ACQUIRING EXPERTISE IN RADIOLOGY

Studies on Development & Assessment of Image Interpretation Skills

Anouk van der Gijp

ISBN: 978-94-92683-68-7
Layout and printed by: Optima Grafische Communicatie, Rotterdam, the Netherlands (www.ogc.nl)

Acquiring Expertise in Radiology
Studies on Development & Assessment of Image Interpretation Skills

Verwerven van Expertise in Radiologie
Studies over de Ontwikkeling en Toetsing van Beeldinterpretatievaardigheden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 19 september 2017 des middags te 4.15 uur

door

Anouk van der Gijp
geboren op 18 oktober 1983 te Apeldoorn

Promotoren:	Prof.dr. J.P.J. van Schaik
	Prof.dr. Th.J. ten Cate
Copromotoren:	Dr. M.F. van der Schaaf
	Dr. I.C. van der Schaaf

CONTENTS

Chapter 1

General introduction

INTRODUCTION

Background

Diagnostic errors can have significant consequences for patients and are an important source of patient harm[1, 2]. A substantial number of diagnostic errors is due to misinterpretation of medical images[3] and high error rates are repeatedly reported in radiology[4-6]. The use of radiological images in clinical decision making is increasing[7, 8] and medical imaging techniques have changed tremendously in recent years[9]. Radiologists traditionally interpreted CT and MRI scans viewing two-dimensional hard-copy images on a light box. Nowadays, they interpret volumetric images displayed on a computer monitor, and they can navigate through the images, change viewing direction, and adjust contrast settings. These advancements have also altered the visual diagnostic process. The amount of visual information processing has tremendously increased, and the possibilities to approach or manipulate the images have expanded[9]. How these advancements affect the visual diagnostic reasoning process and how we should align our training and assessment methods remains unknown and is the subject of this thesis.

In this dissertation, we investigate which knowledge and skills are important for the development of image interpretation expertise in the era of advanced medical imaging techniques, and evaluate training and assessment methods to stimulate image interpretation expertise development.

Concepts and Definitions

Radiology is a medical specialty characterized by obtaining image representations of the human body with various imaging techniques. The medical images are used to diagnose and treat illnesses. The part of the profession that primarily deals with treating illnesses, or diagnosing illnesses with invasive procedures, is called interventional radiology. The part primarily dealing with noninvasive diagnostic imaging is called diagnostic radiology. The studies in this thesis concern diagnostic radiology, and focus on the development of diagnostic radiological skills. In diagnostic radiology, image interpretation is a central skill: radiologists are trained to diagnose or exclude diseases based on images. We primarily use the term *image interpretation* skill to refer to the skill of diagnosing or excluding diseases based on images.

Medical image interpretation refers to diagnostic reasoning mainly based on visual input from *medical images*. Diagnostic reasoning based on visual information is not limited to radiology and occurs in many medical professions, using various types of visual information. A variety of expertise research has been conducted in other visual medical domains such as dermatology[10] and pathology[11, 12]. Even a physical examination has visual components and can be considered diagnostic reasoning based on visual information. An example in the literature is the skill of seizure characterization in pediatric neurology[13]. *Visual diagnostic rea-*

soning refers to diagnostic reasoning based on the processing of mainly visual information. Resources of visual information may vary and may be image or non-image based. Advances in knowledge about radiological image interpretation development can be valuable for other medical professions as well. To connect to these other domains, we sometimes refer to the image interpretation process as *visual diagnostic process* instead of using the more narrow definition of image interpretation.

The thesis concerns 2D and volumetric image interpretation. In radiology education, 2D images may be X-ray images or single cross-sections. X-ray images are projections of anatomical structures in one direction. A cross-section is a virtual cut through of a human body part which is a computational representation of x-ray projections in multiple directions. A volumetric image is a set of two-dimensional cross-sections of a human body part. Observers scroll through thin cross-sections (one or some millimeters thick slices) to visualize the image data. Scrolling through the dataset provides extensive and detailed tissue information. A mental representation of the organs, body fluids and other structures and how they relate to each other can be formed by connecting these separate cross-sections to each other in a virtual three-dimensional space. The technique is called volumetric, because the set of cross-sections cover the complete volume of the body part, and observers can scroll through the volume, frequently in many viewing directions. For the interpretation of the results of our studies, it is important to realize that volumetric images and 3D images are not the same. When volumetric images are reconstructed into three-dimensional renderings, a true 3D representation of the image can be obtained. This type of image can be rotated to view the surface from different perspectives. These reconstruction methods can be useful, for example for the 3D representation of the bony structures of the skull in patients with a skull fracture to assist surgeons in planning their operation strategy. The counter side of this technique is that much information gets lost, because only the surface of the body part is visible. In the example of a 3D reconstruction of the skull, the structures in the skull such as brain tissue and blood vessels are not visible in a 3D reconstruction.

Expertise in Image Interpretation

From visual expertise research in radiology we know that in some types of imaging expert radiologists are able to diagnose diseases in a split second[14, 15]. This short time span does not allow for eye movements or a deliberate reasoning process. It is a nonanalytic process based on a global impression of the image and immediate pattern recognition, based on memorization of previous examples. Novices do not have an adequate repository of disease examples in their minds and cannot rely on pattern recognition. They typically take more time to interpret images[16-25] and deliberately search the image to find and analyze abnormalities[16, 26, 27]. Kundel and Nodine have extensively studied the visual search behavior of radiologists and distinguish three phases in the Visual Search and Detection Model: (1) global impression, (2) focal search and attention, and (3) diagnostic decision making[28]. The first phase entails the first glance on the image, providing an overall, global orientation. In this fast initial impression, radiologists are thought to compare the image with a cognitive scheme of anatomy and pathology. Potential deviations are flagged for the focal analysis in the second phase. In the second phase, radiologists start moving their eyes over the image to examine potential targets and perturbations that have been flagged in the first phase, and to search for other abnormalities not yet detected in the first phase. Finally, in the decision making phase potential targets are inspected in detail to inform decisions about the presence or absence of a lesion. This model is the basis of more recent visual search theories, that distinguish global and focal components of visual search[15, 27, 29].

The current theories of visual expertise are primarily based on studies that concern 2D image interpretation, while volumetric images are commonplace in radiology practice. Volumetric images contain much more visual information than 2D images. Scrolling through the images is necessary to visualize potential relevant information. Abnormalities can be missed not only because of a faulty visual search, but also due to the combined actions of scroll behavior and eye movements. Some early studies that investigated expertise in volumetric image interpretation discerned different types of scroll behavior [30, 31]. Drew et al. distinguish 'drillers' and 'scanners': drillers tend to focus on one area of the image and scroll up and down, before moving on to the next area, whereas scanners scan the complete image before scrolling to the next cross-section[30]. Venjakob et al. described differences in scroll behavior based on the length of the scroll movements: long movements were qualified as 'runs' and short movements as 'oscillations'[31].

In this thesis, we study expertise development in both 2D and volumetric image interpretation. First, the literature on expertise is extensively reviewed to identify knowledge and skills that are required to develop expertise in both 2D and volumetric image interpretation. Next, differences between 2D and volumetric image interpretation expertise are explored. Then, the effect of scroll behavior on perceptual performance is investigated.

Visual Diagnostic Errors

A substantial part of diagnostic errors is due to mistakes in diagnostic reasoning[1, 32]. Visual diagnostic errors can be caused by many factors, observer-related or related to other factors such as technical or image-related issues[33]. A well-known observer-related factor is the effect of cognitive biases[34]. Cognitive biases that frequently occur in radiology are availability bias, confirmation bias and satisfaction of search[35-39]. However, the evidence that cognitive biases cause errors in clinical practice is sparse[40] and interrater agreements for classifying biases are poor[41]. An established classification for radiological errors is the error type classification of Kundel et al.[42], that distinguishes scanning, recognition and decision-making errors based on the amount of time observers look at abnormalities. A drawback of this classification is that it only categorizes the perceptual part of visual expertise (decision making whether there is an abnormality or not) without evaluating the diagnostic reasoning process, including the observer`s thoughts on what is the diagnosis. Therefore other reliable error type classifications that have the potential to direct learning activities should be investigated. In this thesis we explore an error classification approach that covers the visual diagnostic reasoning process and is based on an assessment method that is easy to implement in an educational program.

