Title: The influence of image interpretation training on the accuracy of abnormality detection and written comments on musculoskeletal radiographs by South African radiographers.

Article Type: Quantitative Research article

Corresponding Author: Mrs. Lynne Janette Hazell, M tech

Corresponding Author's Institution: University of Johannesburg

First Author: Lynne Janette Hazell, M tech

Order of Authors: Lynne Janette Hazell, M tech; Jenny Motto, Mtech; Lucky Chipeya, MPhil

Abstract: ABSTRACT
Purpose: The study was conducted in two government hospitals in Johannesburg, South Africa. The current study investigated the extent to which image interpretation and relevant terminology training would improve the accuracy and descriptive comments provided on musculoskeletal images by South African radiographers.

Methods: Nine radiographers interpreted an image bank comprised of 100 skeletal radiographs (50% abnormal) both prior and on completion of a tailored education programme in image interpretation. Radiographer comments were compared to the reference standard diagnosis (single experienced radiologist) and deemed to be correct, partially correct or incorrect. The radiographers were assessed for sensitivity, specificity and accuracy on the image bank pre and post intervention. After testing for normality of the data a Wilcoxon sign rank test was used for non-parametric paired data.

Results: Radiographer accuracy (71.04% to 78%), sensitivity (83.73% to 87.28%) and specificity (59.62% to 70.34%) all improved post education programme. The accuracy of radiographer comments demonstrated a statistically significant improvement from to (Wilcoxon value, z= -2.66 p= 0.008). Incorrect radiographer comments also decreased (24.1% to 17.78%, Wilcoxon value, z= -1.96, p= 0.05). Radiographer vocabulary used when describing abnormalities was more in line with the reference standard diagnosis following training.

Conclusion: This cohort of radiographers demonstrated increased accuracy when commenting on skeletal radiographs with a significant reduction in incorrect comments. Future work should include assessing accuracy and commenting in the clinical environment and whether the improvement in commenting is maintained over time.
3rd December 2014.

To whom it may concern,

I, Lynne Hazell declare there is no potential conflict of interest.

Yours truly,

[Signature]

L. J. Hazell,

Lecturer,

Department of Radiography.

Faculty of Health Science.

University of Johannesburg.
Ms. Ref. No.: JMIRS-14-082R1

Title: The influence of image interpretation training on the accuracy of abnormality detection and written comments on musculoskeletal radiographs by South African radiographers.

Journal of Medical Imaging and Radiation Sciences

Dear Mrs. Lynne Janette Hazell,

The reviewers have commented on your above paper and asked for further revisions. Please address the reviewers' comments (included below), and I invite you to revise and resubmit your manuscript.

When submitting a revised manuscript, please also: a) outline each change made (point by point) as raised in the reviewer comments AND/OR b) provide a suitable rebuttal to each reviewer comment not addressed.

To submit your revision, please do the following:

1. Go to: http://ees.elsevier.com/jmirs/
2. Enter your login details
3. Click [Author Login]
   This takes you to the Author Main Menu.
4. Click [Submissions Needing Revision]

I look forward to receiving your revised manuscript. If you have any questions about any of the comments or about the resubmission process, please contact me anytime at editor@camrt.ca.

Yours sincerely,

Carly McCuaig
Managing Editor
Journal of Medical Imaging and Radiation Sciences

Associate Editor's Comments

Improvements are noted in this revised manuscript. However, the reviewers indicated it is still inadequate for publication at this stage. The major weaknesses include data analysis and results, inappropriate use of references and lack of clarity in writing. Please address all comments provided carefully as a final decision will be made after the second round of revisions. I look forward to seeing this article published in our journal.

Reviewers' Comments

Reviewer # 1

There has been improvement since the first version. As there were no page numbers on my latest version, I have taken page 1 as the Introduction page.

Introduction
1. Page 2, 3rd paragraph grammar - should read either "shorter courses" or "short courses"
Corrected

2. Page 3 4th para, suggest it reads... "The current study investigated the extent to which....improve the accuracy and descriptive comments..." Please review the study purpose of Abstract as well.
Corrected and addressed in the purpose

Methods

3. Page 4, 1st para, suggest it reads ... "Nine volunteer radiographers were recruited.." Corrected

4. Page 6 1st paragraph last sentence suggest this is put in the discussion and reworded slightly to say "....becoming the norm, image interpretation would be assessed with high quality images which could have an impact on the accuracy" Corrected

5. Should also be noted as a limitation that comments were not made available for the images that radiographers incorrectly called. Added as a limitation

6. Page 7 2nd para last sentence is not needed. removed

Results

7. Page 7 table one. I am still not convinced of the benefit of this table as it is not a recognised measure of image interpretation. Have added the table with sensitivity, specificity and accuracy for individuals and the mean value

8. You have not done any statistical analysis of the sensitivity and specificity which is crucial. See above not all data was statistically significant although the researcher felt it was important when compared with previous studies Hargreaves and Mackay also statistically not significant, however, in- line with the results from the literature Brealey et al.

