Title: Computing Eye Gaze Metrics for the Automatic Assessment of Radiographer Performance during X-ray Image Interpretation

Keywords: radiography, eye tracking, interpretation, musculoskeletal, chest,

Authors:

Laura McLaughlin BSc [1] (mclaughlin-l16@email.ulster.ac.uk) (corresponding author)

Raymond Bond PhD 2 (rb.bond@ulster.ac.uk)

Ciara Hughes MSc PhD[1] (cm.hughes@ulster.ac.uk)

Jonathan McConnell PhD [3] (jonathan.mcconnell@ggc.scot.nhs.uk)

Sonyia McFadden PhD [1] (s.mcfadden@ulster.ac.uk)

Affiliations:

[1]Centre for Health and Rehabilitation Technologies, Institute of Nursing and Health, School of Health Sciences, Ulster University (Northern Ireland)

[2]Computer Science Research Institute, School of Computing and Mathematics, Ulster University (Northern Ireland)

[3]Queen Elizabeth University Hospital, NHS Greater Glasgow and Clyde, (Scotland)

Authors: Laura McLaughlin, Dr. Raymond Bond, Dr. Ciara Hughes, Dr. Jonathan McConnell, Dr. Sonyia McFadden

Abstract

Aim: To investigate image interpretation performance by diagnostic radiography students, diagnostic radiographers and reporting radiographers by computing eye gaze metrics using eye tracking technology.

Methods: Three groups of participants were studied during their interpretation of digital 8 radiographic images including the axial and appendicular skeleton, and chest (prevalence of normal images was 12.5%). A total of 464 image interpretations were collected. Participants consisted of 21 radiography students, 19 qualified radiographers and 18 reporting radiographers who were qualified to report on the musculoskeletal (MSK) system.

Outcome measures: Eye tracking data was collected using the Tobii X60 eye tracker and subsequently eye gaze metrics were computed. Voice recordings, confidence levels and diagnoses provided a clear demonstration of the image interpretation and the cognitive processes undertaken by each participant. A questionnaire afforded the participants an opportunity to offer information on their experience in image interpretation and their opinion on the eye tracking technology.

Results: Reporting radiographers demonstrated a 15% greater accuracy rate ($p \leq 0.001$), were more confident ($p \leq 0.001$) and took a mean of 2.4s longer to clinically decide on all features compared to students. Reporting radiographers also had a 15% greater accuracy rate ($p \leq 0.001$), were more confident ($p \leq 0.001$) and took longer to clinically decide 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. Eye tracking patterns, presented within heat maps, 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 radiographer's competency.

Conclusion: The accuracy and confidence of each group 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. This is the first study to use eye tracking technology to assess image interpretation skills between various groups of different levels of experience in radiography, especially on a combination of the MSK system, chest cavity and a variety of pathologies.

1.1 Introduction

In the mid 1990's role progression in radiography allowed appropriately trained radiographers to undertake reporting of radiographic images within a specialised field [1, 2]. Since then there has been a growing body of evidence to support the radiographer's ability to interpret images of the MSK skeleton [3, 4, 5, 6]. Chest radiographic image interpretation does not have the same supporting evidence and is regarded as a more difficult skill due to the number of overlying organs present within the chest cavity and multiple pathologies which can be demonstrated within a chest radiographic image [7, 8]. However, there have been studies completed investigating the interpretation skills of radiographers in accident and emergency images and a combination of both the appendicular and axial skeleton, [4, 9].