Expertise in Clinical Reasoning

Although there are differences between image interpretation and diagnostic reasoning without images, there are also similarities. A substantial part of image interpretation concerns clinical reasoning. The view on expertise in clinical reasoning has changed over the last decades. Initially, diagnostic reasoning was thought to be a general problem-solving skill, characterized by a process of hypothesis testing: the ability to quickly generate high-quality diagnostic hypotheses, and collect data to confirm or rule out these possibilities[43-45]. This hypothetico-deductive model assumed that diagnostic reasoning was a domain-independent skill, and mastery of this skill would enable solving clinical problems one had never encountered before. However, the clinical problem-solving process was found to vary little across different levels of expertise[46]. Besides, individual clinical problem-solving performance was found to vary considerably across tasks[43, 47] and diagnostic accuracy in one patient scenario could not predict performance in another one. This phenomenon is known as content-specificity[43] and indicates that clinical problem solving cannot be considered a general skill separated from knowledge. Solving a clinical case requires domain-specific knowledge.

This raises the question of what types of knowledge are required for diagnostic reasoning and how this knowledge should be organized. Experiments in several clinical areas have shown that experts, in contrast to novices, use little to no knowledge of basic sciences in their routine clinical reasoning[48-50]. Experts tend to primarily rely on clinical knowledge and experience. However, there is some evidence that, in very difficult cases, experts use more basic science knowledge than novices[51]. Another type of knowledge is knowledge concerning manifestations of diseases, which can be organized in 'illness scripts'[52-54]. These are knowledge representations that contain patient-oriented clinical knowledge[53], for example a list of symptoms that are likely to occur in a certain disease. Pattern recognition based on illness scripts has been related to higher diagnostic accuracy rates than the hypothetico-deductive approach[55].

Still, expertise in diagnostic performance cannot be explained only by differences in the amount, use or organization of formal knowledge. Expert clinicians are even found to use less formal knowledge[48-50] and analyze less clinical patient features than less proficient physicians[56]. There are several indications that familiarity plays an important role in diagnostic decision-making[57-59]. This relates to psychological theories of categorization[60]; a cognitive process of everyday life used to identify objects. For example, we can identify different types of chairs without analyzing their features because we have seen many previous examples. Experts have encountered many different representations of diseases through experience.

This enables them to intuitively solve new patient cases, by matching the case with mental representations of the disease[61]. This non-analytical way of thinking may also explain why experts reach the correct diagnosis much faster than novices[10, 61, 62]. In case the diagnostic problem stays unsolved, an inverse relationship can be found and experts take more time than novices[10]. This suggests that experts shift to a slower, more analytical way of thinking in difficult or ambiguous cases.

Theories of Expertise

The difference between analytical and non-analytical approaches are also described in the Dual Process Theory. This theory offers a model for information processing and discerns two ways of 'thinking': System 1 and System 2[63, 64]. System 1 is an intuitive, non-analytical process, while System 2 is a conscious, analytical process. System 1 is mainly based on fast pattern recognition and is an efficient way of information processing. It could be vulnerable to mistakes, especially in complex or atypical situations[65]. However, it is unclear which of the two systems is superior in general or in specific situations[66-68] Besides, the existence of a truly clear distinction between these two systems is debatable and some authors suggest that the systems may run in parallel[67, 69] or may represent the extremes of a cognitive continuum[70].

Experts and novices clearly differ in their approach to clinical problems. The question remains how people become experts. The Expertise Theory of Anders Ericsson[71] deals with the development of expert performance. This theory advocates that expert performance is not primarily dependent on personal characteristics, but rather the result of knowledge and skill acquisition based on extensive experience and practice. To reach high levels of expertise, the learner should be engaged and reflect upon his or her performance, and the activity should be accompanied by coaching or feedback of an expert to refine performance[72]. This cognitively effortful practice is called 'deliberate practice'.

Developing visual diagnostic expertise requires extensive experience and practice[20, 73,74]. There is still a considerable variation in performance of trained radiologists[75-77]. According to the Expertise Theory, deliberate practice dedicated to radiology will improve image interpretation performance. The question remains what exactly should be trained. The knowledge and skills that are required to become an expert in radiology are still largely unknown. There are some training methods that have proven to stimulate learning in radiology, i.e. learning by comparing normal and abnormal cases[78] and learning from Eye Movement Modeling Examples[79]. Training learners to adhere to a specific search strategy (systematic search) has not proven to be effective so far[23]. However, these before mentioned methods were all applied to 2D image interpretation tasks and results cannot be directly generalized to volumetric image interpretation. Learners may benefit more from an efficient search strategy when processing the much larger amount of visual information of volumetric images. In our studies, we search for methods and tools to support learners and teachers in their deliberate practice to develop expertise in 2D and volumetric image interpretation.

Methods To Investigate Visual Diagnostic Expertise

Both perceptual and cognitive processes interactively contribute to the complex skill of visual diagnostic reasoning. To investigate the nature of visual expertise, the perceptual component is mainly investigated by eye tracking. The location, time span and order of eye movements reflect the observer's attention[80]. Inferences can be drawn from differences in attention, for example between different expertise levels. Although eye tracking technology has emerged in the past decades and is an established research method in 2D images, eye tracking in volumetric images remains challenging[81]. Cognitive processes in visual expertise are primarily investigated with verbal protocol studies. Verbal protocols can be obtained during or after the activity, concurrent or retrospective protocols[82] respectively. Concurrent protocols are thought to give a more complete representation of the thoughts of participants[83], but may interfere with the task at hand, especially in particular task, such as reading a text[82]. In digital image interpretation, logfiles of mouse movements or keyboard activities can be analyzed. In visual diagnostic expertise, logfile analysis can be used to track and analyze how learners use images[84] or investigate scroll patterns (chapter 5, this thesis).

AIM

The studies in this thesis aim: 1) to investigate which knowledge and skills are important for the development of expertise in radiological 2D and volumetric image interpretation and 2) to contribute to the development and evaluation of training and assessment methods to stimulate expertise development in radiological 2D and volumetric image interpretation. Chapter 2 and 3 provide insight into the domain-specific knowledge and skills that are required for 2D and volumetric image interpretation. In chapter 4 the visual component of image interpretation is examined more closely with a synthesis of the existing eye tracking literature. Chapter 5 investigates whether learners can be taught to use expert search patterns in volumetric image interpretation and how this affects performance. Chapter 6 evaluates an assessment method developed to reveal and classify errors in the reasoning process of image interpretation. Chapter 7 evaluates an assessment method that was developed to improve the authenticity of image interpretation testing.

RESEARCH QUESTIONS

The research questions are twofold.
Research Question 1. Which knowledge and skills contribute to expertise development in radiological image interpretation?

Sub-questions are:

1.1 Which domain-specific knowledge and skills are needed for radiological image interpretation? (Chapter 2)
1.2 How do these skills differ between 2D and volumetric image interpretation? (Chapter 3)
1.3 How does visual search relate to diagnostic performance? (Chapter 4)

Research Question 2. How can we improve training and assessment to stimulate radiological image interpretation expertise development?

Sub-questions are:

2.1 What is the effect of teaching search strategies on perceptual performance? (Chapter 5)
2.2 How can we identify different types of errors that occur in the visual diagnostic process? (Chapter 6)
2.3 How can digital simulation-based assessment increase the authenticity of volumetric image interpretation skill testing? (Chapter 7)

THESIS OUTLINE

Chapter 2 is a mixed-method study for developing a conceptual framework of knowledge and skills required for image interpretation. The study consists of three phases: a literature study, expert interviews and a think-aloud experiment. In an iterative developmental process, each phase was used to create or adjust preliminary versions of the framework, and ultimately resulted in a final version of the framework.