9. Table 3 no units included in the pre-test/post-test columns. Presumably these are percentages Now Table 4 % added

Discussion

10. Page 10 discussion section half way down re Coleman and Pipers study I don't think you are using a standard measure of accuracy which is defined as Accuracy = (TN + TP)/(TN+TP+FN+FP) = (Number of correct assessments)/Number of all assessments). I suggest you redo your table 1 to reflect the accuracy measure defined here and use the outcomes in your discussion point re Coleman and Piper. Otherwise you are not comparing like with like. Have changed the calculation of accuracy in- line with Coleman and Piper see Table 3

11. Page 10 last paragraph what does the 85% refer to? "Pre-test 89.9%, post-test 93% and 85%" Should be 95% from Mc Connell study

12. Page 12 3rd para incorrect last line should read "sensitivity declined" but the specificity did not change and the accuracy was not measured in that paper. corrected

Conclusion

13. Page 12 conclusion section 1st paragraph why would 85% accuracy be required? removed
Reviewer # 2

The authors have improved this manuscript. There are some further improvements required however before this is suitable for publication.

Abstract

1. The abstract is a concise study summary, ideally 200-250 words; indeed some journals specify a maximum length. An alternate could be:

Purpose: Please see my comments for the Introduction below.
Methods: Nine radiographers interpreted an image bank comprised of 100 skeletal radiographs (50% abnormal) both prior to and upon completion of a tailored education programme in image interpretation. Radiographer comments were compared to the reference standard diagnosis (single experienced radiologist) and deemed to be either correct, partially correct or incorrect. Sensitivity, specificity and accuracy were calculated and a Wilcoxon sign rank test used.
Results: Radiographer accuracy (a to b%), sensitivity (x to y%) and specificity (m to n%) improved with training. The accuracy of radiographer comments demonstrated a statistically significant improvement from c% to d% (Wilcoxon value, p=x). Incorrect radiographer comments also decreased from e% to f% (Wilcoxon value, p=x). Radiographer vocabulary used when describing abnormalities was more in line with the reference standard diagnosis following training.
Conclusion: This cohort of radiographers demonstrated increased accuracy when commenting on skeletal radiographs with a significant reduction in incorrect comments. Future work should include... etc. Has been edited in-line with suggestions 206 words

Introduction

2. P1 lines 12-15 The structure could be improved. The first statement can be a standalone sentence. The red dot system is a method of abnormality detection by radiographers, who highlighted abnormal images to the treating clinician by placing a red dot on the film. Corrected

3. P1 line 25 the SCoR published guidance rather than guidelines. Corrected

4. P1 lines 26 - 35 These sentences are repetitive; they both say that the SCoR advocates replacing the ambiguous red dot system with preliminary clinical evaluation/radiographer commenting. Deleted sentence

5. P1 lines 42 would this not be the preferred method rather than role? Agreed

6. P1 line 44 rather than an ambiguous abnormality detection system? Corrected

7. P1 lines 48-52 and P2 lines 1-10 The evidence in this section needs to be synthesised. Revised to read better

8. P2 line 2 is an unnecessary four word sentence. Recent work has demonstrated comparable accuracy of Australian radiographers and emergency department doctors (88.6% vs. 89.5%), with collaboration suggested as a method of improving patient care in the absence of a radiologist (ref). Corrected
9. P2 lines 15-17 Does this statement have a reference? Is it a standard practice in Australia? Or it is just the setting of a particular study? If it is just the setting of a particular study, this should be made clear. **Found in two studies**

10. P2 lines 22-28 This could be rephrased to one sentence - this would improve flow. **revised**

11. P2 line 30 Suggest changing this to "The ideal radiographer comment should be ..." **agreed**

12. P2 lines 32-38 This should be rephrased and collated into a single coherent statement. Formal postgraduate qualifications in image interpretation are offered by universities in Australia, however Smith et al also suggest that shorter courses and online learning be utilised in the delivery of radiographer image interpretation training. **Corrected**

13. P2 lines 38-46 This needs to be restructured. Current radiographer undergraduate education in South Africa is significantly different to Australia and the United Kingdom as the majority of programmes are a diploma rather than a degree. Undergraduate training is evolving however, with a four year degree programme in development. The postgraduate training models in South Africa are also different; there are currently no image interpretation training opportunities, either formal postgraduate or short courses. **Corrected**

14. P2 line 56 - p3 line 3. This section could be revised for improved flow; the final statement (12 of 18 HEI) could form a separate sentence. **revised**

15. P3 lines 5-7 Suggest changing these to "The image interpretation training offered at South African institutes has not been established, although pattern recognition is within the scope of practice of radiographers [personal communication] [14,15]". **changed**

16. P3 lines 10-14 Suggest changing these to "... identify normal, normal variants and abnormal patterns". **Changed**

17. P3 lines 20-27 This section needs to be rephrased. **revised**

18. P3 line 30 - The aim of the study states that this is an investigation of pattern recognition training, and as stated in the methods (description of intervention). The results and discussion highlight improvements in vocabulary - was any training in the use of appropriate terminology given? It is implied in the discussion that is has. This is important and needs to be adequately addressed throughout to ensure consistency and to support some of the study conclusions. **Have changed to highlight structure and vocabulary required to provide a comment in the training programme**

 Methods

19. P3 line 40 - The first statement would benefit from the inclusion of "educational intervention". **done**

20. P3 line 41 - P4 line 5 These sections could be amalgamated for improved clarity and to remove some repetition. Convention is to give the total sample (n=75) and then reasons for exclusion (workload etc.) to then be left with the final number of participants (n=9). **improved**
21. P4 lines 5-6 The last sentence could be incorporated into the previous statement, "Nine participant radiographers also provided sufficient number of interpretations to ensure the study was adequately powered." revised

