ABSTRACT

BACKGROUND: Despite the importance of radiographic interpretation in orthopaedics, there is not a clear understanding of the specific visual strategies used while analyzing a plain film. Eye-tracking technology allows for the objective study of eye movements while performing a dynamic task, such as reading X-rays. Our study looks to elucidate objective differences in image interpretation between novice and experienced orthopaedic trainees using this novel technology.

METHODS: Novice and experienced orthopaedic trainees (N=23) were asked to interpret AP pelvis films, searching for unilateral acetabular fractures while eye-movements were assessed for pattern of gaze, fixation on regions of interest, and time of fixation at regions of interest. Participants were asked to label radiographs as “fractured” or “not fractured.” If “fractured”, the participant was asked to determine the fracture pattern. A control condition employed Ekman faces and participants judged gender and facial emotion. Data were analyzed for variation in eye movements between participants, accuracy of responses, and response time.

RESULTS: Accuracy: There was no significant difference by level of training for accurately identifying fracture images (p=0.3255). There was a significant association between higher level of training and correctly identifying non-fractured images (p=0.0155); greater training was also associated with more success in identifying the correct Judet-Letournel classification (p=0.0029).

Response Time: Greater training was associated with faster response times (p=0.0009 for fracture images and 0.0012 for non-fractured images). Fixation Duration: There was no correlation of average fixation duration with experience (p=0.9632). Regions of Interest (ROIs): More experience was associated with an average of two fewer fixated ROIs (p=0.0047). Number of Fixations: Increased experience was associated with fewer fixations overall (p=0.0007).

CONCLUSIONS: Experience has a significant impact on both accuracy and efficiency in interpreting plain films. Greater training is associated with a shift toward a more efficient and thorough assessment of plain radiographs. Eyetracking is a useful descriptive tool in the setting of plain film interpretation.

CLINICAL RELEVANCE: We propose further assessment of eye movements in larger populations of orthopaedic surgeons, including staff orthopaedists. Describing the differences between novice and expert interpretation may provide insight into ways to accelerate the learning process in young orthopaedists.

Keywords: orthopaedic resident, orthopaedic surgery, technology, radiographs, eyetracking, resident education

INTRODUCTION

Orthopaedic graduate medical education is rapidly changing and evolving. One key to improving orthopaedic education is to promote efficient resident education and surgical skill development. The established apprenticeship model of orthopaedic training is undergoing a transformation with new expectations regarding didactic knowledge, surgical skills, and professionalism in the challenging context of resident work-hour restrictions. Moreover, there are concerns that the current methods of evaluation are not adequately meeting changes in orthopaedic curricula.

In recent years, the American Academy of Orthopaedic Surgery (AAOS), the American Board of Orthopaedic Surgery (ABOS), the Resident Review Committee for Orthopaedic Surgery (RRC-OS), and orthopaedic residencies have all recognized the importance of focused
orthopaedic training. Efforts have been made by several groups to improve surgical skills using stimulation and cadaver models to enhance preparation for the Orthopaedic In-Training Examination (OTTE), and meet the Accreditation Council for Graduate Medical Education (ACGME) core competencies. These efforts insinuate that there is significant room for improvement.

Orthopaedic surgery demands an in-depth knowledge of anatomy and ability to interpret three-dimensional (3-D) anatomy based on two-dimensional (2-D) images, for example, to interpret a fracture pattern based on a radiographic plain film (X-rays). Currently, orthopaedic residency programs employ traditional teaching methods on interpretation of X-rays, but there is often no explicit strategy for interpreting a plain film. Improvement in interpretation of plain films is generally based on repetitive experience, more than directed teaching. Essentially, training programs and educators utilize a “more practice” approach of “more practice”. Subsequent improvement is multifactorial, but is likely a combination of didactic knowledge and improved efficiency of interpretation. These two factors would be consistent with established theories of dissociable memory systems specialized for learning facts versus learning skills. Despite the importance of this skill for clinical practice, to our knowledge no study has evaluated how orthopaedic residents develop or improve their interpretation of radiographs.

Eye-tracking technology allows experimenters to record an individual’s eye movements, including stable points of fixation and rapid jumps between fixations (saccades). Eye movements can be tracked using non-invasive, covert methods while participants are performing dynamic tasks, as well as during apprehension of static, complex information. This methodology has previously been applied in the field of radiology, and has even been used to measure diagnostic decisions and error in human interpretation of medical images. However, eye-tracking methods have been rarely used in the realm of orthopaedics and have not been used to aid understanding of image interpretation to advance the education of residents.

