
Article:
Thirkettle, M. orcid.org/0000-0002-6200-3130, Thyoka, M., Fernandes, N. et al. (3 more authors) (2019) Internet-based measurement of visual assessment skill of trainee radiologists: developing a sensitive tool. British Journal of Radiology, 92 (1097). ISSN 0007-1285

https://doi.org/10.1259/bjr.20180958

Title: Internet-based Measurement of Visual Assessment Skill of Trainee Radiologists: Developing a Sensitive Tool

Shortened Title: Web Measurement of Visual Assessment Skill of Trainee Radiologists

Type of Manuscript: Full Paper

Author names:

Martin Thirkettle, PhD, Centre for Behavioural Science and Applied Psychology, Department of Psychology, Sociology & Politics, Sheffield Hallam University, UK

Mandela Thyoka, Department of Radiology, Sheffield Teaching Hospitals, Sheffield, UK

Nadiah Fernandes, Imperial College London School of Medicine, London, UK

Padmini Gopalan, Department of Radiology, Sheffield Teaching Hospitals, Sheffield, UK

Tom Stafford, PhD, Department of Psychology, University of Sheffield, Sheffield, UK

Amaka C Offiah, BSc, MBBS, MRCP, FRCR, PhD, FHEA, Academic Unit of Child Health, University of Sheffield, Sheffield, UK, Department of Radiology, Sheffield Children's NHS Foundation Trust, Sheffield, UK

Author Conflicts of interest / funding: This work was supported by The Sheffield Children's Hospital Charity. The authors certify that there is no actual or potential conflict of interest in relation to this article.

Abstract:

Objective:

Expert radiologists exhibit high levels of visual diagnostic accuracy from review of radiological images, doing so after accumulating years of training and experience. To train new radiologists, learning interventions must focus on the development of these skills. By developing a web-based measure of image assessment, a key part of visual diagnosis, we aimed to capture differences in the performance of expert, trainee and non-radiologists.

Methods:

Twelve consultant paediatric radiologists, twelve radiology registrars, and thirty-nine medical students were recruited to the study. All participants completed a two-part, online task requiring them to visually assess 30 images (25 containing an

abnormality) drawn from a library of 150 paediatric skeletal radiographs assessed prior to the study. Participants first identified whether an image contained an abnormality, and then clicked within the image to mark its location. Performance measures of identification accuracy, localisation precision, and task time were collected.

Results:

Despite the difficulties of web-based testing, large differences in performance, both in terms of the accuracy of abnormality identification and in the precision of abnormality localisation were found between groups, with consultant radiologists the most accurate both at identifying images containing abnormalities ($p < 0.001$) and at localising abnormalities on the images ($p < 0.001$).

Conclusions:

Our data demonstrate that an online measurement of radiological skill is sufficiently sensitive to detect group level changes in performance consistent with the development of expertise.

Advances in knowledge:

The developed tool will allow future studies assessing the impact of different training strategies on cognitive performance and diagnostic accuracy.

Title: Internet-based Measurement of Visual Assessment Skill of Trainee Radiologists: Developing a Sensitive Tool

Shortened Title: Web Measurement of Visual Assessment Skill of Trainee Radiologists

Type of Manuscript: Full Paper

Author names:

Author Conflicts of interest / funding:

Internet-based Measurement of Visual Assessment Skill of Trainee Radiologists: Developing a Sensitive Tool

Abstract

Objective:

Expert radiologists exhibit high levels of visual diagnostic accuracy from review of radiological images, doing so after accumulating years of training and experience. To train new radiologists, learning interventions must focus on the development of these skills. By developing a web-based measure of image assessment, a key part of visual diagnosis, we aimed to capture differences in the performance of expert, trainee and non-radiologists.

Methods:

Twelve consultant paediatric radiologists, twelve radiology registrars, and thirty-nine medical students were recruited to the study. All participants completed a, two-part, online task requiring them to visually assess 30 images (25 containing an abnormality) drawn from a library of 150 paediatric skeletal radiographs assessed prior to the study. Participants first identified whether an image contained an abnormality, and then clicked within the image to mark its location. Performance measures of identification accuracy, localisation precision, and task time were collected.

Results:

Despite the difficulties of web-based testing, large differences in performance, both in terms of the accuracy of abnormality identification and in the precision of abnormality localisation were found between groups, with consultant radiologists the most accurate both at identifying images containing abnormalities ($p < 0.001$) and at localising abnormalities on the images ($p < 0.001$).

Conclusions:

Our data demonstrate that an online measurement of radiological skill is sufficiently sensitive to detect group level changes in performance consistent with the development of expertise.

Advances in knowledge:

The developed tool will allow future studies assessing the impact of different training strategies on cognitive performance and diagnostic accuracy.

Keywords

visual search, visual cognition, expertise, paediatric, radiology

Introduction

The accurate interpretation of radiological images in order to reach a correct diagnosis is at the heart of the expertise of a radiologist (1). Because radiographic images are part of the 'first-line' of diagnosis for traumatic medical conditions, the identification and localisation of abnormalities is a highly valuable area of clinical expertise. Accordingly, understanding the development of this expertise has attracted significant interest not only from radiologists striving to improve performance in the field (2–5), but also from psychologists, for whom radiology acts as an excellent, real-world assay of visual cognition and expertise more generally (6–10).