22. P4 lines 11 - 23 Punctuation needs to be improved, a few commas are missing. Punctuation corrected

23. P4 line 18 Each what? Each tutorial was offered... yes

24. P4 line 40 - Delete the "Two tutorials were offered" and Begin with "The first tutorial..." agreed

25. P5 lines 30-37 This would read better as a series of bullet points. Have done so

Test Bank selection

26. P5 line 44-P6 line 8 This section could be rephrased to improve clarity. P5 lines 44-48 delete "to provide a statistically viable number of results. The two test banks were numerically the same" and consider "A pre and post training image test bank (n= 100, 50% abnormal) were given to the participants, with no duplication of cases between banks. The order of the images within each bank was randomised using a computer generated algorithm. Participants were blinded to the disease prevalence. Images used in the study were hard copy films, the most familiar method of viewing images by the radiographers." P6 lines 4-8 This is a limitation and should be moves to the discussion section. Much improved structure now

27. P6 line 56 - P7 line 4 This section should be rephrased to improve clarity. revised

28. P7 lines 4-6 Delete the last sentence deleted

29. P7 lines 8-25 This paragraph needs to be restructured. Suggest it reads "Data was analysed for normality using the Kolmogorov-Smirnov statistic and found not to be normally distributed. Wilcoxon signed rank test was used to compare the radiographers performance pre and post training (paired, non-parametric) with p values less than 0.05 deemed significant." revised

Results

30. There is some inconsistency here regarding the data presented in Tables 1 and 3. Although Table 1 and the accompanying text discusses correct/incorrect responses it appears that this may actually be accuracy, as correct, partially correct and incorrect responses are discussed in Table 3 and accompanying text. It is essential that this is addressed Have addressed this

31. Table 1 - this should indicate that this is average/mean performance (is it accuracy?)

32. P7 lines 45-52 This paragraph could be restructured to improve clarity. For example: "The image interpretation performance of the radiographers demonstrated a significant improvement following training, with correct responses increasing from 70.1% to 77.6% and incorrect responses decreasing from 28.0% to 21.0% (Wilcoxon Z=-2.66, p=0.008). revised

33. Table 2. It would be useful to have the pre and post-test sensitivity, specificity and accuracy for each participant included, or possibly as another Table. Table 4
34. P8 lines 18-24 If you wish to include definitions of TP etc. then these need to be moved to the methods section.  
Moved to methods

Table 3 (see above section).

35. P8 line 43 - should the text refer to Table 3? yes

36. P8 lines 43-58 This section should be revised to improve clarity (see previous suggestion above). As mentioned previously, the study aim and description of the intervention appear to relate to pattern recognition; however p9 of the results discusses improvement in the vocabulary used by radiographers when commenting/describing abnormal x-rays. Was training on this given? changed

37. P9 line 1 Suggest replacing the beginning of this sentence with "When the post-test training comments were assessed, radiographers .... corrected

38. P10 line 1 All radiographers were "asked" to submit a log book when completing the post training image bank, with a very low completion rate (2 of 9, 22%). revised

39. P10 lines 4-8 This should be moved to the discussion/limitations section moved

Discussion

40. This section has been improved, with the current study placed into context by comparing these results to the relevant literature. Further work is required however to improve clarity and flow. Brealey and colleagues performed a systematic review on the abnormality detection accuracy of radiographers (Brealey et al Clin Radiol 2006;61:604-15). This is one of the seminal pieces of work in this area - how does performance compare? Reference to Brealy et al

41. P11 lines 20-28 this paragraph should be rephrased and incorporated with the next paragraph (lines 31-47). Is there a reference for that statement regarding South African undergraduate training? Gqweta
Suggest it reads "Several authors have identified that radiographers do not have the vocabulary required to accurately comment on abnormal x-rays, with the lack of training during undergraduate education identified as a barrier..." Corrected

42. P11 lines 38-44 Given that the authors suggest adopting an 85% accuracy threshold for radiographer commenting, can they then claim that a 7% improvement in performance to 77% demonstrates that this training and the results of this study are sufficient to justify the implementation of radiographer commenting by South African radiographers? Would a more appropriate suggestion be that this study has shown that with training, radiographer accuracy and vocabulary has improved, but further work is needed to ensure that performance is (a) improved to 85% and (b) doesn't decline with time? Have added this in

43.P12 lines 17-25 This paragraph needs rephrasing. Robinson and Tudor have also published on variation and agreement between consultant radiologists. A robust reference standard diagnosis often taken to be 2-3 independent radiologist interpretations which are in complete agreement. Included Robinson

44. P12 lines 27-31. One sentence paragraph should not be used. This is an important section, especially as the authors are suggesting that this study provides the evidence to support
radiographer commenting in clinical practice, which will occur some time after the training has been given. **Added to the paragraph**

**Conclusion**

45. P12 lines 43-49 The presented evidence does not support the statement regarding the need for a postgraduate programme for radiographer commenting. This should be removed. **removed**

46. P12 lines 51-60 The future work/suggested research section needs to be revised. Suggest it reads "Future work could examining the role of clinical mentorship on radiographer commenting, performance of radiographers in a clinical setting and assess for declining performance after training has been completed." **revised**

******************************************************

Please note that the editorial process varies considerably from journal to journal. To view a sample editorial process, please click here:  

For further assistance, please visit our customer support site at [http://help.elsevier.com/app/answers/list/p/7923](http://help.elsevier.com/app/answers/list/p/7923). Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EES via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.
The influence of image interpretation training on the accuracy of abnormality detection and written comments on musculoskeletal radiographs by South African radiographers.