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 [10,11]. Studies have used participants with various levels of expertise to try establish the differing image interpretation patterns shown by different groups. Eye tracking technology provides an insight into the subconscious cognitive processes during their image

interpretation. Donovan et al. 2008 [10] and Manning et al. 2006a [11] 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 Receiver Operating Characteristic (AFROC) methodology. Donovan et al. (2008) [10] 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% ($p<0.05$). Jackknife alternative free-response receiver operating characteristic (JAFROC), the analysis software, 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’ [10]. There was less of an effect noted in the performance of naïve (students and staff from other disciplines) and expert (radiologists and reporting radiographers) participants leading to the conclusion that perceptual feedback of eye tracking may be beneficial to student radiographers earlier in their training. Manning et al. (2006a) [11] utilised eye tracking technology to monitor performance measures; results were significantly better in the expert/trained radiographers in comparison to the rest of the studied cohort ($p=0.046$). Manning et al. 2006a [11] 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 also assumed this method after receiving their training. Donovan et al. (2008) [10] noted that the eye tracking data of Level 1 (first year) and Level 2 (second year) undergraduate student radiographers displayed a great deal of variability. Eye tracking technology provided valuable information on how the participant groups viewed images [10, 11]. The use of a range of radiographic images within differing body areas combined with wider pathologies could challenge the participant and stimulate the radiographer to interpret the image using a different search strategy. 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. Therefore, our study was completed

by using eye tracking software to examine the image interpretation process for a range of participants using a range of pathologies and anatomical areas.

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 [12, 13]. The 'think-aloud' technique is when the participant verbalises their thought processes during their interpretation. As a result, the think-aloud protocol 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, we can further understand the participant's image interpretation process which complements the eye gaze data [13]. Although an anticipated higher accuracy and confidence level from the experienced and qualified reporting clinicians was to be expected, we were interested in 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.

1.1.1 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. 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 methods and diagnostic accuracy
- inter-rater reliability amongst all participants and the common interpretation errors and pitfalls.

1.2 Materials and methods

1.2.1 Ethics: Ethical approval for this study was granted by the Ulster University Research and Ethics Filter Committee. Written informed consent was obtained from each participant prior to the study.

1.2.2 Inclusion/Exclusion criteria: Students, novices and expert radiographers in recording or interpreting radiographic images were included. Participants were included if they were willing to dedicate their time to the study and supply written informed consent for their agreed participation.

Participants were excluded if they did not have training or experience in radiography/medical image interpretation or if they chose to withdraw participation in the study.

1.2.3 Participants: Radiography students (n=21), with at least one year undergraduate education in diagnostic radiography and imaging, were recruited within Ulster University. Experienced radiographers (n=19) of various specialities and years of experience were recruited through their attendance of the UK Society and College of Radiographers conference within Northern Ireland. Experienced reporting radiographers (n=18) trained to interpret images of the MSK skeleton were recruited at a Reporting Radiographers Interest Group Scotland meeting.

1.2.4 Images: The study included 8 Joint Photographic Experts Group (JPEG) images of the appendicular skeleton, axial skeleton and chest cavity; 6 MSK and 2 chest X-ray images were used. The same 8 digital 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. 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, facture, pneumothorax, lung

mass etc. We chose the images to represent and test participants on a range of abnormalities. Participants were asked to form a diagnosis solely on the image they were presented.

1.2.5 Reference standard: The digital 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. Ethical approval was therefore not required for the use of patient images. For completeness, a senior reporting radiographer (member of the research team) was asked to provide a written diagnosis and the 'gold standard' diagnosis for each image was then agreed by considering both diagnoses.

1.2.6 Equipment: The Tobii Studio X60 eye tracker and the Tobii studio software© were utilised for data collection and for computing eye gaze metrics ([14] 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. Equipment calibration was completed prior to the study. Care and stringent checks were 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" [15]. Measuring eye tracking quality allows the collection of eye tracking data from the participant's performance to be monitored.

1.2.7 Prior to the study: Participant distance from the viewing monitor and chair height was altered to complete calibration successfully and receive optimum eye tracking data. As the 'think aloud' method was incorporated within the study, all participants were reminded to verbalise their thought processes as much as they could.

1.2.8 Outcome measures:

Participant accuracy was measured within the study. Images were marked as correct (1) if the reference diagnosis was stated or similar to what the participant described and (0) if incorrect.

Once the diagnosis was provided, a confidence level on the given diagnosis was requested. 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 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).