Previous research has divided the visual expertise of the radiologist into two constituent parts; visual search expertise and cognitive or analytical skills (11,12). The first step involves perceptual interrogation of the medical image, noting any abnormalities. The second step analyses what has been noted within the clinical context of the patient's presentation. While there is mixed support for performance differences in the first of these two steps when radiologists and non-radiologists are compared in tasks using non-clinical images (13,14), trainee radiologists must acquire both abilities to develop expertise in diagnostic radiology (15).

Previous studies in visual search expertise have demonstrated that expert radiologists spend less time scrutinising each image than novices or trainees (12,14) and visually explore images in a different manner to trainees (7,14), suggesting that a more profound, strategic change rather than the simple accumulation of knowledge is the foundation of expert radiological skill. Experts develop a robust memory structure that forms the basis of an extensive knowledge base and devise analytical strategies to assist them in correlating the clinical information and image data, allowing for superior information processing. Indeed, experts demonstrate recall superior to that of novices, especially when time is limited (16).

The present study describes the development of a web-based behavioural measure of visual assessment using a library of paediatric skeletal radiographs. We aimed to assess the feasibility of the on-line and pragmatic interpretation of radiographs and to determine whether the collection of radiographs was of sufficient quality and sensitivity to allow future longitudinal visual tracking experiments. Our hypothesis, in line with previous literature (1,2,16,17), is that the consultant radiologists would perform more accurately across all measures, and do so while spending less time assessing a given image than the radiology registrars, who in turn would be better and faster than the medical students.

Materials and Methods

Study Design

This was a web-based study using a bank of paediatric radiographs predominantly of fractures, but also including normal variants and congenital abnormalities. A computer-based task was developed to quantify the ability of radiologists of varying experience. The dedicated library of 150 skeletal radiographs was selected from 3000 radiographs obtained from XXX Hospital on children presenting to the Emergency Department over a six-year period (2008-2013), following trauma. The images were assessed separately by a consultant paediatric radiologist and a radiology trainee, each of whom made their assessment with access to the radiology reports – only radiographs in which there

was no discrepancy between initial reporter, the consultant paediatric radiologist and the trainee were included (where available, follow-up radiographs were also assessed to help with the decision-making). For each image, the veridical location of any abnormalities as documented on the picture archiving and communications system (PACS) was recorded for comparison against participant responses and (for the purposes of computing diagnostic accuracy), was taken as the gold standard. The two assessors, through agreement, subjectively graded each image into one of three categories of difficulty from the perspective of the second-year radiology trainee: easy, intermediate and difficult. Sixteen normal radiographs were included. All identifying details were removed from the images. Such a large library of images was curated in order that in future the same participant could revisit the task multiple times and be faced with a different sample of pre-assessed images, avoiding the possibility in a longitudinal study that specific image familiarity could explain any improvement in performance.

To balance the desire to replicate clinical practice as far as possible within the task against the need to quantify performance as accurately as possible across the various cognitive demands of radiology, the task was split into two stages for each image – identification of the presence of an abnormality, and, if an abnormality was detected, its localisation.

Delivery of the survey, image display and collection of participants' demographic information were all managed by the on-line survey platform Qualtrics, augmented with custom Javascript for abnormality localisation measures.

Participant selection and recruitment

Twelve consultant paediatric radiologists (referred to from this point on as consultant radiologists or "CR"), all members of the European Society of Paediatric Radiology Child Abuse Taskforce, responded to an open invitation to participate in the study. Consultants reported having between 6 and 31 years of radiology experience (mean 15, SD 6.7) and all were practicing within the UK or EU. Concurrently, twelve radiology registrars from across the five years of the XXX Radiology Training

Scheme (referred to from this point on as trainee radiologists or “TR”), and thirty-nine medical students from The University of XXX Medical School’s five-year degree, who had not received radiology training (referred to from this point on as “MS”), were recruited. Power analysis was conducted to ensure that sufficient participants had been recruited, for testing at an $\alpha = 0.05$ level of significance. All participants were recruited through email, ensuring that only those invited to participate could access the test web page. The reading environment, computer screen and time(s) of reading, were all left to the reader’s discretion. All aspects of the study were conducted in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki. The study was approved by the University of XXX Psychology Department Research Ethics board. Participants were exposed to a full ethics and consent statement and provided explicit consent before completing the task. Participants were given the opportunity to withdraw from the study at any time by contacting the research team. The consultants were remunerated for their time, while the registrars and medical students were entered into separate draws to win a £50 book voucher.

Procedure

After consenting to the study, participants completed a short demographic section including years of experience. Via on-screen instruction, they were then briefed on the experimental task and encouraged to ensure they were not using a particularly small screen, before completing a practice set of image responses. The instructions were repeated at this point, and the participant then began the testing session proper.

All participants completed responses to the same 30 images, 10 each of those previously ranked as easy, intermediate and difficult. Five of these 30 images had been assessed as not containing an abnormality, and these acted as ‘target absent’ images in the test, included to detect ‘false positive’ responses from the participants. This gave our test image set an abnormality prevalence of 83.3%, far greater than that found in a clinical setting, but not uncommon in psychological studies (10). All images were resized prior to data collection to maintain a constant image height (600 pixels) to suit

web viewing. The order of image presentation was randomised across participants to avoid any order effects. Date, time, computer, screen resolution and other related information was recorded.

Participants were asked to identify both pathology and normal variants. First, participants were asked to rank how likely they thought there was an abnormality on the given radiograph on a 6-point Likert scale (definitely yes, probably yes, possibly yes, possibly no, probably no, definitely no). The time taken for this was labelled the "decision time". If participants clicked any of the first three options indicating they thought an abnormality likely, then they were asked to indicate location by clicking the point(s) on the radiograph where they believed there was abnormality/abnormalities. Once satisfied, they clicked the submit button to move onto the next radiograph (the time taken to click the submit button was labelled the "localisation time"). Some images in the bank contained two projections of the same area (e.g. AP and lateral knee); in these cases, participants were instructed to identify and click to locate abnormalities visible on either or both projections.