Chapter 3 describes a think-aloud experiment that compares knowledge and skills used for interpretation of volumetric versus 2D images. We hypothesized that interpretation of volumetric images demands different knowledge and skills than the interpretation of 2D images. Participants were asked to think out loud while reading volumetric and 2D CT images. The framework from chapter 2 was used to categorize knowledge and skills. A within-subject analysis was performed to compare cognitive processes during volumetric versus 2D image interpretation.

Chapter 4 is a review study of eye tracking literature in the radiology domain that aims to identify visual search patterns associated with high perceptual performance. Six electronic literature databases were searched using 'visual perception' OR 'eye tracking' AND 'radiology' and synonyms. A thematic analysis was conducted to extract and arrange study results, and a textual narrative synthesis was applied for data integration and interpretation.

Chapter 5 is an experimental study investigating the effect of teaching search strategies on perceptual performance for lung nodule detection at CT. Two search strategies, 'scanning' and 'drilling', were tested with a randomized cross-over design. Nineteen junior radiology residents were randomized into two groups. Both groups first completed a baseline lung nodule detection test allowing a free search strategy, followed by a test after scanning instruction and drilling instruction or vice versa. True and false positive scores and scroll behavior were registered. A mixed design ANOVA was applied to compare the three search conditions.

Chapter 6 addresses the use of a radiology test with stepwise reasoning questions to distinguish error types in the visual diagnostic process. Radiology test results of 109 medical students were evaluated and error types, latent errors or partial knowledge were identified based on the stepwise reasoning approach.

Chapter 7 describes the implementation and development process of a digital simulation-based test method to increase the authenticity of image interpretation skill assessment. A digital application, allowing volumetric image viewing and manipulation, was used for three test administrations of the national Dutch Radiology Progress Test (DRPT) for residents. We longitudinally describe the implementation and development process of the three digital tests, focusing on methods, challenges and lessons learned. To assess the authenticity of the digital tests, perceived image quality and correspondence to clinical practice was evaluated and compared with previous paper-based tests.

REFERENCES

1. Zwaan L, de Bruijne M, Wagner C, Thijs A, Smits M, van der Wal G, and Timmermans DR: Patient Record Review of the Incidence, Consequences, and Causes of Diagnostic Adverse Events. Arch Intern Med 2010; 170: 1015-21.
2. Balogh EP, Miller BT, and Ball JR: Improving Diagnosis in Medicine. Washington, National Academy of Sciences, 2015.
3. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, and Studdert DM: Missed and Delayed Diagnoses in the Ambulatory Setting: A Study of Closed Malpractice Claims. Ann.Intern.Med. 2006; 145: 488-96.
4. Berlin L: Radiologic Errors, Past Present and Future. Diagnosis 2014; 1: 79-84.
5. Garland L: On the Scientific Evaluation of Diagnostic Procedures. Radiology 1949; 52: 309-28.
6. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. Journal of Medical Imaging and Radiation Oncology 2012; 56: 173-78.
7. Bhargavan M, and Sunshine JH: Workload of Radiologists in the United States in 2002-2003 and Trends since 1991-1992. Radiology 2005; 236: 920-31.
8. Bhargavan M, Kaye AH, Forman HP, and Sunshine JH: Workload of Radiologists in United States in 2006-2007 and Trends since 1991-1992. Radiology 2009; 252: 458-67.
9. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, and Seltzer SE: Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section Ct Scan Can Ruin Your Whole Day. Radiology 2011; 259: 346-62.
10. Norman GR, Rosenthal D, Brooks LR, Allen SW, and Muzzin LJ: The Development of Expertise in Dermatology. Arch Dermatol 1989; 125: 1063-8.
11. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, and Boshuizen HP: Expertise in Clinical Pathology: Combining the Visual and Cognitive Perspective. Adv Health Sci Educ Theory Pract 2015; 20: 1089-106.
12. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, and Boshuizen HP: Expertise under the Microscope: Processing Histopathological Slides. Med Educ 2014; 48: 292-300.
13. Balslev T, Jarodzka H, Holmqvist K, de Grave W, Muijtjens AM, Eika B, van Merrienboer J, and Scherpbier AJ: Visual Expertise in Paediatric Neurology. Eur J Paediatr Neurol 2012; 16: 161-6.
14. Kundel HL, and Nodine CF: Interpreting Chest Radiographs without Visual Search. Radiology 1975; 116: 527-32.
15. Drew T, Evans K, Võ MLH, Jacobson FL, and Wolfe JM: Informatics in Radiology: What Can You See in a Single Glance and How Might This Guide Visual Search in Medical Images? Radiographics 2013; 33: 263-74.
16. Alzubaidi M, Black Jr JA, Patel A, and Panchanathan S: Conscious Vs. Subconscious Perception, as a Function of Radiological Expertise. Albuquerque, NM, 2009.
17. Cooper L, Gale A, Saada J, Gedela S, Scott H, and Toms A: The Assessment of Stroke Multidimensional Ct and Mr Imaging Using Eye Movement Analysis: Does Modality Preference Enhance Observer Performance? Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment 2010; 7627: 76270B.
18. Kok EM, de Bruin ABH, Robben SGF, and van Merriënboer JJG: Looking in the Same Manner but Seeing It Differently: Bottom-up and Expertise Effects in Radiology. Applied Cognitive Psychology 2012; 26: 854-62.
19. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Academic radiology 1996; 3: 137-44.

20. Manning D, Ethell S, Donovan T, and Crawford T: How Do Radiologists Do It? The Influence of Experience and Training on Searching for Chest Nodules. Radiography 2006; 12: 134-42.
21. Giovinco NA, Sutton SM, Miller JD, Rankin TM, Gonzalez GW, Najafi B, and Armstrong DG: A Passing Glance? Differences in Eye Tracking and Gaze Patterns between Trainees and Experts Reading Plain Film Bunion Radiographs. J Foot Ankle Surg 2014; 54: 382-91.
22. Wood G, Knapp KM, Rock B, Cousens C, Roobottom C, and Wilson MR: Visual Expertise in Detecting and Diagnosing Skeletal Fractures. Skeletal Radiology 2013; 42: 165-72.
23. Kok EM, Jarodzka H, de Bruin AB, BinAmir HA, Robben SG, and van Merrienboer JJ: Systematic Viewing in Radiology: Seeing More, Missing Less? Adv Health Sci Educ Theory Pract 2016; 21: 189-205.
24. Voisin S, Pinto F, Xu S, Morin-Ducote G, Hudson K, and Tourassi GD: Investigating the Association of Eye Gaze Pattern and Diagnostic Error in Mammography. In Proceedings of SPIE - The International Society for Optical Engineering, 2013.
25. Rubin GD, Roos JE, Tall M, Harrawood B, Bag S, Ly DL, Seaman DM, Hurwitz LM, Napel S, and Roy Choudhury K: Characterizing Search, Recognition, and Decision in the Detection of Lung Nodules on Ct Scans: Elucidation with Eye Tracking. Radiology 2015; 274: 276-86.
26. Leong JJH, Nicolaou M, Emery RJ, Darzi AW, and Yang GZ: Visual Search Behaviour in Skeletal Radiographs: A Cross-Speciality Study. Clinical Radiology 2007; 62: 1069-77.
27. Kundel HL, Nodine CF, Conant EF, and Weinstein SP: Holistic Component of Image Perception in Mammogram Interpretation: Gaze-Tracking Study. Radiology 2007; 242: 396-402.
28. Nodine CF, and Kundel HL: Using Eye Movements to Study Visual Search and to Improve Tumor Detection. Radiographics : a review publication of the Radiological Society of North America, Inc 1987; 7: 1241-50.
29. Swensson RG: A Two-Stage Detection Model Applied to Skilled Visual Search by Radiologists. Perception and Psychophysics 1980; 27: 11-16.
30. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13.
31. Venjakob A, Marnitz T, Mahler J, Sechelmann S, and Roetting M: Radiologists' Eye Gaze When Reading Cranial Ct Images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment. Edited by Abbey CK and MelloThoms CR. 2012.
32. Zwaan L, Thijs A, Wagner C, van der Wal G, and Timmermans DR: Relating Faults in Diagnostic Reasoning with Diagnostic Errors and Patient Harm. Acad Med 2012; 87: 149-56.
33. Pinto A, Acampora C, Pinto F, Kourdioukova E, Romano L, and Verstraete K: Learning from Diagnostic Errors: A Good Way to Improve Education in Radiology. Eur J Radiol 2011; 78: 372-6.
34. Tversky A, and Kahneman D: Judgment under Uncertainty: Heuristics and Biases. Science 1974; 185: 1124-31.
35. Jager G, Futterer JJ, and Rutten M: Cognitive Errors in Radiology: "Thinking Fast and Slow". In European Congress of Radiology, Vienna, Austria, 2014.
36. Berbaum KS, Schartz KM, Caldwell RT, Madsen MT, Thompson BH, Mullan BF, Ellingson AN, and Franken EA: Satisfaction of Search from Detection of Pulmonary Nodules in Computed Tomography of the Chest. Academic Radiology 2013; 20: 194-201.
37. Berbaum KS, Franken EA, Jr., Dorfman DD, Miller EM, Krupinski EA, Kreinbring K, Caldwell RT, and Lu CH: Cause of Satisfaction of Search Effects in Contrast Studies of the Abdomen. Acad Radiol 1996; 3: 815-26.
38. Nodine CF, Krupinski EA, Kundel HL, Toto L, and Herman GT: Satisfaction of Search (Sos). Invest Radiol 1992; 27: 571-3.