This study was designed to examine the feasibility of the use of eye-tracking methods in orthopaedic plain film interpretation. Furthermore, we believe that analysis of eye movements during radiographic interpretation has the potential to support improved characterization and understanding of the differences in attentional allocation during assessment of plain films. We expected that experienced orthopaedic residents would have more organized, more efficient patterns of image inspection than those with less training. Further, we hypothesized that diagnostic accuracy and speed would be different with experienced trainees being faster and more accurate.

METHODS

Task Stimuli

Forty anterior-posterior (AP) pelvis films were gathered from a single institution’s electronic medical record after Institutional Review Board (IRB) exception was granted. AP pelvis films of skeletally mature patients, ages 18 and older, were retrospectively obtained from the electronic medical record using ICD-9 code 808.0 (closed fracture of acetabulum). Images were obtained from both the emergency room and clinical settings. Images from 2005-2015 were carefully analyzed by two orthopaedic surgeons (authors JMH and JAB) for isolated unilateral acetabular fractures. Twenty fractured images and twenty non-fractured or “normal” AP pelvis X-rays were utilized. Exclusion criteria for a “normal” X-ray included: osteoarthritis, other fracture, CAM or Pincer impingement, hip dysplasia, tumor, AVN, incomplete or poor visualization of all bony architecture, and open physes. Acetabular fractures on the abnormal X-rays were classified using the Judet-Letournel classification system. Exclusion criteria for “fractured” pelvis images included: other fracture, incomplete or poor visualization of bony architecture and images that did not fit the Judet-Letournel classification system of: (1) both column (2) posterior wall, (3) transverse or (4) T-type. These fracture patterns were selected based on a higher reported frequency in the population. All fracture types were confirmed with Computed Tomography (CT) scan. Each plain film was then standardized for size, orientation and image intensity.

Participants

Orthopaedic residents and fourth year orthopaedic sub-interns from a single institution participated in the study. Consent was verbalized after a description of and prior to beginning the experimental task. Participants were given the opportunity to remove themselves from the study at any time for any reason. Basic demographic data was collected, including: gender, age, and level of training. Over a 3-month period, 23 orthopaedic residents and medical students were tested: 2 fourth-year medical students, 4 first-year residents, 4 second-year residents, 3 third-year residents, 6 fourth-year residents and 4 fifth-year residents. Average age was 29.1±2.9 years (25 to 38 years) with 5 (21.7%) females and 18 males (78.3%) (Table 1). All participants were screened for visual acuity during image calibration.

Experimental Task

Participants were asked to interpret AP pelvis films by looking for unilateral acetabular fractures while their eye movements were monitored. Specifically, participants were asked to interpret radiographs as fractured or
non-fractured and, if a fracture was present, to identify the type of fracture to the extent possible with the 2D radiograph. When the participant felt ready to generate an answer, s/he clicked the mouse button to advance to the first response phase.

If “not fractured” was selected, the participant was advanced to the next question. If the participant chose “fractured”, s/he was shown a second display to select the specific pattern of fracture using the Judet-Letournel classification. The response options were: (1) both column; (2) posterior wall; (3) transverse; or (4) T-type (Figure 1). The participants were informed of the total number of images, however were not given any information regarding the relative proportion of fractured versus non-fractured stimuli.

During the task, participants were seated comfortably with their head positioned in a chinrest, facing the computer display at a distance of 50 cm. Eye movements were recorded with a temporal resolution of 1000 Hz using an EyeLink 1000 infrared camera system (SR Research, Ltd., Kanata, Ontario, Canada). Visual stimuli were presented on a 598 x 337-mm liquid crystal display monitor (model W2753VC; LG Electronics, Slough, Berkshire, UK) with a vertical refresh rate of 60 Hz and a resolution of 1920 x 1080 pixels. Visual stimuli were presented using MatLab (R2007b; The MathWorks, Inc., Natick, MA) with the Psychophysics Toolbox extension.

Prior to beginning the eye-tracking task, the EyeLink system was calibrated: participants fixated a visual stimulus at five locations on the screen (center, 15.98° up, 15.98° down, 29.58° left, and 29.58° right). Calibration accuracy was immediately validated by having participants fixate stimuli at the same five locations. If the calibrated eye position was not less than or equal to 0.58° from the stimulus position at each location, the calibration was repeated. If calibration was determined to have been disrupted during the task, the system was recalibrated between trials.

Control Task

In order to control for task understanding and an equivalent ability of participants to fixate upon and process visual stimuli, recognition of emotion and gender was used as a control condition. As suggested by P. Ekman, facial expressions of basic, primary emotions are universally recognized across cultures. Using a set of pictures of facial affect, a control task was designed. Participants were first asked to identify gender: “male” or “female” after being shown an individual face demonstrating a generally accepted and recognizable emotion. Once participants identified each image as “male” or “female”, they were asked to identify which of six emotions was displayed by the face: (1) happiness, (2) disgust, (3) surprise, (4) sadness, (5) fear or (6) anger.