Lynne Hazell, MTech (Radiography), Jenny Motto, MTech(Education) and Lucky Chipeya, MPhil HE

Department of Medical Imaging and Radiation Sciences, Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa.

Corresponding author: Lynne Hazell, MTech, Department of Radiography, Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa.

E-mail address: lynneh@uj.ac.za
Introduction

The development of the role of the radiographer into image interpretation has been established in the United Kingdom (UK) since in the 1980s [1, 2]. However, in South Africa, research has not established if diagnostic radiographers would be able to undertake this role; and if so, what training would be required to ensure their accurate interpretation of images.

The “red dot” system is a method of abnormality detection by radiographers, who highlight abnormal images to the treating clinical by placing a red dot on the film [3]. The “red dot” system had proved to be an ambiguous system, as the referring physician did not know if a lack of a “red dot” was because there was no abnormality, if the radiographer did not see the abnormality, or did not wish to provide an opinion on the image [4, 5, 6, 7]. In 2006, the Society and College of Radiographers (SCoR) published guidance in the UK recommending that by 2010 radiographers in the UK would provide an initial written or verbal comment on all musculoskeletal trauma images [4]. In a recent SCoR document the Society has stated the “red dot” system should be replaced by a preliminary clinical evaluation previously called a clinical comment to overcome the ambiguity and optional participation in the “red dot” system [5]. Therefore when considering the optimum form of role extension for South African radiographers, a commenting system would be the preferred method for radiographers rather than an ambiguous abnormality detection system [5].

As has been stated previously, a great deal of research has been performed in the UK demonstrating the accuracy of radiographers in musculoskeletal image interpretation [1, 2, 3]. Unfortunately this has not translated into radiographers being able to perform these tasks in other countries without further research specific to the country. South Africa and Australia have similar challenges regarding the introduction of radiographers providing image interpretation. In rural areas of Australia there is a lack of radiologists, and therefore delays in reporting trauma images is common [8, 9]. Radiographers in rural areas of Australia are often asked for their opinion on images, and radiographers in South Africa experience a similar expectation from clinicians in rural areas. Research has demonstrated that Australian radiographers and emergency department doctors have similar accuracy in image interpretation (88.6% vs 89.5%). In the absence of a radiologist, it has been suggested that collaboration in image interpretation would be beneficial to the patient [9]. However, it appears the Australian Government and radiologists in Australia are not supportive of role extension for radiographers into initial image reporting [8, 9].

Radiographers in Australian studies have used a radiographic opinion form (ROF) to assist in the identification of the abnormality and to select a type of abnormality from a check list. The radiographers are provided with an opportunity to add a comment once they have established there is an abnormality; however, it was found that they did not have the vocabulary to provide an accurate comment [8, 9]. When considering the content for the training the researcher felt that training in the descriptive vocabulary and structure of a report was required, this is supported by McConnell et al in their research [9]. The ideal radiographer comment should be clear, brief and specific [10, 11].

Formal postgraduate qualifications in image interpretation are offered by universities in Australia; however, Smith et al, also suggest short courses and online learning be utilised in the delivery of radiographer image interpretation training [8]. Current radiographer undergraduate education in South Africa is significantly different to that offered in Australia and the United Kingdom. In South Africa, the majority of programmes provide diploma rather than degree courses. Undergraduate training is evolving however, with a four year degree programme in development. The postgraduate training models are also different in South Africa; there are currently no
image interpretation training opportunities, either formal postgraduate or short courses [11, 12].

A survey of Higher Education Institutes (HEIs) in the UK had a response rate of 76% (25 out of 33 HEIs). From the 25 HEIs who responded, 19 indicated that image interpretation is offered at undergraduate level in the form of lectures and tutorials, with 12 institutions having image interpretation as a clinical learning outcome. Twelve out of eighteen HEIs offer postgraduate image interpretation modules [6].

“The image interpretation training offered at South African institutes has not been established, pattern recognition is within the scope of practice radiographers [personal communication] [14, 15].” Pattern recognition provides undergraduate students in South Africa with the ability to identify normal, normal variants and abnormal patterns; however, there is no expectation that they would comment on images once qualified as has been stated in the UK [4]. Kumar [16] found that radiographers who had postgraduate education became more confident in reporting and the wording of the reports improved.

At present there is a shortage of radiologists in South Africa [17], which leads to delayed reporting on images. Radiographers may be able to provide an initial comment on images to alleviate this problem. However, there have been no studies to investigate the type of training necessary to enable diagnostic radiographers to gain the skills to accurately comment on images. The current study investigated the extent to which training in pattern recognition and how to construct a comment could improve the accuracy and descriptive comments on musculoskeletal images by diagnostic radiographers in two Government hospitals in Johannesburg, South Africa.