39. Lee CS, Nagy PG, Weaver SJ, and Newman-Toker DE: Cognitive and System Factors Contributing to Diagnostic Errors in Radiology. AJR Am J Roentgenol 2013; 201: 611-7.
40. Blumenthal-Barby JS, and Krieger H: Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy. Med Decis Making 2015; 35: 539-57.
41. Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, and Norman G: Is Bias in the Eye of the Beholder? A Vignette Study to Assess Recognition of Cognitive Biases in Clinical Case Workups. BMJ Qual Saf 2016.
42. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Investigative Radiology 1978; 13: 175-81.
43. Elstein AS, Shulman LS, and Sprafka SA: Medical Problem Solving: An Analysis of Clinical Reasoning. . Cambridge, MA, Harvard University Press, 1978.
44. Barrows HS, Norman GR, Neufeld VR, and Feightner JW: The Clinical Reasoning of Randomly Selected Physicians in General Medical Practice. Clin Invest Med 1982; 5: 49-55.
45. Gale J: Some Cognitive Components of the Diagnostic Thinking Process. Br J Educ Psychol 1982; 52: 64-76.
46. Neufeld VR, Norman GR, Feightner JW, and Barrows HS: Clinical Problem-Solving by Medical Students: A Cross-Sectional and Longitudinal Analysis. Med Educ 1981; 15: 315-22.
47. Norman GR, Tugwell P, Feightner JW, Muzzin LJ, and Jacoby LL: Knowledge and Clinical Problem-Solving. Med Educ 1985; 19: 344-56.
48. Evans DA, Patel VL, and Groen GJ: Biomedical Knowledge and Clinical Reasoning. Cambridge, MA, MIT Press, 1989.
49. Boshuizen HPA, and Schmidt H: On the Role of Biomedical Knowledge in Clinical Reasoning by Experts, Intermediates and Novices Cognitive Science 1992; 153-84
50. Patel VL, and Groen GJ: Knowledge Base Solution Strategies in Medical Reasoning. Cognitive Science 1986; 10: 91-116.
51. Norman GR, Trott AD, Brooks LR, and Smith EKM: Cognitive Differences in Clinical Reasoning Related to Postgraduate Training. Teaching and Learning in Medicine 1994; 6: 114-20.
52. Barrows HS, and Feltovich PJ: The Clinical Reasoning Process. Med Educ 1987; 21: 86-91.
53. Custers EJ: Thirty Years of Illness Scripts: Theoretical Origins and Practical Applications. Med Teach 2015; 37: 457-62.
54. Feltovich PJ, and Barrows HS: Issues of Generality in Medical Problem Solving. Assen / Maastricht, Van Gorcum 1984.
55. Coderre S, Mandin H, Harasym PH, and Fick GH: Diagnostic Reasoning Strategies and Diagnostic Success. Med Educ 2003; 37: 695-703.
56. Groves M, O'Rourke P, and Alexander H: The Clinical Reasoning Characteristics of Diagnostic Experts. Med Teach 2003; 25: 308-13.
57. Brooks LR, Norman GR, and Allen SW: Role of Specific Similarity in a Medical Diagnostic Task. J Exp Psychol Gen 1991; 120: 278-87.
58. Hatala R, Norman G, and Brooks LR: Influence of a Single Example on Subsequent Electrocardiogram Interpretation Teaching and Learning in Medicine 1999; 11: 110-17.
59. McElree B, Dolan PO, and Jacoby LL: Isolating the Contributions of Familiarity and Source Information to Item Recognition: A Time Course Analysis. Journal of Experimental Psychology. Learning, Memory, and Cognition 1999; 25: 563-82.
60. Murphy GL, and Ross BH: Predictions from Uncertain Categorizations. Cognitive psychology 1994; 148-93.
61. Norman G, Young M, and Brooks L: Non-Analytical Models of Clinical Reasoning: The Role of Experience. Med Educ 2007; 41: 1140-5.

62. Nodine CF, Mello-Thoms C, Kundel HL, and Weinstein SP: Time Course of Perception and Decision Making During Mammographic Interpretation. AJR Am J Roentgenol 2002; 179: 917-23.
63. Kahneman D: Thinking, Fast and Slow. New York, Farra, Straus and Giroux, 2011.
64. Kahneman D: A Perspective on Judgment and Choice: Mapping Bounded Rationality. American Psychologist 2003; 58: 697-720.
65. Croskerry P: A Universal Model of Diagnostic Reasoning. Acad Med 2009; 84: 1022-8.
66. Croskerry P, Petrie DA, Reilly JB, and Tait G: Deciding About Fast and Slow Decisions. Acad Med 2014; 89: 197-200.
67. Norman G, Sherbino J, Dore K, Wood T, Young M, Gaissmaier W, Kreuger S, and Monteiro S: The Etiology of Diagnostic Errors: A Controlled Trial of System 1 Versus System 2 Reasoning. Acad Med 2014; 89: 277-84.
68. Sherbino J, Dore KL, Wood TJ, Young ME, Gaissmaier W, Kreuger S, and Norman GR: The Relationship between Response Time and Diagnostic Accuracy. Acad Med 2012; 87: 785-91.
69. van Merrienboer JJ: Teaching Based on Thinking Fast and Slow. Acad Med 2014; 89: 8.
70. Custers EJ: Medical Education and Cognitive Continuum Theory: An Alternative Perspective on Medical Problem Solving and Clinical Reasoning. Acad Med 2013; 88: 1074-80.
71. Ericsson KA, and Lehmann AC: Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints Annual Review of Psychology 1996; 273-305.
72. Ericsson KA: Why Expert Performance Is Special and Cannot Be Extrapolated from Studies of Performance in the General Population: A Response to Criticisms. Intelligence 2014; 45: 81-103.
73. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, and Conant EF: How Experience and Training Influence Mammography Expertise. Acad Radiol 1999; 6: 575-85.
74. Nodine CF, Kundel HLMD, Lauver SC, and Toto LC: Nature of Expertise in Searching Mammograms for Breast Masses. In Proceedings of SPIE - The International Society for Optical Engineering, 1996, 89-94.
75. Berlin L: Radiologic Errors, Past, Present and Future. Diagnosis 2014; 1: 79–84.
76. Krupinski EA: The Role of Perception in Imaging: Past and Future. Seminars in Nuclear Medicine 2011; 41: 392-400.
77. Robinson PJ: Radiology's Achilles' Heel: Error and Variation in the Interpretation of the Rontgen Image. Br J Radiol 1997; 70: 1085-98.
78. Kok EM, de Bruin AB, Leppink J, van Merrienboer JJ, and Robben SG: Case Comparisons: An Efficient Way of Learning Radiology. Acad Radiol 2015.
79. Litchfield D, Ball LJ, Donovan T, Manning DJ, and Crawford T: Viewing Another Person's Eye Movements Improves Identification of Pulmonary Nodules in Chest X-Ray Inspection. J Exp Psychol Appl 2010; 16: 251-62.
80. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, and van de Weijer J: Eye Tracking: A Comprehensive Guide to Methods and Measures. New York, Oxford University Press, 2011.
81. Venjakob A, and Mello-Thoms C: Review of Prospects and Challenges of Eye Tracking in Volumetric Imaging. Journal of Medical Imaging 2015; 3.
82. Ericsson KA, and Simon HA: Protocol Analysis: Verbal Reports as Data. London, Massachusetts Institute of Technology, 1993.
83. van Gog T, Paas F, van Merriënboer JJG, and Witte P: Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. Journal of Experimental Psychology: Applied 2005; 11: 237-44.
84. Pecaric M, Boutis K, Beckstead J, and Pusic M: A Big Data and Learning Analytics Approach to Process-Level Feedback in Cognitive Simulations. Acad Med 2017; 92: 175-84.