1.2.9 Data analysis

Descriptive statistics and tests of normality were completed before deciding on which hypothesis tests to use. Spearmans *rho* coefficient was then used to investigate correlations. A one way analysis of variance (ANOVA), Kruskal-Wallis and Mann-Whitney U tests were completed to investigate significant differences.

1.3 Results

1.3.1 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 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%

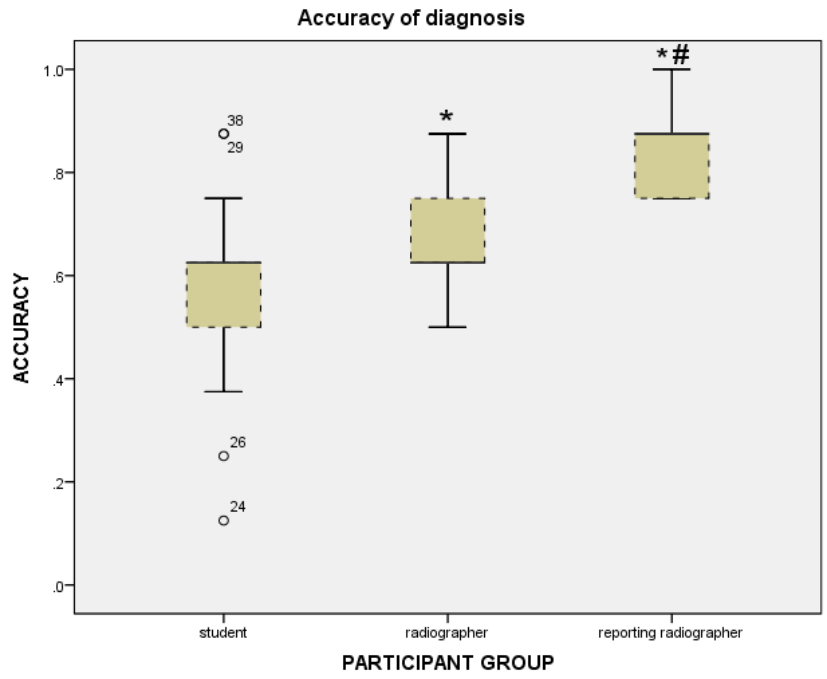
1.3.2 Confidence and accuracy

Table 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 unconfident and 10 being confident in their given diagnosis). *indicates significantly different to students (P<0.05) # indicates significantly different to radiographers (P<0.05)*

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



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

Reporting radiographers were more confident in their given diagnosis than radiographers (p<0.001) and students (p<0.001) (Table 2). Reporting radiographers had a greater median confidence level of 2.2 (on a scale of 1-10 with 1 being not confident and 10 being very confident in their given diagnosis) 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≤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 1).

1.3.3 Eye tracking

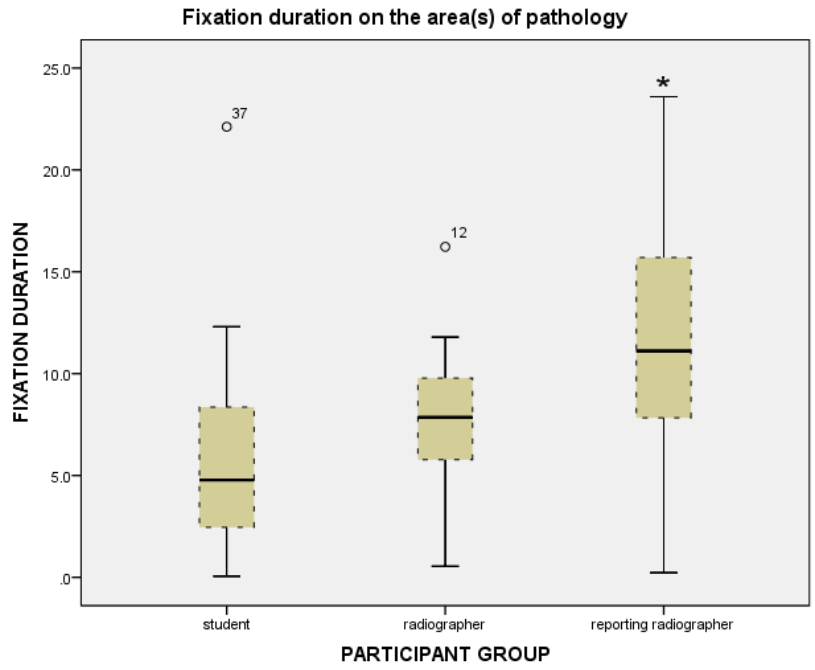
Table 3: Eye tracking data for each participant group which was collected from the area of pathology within each image