Methods

A single group pre- and post-test study was undertaken with an educational intervention [3, 17, and 18]. There was a total sample (n= 75) of radiographers employed at the two hospitals in Johannesburg, South Africa who were eligible to participate in the study. Due to the work commitments and staffing requirements of the departments it was not possible for there to be a larger sample, and the final number of participants who volunteered for the study were small (n=9).

All radiographers in South Africa are registered with the Health Professions Council of South Africa. Both hospitals gave permission for the participants to take part in the research. The Academic Ethics Committee of the Faculty of Health Sciences gave ethical approval – Clearance reference number: 47/08 on the 26th September 2008. The nine participants completed informed consent and could withdraw at any point during the research.

Nine participant radiographers also provided sufficient number of interpretations to ensure the study was adequately powered.

The Structure of the Intervention

A combination of six lectures and tutorials were offered, approximately two hours duration, providing pattern recognition training for musculoskeletal images of the appendicular and axial systems, excluding the skull. Attendance registers were taken for each lecture/ tutorial, to monitor that every participant attended the scheduled training opportunities. Each tutorial was offered on two separate occasions over a period of four months [19]. After each tutorial the radiographers could spend time independently looking at images and applying the techniques learnt in the lectures with the lecturer available for consultation.

Four lecture topics included:
An introduction to pattern recognition
Application of pattern recognition skills to the upper extremity
Application of pattern recognition skills to the lower extremity
Pattern recognition interpretation for the spine and pelvis.

The first tutorial introduced the radiographers to a checklist to assist with pattern recognition and interpretation of musculoskeletal images. The tutorial assisted the radiographers in how to construct a written comment and a systematic method of analysing the images for commenting. The second took place after the radiographers had used the checklist in the clinical environment. This took the form of feedback on plain film commenting after utilising a checklist.

The lectures and tutorials were a combination of PowerPoint presentations and hands-on workshops where the participants used the skills and techniques of pattern recognition provided in the presentations to interpret images that demonstrated normal and abnormal patterns. These PowerPoint presentations highlighted a variety of abnormalities to ensure the participants had exposure to unusual and subtle cases, as it is not always possible to have hard copies of all the pathologies. The participants also had tutorials with banks of images to assist them in their ability to accurately comment on images. The banks of images that were used for the tutorials, pre and post-tests were all of different radiographs.

The PowerPoint presentations also provided points of reference for the participants to use in their clinical environment. The participants had access to the PowerPoint presentations via email and were able to print these for ease of reference whilst working. Log books were introduced to provide the opportunity for the students to apply their training in the clinical environment and have a mentor in the workplace. The participants were required to comment on an image in the clinical situation and then discuss their comment and the image with a radiologist at their institution. The radiologist was required to sign the log book to acknowledge the discussion.

The log book included five areas of the skeleton:
- Lower extremity (knee, tibia/ fibula, ankle and foot)
- Pelvis, hip and femur
- Upper extremity (elbow, forearm, wrist and hand)
- Shoulder girdle
- Spine.

Some of the images they interpreted could be normal, as it was essential to understand normal variants in conjunction with the abnormal patterns. The minimum number of comments was twenty per area.

**Test Bank Selection**

A pre- and post-training image test bank (n=100, 50% abnormal) were given to the participants. The order of the images within each bank was randomised using a computer generated algorithm [3]. The participants were blinded to the disease prevalence. Images used in the study were hard copy films, the most familiar method of viewing images by the radiographers. The images were taken from teaching files, all images were anonymous and the images were not used for training purposes. The images were not from the hospitals where the radiographers worked.

Included in the 100 images were images of the upper limb including the shoulder; the lower limb; including the hip and the cervical; and thoracic and lumbar spine and pelvis. The images included epiphyseal growth plates of the hand, wrist, forearm, elbow, foot, ankle, knee and hip; thus testing the participants extensively on their ability to recognize the epiphyseal lines, which are important normal variants that a diagnostic radiographer should be able to recognize. In the study the participants were assessed not only on
trauma images, but also on pathology—such as change in shape of bone and joint spaces, and whether lytic and sclerotic areas were recognized—which supported the need to use 100 images in the test bank [3,18]. These were pattern recognition criteria taught in the training and would by necessity need to be included in the tests.

A single experienced consultant radiologist was used as the reference standard in the study [20]. The radiologist is considered the “gold standard” for image interpretation [21]. The participants and the radiologist were provided with data sheets where they initially identified whether the image was normal or abnormal by a cross in the appropriate column. The results were classified True Negative (TN) correctly identifying a normal image and a True Positive (TP) correctly identifying an abnormal image. The False Negative (FN) incorrectly identifying the image as normal when there is an abnormality and False Positive (FP) incorrectly identifying the normal image as abnormal and then for all the abnormal images they were asked to supply a comment on the abnormality. The participants were given 150 minutes to complete the pre- and post-tests. This was in accordance with previous research conducted that, allowed 30 minutes for 20 images, the equivalent of 90 seconds per image [3, 19]. The responses were assessed by the researcher and judged as incorrect, partially correct or correct against the reference standard.

The researcher identified that diagnostic radiographers should not only identify the normal and the abnormal images, but that they should also provide a comment on the image. The radiographers provided a comment on the images they identified as abnormal, stating the type of abnormality and its location. The comments were categorised into three categories: incorrect, partially correct and correct.