Chapter 2

Interpretation of Radiological Images: towards a Framework of Knowledge and Skills

This chapter is based on:
Van der Gijp, A., Van der Schaaf, M.F., Van der Schaaf, I.C., Huige, J.C.B.M., Ravesloot, C.J., Van Schaik, J.P.J., Ten Cate, Th. J. Interpretation of radiological images: towards a framework of knowledge and skills. Advances in Health Sciences Education 2014; 19(4): 565-80.

ABSTRACT

Objective The knowledge and skills that are required for radiological image interpretation are not well documented, even though medical imaging is gaining importance. This study aims to develop a comprehensive framework of knowledge and skills, required for two-dimensional and volumetric image interpretation in radiology.

Methods A mixed-method study approach was applied. First, a literature search was performed to identify knowledge and skills that are important for image interpretation. Three databases, PubMed, PsycINFO and Embase, were searched for studies using synonyms of *image interpretation skills* or *visual expertise* combined with synonyms of *radiology.* Empirical or review studies concerning knowledge and skills for medical image interpretation were included and relevant knowledge and skill items were extracted. Second, a preliminary framework was built and discussed with nine selective experts in individual semi-structured interviews. The expert team consisted of four radiologists, one radiology resident, two education scientists, one cognitive psychologist and one neuropsychologist. The framework was optimized based on the expert's comments. Finally, the framework was applied to empirical data, derived from verbal protocols of ten observers interpreting two-dimensional and volumetric radiological images. In consensus meetings adjustments were made to resolve discrepancies of the framework with the verbal protocol data.

Results We designed a framework with three main components of image interpretation: perception, analysis and synthesis. The literature study provided four knowledge and twelve skill items. As a result of the expert interviews, one skill item was added and formulations of existing items were adjusted. The think-aloud experiment showed that all knowledge items and three of the skill items were applied within all three main components of the image interpretation process. The remaining framework items were apparent only within one of the main components. After combining two knowledge items, we finally identified three knowledge items and thirteen skills, essential for image interpretation by trainees.

Conclusion The framework can serve as a guideline for education and assessment of two- and three-dimensional image interpretation. Further validation of the framework in larger study groups with different levels of expertise is needed.

INTRODUCTION

Clinical reasoning with the use of images differs from clinical reasoning without[1]. In addition to the cognitive component of the diagnostic reasoning process, there is a substantial perceptual component[1-3]. The knowledge and skills required for medical image interpretation are not well documented, even though medical imaging is gaining importance in many medical domains.

Expert image interpretation is, in contrast to interpretation by novices, characterized by a holistic approach and efficient search strategies. Experienced observers fixate and recognize abnormalities faster[4, 5] and terminate their search earlier[6]. Their fast holistic approach results in higher accuracy than the relatively slow search-to-find approach of less proficient observers.

Most studies in image interpretation focus on expert search strategies. However, research in clinical reasoning suggests that the teaching of a strategy without focusing on the related content knowledge base, is unlikely to be effective[7]. To improve education and assessment in image interpretation, knowledge and skills needed to interpret images must be identified.

The literature on radiological expertise builds upon two dominant investigational approaches: a perceptual study approach, using eye tracking methods[2, 4, 8-12] and a cognitive approach[13], usually employing a verbal protocol technique[14-18]. Eye tracking studies demonstrate that differences in image interpretation expertise rely on perceptual skills, such as visual search patterns. Think-aloud studies report cognitive skills in image interpretation, such as characterizing radiological findings and using anatomical knowledge. Since perceptual and cognitive skills contribute to image interpretation[18], both deserve attention in radiology education. The two broad notions include several more subtle sub-skills that have not been systematically described. This study aims to create a framework that specifies the knowledge and skills that are required for an adequate understanding and interpretation of radiological images.

Radiological image interpretation is a complex process. Many authors identify perceptual and cognitive components in radiological image interpretation[1, 18], but the terminologies to classify these skills vary considerably[1]. Morita et al. distinguish perceptual and conceptual processing, based on Neisser's perceptual cycle theory (1976) and identify perceptual processing, resembling extracting features from environments with visual search and object-recognition, and conceptual processing referring to decision-making and hypothesis-testing based on extracted features[18]. Kundel and Nodine however classify these processes based on error types: scanning or sampling error, pattern recognition error, and decision making error[2, 19] as do Rogers et al., distinguishing detection error, labelling error and integration error[17].

For educational purposes we have chosen to emphasize learning and skills, instead of possible error types. We distinguish three components of image interpretation: 'perception', 'analysis' and 'synthesis'. *Perception* is defined as the 'identification of radiological findings', which aligns with the definition of the Oxford dictionary: 'the awareness of something through the senses'*; analysis* is defined as 'the examination of the features of radiological findings' and *synthesis* comprises 'the combination of radiological and clinical findings into a conclusion about the differential diagnosis and patient management'. These three components integrate cognitive and perceptual processes. The three components are not necessarily successive, as they may alternate during the image interpretation process[18].

In the past decades radiological image interpretation has changed substantially[20, 21]. Cross-sectional imaging techniques such as multidetector computed tomography (MDCT) and magnetic resonance (MR) imaging currently require radiologists to interpret large sets of consecutive images. While traditionally they had to interpret two-dimensional CT and MR cross-sections of the body, they now must navigate through three-dimensional datasets in various section planes. These three-dimensional datasets are also called volumetric images. Until now, research in the acquisition of radiologic expertise has focused on conventional techniques, particularly chest X-rays or mammography. Despite this tremendous development in radiology imaging techniques and image reading possibilities, no studies have appeared on the required skills for full navigation and interpretation of volumetric images. Among the studies investigating cross-sectional image interpretation strategies prior to 2013[12, 18, 22-25], in three studies participants were allowed to freely navigate through the images[18, 22, 25] but not in multiple planes. Such studies do not reveal knowledge and skills needed to apply these strategies. Both two-dimensional and volumetric image interpretation should be reflected in the training and assessment of radiology trainees, because it is expected that volumetric image interpretation requires additional skills, compared to 2D image interpretation[26].

The aim of our study is to develop a comprehensive framework of skills and knowledge, required for two-dimensional and volumetric image interpretation ability in radiology. Translated to a research question: Which knowledge and skills are required for radiological image interpretation of trainees? Such a framework should be conducive to build focused skills training and to research enhancing our understanding of the development of radiology skills.

Study Design

The creation of a framework of knowledge and skills in image interpretation was an iterative process, based on a mixed-method study approach. First, a literature survey on knowledge and skills for image interpretation was conducted. Second, a preliminary framework was built and discussed with selective experts in semi-structured interviews to optimize the framework. Finally, the framework was applied to empirical data, derived from verbal protocols of observers interpreting radiological 2D and volumetric images. Figure 1 shows a graphical presentation of the study design.