	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) # significantly different to radiographers (P<0.05)*

Figure 2: 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

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 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). However, there were no significant differences between the total times to first fixate (Table 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; Figure 2). 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). However, a trend was evident 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/competency increased (Table 3; Figure 2).

Table 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)*

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 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 4), before coming to a decision on the diagnosis. Yet students had the lowest median accuracy (%) and reporting radiographers had the highest accuracy (%) (Figure 1).

1.3.4 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 MSK system images (Figure 1). Reporting radiographers only incorrectly diagnosed chest images, with 10/18 identifying a pneumothorax in image 1 and only 5/18 identifying a round opacity lesion in image 6.

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.

1.3.4 Fixation frequency

The fixation frequency is the number of fixations per second (hertz). 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 and were possibly more 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 difference between these two groups was not as great as that observed between students and reporting radiographers ($p\leq0.001$) or reporting radiographers and radiographers ($p\leq0.001$). 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.

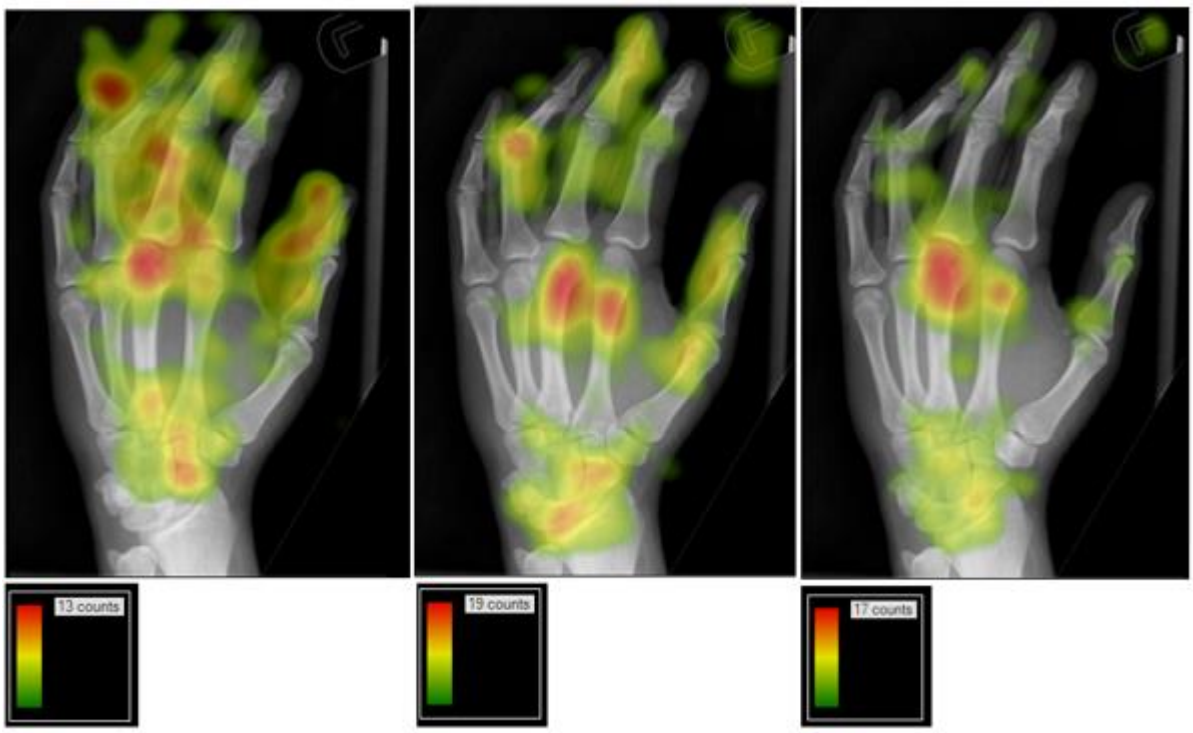
There was a positive correlation between total confidence and total fixation frequency for students ($r=0.21$, $P<0.001$). The more sporadic the students were in their interpretation, the more confident they were in their diagnosis. However, there was a negative correlation between total confidence and total fixation frequency for radiographers ($r=-0.62$, $P<0.001$) and for reporting radiographers ($r=-0.20$, $P<0.001$). As the participants with greater experience became more sporadic their confidence levels decreased (Table 2).

