Case Comparisons:

An Efficient Way of Learning Radiology

Ellen M. Kok, MSc, Anique B. H. de Bruin, PhD, Jimmie Leppink, PhD, Jeroen J. G. van Merriënboer, PhD, Simon G. F. Robben, PhD

Rationale and Objectives: Radiologists commonly use comparison films to improve their differential diagnosis. Educational literature suggests that this technique might also be used to bolster the process of *learning* to interpret radiographs. We investigated the effectiveness of three comparison techniques in medical students, whom we invited to compare cases of the same disease (same-disease comparison), cases of different diseases (different-disease comparison), disease images with normal images (disease/normal comparison), and identical images (no comparison/control condition). Furthermore, we used eye-tracking technology to investigate which elements of the two cases were compared by the students.

Materials and Methods: We randomly assigned 84 medical students to one of four conditions and had them study different diseases on chest radiographs, while their eye movements were being measured. Thereafter, participants took two tests that measured diagnostic performance and their ability to locate diseases, respectively.

Results: Students studied most efficiently in the same-disease and different-disease comparison conditions: test 1, $F(3, 68) = 3.31$, $P = .025$, $\eta_p^2 = 0.128$; test 2, $F(3, 65) = 2.88$, $P = .043$, $\eta_p^2 = 0.117$. We found that comparisons were effected in 91% of all trials (except for the control condition). Comparisons between normal anatomy were particularly common (45.8%) in all conditions.

Conclusions: Comparing cases can be an efficient way of learning to interpret radiographs, especially when the comparison technique used is specifically tailored to the learning goal. Eye tracking provided insight into the comparison process, by showing that few comparisons were made between abnormalities, for example.

Key Words: Case comparison; eye movements; education; learning; radiology.

It is common practice for radiologists to compare films of a particular patient over time. This practice is taught to radiologist in training (1). It was found that, especially in the case of junior radiology residents, abnormalities are more easily detected when a prior image with no abnormalities (normal image) is presented alongside the case to be diagnosed (2). Hence, comparison can help to differentiate abnormalities from normal anatomy (3).

In a context of radiology education, it is of paramount importance that students *learn* to recognize common abnormalities on radiographs (4). Educational literature suggests that the use of comparison can bolster this learning process (5–8). The web-based training program COMPARE (University of Erlangen-Nuremberg, Erlangen, Germany) (5,7), for example, uses a page format in which a normal image flanks a pathologic image, and students are prompted to compare these. As much as 91% of the students and 88% of the residents who used this program valued the technique as useful or very useful (7). In addition, it was found that students learned more effectively when comparing focal diseases (ie, lesions in one location) to normal images than when comparing two pathologic images (6).

What the aforementioned studies did not probe, however, is whether such a pathologic/normal comparison technique still holds superiority in the face of a no-comparison/control condition. Besides this alternative, two other comparison options have been left uninvestigated: comparison of two images of patients with different diseases and comparison of two images of patients with the same disease. The extent to which these different comparison techniques can be effective for learning, to date, has not been investigated.

Arguably, case comparisons could be more time-consuming than a simple review of individual cases; therefore, it is important that the time spent on learning be recorded. In addition, caution should be exercised that learning materials are not presented in a suboptimal way, as this can impose an extraneous cognitive load on students' minds, that is, a cognitive load that does not contribute to learning but may hamper learning (9). Therefore, it is critical to check that the addition of a second case for comparison purposes does not inflate extraneous cognitive load. These two factors could influence

Acad Radiol 2015; ■:1–10

From the Department of Educational Development and Research, School of Health Professions Education, Maastricht University, PO Box 616, Maastricht, MD 6200, The Netherlands (E.M.K., A.B.H.d.B, J.L., J.J.G.v.M.); and Department of Radiology, Maastricht University Medical Center, Maastricht, The Netherlands (S.G.F.R.). Received March 5, 2015; accepted April 27, 2015. **Address correspondence to:** E.M.K. e-mail: e.kok@maastrichtuniversity.nl

http://dx.doi.org/10.1016/j.acra.2015.04.012

the extent to which case comparisons can be effective techniques for learning to interpret chest radiographs.