![Table 1. The table demonstrates participant demographics (age, gender and level of training).](image)

Table 1. The table demonstrates participant demographics (age, gender and level of training).

<table>
<thead>
<tr>
<th>Training Level</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>4</td>
<td>17.39</td>
</tr>
<tr>
<td>R2</td>
<td>4</td>
<td>17.39</td>
</tr>
<tr>
<td>R3</td>
<td>3</td>
<td>13.04</td>
</tr>
<tr>
<td>R4</td>
<td>6</td>
<td>26.09</td>
</tr>
<tr>
<td>M</td>
<td>18</td>
<td>78.26</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>21.74</td>
</tr>
</tbody>
</table>

Figure 1. This image shows an example of a fractured AP pelvis x-ray, with subsequent task screens.
In order to assess and analyze eye movements, each X-ray image was divided into a square grid. The authors JMH and JAB then specified ROIs that one would need to have perceived in order to correctly identify the image and answer the questions at hand. These were determined using commonly accepted radiographic landmarks on a pelvic radiograph: the teardrop, the ilioischial line, the ilioinguinal line, the anterior rim of acetabulum and the posterior rim of the acetabulum.

From these data, we were able to analyze the following variables: (1) the number of fixations per test display, which is defined as the number of discrete pauses in eye movements according to criteria including eye velocity and eye acceleration; (2) fixation duration, which is the length of time in which the eye pauses on the image (typically between 200–300 ms long); and (3) the number of regions of interest fixated, which is defined as the number of discrete regions of interest viewed within the image. Eye-tracking data were analyzed separately in MATLAB, Python, and R (version 2.15.1; http://www.r-project.org) using custom software written by author DEW.

The association between level of training and correctly identified fractures as well as fracture type was analyzed using logistic regression adjusted for correlation between responses within participants. The association between speed of interpretation and level of training was evaluated using a linear mixed model. The accuracy and speed of the interpretation of Ekman faces of emotion and gender were utilized as control stimuli and analyzed using similar methods. These data were analyzed using SAS software (V9.4, SAS Institute Inc., Cary, NC). For all statistical tests, alpha was set at p<0.05.

**RESULTS**

**Accuracy:** There was no significant difference by level of training for accurately identifying fracture images (OR=0.92, 95%CI=0.79-1.08, p=0.3255) (Figure 2). However, there was a significant association between higher level of training and correctly identifying non-fractured images (OR=2.05, 95%CI=1.15-3.66, p=0.0155) (Figure 3). Those with more experience were also more successful in identifying the correct Judet-Letournel classification that best fit each fractured image (OR=1.17, 95%CI=1.05-1.29, p=0.0029) (Figure 2). For the control task, there
was no difference related to training duration in accuracy for identifying emotion (p=0.3547) or gender (p=0.9022) for Ekman faces (p<0.05).

**Response Time:** When controlling for age, participants with more training required less time to decide if a fracture was present or not (p=0.0009 for fracture images and 0.0012 for normal images). For the control task, there were no significant associations between response time and level of training (p<0.05). If Ekman face responses were controlled for age, p=0.1853 for gender identification, p=0.3541 for emotion identification. If age was not controlled for, p=0.2059 for gender identification and p=0.6447 for emotion response.

**Fixation Duration:** There was no difference of average duration of fixations between images (p=0.2526). Additionally, there was no correlation of average fixation duration with experience (p=0.9632).

**Regions of Interest:** Across groups, an average of one more grid ROI was fixated for fractured images when compared to non-fractured images (p=0.0005) (Figure 4). More experience was associated with an average of two fewer ROIs fixated (p=0.0047).

**Number of Fixations:** Increased experience was associated with fewer fixations (p=0.0007) (Figure 5). On average, each year of experience was associated with 9.4 fewer image fixations. In addition, there was a significant difference in the number of fixations between the fractured and non-fractured images (p=0.0005).

**DISCUSSION**

Eye-tracking is a feasible technique of identifying individual differences in image interpretation skill among orthopaedic residents and students of variable levels of training and experience. As one would expect, diagnostic accuracy and efficiency were generally better in students with more experience. However, these results also highlight specific and potentially modifiable elements of image interpretation behavior between experienced and novice orthopaedic trainees.