1.3.5 Heat map results:

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

Figure 5: 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.

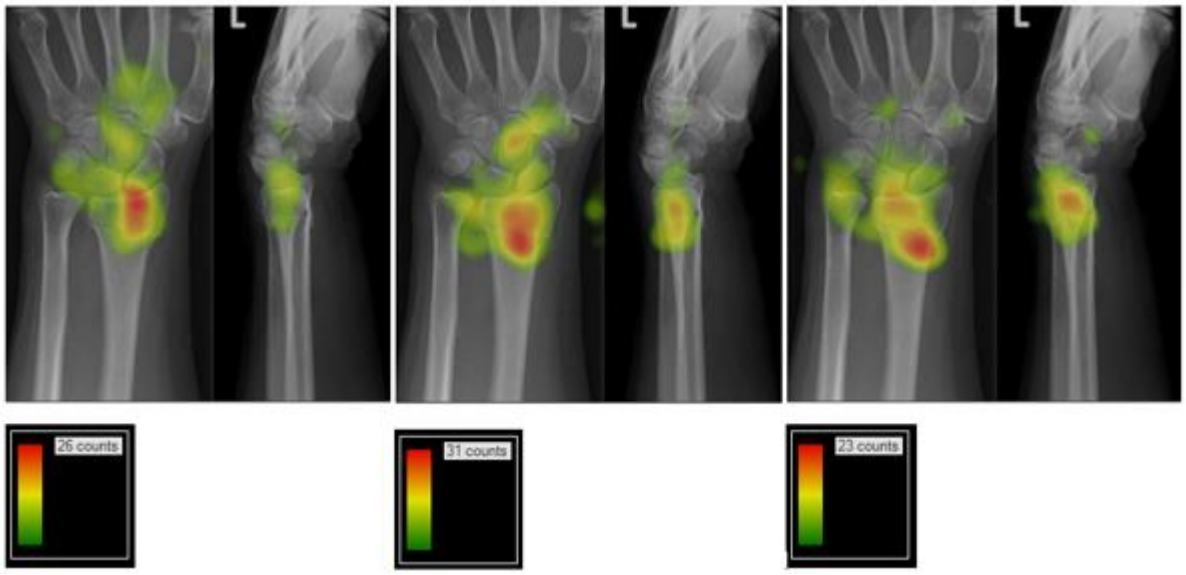
(a) **(b)** **(c)**



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 was shown by the reporting radiographers in their fixation areas, as they began to "zone in" on the areas of concern. However, the students demonstrate a large variability in their gaze.

Figure 6: 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.