Data was collected on the accuracy of the participant's identification of an abnormality – as compared to the reference answer for each image – the participants' decision time, localisation time and accuracy of the participants' localisation of each abnormality. Localisation error was calculated as the distance in pixels of the participant's click from the reference location of the abnormality. A mixed ANOVA was used to test participant performance. Post-hoc testing was used to further interrogate significant differences – significance was defined as ($p < 0.05$). Statistical analysis was performed using SPSS version 20.

Results

Results of two observers were excluded for low task engagement (1 MS) and excessively long response times (1 TR). Therefore, results presented are for 12 CR, 11 TR and 38 MS.

Identification task accuracy and inaccuracies

Participants' responses (Figure 1) were categorised as positive or negative with respect to an abnormality and the corresponding sensitivity (proportion of images containing abnormalities correctly identified) and specificity were calculated for each group of observers and for each image difficulty level.

D' ("d prime") for each group was then calculated to combine specificity and sensitivity (18), collapsing across image difficulty. Figure 2 shows the d' across the three groups. This combined measure was significantly affected by group $F(2,58) = 29.698$, $p < 0.001$, $\eta_p^2 = 0.506$) and post hoc testing showed that only the sensitivity results of CR and TR were not significantly different ($p = 0.27$), all other post-hoc comparisons were significant at the $p = 0.001$ level.

An unpaired t-test showed that the sensitivity of MS was not significantly different from zero ($t(37) = 1.90$, $p = 0.065$, Mdiff = 0.16, 95% CI = -0.01 – 0.33) suggesting a discriminative ability on the identification task which was indistinguishable from guessing.

There were significant group differences ($p<0.001$) in true and false positive and negative rates as summarised in Table 1.

Localisation error

There was no significant effect of image difficulty on localisation error ($F(2,116) = 1.705$ $p = 0.153$, $\eta_p^2 = 0.056$). Therefore, localisation error performance was analysed after collapsing across difficulty levels. CR and TR were far more accurate than MS in locating abnormalities, clicking on the image far closer to the reference location (mean location error CR = 46.26 (SD 18.8), TR = 43.27 (SD 22.99), MS = 97.98 (SD 35.38) pixels, $F(2,58) = 21.185$, $p < 0.001$, $\eta_p^2 = 0.422$) . Post hoc testing showed that MS were significantly less accurate than CR ($p < 0.001$) and TR ($p < 0.001$), while CR and TR were not significantly different from each other ($p = 1.0$) – Figure 3A.

Task Time

On average, MS spent 15.8 sec (SD 8.5 sec) completing the identification and localisation tasks for each image making them faster than both TR (M = 20.7 sec, SD = 8.1 sec) and CR (M = 36.9 sec, SD = 18.8 sec). There was a significant effect of group on total time spent per image ($F(2, 58) = 16.383$, $p < 0.001$, $\eta_\rho^2 = 0.361$), which post-hoc testing confirmed was driven by the speed of the CR, who were significantly slower than both TR ($p = 0.003$) and MS ($p < 0.001$). TR and MS were not significantly faster or slower than each other ($p = 0.6$) – Figure 3B.

As Figure 3B shows, splitting the total time per image into the two component tasks – identification and localisation - there was a significant difference between groups in time taken on the first part of the task, deciding if an image contained an abnormality or not ($F(2,58) = 19.75$, $p < 0.001$, $\eta_\rho^2 = 0.405$), but no difference between groups on time taken to localise the abnormality on the image ($F(2,58) = 2.27$, $p = 0.122$, $\eta_\rho^2 = 0.073$).

Discussion

We present data from consultant radiologists, trainee radiologists and medical students to demonstrate that on-line testing is sensitive enough to meaningfully capture observer differences in diagnostic accuracy from radiographs, despite the near total lack of control over the conditions within which the task was performed and the hardware used by each participant to access the study. One important aspect of the hardware used for the task is screen resolution, which varied significantly between participants, and it is reasonable to suggest this may have impacted performance. However, at the group level, differences between the resolutions used do not predict task performance. On average, consultants used higher screen resolution than both the trainees and medical students, but there was no difference between trainee radiologists and medical students. This means that it is difficult to attribute differences in task performance entirely to screen resolution, either in terms of detection accuracy - where consultants did not outperform trainees despite the higher resolution of their screen, or localisation error - where trainees significantly

outperformed the medical students on similar resolution hardware. Moreover, we are not able to comment whatsoever on the ambient lighting or broader testing environment each participant chose for their participation, which potentially may also have affected their performance to some degree. While caution should be taken, particularly in cross-sectional designs, when interpreting data from internet-based studies where so much is left uncontrolled, the results of this study show that overall accuracy followed the expected pattern based on the level of radiological expertise with the caveat that our task was, unusually, also completed more slowly by the most expert participants.

It has previously been shown that the diagnostic accuracy of relatively senior radiologists for the detection of the subtle fractures of child abuse was low and that there was no correlation between years of experience and expertise, in spite of many experts being experienced (19). A further study demonstrated that UK radiologists perceive they would benefit from improved training in this field (20). However, to design and assess any such training, it is imperative to fully understand what makes one individual more “expert” than another and this study is the first step towards developing that understanding.