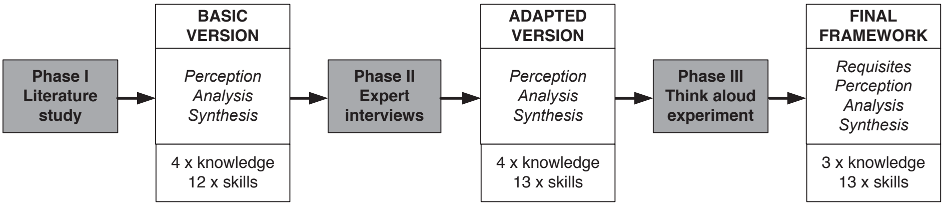

Figure 1. Flow chart of the study design: the framework was constructed in three phases.

Ethical approval

This study was approved by the Ethical Review Board of the Netherlands Association for Medical Education. All participants gave written informed consent to participate in the study.

Phase I. Literature Study

METHODS

Literature Search

A systematic literature search was conducted using PubMed, PsycINFO and Embase. Search terms included synonyms of *image interpretation skills* or *visual expertise* combined with synonyms of *radiology*. The search syntax is presented in Appendix 1. The references of relevant articles from the search were checked. References of relevant chapters in books on visual expertise and medical image interpretation were searched for additional sources. Studies published between 1969 and January 2012 were screened. Inclusion criteria were: (a) empirical or review studies, (b) English language articles and (c) studies concerning knowledge and skills for medical image interpretation by radiologists, residents, medical students or non-radiology physicians. Exclusion criteria were (a) studies concerning qualities not considered knowledge or skills and (b) studies concerning non-medical domains.

Procedure

The first author screened title and abstract of the resulting articles for relevance, based on in- and exclusion criteria. Relevant articles were included and read by the first author. Reported knowledge and skills were included when suggestive of importance to medical image interpretation, and excluded when either they were not considered helpful for it or when there were conflicting opinions. Decisions about extraction of the items were discussed with at least three other members of the research team until consensus was reached. The resulting knowledge and skill items were used to construct a first, basic version of the framework. As terminology varies widely, the framework terminology was determined in a consensus meeting of the research team. In a second consensus meeting, items were categorized into the three predetermined components of image interpretation, i.e. perception, analysis or synthesis.

RESULTS

After applying the in- and exclusion criteria, the literature search yielded eighteen studies with evidence for certain knowledge or skills in medical image interpretation.

Within the framework of the three predetermined components (perception, analysis and synthesis) the studies yielded sixteen literature-based subcomponents (twelve skill items and four knowledge items). Perceptual subcomponents found were *detecting radiological findings*[2, 12], *knowledge of radiological anatomy*[15], *knowledge of radiological imaging techniques*[15], *spatial abilities*[27, 28], *image manipulation*[28], *using search strategies*[9, 29, 30], and *pattern recognition*[31]. Analysis subcomponents were *knowledge of radiologic pathology*[32, 33], *using clinical information*[15, 34], *comparing with previous images*[35-37], *evaluating radiological findings*[17] and *distinguish relevant from irrelevant findings*[16]. The subcomponents of synthesis were *consider differential diagnoses*[15], *decision making*[2], *epidemiological knowledge*[38] and *clinical-radiological reasoning*[15, 16].

The first, basic version of the framework consisted of the abovementioned categorized knowledge and skill items and served as input for the expert interviews.

Phase II. Expert Interviews

METHODS

Participants

We invited nine experts with established radiological, educational, psychological or combined expertise. The expert group consisted of four radiologists, one radiology resident, two education scientists, one cognitive psychologist and one neuropsychologist. The experts were chosen based on their experience and expertise in their domain and their affection with radiology education. All radiologists had extensive experience with educating image interpretation to residents and medical students. One of the radiologists was also a program director. The non-radiologist experts were involved in research in radiology education. These selection criteria were chosen to enable discussion of the image interpretation task from several important viewpoints.

Data Collection

The nine experts were interviewed individually by two members of the research team in semi-structured interviews of 60-90 minutes. All interviews were audio recorded. The experts were asked to comment on: (i) the knowledge and skill items included in the first version of the framework, i.e., on the formulation of the items and definitions, (ii) the relevance of the items, (iii) the location of the items in the framework, (iv) overlap between items and (v) coverage of the framework.

Data Analysis

Interviews were independently categorized by the interviewers. Four categories were distinguished on item level: 'irrelevant item', 'incorrect formulation', 'incorrect location' and remaining comments. On framework level, the categories 'overlapping items', 'missing items' and 'remaining comments' were scored. It was predetermined that two (or more) similar comments of at least one radiologist and one non-radiologist should lead to an adjustment in the framework. If two or more similar comments were made by just radiologists or just non-radiologists, only small adjustments were made in the formulation of an item, while preserving the content and meaning of the item. Final decisions about the adjustments of the framework were made in consensus meetings of the research team. To check if the interviews were interpreted correctly, the experts were asked to review and comment on the adjusted framework (member check).

RESULTS

The interviewers identified and categorised 146 expert comments, 87 comments at item level and 59 at framework level, with a high inter-observer reliability (Cohen's κ =.83). Radiologists made fewer comments than other experts (66 versus 80). The categories 'incorrect formulation' and 'remaining comments' were scored most.

All experts recognized the three components of image interpretation of the framework and found the model to be important, representative and useful for education in medical image interpretation. The expert comments led to nine adjustments in the basic framework. The formulation of eight items was modified. The item 'information retrieval' was added, because two experts introduced it as a crucial part of image interpretation. It refers to the ability to find relevant information to help interpret the image correctly. Relevant information could be background information about a disease, visual information[39], evidence-based information or additional clinical information. Major modifications are illustrated in Table 1, using examples of the experts' comments. None of the experts reported any misinterpretations when asked during the member check.

Table 1. Examples of expert comments, leading to major framework adjustments

Framework component	Item basic version	Examples of experts' comments	Item adapted version
Perception	Detecting radiological findings	"To have datasets of normality in your mind is a prerequisite for recognizing abnormalities" (cognitive psychologist) "Detection of radiological findings is actually the end goal of perception instead of a perceptual component" (neuropsychologist)	Discriminating normal from abnormal findings
Analysis	Evaluate findings	"This formulation is unclear and to broad. I prefer the term characterizing findings" (radiologist)	Characterizing findings
Synthesis	-	"Radiologists need to know where to find relevant and accurate information to guide the image interpretation process" (education scientist)	Information retrieval
	Clinical-radiological reasoning	"…as a matter of fact, the presented process *is* clinical radiological reasoning" (cognitive psychologist)	Integrating radiological findings
	Consider differential diagnoses	"decision making refers to generating a differential diagnosis" (radiologist)	Generating a differential diagnosis
	Decision making	"this term is very broad. It should specify what kind of decision making it refers to" (education scientist) "decision making refers to generating a differential diagnosis, but it is also involved in giving advice for follow up examinations or in deciding for action, for example giving a call to the requesting physician immediately" (radiologist)	Deciding about advice or action

Adjustments led to a second version of the framework, with four knowledge items and thirteen skill items. This adapted framework served as a coding scheme to analyze think-aloud protocols in the think-aloud study.

Phase III. Think-Aloud Study

METHODS

Participants

We investigated image interpretation in advanced medical students, because radiologists were expected to verbalize less knowledge and skills they use, as experts tend to have difficulty in verbalizing how they solve problems in contrast to intermediates[40, 41]. We consecutively invited 23 medical students rotating at the radiology department at the University Medical Center Utrecht (UMCU). They were all fourth, fifth or sixth-year medical students. We call those advanced medical students 'clerks'. At the UMCU, radiology education is intensively integrated into medical school. Medical students start with volumetric CT image interpretation in the first year. In the first two years they study CT anatomy in volumetric images of the brain, chest and abdomen and basic pathologic 2D images. This is examined at the end of the second year. In the third and fourth year, as a preparation for each clerkship, radiologic image interpretation of pathology related to the upcoming clerkship is being taught. In addition to the regular radiology education in medical school, the invited clerks successfully completed an intensive six or twelve-week clinical radiology clerkship less than one week preceding the think-aloud session. They can be considered “intermediates” in image interpretation. The learning objectives of the clerkship included acute or sub acute pathology in neuroradiology, musculoskeletal radiology, abdominal radiology and chest radiology. Ten clerks (response rate 43%) participated in the study. All participants gave written informed consent.