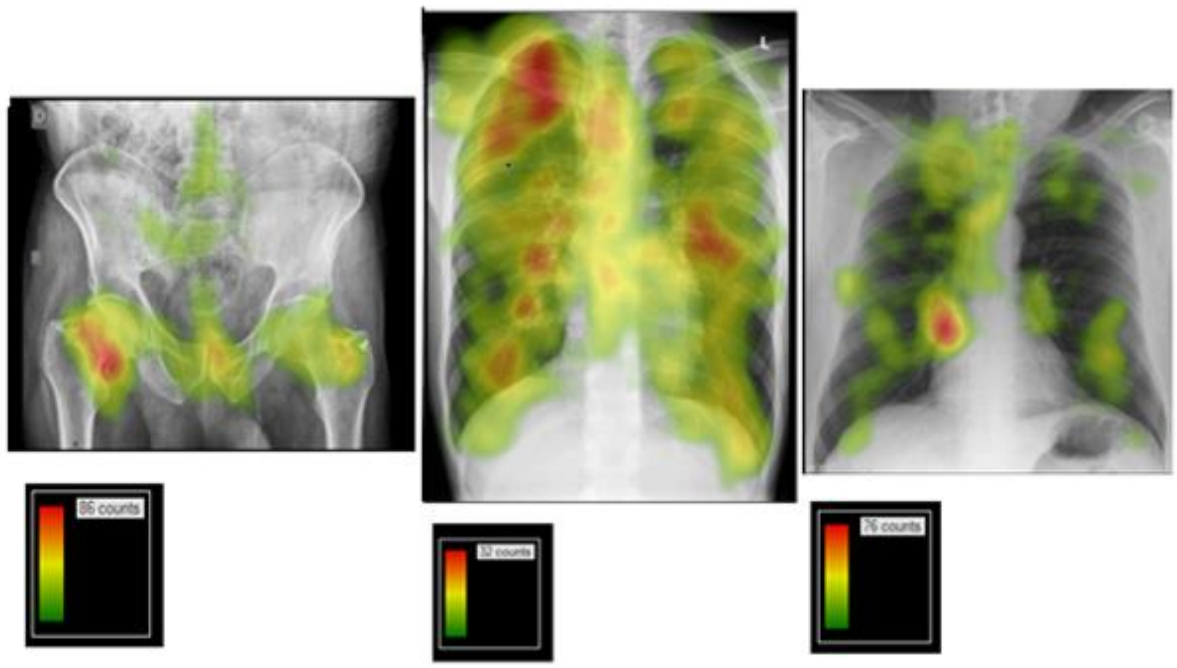
(a) (b) (c)



Radiographers and students demonstrated fewer fixations on the second pathology of the fractured ulna styloid than the reporting radiographers. 13/19 radiographers and 11/21 students failed to report the secondary pathology of the fractured ulna. All of the reporting radiographers identified the second pathology.

Figure 7: 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.