The radiographers’ performances were rated as true or false, and positive or negative; from these results the sensitivity, specificity and accuracy could be calculated. Data was analysed for normality using Kolmogorov- Smirnov statistic and found not to be normally distributed. The Wilcoxon signed ranks test was used to compare the radiographers performance pre- and post- training (paired, non-parametric) with p-value less than 0.05 deemed significant. The p-value of 0.05 is generally a standard value in social science research [22].

**Results**

The ability of the participants to identify normal and abnormal images, improved in the post-test.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct responses</td>
<td>70.1%</td>
<td>77.6%</td>
<td>-2.66</td>
<td>.008</td>
</tr>
<tr>
<td>Incorrect responses</td>
<td>28.0%</td>
<td>21%</td>
<td>-2.66</td>
<td>.008</td>
</tr>
</tbody>
</table>

The image interpretation performance of the radiographers (Table 1) demonstrated a significant improvement following training with correct responses increasing (70.1% to 77.6%) and incorrect responses decreasing (28.0% to 21.0%, Wilcoxon z=2.66, p=0.008).
Table 2: Individual results for true negative, true positive, false negative and false positive pre and post-test.

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Pre Test</th>
<th>Post test</th>
<th>Pre Test</th>
<th>Post test</th>
<th>Pre Test</th>
<th>Post test</th>
<th>Pre Test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TN</td>
<td>TP</td>
<td>FN</td>
<td>FP</td>
<td>TN</td>
<td>TP</td>
<td>FN</td>
<td>FP</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>45</td>
<td>4</td>
<td>27</td>
<td>29</td>
<td>43</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>39</td>
<td>10</td>
<td>8</td>
<td>40</td>
<td>47</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>36</td>
<td>13</td>
<td>16</td>
<td>38</td>
<td>42</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>41</td>
<td>8</td>
<td>20</td>
<td>38</td>
<td>41</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>44</td>
<td>4</td>
<td>23</td>
<td>39</td>
<td>39</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>39</td>
<td>9</td>
<td>20</td>
<td>35</td>
<td>46</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>40</td>
<td>7</td>
<td>27</td>
<td>37</td>
<td>42</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
<td>39</td>
<td>8</td>
<td>21</td>
<td>24</td>
<td>47</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>42</td>
<td>8</td>
<td>19</td>
<td>30</td>
<td>45</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>268</td>
<td>365</td>
<td>71</td>
<td>181</td>
<td>310</td>
<td>392</td>
<td>57</td>
<td>131</td>
</tr>
</tbody>
</table>

Table 2 demonstrates the individual results for the participants their identification of True Negative (TN) True Positive (TP) False Negative (FN) and False Positive (FP) images.

Table 3: Individual results for sensitivity, specificity and accuracy.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Specificity</td>
<td>Accuracy</td>
<td>Accuracy</td>
</tr>
<tr>
<td>1</td>
<td>91.84</td>
<td>86.00</td>
<td>46.00</td>
<td>58</td>
<td>68.69</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>79.59</td>
<td>94.00</td>
<td>84.00</td>
<td>83.33</td>
<td>81.82</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>73.47</td>
<td>85.71</td>
<td>68.00</td>
<td>77.55</td>
<td>70.71</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>83.67</td>
<td>82.00</td>
<td>60.00</td>
<td>77.55</td>
<td>71.72</td>
<td>79</td>
</tr>
<tr>
<td>5</td>
<td>91.67</td>
<td>81.25</td>
<td>54.00</td>
<td>76.47</td>
<td>71.72</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>81.25</td>
<td>92.00</td>
<td>60.00</td>
<td>72.92</td>
<td>69.70</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>85.11</td>
<td>82.35</td>
<td>43.75</td>
<td>77.08</td>
<td>61.62</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>82.98</td>
<td>94.00</td>
<td>59.62</td>
<td>48.98</td>
<td>70.71</td>
<td>71</td>
</tr>
<tr>
<td>9</td>
<td>84.00</td>
<td>88.24</td>
<td>61.22</td>
<td>61.22</td>
<td>72.73</td>
<td>75</td>
</tr>
<tr>
<td>Mean value</td>
<td>83.73</td>
<td>87.28</td>
<td>59.62</td>
<td>70.34</td>
<td>71.04</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 3 demonstrates the mean group values for sensitivity pre-test and post-test (83.73% and 87.28%). The mean group value for specificity pre-test and post-test (59.62% and 70.34%). The group mean accuracy pre-test and post-test (71.04% and 78%). The accuracy was statistically significant p=0.008, however, the changes in sensitivity and specificity were not statistically significant.

Table 4: Accuracy of comments and Wilcoxon Signed Ranks Test

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect comments</td>
<td>24.11%</td>
<td>17.78%</td>
<td>-1.960</td>
<td>0.05</td>
</tr>
<tr>
<td>Partially correct comments</td>
<td>16.78%</td>
<td>21.78%</td>
<td>-1.305</td>
<td>0.192</td>
</tr>
<tr>
<td>Correct comments</td>
<td>7.78%</td>
<td>10.33%</td>
<td>-0.563</td>
<td>0.574</td>
</tr>
</tbody>
</table>
From Table 4, the incorrect comments on images decreased (24.11% to 17.78%, Wilcoxon z=-1.96, p=0.05). The partially correct comments improved (16.78% to 21.78%, Wilcoxon z=-1.30, p=0.19), a five percent (5%) improvement. The correct comments provided for the abnormal images had improved (7.78% to 10.33%, Wilcoxon z=-0.56, p=0.57), a 2.55% increase.