Recent research has focussed on classification of the errors made by paediatric radiologists in visual assessment of radiographs (21). Our approach is to develop an easily deliverable, sensitive, and repeatable measure of the skill being acquired, by collecting image ratings from a wide set of experts and trainees and comparing behavioural markers of performance across levels of training and experience. This approach allows the development of rapid assessment and quantification of the underlying skill differences between the expert and trainees. In turn, results from such tasks can be used to drive improvements in training interventions. Such improvements to the design and delivery of teaching during the training of trainee radiologists would potentially allow the faster development of the skills involved in the visual assessment of radiographs, leading to better diagnosis of trauma and disease and allow expert levels of performance to be attained earlier in a trainee’s career, with resulting improvements in clinical performance and patient care.

Consultant radiologists, experts in their field, outperformed their intermediary and novice counterparts. Medical students achieved the poorest accuracy score of the three groups. The d' measure is a common and reliable statistic within psychological studies of visual search tasks (18,22). A higher *d'* value suggests a higher sensitivity for the task, with both lower false positive and higher true positive responses required for higher *d'* values. A *d'* of zero reflects chance performance and would indicate the participant to be wholly insensitive to the task and just guessing. In this study, MS performed so poorly as to be indistinguishable from chance. The MS had not received radiology training, so we are not surprised by this result, which does however support the use of the tool for future studies. Further support for the tool is found in disaggregated analysis, not reported here, which showed that both trainee and consultant radiologists' abnormality identification performance reduced as image difficulty rating increased.

One unusual result was the finding that consultant radiologists performed this task significantly slower than novice participants. This contradicts previous research findings (12) and runs counter to perceived wisdom on expertise development; as skill develops, both speed and accuracy improve (1,7). However, rather than propose our results genuinely suggest a revision to this position, it seems more likely that the open nature of our task left the participants free to perform the task at different levels of meticulousness. The near chance level of accurate performance in the novice group supports a view of these participants clicking through the task without the level of diligence shown by the consultants, resulting in far faster task performance. Future studies will need to address task engagement to ensure that participants at all levels of expertise engage with the task sufficiently to provide a reliable measure of their ability. That said, the current results support the potential of web-based assessment protocols, particularly either within a structured training program, where tutors can use the measures described here to evaluate trainees' performance or as a self-assessment program.

The next step in this project is to use the library of validated radiographs in longitudinal studies of cohorts as they complete their training, and to add eye tracking experiments to examine changes in participants' search strategies with increasing experience. This has been done in other contexts (23) and we will adapt these published methods to our needs in paediatric radiology.

Conclusion

The present study demonstrates that consultant radiologists performed far better than trainee radiologists or medical students in correctly detecting the presence of an abnormality on paediatric musculoskeletal radiographs. We show that a web-based delivery of the experimental task is sensitive enough to detect between group differences in performance. Previous work has reported similar findings to these under laboratory conditions, and our results add to this literature. The shift to web-based testing, and the task design - which attempted to resemble clinical practice - may explain the discrepancies between the current results and those from previous studies, both in terms of variation in task performance within groups and in the pattern of performance between levels of expertise.

Future studies will refine the testing platform and provide insight into the development of expert visual diagnostic abilities by radiology trainees through both additional physiological methodologies (e.g. visual tracking) and longitudinal studies of trainee cohorts.

References

1. Norman GR, Coblentz CL, Brooks LR, Babcook CJ. Expertise in visual diagnosis. Acad Med [Internet]. 1992 Oct 1 [cited 2015 Mar 23];67(10):S78-83. Available from: http://europepmc.org/abstract/med/1388563
2. Wood G, Knapp KM, Rock B, Cousens C, Roobottom C, Wilson MR. Visual expertise in detecting and diagnosing skeletal fractures. Skeletal Radiol [Internet]. 2013 Feb [cited 2015 Mar 23];42(2):165–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22940835
3. Wood BP. Visual expertise. Radiology [Internet]. Radiological Society of North America; 1999 Apr 1 [cited 2015 Mar 23];211(1):1–3. Available from: http://pubs.rsna.org/doi/full/10.1148/radiology.211.1.r99ap431
4. Potchen EJ. Measuring Observer Performance in Chest Radiology: Some Experiences. J Am

Coll Radiol [Internet]. 2006 Jun [cited 2017 Jun 30];3(6):423–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17412097

5. Kok EM, van Geel K, van Merriënboer JJG, Robben SGF. What We Do and Do Not Know about Teaching Medical Image Interpretation. Front Psychol [Internet]. 2017 Mar 3;8. Available from: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00309/full

6. Evans KK, Cohen MA, Tambouret R, Horowitz T, Kreindel E, Wolfe JM. Does visual expertise improve visual recognition memory? Attention, Perception, Psychophys [Internet]. 2011 Jan 9 [cited 2017 Jul 3];73(1):30–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21258906

7. Bourne LE, Kole JA, Healy AF. Expertise: Defined, described, explained. Front Psychol [Internet]. Frontiers; 2014 Mar 4 [cited 2017 Jun 30];5(MAR):186. Available from: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00186/abstract

8. Nakashima R, Kobayashi K, Maeda E, Yoshikawa T, Yokosawa K. Visual search of experts in medical image reading: the effect of training, target prevalence, and expert knowledge. Front Psychol [Internet]. Frontiers Media SA; 2013 [cited 2017 Jun 30];4:166. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23576997

9. Drew T, Võ ML-H, Wolfe JM. The invisible gorilla strikes again: sustained inattentional blindness in expert observers. Psychol Sci [Internet]. NIH Public Access; 2013 Sep [cited 2017 Jul 3];24(9):1848–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23863753