Procedure

All participants (n=10) were asked to think aloud while reading digital radiologic cases. The first five participants read 23 cases (nine volumetric CT images, nine 2D CT images and five X-ray studies). This turned out to be quite intensive for the participants. As saturation within-subject was repeatedly reached during the first half of the protocol analysis, we decided to reduce the reading set to 11 cases (four volumetric CT images, four 2D CT images and three X-ray studies). Both case sets represented the four clerkship domains, equally distributed to the image modalities (volumetric CT, 2D CT and X-ray). The cases were in alignment with the learning objectives of the clerkship and consisted of common diseases in neuroradiology (i.e. cerebral bleeding), musculoskeletal radiology (i.e. wrist fracture), abdominal radiology (i.e. ruptured aortic aneurysm) and chest radiology (i.e. pulmonary embolism). In the case of volumetric CT images, participants could navigate through the images and change imaging plane or image contrast. Prior to the session the participants received a standardized instruction and a short training in thinking out loud[42]. Verbalisations were recorded and the computer display was videotaped simultaneously.

Instrumentation

Protocol analysis was performed[43], using the framework as a coding scheme. This means that elements of the framework were identified in the vocabulary of the verbal protocols.

Data Analysis

Two investigators independently scored the verbal protocols of the first five participants using the coding scheme. Because additional information as books and the internet were not available during the experiment, the item ‘information retrieval skills’ was scored only if participants verbalized the intention to search for additional information. After scoring verbalisations of each participant, the inter-observer reliability was calculated. Discrepancies were discussed and resolved. The remaining five participants were scored by one investigator. The videotape was used to clarify the verbal protocols. In case an item of the coding scheme did not apply to the verbal data, adjustments were made during consensus meetings of the research team. Remaining utterances were collected and examined to find out if additional knowledge and skills had to be included in the coding scheme.

RESULTS

The coding scheme covered 86% of the verbal data. The remaining utterances mainly consisted of reading aloud, of metacognitive statements or of remarks on the functionality of the imaging software and did not contain any adjacent knowledge or skills related to image interpretation. All knowledge and skills in the framework could be identified in the verbal protocols, except for spatial ability. Mean inter-observer reliability was calculated in half of the participants (Cohen's κ = .76). Examples of each item are summarized in Table 2.

Adjustments mostly concerned the position of knowledge and skills in the framework. Seven items seemed to be present in more than one component of the image interpretation process and were detached from the three components of the framework. For example, clinical information use was identified as a subcomponent of perception, analysis and synthesis.

> "Let's see. *It was a left sided hemi paresis*, so I expect something on the right." (perceptual, participant 2, cerebral infarction)

> "I think this might be a subdural bleeding, because of the sickle-shaped lesion, hyperdense. *Together with the clinical information this means it is a subdural bleeding*" (synthesis, participant 1, subdural hematoma)

Because 'knowledge of pathology' and 'epidemiological knowledge' were not attached to a component of the image interpretation process in the framework anymore, it was decided to combine 'knowledge of epidemiology' and 'knowledge of pathology' into 'knowledge of pathology/epidemiology'.

Besides these major adjustments in the framework, there were multiple minor adjustments in the definitions of the framework items. For instance, the definition of 'generate a differential diagnosis' was broadened, in order to include the classification of the diagnosis.

> "How do you call it, a dissection I believe. And where does it start, before or after the vessels. No, after. So, it is a type B. Yes, type B." (participant 4, aortic dissection)

Table 2. All resultant knowledge and skill items of this study, with examples of verbalisations by clerks during interpretation of radiological images.

Framework component	Item	Example
Requisites*	Knowledge of anatomy	"I see hypodense structures, *they are localised in the liver*" (analysis, participant 2, liver laceration)
	Knowledge of pathology / epidemiology	"What is this area here? (points to lung lesion, which is a lung infarction)… Let`s see he has prostate carcinoma. *It is possible, because of a carcinoma, to get pulmonary embolism as a result of clotting problems.* But he could also have a lung tumour. I think it is a lung tumour." (synthesis, participant 2, pulmonary embolism with lung infarction"
	Knowledge of radiological imaging techniques	"This white thing here, is really attached to the falx, how was it again, is blood white? *Only when a contrast agent is used, I believe.* I don`t think contrast has been used. Then, blood shouldn`t be white. (analysis, participant 4, cerebral infarction)
	Spatial abilities**	-
	Image manipulation skills (navigating through, changing views or contrast)	"Here I see a somewhat hypodense area, is that in the liver or beneath the liver? *Let`s look in another view.* (changes to coronal view). Or is it just the gallbladder? No, it is something. It looks like it is situated in the liver" (analysis, participant 2, liver laceration)
	Acquaintance of clinical information and context	"Hmm let`s see. *It was a left sided hemi paresis*, so I expect something on the right." (perception, participant 2, cerebral infarction)
Perception	Using efficient search strategies	"Oh the spleen does not look normal as well as the liver. *Well, systematic assessment*. First the liver is not homogeneous"(perception, participant 4, spleen laceration)
	Discriminating normal from abnormal findings	"Compression of the main bronchus and *the lungs do not look normal."*(perception, participant 3, pulmonary embolism with lung infarction)
	Pattern recognition	"Where I directly notice a huge spleen laceration grade 4" (perception, participant 1, spleen laceration)
A**nalysis**	Comparing with previous images	"Hmm I find the pulmonary markings to be very pronounced. *It is too bad that I don't have an earlier one to compare with*." (analysis, participant 4, malposition gastric feeding tube)
	Characterizing findings	"I find this difficult. Let's see, I see a *well circumscribed*…no at least it does not leak" (analysis, participant 2, aneurysm of the abdominal aorta)
	Discriminating relevant from irrelevant findings	"There is a kidney cyst but *this is not a potentially life-threatening disease*." (analysis, participant 4, aneurysm of the abdominal aorta)
Synthesis	Information retrieval	"When was the ileocecal resection? Was it recent or not?" (synthesis, participant 2, abdominal abscess)
	Integrating radiological findings	"I see a…what was it called again that shape not sickle-shaped but a lenticular white abnormality right. This fits well with an epidural bleeding. *I see soft tissue swelling right that also all fits in*." (synthesis, participant 4, epidural hematoma)
	Generating a (differential) diagnosis	"Is that a *hemothorax* or is it just a *pneumothorax with serous fluid*?...no *pleural fluid*. It could also be a little bit of *blood*. No it is a *pneumothorax*….air is present." (synthesis, participant 2, pneumothorax)
	Deciding about advice or action	"Free air in the abdomen, with wide…which depicts an ileus. *I would like a CT*."(synthesis, participant 1, free air in the abdomen) "I will call it a spleen rupture, so that is urgent. *Call the trauma doctor*."(synthesis, participant 5, spleen laceration)

*The requisite category contains knowledge and skill items that could be found during perception, analysis and synthesis. Given examples originate from one of the three components.

**No example of spatial abilities could be given, as it could not be objectively identified in the verbalisations of the participants.