When the post-test training comments were assessed, radiographers were able to describe the pathology using more accurate medically accepted language; for example, mentioning soft tissue changes and more accurately identifying the location of the abnormality these requirements were taught during the training. An example on an image of a humerus:

- Pre-test comment, “abnormal bone trabecular, bones abnormal.”
- Post-test comment: “Loss of soft tissue of the humerus. Increased bone density of the humerus. Loss of trabecular pattern of the bones and shoulder joint space is reduced.”

The comment improved from partially correct to in-line with the reference standard; thus it had improved to a correct comment. The participant demonstrated a better understanding of and use of medically accepted language post-training. As an example, participant three (3) interpreted an image of the elbow correctly as abnormal in the pre-test; however, no comment was provided. In the post-test, the image was correctly identified as abnormal and the comment was:

- “There is soft tissue swelling on the left arm and a fat pad is seen displaced on the posterior aspect of the elbow.”

The comment was in-line with the reference standard and provided information pertaining to location and soft tissue changes.

An example of a partially correct comment would be:

- “A radiolucent line on the distal radius of the left wrist was seen. Discontinuation of the cortical outline of the radius.”

No mention was made of a fracture of the styloid process of the ulna.

A paediatric image of the elbow was correctly identified by the nine radiographers as abnormal; however, six radiographers provided an incorrect comment or no comment at all on the image.

There was one image in the pre-test which was not considered as the radiologist felt it could be ambiguous so was excluded.

All radiographers were asked to submit a log book when completing the post-training image bank, with a very low completion rate (2 of 9, 22%).

**Discussion**

When compared to previous studies, there are similar results to those in the current study. In one study where an ROF form was used, there was a significant change in accuracy in the general opinions and observations following a short continuous education programme [8]. This would be similar to the findings in Table 3 for the improvement in accuracy after the intervention. Even though the results were not significant when comparing vocabulary, the understanding and content of the comment has improved in the Australian study; again, similar to the present study. Researchers in the UK found radiographers could provide a written comment as an initial image interpretation in the Accident and Emergency environment post training [20]. Radiographers’ commenting was compared with commenting by Casualty Officers and Accident and Emergency nurses and radiographers had significantly higher accuracy scores [20].
radiographers in the South African research had an accuracy pre-test and post-test (71.04% and 78%) (Table 3) and in Coleman and Piper’s study the radiographers had 71.5% accuracy [20]. The results would appear to be comparable as South African radiographers pre-test with no training had similar accuracy as in Coleman and Piper’s study. Coleman and Piper suggested radiographers should use a specific tick-box form when providing a comment with the option to add a descriptive comment, because radiographers lacked the vocabulary to comment and this would improve their confidence levels. [20]. For future research in South Africa, a tick-box or radiographer opinion form could be used and assessed in the clinical situation.

Hargreaves and Mackay found in their research that radiographers have accuracy of pre-test 89.9%, post-test 93%; and in and McConnell et al, the mean accuracy was 95%—far more accurate than the South African radiographers [18, 9]. The training has improved accuracy and vocabulary; however, further work will be needed to reach a performance of 85% accuracy, in-line with the recommendation in the Australian study [8]. Whether the improvement in performance is maintained would also need further research. When considering sensitivity, the improvement was similar in this study to that of Hargreaves and Mackay with an increase of 3.55% in this study and 5.1%; however, the specificity is not in-line with the previous literature. Although the training has improved the participants’ specificity by 10.72%, it is only 70%—far below an acceptable level [18]. The radiographers appear to lack an understanding of the normal variants and having the confidence to identify an image as normal. In the study radiographers were asked to interpret all musculoskeletal images, not just trauma images. Many of the studies used for comparison often only used trauma images, which may provide a reason for the participants not achieving such high levels of specificity and sensitivity. When placing the current study in the context of previous studies, the performances vary in that the values for sensitivity are higher in the study than the specificity, whereas in previous research the opposite is true. Accuracy improved, although not to the extent of previous studies [25].

The radiographers’ changes in performance were not consistent, and this is similar to the findings of Hargreaves and Mackay. For sensitivity, four radiographers had values which fell, although by only 2% for two radiographers. The values for specificity were more uniform, with only number eight being dramatically less than the pre-test. The differences could possibly be attributed to the clinical application and the completion of the log book, which appeared to assist the values for participant 3.

Several authors have identified that radiographers do not have the vocabulary required to accurately comment on abnormal x-rays, with the lack of training during undergraduate education identified as a barrier to providing an initial well-structured accurate comment on qualification. Radiographers do not learn the language required for commenting on images in their undergraduate qualification in South Africa [23]. Therefore an important aspect of the training was to provide guidance for writing an accurate comment [24]. Radiographers often find it difficult to identify the important findings and communicate them in a clear and concise manner [3, 15, 18]. The training emphasized the importance of communicating all findings rather than “a fracture is seen,” which would have been the pre-training response. Therefore the need for training in the structure of the comment provided is important in order to provide the referral doctor with the most useful information for optimum diagnosis and treatment of the patient [6, 26, 27]. The accuracy of commenting on images improved by 6.32%, which could be considered a significant result demonstrating that the training had given the radiographers the appropriate skills in image interpretation.