10. Horowitz TS. Prevalence in Visual Search: From the Clinic to the Lab and Back Again [Internet]. Vol. 59, Japanese Psychological Research. 2017 [cited 2018 Jul 3]. p. 65–108. Available from: http://doi.wiley.com/10.1111/jpr.12153

11. Berbaum KS, Smoker WRK, Smith WL. Measurement and prediction of diagnostic performance during radiology training. Am J Roentgenol [Internet]. 1985 Dec [cited 2017 Jun 30];145(6):1305–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3877443

12. Krupinski EA, Graham AR, Weinstein RS. Characterizing the development of visual search expertise in pathology residents viewing whole slide images. Hum Pathol [Internet]. 2013 Mar [cited 2017 Jun 30];44(3):357–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22835956

13. Nodine CF, Krupinski EA. Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. Acad Radiol [Internet]. 1998 Sep [cited 2015 Mar 23];5(9):603–12. Available from: http://www.sciencedirect.com/science/article/pii/S107663329880295X

14. Schuster D, Rivera J, Sellers BC, Fiore SM, Jentsch F. Perceptual training for visual search. Ergonomics [Internet]. 2013 Jul [cited 2017 Jun 30];56(7):1101–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23650877

15. Donovan T, Manning DJ. The radiology task: Bayesian theory and perception. Br J Radiol [Internet]. 2007 Jun [cited 2017 Jun 30];80(954):389–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17510249

16. Drew T, Evans K, Võ ML-H-H, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? Radiographics [Internet]. 2012 Jan [cited 2015 Aug 11];33(1):263–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23104971

17. Reingold EM, Sheridan H. Eye movements and visual expertise in chess and medicine. In: Oxford handbook on eye movements [Internet]. Oxford University Press; 2011 [cited 2015 Mar 23]. p. 528–50. Available from: http://eprints.soton.ac.uk/367506/1/Reingold %26 Sheridan %282011%29 expertise chapter.pdf

18. Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behav Res Methods, Instruments, Comput [Internet]. Springer-Verlag; 1999 Mar [cited 2018 Jul 5];31(1):137–49. Available from: http://www.springerlink.com/index/10.3758/BF03207704

19. Offiah AC, Moon L, Hall CM, Todd-Pokropek A. Diagnostic accuracy of fracture detection in suspected non-accidental injury: the effect of edge enhancement and digital display on observer performance. Clin Radiol [Internet]. 2006 Feb [cited 2017 Jun 30];61(2):163–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16439222

20. Leung RS, Nwachuckwu C, Pervaiz A, Wallace C, Landes C, Offiah AC. Are UK radiologists satisfied with the training and support received in suspected child abuse? Clin Radiol [Internet]. 2009 Jul [cited 2017 Jun 30];64(7):690–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19520213

21. Taylor GA. Perceptual errors in pediatric radiology. Diagnosis [Internet]. 2017 Sep 26 [cited 2018 Jul 6];4(3):141–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29536929

22. Green DM, Swets JA. Signal detection theory and psychophysics. Oxford, UK: John Wiley & Sons, Ltd; 1966. 479 p.

23. Manning D, Ethell S, Donovan T, Crawford T. How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography [Internet]. 2006 May [cited 2015 Jan 16];12(2):134–42. Available from: http://www.sciencedirect.com/science/article/pii/S1078817405000131

Tables

	Medical Students	*Trainee Radiologists*	*Consultant Radiologists*
True positive responses	14.24 (2.81)	17.00 (1.95)	19.83 (2.65)
True negative responses	2.42 (1.13)	3.45 (1.13)	3.41 (0.90)
False positive responses	2.55 (1.10)	1.55 (1.13)	1.58 (0.90)
False negative responses	10.713 (2.72)	7.91 (1.97)	5.16 (2.66)

Table 1 – Group mean (and standard deviations) of performance for responses to the abnormality identification task.