(a) **(b)** **(c)**

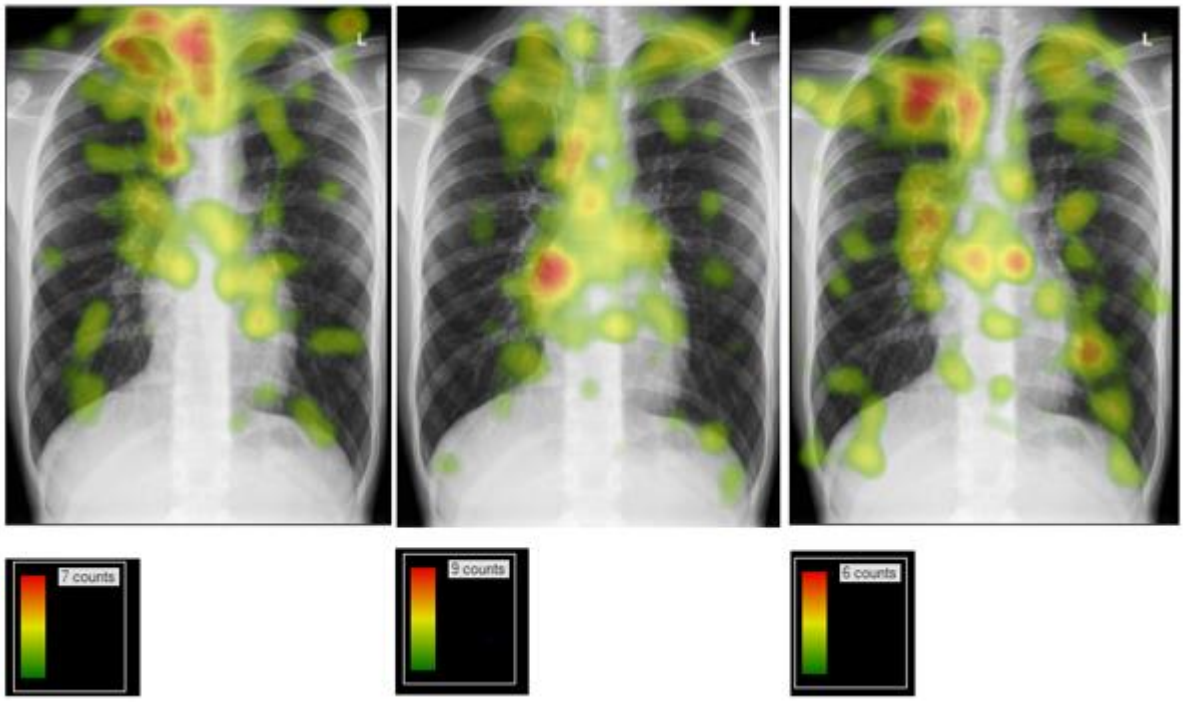


The reporting radiographers had a greater number of fixations at each of the areas of pathology during the first 10 seconds of their interpretation but importantly it is the only group to have a high fixation count on the area of pathology in each image (including chest images in which they have not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

The reporting radiographers demonstrated greater variation in their gaze patterns when viewing the chest images than MSK imaging, which is likely to be due to their training in MSK reporting only.

Figure 8: 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.**

(a) **(b)** **(c)**



The radiographers had fewer and smaller areas of fixations than the students and reporting radiographers and 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.

1.4 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 [5, 7, 8]. Accuracy of axial and appendicular reporting by radiographers was demonstrated to be between 91.8%-93.7% post training 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 [5, 7].The high median accuracy of reporting radiographers (median score of 87.5% of the 6 MSK and 2 chest images within the study) may reflect the higher fixation and visit counts from this group. Reporting radiographers often gave a more detailed explanation of the pathology, possibly increasing their fixations and visits on the areas of pathology to assist their formation of the diagnosis. Reporting radiographers had 100% accuracy in their reporting of the MSK images and this group therefore only incorrectly diagnosed the chest radiographic images in which they had no specific training. 10/18 reporting radiographers identified a pneumothorax and 5/18 identified a round opacity lesion, suggesting the pathology type could have led to misinterpretation. Reporting radiographers spent more time concentrating on the images which they were less familiar with, chest radiographic images, leading to a negative correlation within this group between decision time and accuracy. Due to the reporting radiographers lack of training/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. As decision time increased for this particular group their accuracy decreased. These results are supported by evidence of Manning et al. 2006b [16] whereby incorrect negative decisions were characterised by longer dwell times. However, Manning et al. (2006b) [16] 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. The training which reporting radiographers have received can provide them with confidence in their professional role. Coleman et al. (2009) [17] 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 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 care. 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.

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 [11]. Manning et al. (2006) [11] 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) [11]. 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) [11] permitted a maximum observation time of 40 minutes. This was not exceeded and the longest time taken to complete image interpretation within this study was 14 minutes 51 seconds. However, this methodological feature may again have affected how each cohort of participants approached each study.