Hard copy films were used in this study as the radiographers were familiar with this type of image. In their study from 2007, Hardy and Culpan used PowerPoint images to assess their participants and this was considered a limitation of the study as hard copies were considered the “gold standard” [3]. Although in 2015, if a high definition monitor is used
and with digital radiography becoming the norm, image interpretation could be assessed with high quality digital images.

The log books issued to participants were only submitted by 2/9 radiographers (22%). From the data analysis these were stronger performing radiographers. In future log books could be used to enhance the training. To ensure the application of the training the books could be submitted one week after a lecture for each area to demonstrate the radiographers ability to interpret images in the clinical situation.

**Limitations**

The participants expressed that the time allocated for commenting on images was too short for the number of images and it was difficult to concentrate for an extended period. Therefore using batches of images with rest periods interspersed would possibly have enabled the participants to be tested more successfully.

A combination of the test and the clinical experience using the log book could have provided a more accurate assessment of their ability to provide a comment. The participants did not receive feedback on the pre-test, as the images were different in the post-test, this could have been a beneficial learning opportunity. Participant three was one of only two participants who submitted their log books after the study. The participant demonstrated the greatest improvement in commenting which could be a result of discussing the comments with the radiologist and recording this in the log book [28]. The test conditions have previously been questioned as whether they provide a true reflection of the ability of the radiographers and a clinical setting for the assessment would provide a more authentic result [29].

Only one radiologist was used as the reference standard this has previously been seen as a limitation. Due to substantial observer variation between consultant radiologists and on skeletal radiographs there could be a major disagreement in 10% of the cases [30, 31]. In future studies two or three radiologists would need to be used to eliminate bias. The radiographers performance was only assessed by the post-test training test bank and no follow up test after six months of clinical practice was performed. Previous studies identify that the training only provides an initial improvement and that over time performance decreases and therefore further continuous development programmes are required to ensure the performance standard is maintained [19]. Future work would be necessary to establish the performance in the clinical environment.

**Conclusion**

This study has demonstrated that training in pattern recognition and construction of a comment could enable diagnostic radiographers to improve their accuracy and the ability to provide a descriptive comment on an image.

Future studies could examine the role of clinical mentorship on radiographer commenting, the effect the clinical environment has on performance and assess whether the performance is maintained after completion of the training. The scope of practice may need to be reconsidered and the legal implications on practice for radiographers who undertake image interpretation would require research in the South African Health sector.

**References**


Journal of Medical Imaging and Radiation Sciences (JMIIRS) Authorship Responsibility, Conflicts of Interest and Funding, Ethics and Copyright Transfer/Publishing Agreement

The corresponding author must sign the following statements on behalf of all co-authors, including the Acknowledgement Statement if applicable. Detailed Instructions for Authors are available online at http://www.jmiirs.org/authorinfo.

Article Title: Accuracy of comments provided on radiographic images by diagnostic radiographers can be improved with training

Authorship Responsibility, Criteria, and Contributions

Each author should meet all criteria below (A, B, C and D).

A. I certify that:
   - The manuscript represents original and valid work.
   - Neither this manuscript nor one with substantially similar content has been published or is being considered for publication elsewhere, except as described in an attachment, and copies of closely related manuscripts are provided.

B. I will be the corresponding author and agree to serve as the primary correspondent with the editorial office, and to review the edited typescript and proof.

C. To qualify for authorship, all authors must meet the criteria of at least 1 box for each of the 2 categories of contributions listed below.

   1. □ Conception and design
      □ Acquisition of data
      □ Analysis and interpretation of data
   2. □ Drafting of the manuscript
      □ Critical revision of the manuscript for important intellectual content
   D. All authors have given final approval of the submitted manuscript.

Any changes in authorship after submission of a manuscript must be explained in a letter to the editor-in-chief that is signed by all co-authors to indicate their consent to the change. These changes include the deletion and addition of authors as well as any change in the order of authors.

Ethics Agreement (if applicable)

Manuscripts containing the results of experimental studies on human participants must disclose in the first paragraph of the Materials and Methods section whether informed consent was obtained from patients in the study after the nature of the procedure had been fully explained.

Copyright Transfer/Publishing Agreement

In consideration of the action of the JMIIRS in reviewing and editing this submission (manuscript, tables, figures, video, audio, and other supplemental files for publication), I hereby transfer, assign, or otherwise convey all copyright ownership, including any and all rights incidental thereto, exclusively to the JMIIRS, in the event that such work is published by the JMIIRS.

Acknowledgement Statement (if applicable)

Authors should obtain permission from all individuals named in an Acknowledgment, since readers may infer their endorsement of data and conclusions. The corresponding author must sign the following statement:

□ I certify that all persons who have made substantial contributions to the work reported in this manuscript (e.g., data collection, analysis, or writing or editing assistance) but who do not fulfill the authorship criteria are named with their specific contributions in an Acknowledgment in the manuscript.

□ I certify that all persons named in the Acknowledgment have provided me with permission to be named.