Figures

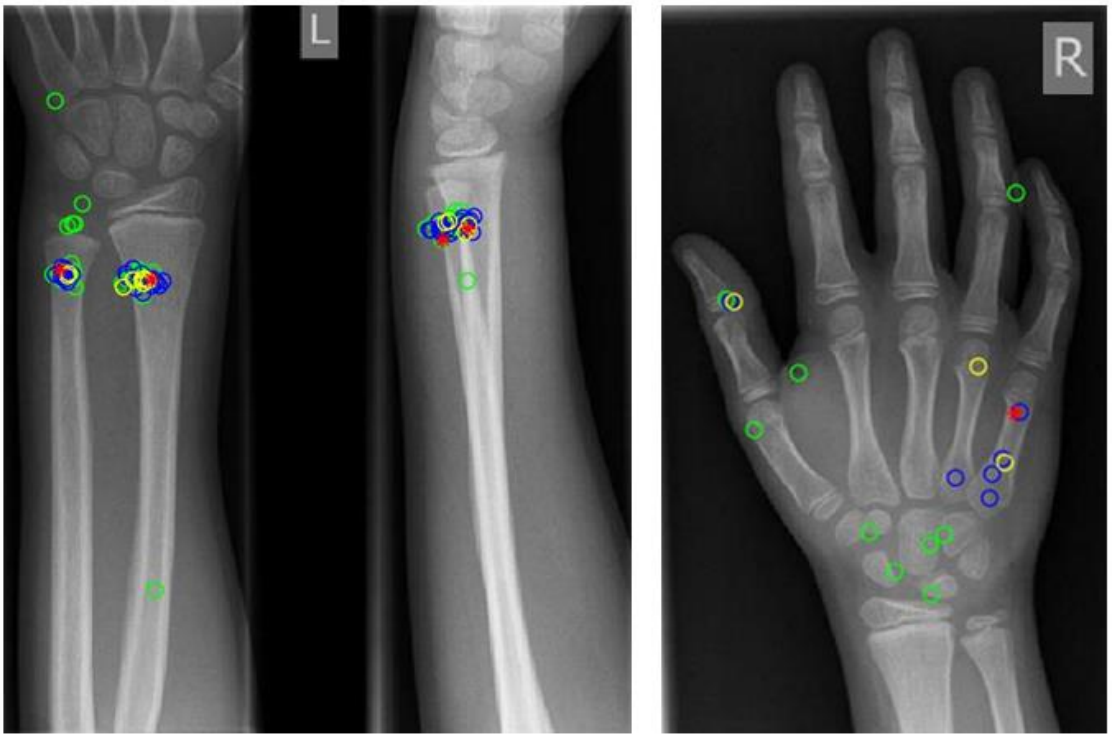

Figure 1 – A. DP and lateral L wrist. B. DP right hand. Example radiographs from the library as presented in the experimental task showing reference locations for abnormalities (red), marks placed by consultants (blue), trainees (yellow), and medical students (green). Figure 1A was graded as easy, and Figure 1b as intermediate.

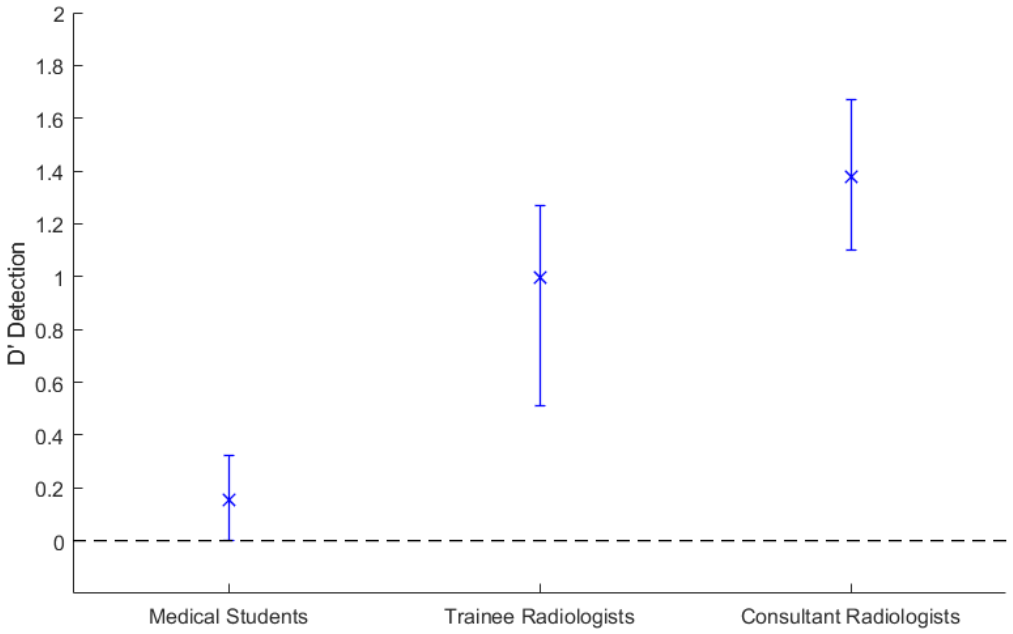

Figure 2 – Average d' for each participant group for abnormality identification. Error bars show bootstrapped 95% confidence intervals.

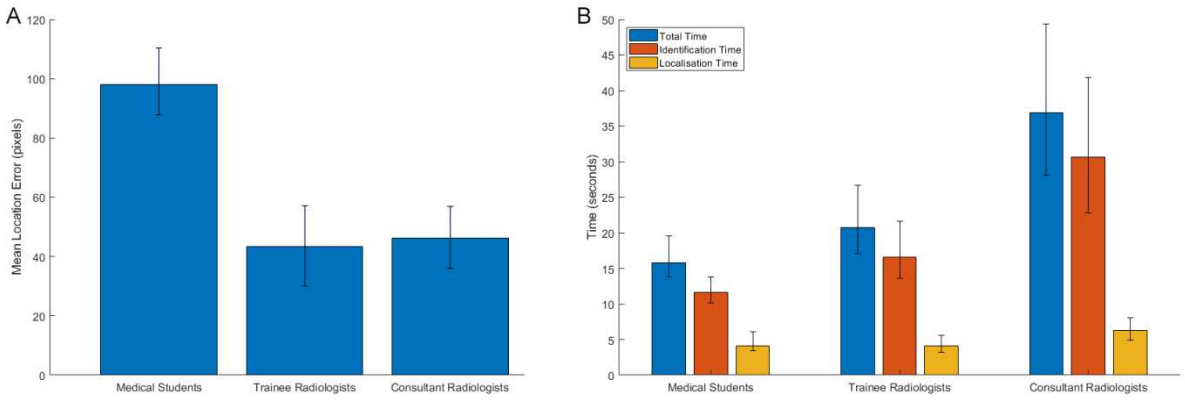

Figure 3 – Group differences in localisation error (A) and time spent per image (B). Medical students were far less precise in their localisation of abnormalities compared to the reference location for each image (A), while responding far quicker than the trainee or consultant radiologists (B). Error bars show bootstrapped 95% confidence intervals