Another question that remains unanswered is how students avail themselves of the opportunity to compare; researchers are still in the dark about what happens during the comparison process. More specifically, we do not even know whether comparisons are actually effected when participants are presented with two or more juxtaposed images. For example, the COMPARE program instructs participants to compare the pathologic image with the normal image, but the researchers have to take it for granted that the participants actually adhere to these instructions. In such cases, eye-tracking technology (10) can provide a solution, as it measures the movements of the eye to see what a person is looking at, for how long, and in what order. As such, it can be deployed to verify and quantify the degree of comparison taking place, as well as to reveal the exact parts of the images that are being compared.

The present study has two aims. The first aim is to assess the effectiveness of three different comparison techniques in relation to a no-comparison/control condition. The second aim is to investigate which parts of the images are being compared by using eye tracking. In particular, we expect two types of comparisons to be effective for learning. First, comparing abnormalities to each other or to normal tissue could help students understand distinguishing features of abnormalities. Second, comparison of the normal tissue between two images (such as the shape of the hila in two patients) could help students learn what normal tissue looks like.

MATERIALS AND METHODS

Procedure

Participants were invited to study a series of 48 chest radiographs that were captioned with a diagnosis each and were always presented in sets of two. Participants were randomly assigned to one of four conditions in which they were asked to compare (1) cases of the same disease (same-disease condition), (2) cases of different diseases (different-disease condition), (3) disease images with normal images (disease/normal condition), and (4) identical images (no-comparison/control condition). The images were paired in accordance with the condition as follows: in the first condition, each disease case was put adjacent to a case of the same disease but pertinent to another patient; in the second condition, each disease case was paired with an image of another disease; in the third condition, each disease case was placed alongside a normal image, that is, an image showing no abnormalities; and finally in the control condition, each case was put beside an identical case, so comparison was pointless. Figure 1 showcases examples of such case pairs for each of these four conditions.

Although the participants in the first three conditions received explicit instructions to compare the two images, those in the control condition were informed about the two images being identical. All case pairs were presented in a random order and had a 30-second time slot each, but moving on to the next case pair was allowed if the participant finished earlier. The 30-second maximum was based on pilot testing.

First, the eye tracker was calibrated by repeating a 9-point calibration until accuracy was less than 1° of visual angle on both the x- and y-axis. As they had their eye movements measured, participants undertook to study the case pairs. As soon as this study phase had ended, the eye tracker was turned off. Participants subsequently indicated the extent to which they had experienced extraneous cognitive load during studying the case pairs. They then proceeded with two tests, which were identical for all participants: (1) a multiple-choice question (MCQ) test of 30 questions, which aimed to measure diagnostic performance; and (2) a region of interest (ROI) test that required participants to indicate which part of the image was abnormal by drawing an ROI around the abnormality (ROI test) to measure their ability to locate the disease. The experiment ended by thanking the participants for participation and presenting them a gift voucher.

Participants

A total of 84 third-year medical students (65 female) were participants, with mean age of 22.06 years (standard deviation, SD = 1.54). Three students were excluded from the analysis outright, as two of them reported a substantial amount of prior experience of radiology (>50 hours), and the third one had accidentally partaken in the study phase of two conditions. The 81 students that remained reported little prior experience of radiology (<2 hours) and were evenly distributed between the four conditions, with 21 participants in the same-disease condition and 20 participants in each of the other conditions. Furthermore, eye-tracking data of nine participants were excluded from the analysis as well because of insufficient data quality (ie, during calibration, the threshold of 1° of visual angle could not be reached). Eventually, the analysis of eye-tracking data included 20 participants in the same-disease condition, 17 in the different-disease condition, 16 in the disease/normal condition, and 19 in the control condition.

Cases

Although the term "case" is usually taken to denote the ensemble of one or more radiographs, patient history, and clinical questions for the purpose of this experiment, we use this term to refer to individual posterioranterior (PA) chest radiographs void of any additional information. For each of eight different diseases, a board-certified radiologist collected nine cases with a typical radiographic manifestation. The final diagnosis was established based on clinical information, clinical course, and other images (eg, computed tomography or chest radiographs made at other moments). Four of these diseases were focal in kind (atelectasis, solitary lung tumor, pneumonia, and pleural effusion), that is, the abnormality was centered in one location with the rest of the lung being normal (11), whereas the other four were diffuse diseases, in which the whole lung was abnormal (cystic fibrosis, lung fibrosis, metastases, and