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. Reporting radiographers are trained to complete satisfaction of search. Satisfaction of search suggests that once a pathology has been identified, the image interpreter applies further diligence to continue in searching the image for more than one pathology, to avoid a clinically significant pathology being missed [18, 19]. 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 [7, 20]. 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). Kok et al. (2015) [21] 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) [10] where Level 1 and Level 2 groups demonstrated a great deal of variability within their eye tracking data is alike 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) [11] noticed fewer film areas were inspected by radiographers following training, this is similar to the lower variability within MSK images interpreted by reporting radiographers. The training delivered within Manning et al. (2006a) [11] was 6 months chest image interpretation training and therefore although the reporting radiographers within our study demonstrated increased variability within chest images, their lower variability MSK imaging is similar to Manning et al. (2006a) [11]. Lower variability was seen both in this study and previous studies within participant groups who were trained to interpret images relevant to their education. 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, reduced 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 clinical and reporting experience is opined to have contributed to how they approached the task of interpreting the chest images. However, given their lack of training in interpreting chest images, it was expected that there would be reduced confidence when viewing these images. Nevertheless, heat maps extracted suggest that the reporting radiographers adopted a systematic approach within the chest images and the greater variability of their eye gazes reveals their aim to achieve this. Some of the reporting radiographers mentioned the increased difficulty performing this task compared to the interpretation of the MSK images.

Radiographers and students demonstrated fewer fixations on the second pathology of the fractured ulnar styloid in image 4 than the reporting radiographers. 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', combined with their knowledge of common mechanisms of injury and patterns of abnormality. The radiographer's low score when identifying the second pathology of the fractured ulna styloid process reflects the 67% of radiographers within the Coleman et al. (2009) [17] study who failed to notice further fractures once having identified one fracture within an image.

In general, the heat maps provided information on the groups' approach to image interpretation. 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 (a chest image), the students had only provided a small number of fixations on the chest pathology within the first 5 seconds, 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 8/21 students correctly identified the chest

pathology, heat maps could therefore be a good indicator of whether the participant group will diagnose accurately in extreme cases.

1.5 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 quality. Completing the study within a test environment, rather than a clinical environment, perhaps caused participants to err on the side of caution.

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. Clinical viewing conditions were replicated as fully as possible and meet the minimum expectations as described by Spigos et al. (1999) [22] and the Royal College of Radiologists (RCR) [23] guidelines.

Prevalence of normal images (12.5%) was a poor representation of the prevalence of normal images 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 [24, 25].

The eye tracking sampling quality collected from the participants varied. This was not ideal however we thought it best to include all of the participants rather than excluding them over the eye tracking quality received, which cannot be completely controlled. Data quality can be influenced by participants, operators, the task, recording environment, geometry or the eye tracking design [15].

1.6 Conclusion:

Reporting radiographers were more confident in their 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. 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) [11]. 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 MSK images by reporting radiographers is an established role progression within the radiographic profession. This study reinforces evidence for the ability of radiographers to complete a role successfully which they have been appropriately trained to complete, owing to their high accuracy and ability to complete interpretation systematically to assess all areas

of the image. It is vitally important that the role progression within reporting radiography is supported, the high standards of their performance are acknowledged and that their work continues to be audited to instil confidence in this role throughout the healthcare system.

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 the MSK system, chest cavity images 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 needs to be undertaken on the effect of training on the image interpretation of participants.

2.2 References

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

[2] 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.

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

[4] 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,* 76(901) 57-61.

[5] 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.

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

[7] 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 6p.

[8] 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.

[9] 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.

[10] Donovan, T., Manning, D.J. and Crawford, T. (2008) Performance changes in lung nodule detection following perceptual feedback of eye movements. *Medical Imaging*, 691703-691709.

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

[12] 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).

[13] 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.

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

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

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

[17] 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.

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

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

[20] 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

[21] 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* 21 189-205.

[22] Spigos, D., Tzalonikou, M., Bennett, W., Mueller, C. and Terrell, J. (1999) Accuracy of digital imaging interpretation on an 1K× 1K PC-based workstation in the emergency department. *Emergency Radiology,* 6(5), 272-275.

[23] The Royal College of Radiologists, (2008) RCR Picture archiving and communication system (PACS) and guidelines on diagnostic displays, London.

[24] 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.

[25] 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.