Figure 1

Click here to access/download;Figure;Figure 1.jpg

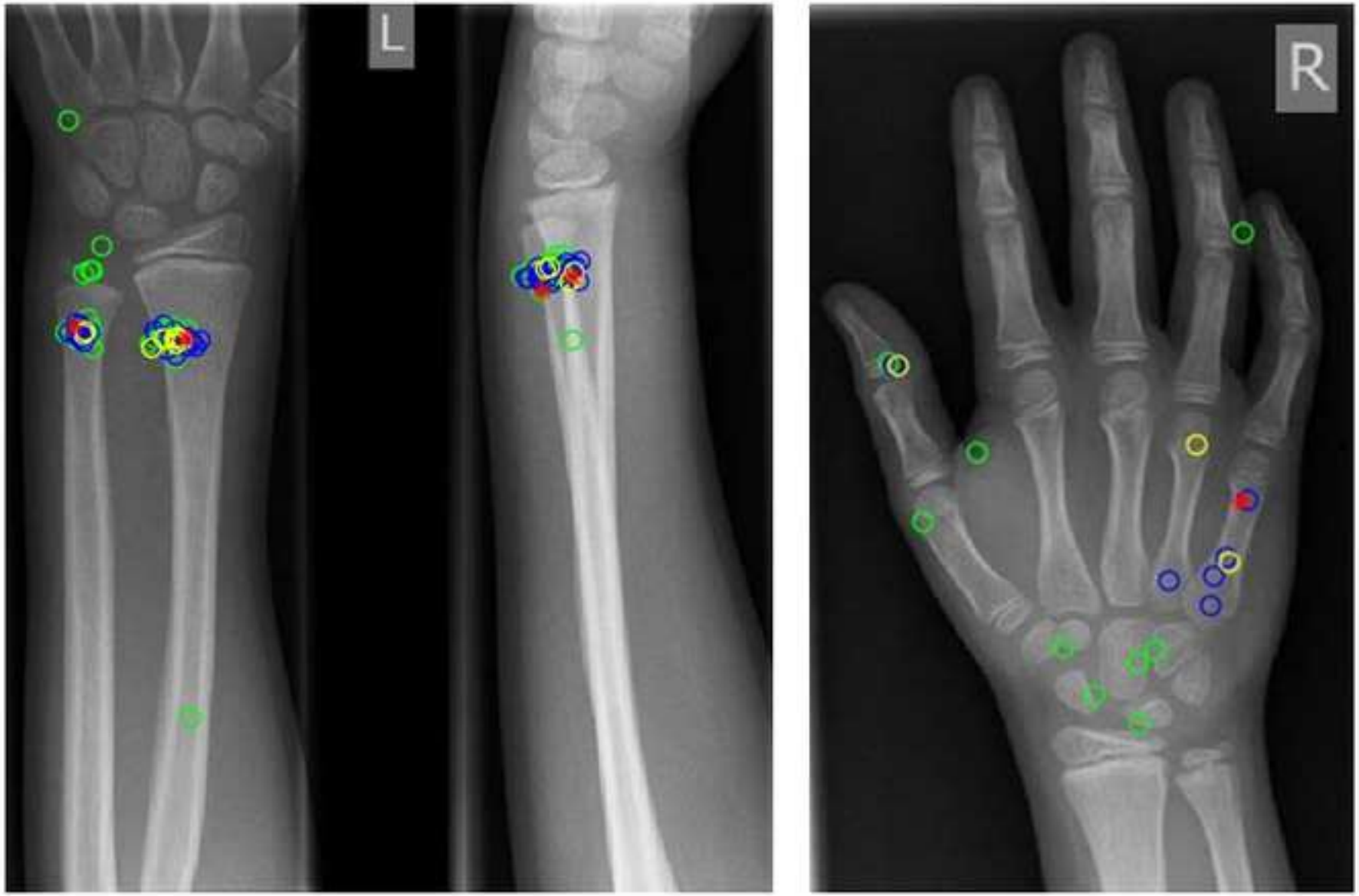

Figure 2

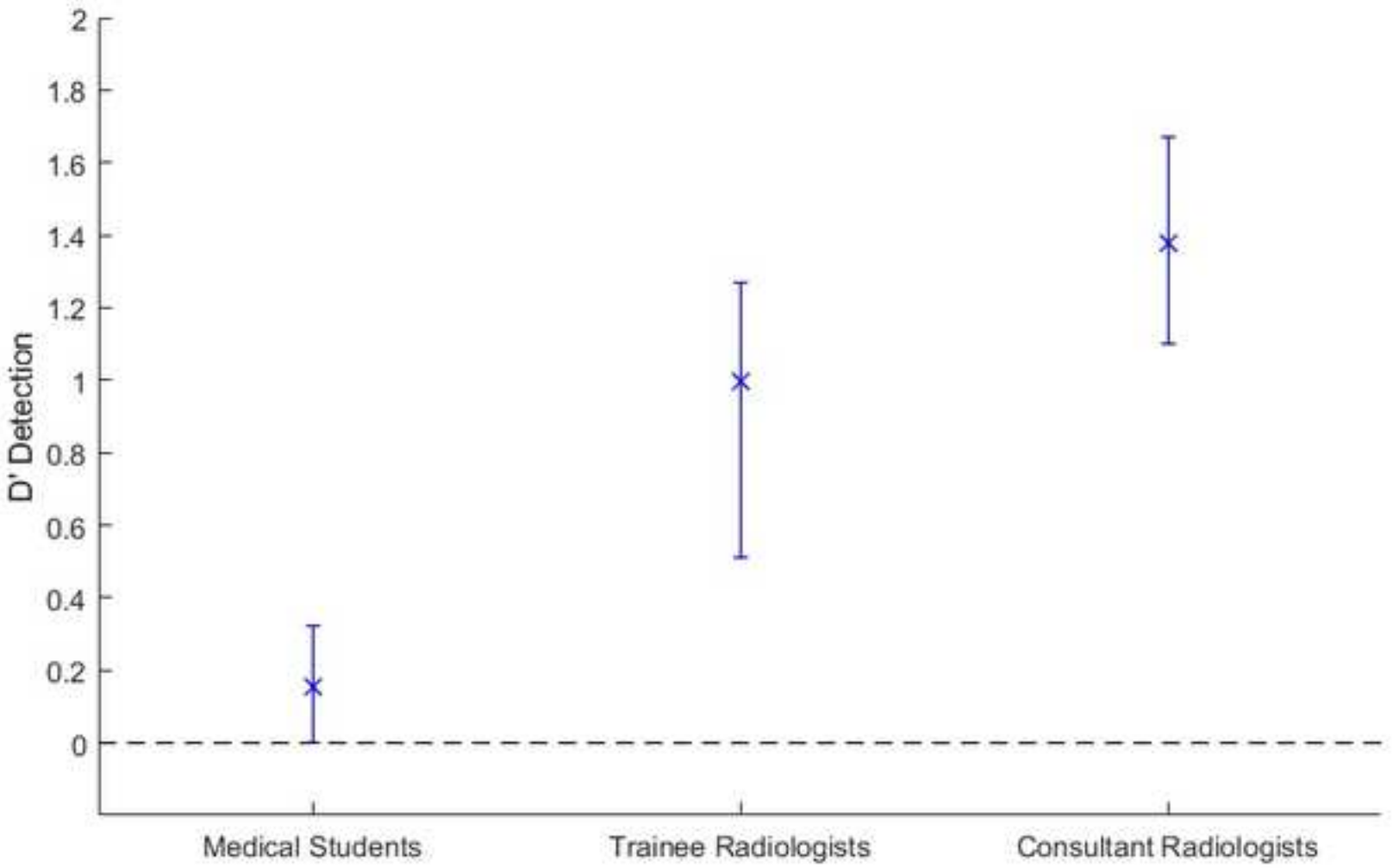

Figure 3

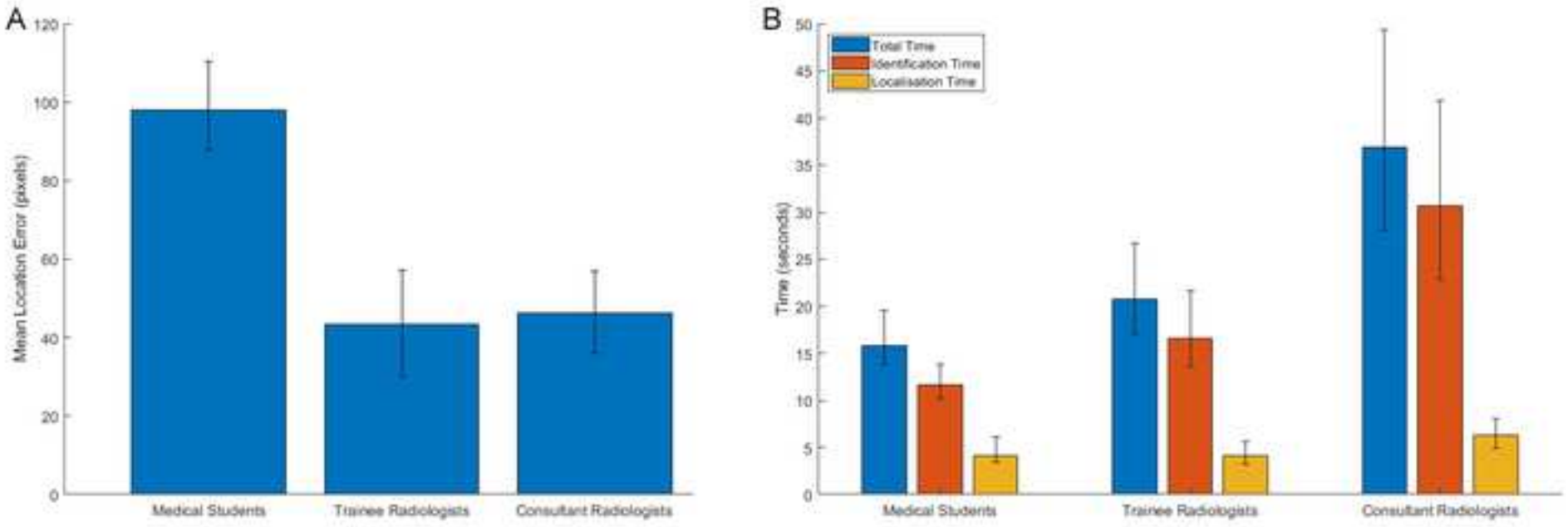

Table 1

	Medical Students	Trainee Radiologists	Consultant Radiologists
True positive responses	14.24 (2.81)	17.00 (1.95)	19.83 (2.65)
True negative responses	2.42 (1.13)	3.45 (1.13)	3.41 (0.90)
False positive responses	2.55 (1.10)	1.55 (1.13)	1.58 (0.90)
False negative responses	10.713 (2.72)	7.91 (1.97)	5.16 (2.66)