Myles-Worsley and colleagues surmised that radiologic expertise depends on two kinds of knowledge: 1) knowledge of the characteristic features of clinically normal exemplars of a class of X-ray films and 2) knowledge of the particular set of uncharacteristic features that signal pathology. Focused study of this knowledge can improve performance measures in early training. Objective analysis of eye movements in novice and more experienced orthopaedic trainees may be sufficient to identify skill gaps based on where eye movements and attention are focused when viewing plain radiographs. Understanding the differences in image interpretation between early and late stages of training could help to formulate directed curriculums with the goal of instructing young residents in more efficient and accurate plain film interpretation.

Accuracy among training levels was similar when fractured images were viewed but there was a significant difference when scrutinizing normal images. The ability to properly identify a non-fractured pelvis was associated with year in residency: those with more training were more likely to correctly identify normal bony anatomy. This would suggest that novice orthopaedists have difficulty with fixating on subtle abnormalities and distinguishing relevant and irrelevant observations. Different angles of the X-ray beam can create subtle shadows that may confuse those with less experience, leading them to interpret overlay of normal three-dimensional anatomy as a fracture.

Although participants with less experience could identify the presence of an acetabular fracture as well as those with more training, novices had a difficult time correctly characterizing fractures using a globally accepted classification system. This is likely related to a lack of familiarity and experience with the Judet-Letournel acetabular fracture classification system, and it underlines the importance of both recognition and interpretation of pathology in image interpretation.

Participants with more experience also tended to be more efficient in formulating a response. Those with more training required fewer fixations when viewing an image, as they presumably were familiar with certain characteristic fracture patterns. Fewer fixations could imply a more gestalt understanding of pelvic anatomy and a holistic approach to image visualization rather than a “search and find” strategy. This would be consistent with the suggestion that the improved performance demonstrated by the experienced orthopaedist involves a shift in the mechanism of image perception from a piecemeal approach to a more holistic methodology. Interestingly, more grid ROIs were fixated when an image was fractured, despite the fact that the entire pelvis would need to be examined to rule out fractures in a “normal” image. This pattern may have emerged because viewers required additional fixations to classify the specific fracture pattern, which generally necessitates the use of several different radiographic markers. A variant of this study in which the task was simply to distinguish fractured from non-fractured images could address this prediction.

Our results are consistent with Wood et al 2013, whose study compared eye-tracking data among radiology students and experts while interpreting fractures in different anatomic locations with varying levels of complexity. Their results suggest that the performance advantage of expert radiologists is underpinned by superior pattern recognition skills, as evidenced by a quicker time to first fixate the pathology, and less time spent searching the image.
The literature suggests that expertise in image interpretation requires the recognition of patterns of abnormality, the interpretation of such patterns and the maintenance of flexibility in interpretation when new information is presented\(^8\). Visual experts believe that in order to improve performance, one must build a mental repertoire of patterns of normality and abnormality\(^9,10\). This should be in concert with specific directed feedback, as novices have limited ability to self-assess\(^11,12\). Dedicated time should be spent teaching the subtleties of interpreting angles from an X-ray beam and radiographic shadows.

With guidance and practice, a novice can develop a more gestalt understanding of three-dimensional anatomy from a 2D image, which in turn, leads to more efficient and accurate detection of the presence and characteristics of a fracture on plain radiographs. Furthermore, a structured curriculum for teaching X-ray interpretation would allow for frequent feedback and affirmation to ensure a novice continues to develop skill and understanding.

There are several limitations to this study. First, this is an experimental situation and the results must be interpreted within this setting. Participants were informed of the goal of our study prior as appropriate for experimental research. This could have potentially introduced study bias and selectively influenced the participant’s eye movements and directed his/her attention toward the acetabulum only. While the images were standardized, the physical space and testing scenario controlled, and the task proctored by a member of the research team, participant attention and effort could have affected our data.

Additionally, this is a small population of participants at a single institution. Future studies should include more participants from other orthopaedic training programs to determine if our findings are generalizable. Lastly, the classification of an acetabular fracture on plain film often depends on the interpretation of Judet views, which were not presented to participants. This might be addressed by using a more sophisticated experimental task using oblique views.

The current work was limited to acetabular fracture identification, but has the potential to be applied to other anatomic locations (i.e. hand or knee radiographs) and broadly applied in the realm of orthopaedic education.

CONCLUSIONS
This study demonstrates that eyetracking is a useful, objective approach to quantifying and describing differences in radiographic image interpretation strategies in orthopaedists with varying levels of training. Our results suggest that experience has a significant impact on radiographic interpretation strategy, accuracy and efficiency. Analysis of differences associated with training identified several modifiable elements of image interpretation. Further work in this area may be useful in formulating directed curriculums to improve plain film interpretation in the young orthopaedist.

SOURCE OF FUNDING
Funding for this project was provided by a development research grant from the University of Iowa Department of Graduate Medical Education.

REFERENCES


