795
IC2	390.994
IC2	266.281
IC2	626.612
IC2	1032.081
IC2	614.072
IC2	837.534
IC2	749.301
IC2	566.377
IC2	293.42
IC2	329.584

IC2	282.033
IC2	501.734
IC2	438.394
IC2	501.528
IC2	524.744
IB1	383.372
IC2	548.194
IC2	639.449
IC2	390.392
Control	23.269
IC2	576.281
IC2	323.811
IC2	7732.196
IC2	470.799
IC2	202.949
IC2	206.657
IC2	393.34
Control	250.092

Appendix E - Chapter 6

Appendix 6.1 Publications, presentations and awards

Location	Authors	Title	Year
Publications			
Methods employed for chest radiograph interpretation education for radiographers: A systematic review of the literature	L McLaughlin, J McConnell, S McFadden, R Bond, C Hughes	2017	Radiography Journal
Computing Eye Gaze Metrics for the Automatic Assessment of Radiographer Performance during X-ray Image Interpretation	L McLaughlin, R Bond, C Hughes, J McConnell, S McFadden	2017	International Journal of Medical Informatics
Digital training platform for interpreting radiographic images of the chest	L McLaughlin, N Woznitza, A Cairns, S McFadden, R Bond, C Hughes, A Elsayed, D Finlay, J McConnell	2018	Radiography Journal
Presentations			
The use of eye tracking	L McLaughlin, S	2017	UK Radiological and

technology to assess radiographer interpretation of X-ray images	McFadden, C Hughes, J McConnell, R Bond		Radiation Oncology Congress
Development of a digital platform to aid chest radiographic image interpretation	L McLaughlin, R Bond, C Hughes, J McConnell, N Woznitza, A Elsayed, A Cairns, D Finlay, S McFadden	2017	European Congress of Radiology
The use of a digital training platform of chest image interpretation to reporting radiographers of the musculoskeletal system	L McLaughlin, C Hughes, R Bond, J McConnell, N Woznitza, A Cairns, A Elsayed, S McFadden	2018	20th International Society of Radiographers and Radiological Technologist World Congress
Advancing clinical practice and education with a digital training platform of chest image interpretation	L McLaughlin[1], CM Hughes, R Bond, J McConnell, SL McFadden	2018	Leading the way: Advancing practice in radiography
Posters			
The effect of a digital chest image interpretation training tool on interpreter performance	L McLaughlin, S McFadden, R Bond, J McConnell, N Woznitza, A Cairns, A Elsayed, D Finlay, C Hughes	2017	8th Annual Translational Medicine Conference

The Potential of Hybridising Interactive Eye Tracking Technology with Decision Support in Medical Image Interpretation	L McLaughlin, R Bond, J McConnell, C Hughes, S McFadden	2018	Human Computer Interaction Conference Belfast
Trial of a digital training tool to support chest image interpretation in radiography	L McLaughlin, C Hughes, S McFadden, R Bond, J McConnell, N Woznitza, A Elsayed	2018	European Congress of Radiology
Awards			
Full bursary for conference attendance		2017	UK Radiological and Radiation Oncology Congress
Finalist of The Belfast HSC Trust Science driving innovation in healthcare delivery award		2017	Advancing Healthcare Northern Ireland awards