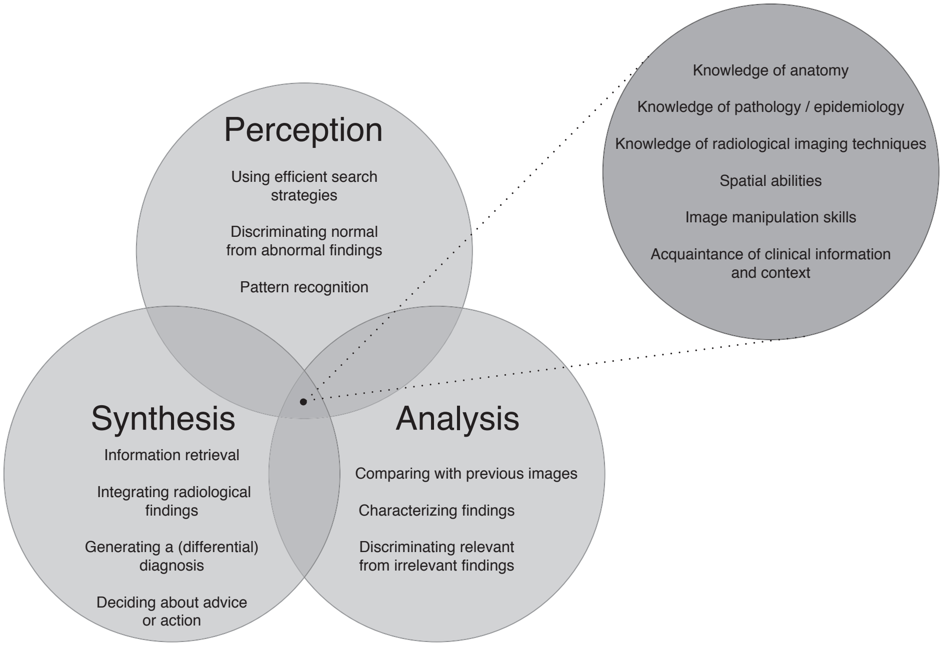

Figure 2. The final framework representing important knowledge and skills for radiological image interpretation.

The modifications of the coding scheme were applied to the framework. The final resulting framework with knowledge and skills contains three knowledge items and thirteen skill items and is presented in Figure 2. Final definitions are listed in Appendix 2.

DISCUSSION

In this study, we constructed a conceptual framework of knowledge and skills required for 2D and volumetric image interpretation in radiology. We identified three knowledge elements and thirteen skills essential to image interpretation of trainees, based on a mixed-method approach with a literature survey, expert interviews and verbal protocols. The think-aloud experiment showed that all three knowledge items and three of the skill items were applied to all three components of the image interpretation process. The remaining framework items were apparent only within one of the components.

Most published studies concerning visual expertise in radiology describe the process or strategy of image interpretation[2, 17, 19, 44], but the knowledge and skills required to apply these strategies are generally unclear. We included the interpretation of volumetric imaging in addition to the conventional imaging techniques, as volumetric imaging now accounts for a major part of radiology practice, but is not included in previous studies[12, 18, 22-25]. We found image manipulation to be an important skill in volumetric image interpretation.

In this study and the resulting framework we did not incorporate interpretation of ultrasound and angiography, while this is an important skill for radiologists as well. These imaging techniques and the concurrent interpretation require specific motor skills and eye-hand coordination, which are crucial elements of the task. Studying the knowledge and skills needed to interpret these modalities should also include the knowledge and skills needed to carry out the techniques. This would require a different study design with for example video recording of (simulated) ultrasound and angiography investigations.

This study has some limitations. The expert panel was selected based on years of experience in radiology, experience in radiological education and general educational expertise and experience. The literature does not provide guidelines for the desired characteristics of a visual perception expert. A different panel could have yielded other items. Further, expert radiologists could be unconscious of the knowledge they apply during image interpretation[45], because of their fast, holistic approach[4]. We anticipated the presence of such tacit, encapsulated knowledge in experienced radiologists[40, 41] and deliberately added advanced medical students to include more fundamental self-observations provided by their verbal protocols. Knowledge and skills verbalized by intermediates could reveal more items, as they appear to verbalize more than both novices and experts do. This is known as 'the intermediate effect'[40]. As it is known that verbalisations during image interpretation differ between expert levels[15, 17, 18], we cannot generalize these results to other expert levels.

We further acknowledge the dynamic status of the framework that we constructed. The research group has elaborated many hours on the construction of it, attempting to capture observed actions into words, suitable for both theory building and application among radiologists with their own content jargon. One aspect particularly has occupied the research group. That is the visual representation of the required knowledge and skills within the radiological image interpretation process. Radiological image interpretation is a complex cognitive process and many knowledge and skill items are partly integrated. We tried to capture this in a graphic presentation of the process in figure 2, but we do realize that it is a simplification of the actual image interpretation process. Given what we know from the literature, from the consultation of a panel of experts and from observing intermediate trainees of radiology at work, we believe the current framework is the best representation of what happens during the complex mental processing of radiology images with current techniques.

The framework can serve as a guideline for radiology education and assessment. In particular, it could be used as a framework for learning objectives or test blueprints. As the study has been focussed on radiologists, we do not know if the resulting knowledge and skills are equally important for other clinicians who interpret radiological images. Nonetheless we think the knowledge and skills might be used to educate other physicians to teach them to interpret radiological images in the way a radiologist would do it. The framework might even serve as a model for image interpretation in other medical domains that draw on visual diagnosis such as pathology and dermatology.

To further validate the framework, image interpretation processes should be researched in larger study groups with different levels of expertise. Eye tracking studies may add insight in the perceptual component of image interpretation, especially in volumetric imaging, which is not largely investigated yet. It would also be helpful to identify modality-specific skills, for example by comparing 2D with volumetric image interpretation. If the implementation of the framework in a course or curriculum will alter the learning process of students should be determined in future research.

Appendix 1. The search syntax of the literature study

Search terms entered into databases	Databases
interpretation skills OR interpretation activities OR cognitive skills OR cognitive activities OR interpretation abilities OR reading skills OR reading abilities OR perceptual activities OR perceptual abilities OR perceptual skills OR expertise AND radiology OR radiologist OR radiologic OR radiologists	Pubmed PsychINFO Embase

Appendix 2. Final definitions of the knowledge and skills in the framework

Requisite knowledge and skills	Knowledge and skills which are used in all components
Knowledge of anatomy	Knowledge of anatomy, in particular radiological anatomical structures, as visualized with radiological imaging methods.
Knowledge of pathology	Knowledge of pathology (diseases) and in particular features of pathology as visualized with radiological imaging methods, but also including knowledge of the clinical features, treatment and prevalence of the disease.
Knowledge of radiological imaging techniques	Knowledge of the acquisition of the images and the effect of radiological techniques on the image representation[15].
Spatial abilities	The ability to imagine a 2D representation as a 3D structure, to mentally rotate this structure and to analyze the relationship between different spatial representations[46].
Image manipulation skills (navigating through, changing views or contrast)	To have knowledge of the effects of image manipulation on image representation and the ability to choose the optimal representation for a task, e.g. by navigating through the image, changing contrast or changing views (meta-representational competence)[26].
Acquaintance of clinical information and context	Using clinical information (from the request form, the patient status and from patient interaction) or clinical context (the type of hospital, requesting clinician etc.) for interpretation of the radiological examination.

Perception	Identification of radiological findings
Using efficient search strategies	Apply efficient search strategies, such as global search, systematic search and hypothesis-guided search. Search strategies might be successively applied during interpretation of a radiological image[47].
Discriminating normal from abnormal findings[48]	Discriminate normal (and normal variance) from abnormal findings[48].
Pattern recognition	Recognizing the diagnosis directly and unconsciously, based on similar patterns in visual memory.

Analysis	Examination of the features of radiological findings
Comparing with previous images	Comparing radiological findings with findings on previous examinations of the patient.
Characterizing findings	Evaluating features (e.g. density, shape, contour) of the findings, if necessary by using post processing.
Discriminating relevant from irrelevant findings	Discriminate clinically relevant findings from findings which are not clinically relevant.

Synthesis	**The synthesis of radiological and clinical findings into a conclusion about the differential diagnosis and patient management.**
Information retrieval	Skills for a purposeful search for information in a system, in which information is stored and represented (e.g. information in books, the internet). For the purpose of the verbal protocols: to realize that more information (clinical information, information out of books) is necessary to interpret the image properly
Integrating radiological findings	Connecting radiological findings with each other (e.g. a lung nodule and mediastinal lymph nodes)
Generating a (differential) diagnosis	Generating a list of possible diagnoses in order of probability. This incorporates generating as well as rejecting diagnoses and the classification of diseases, such as type or stage.
Deciding about advice or action	Decide if it is necessary to give an advice (e.g. further examinations, follow-up exams) or take action (e.g. calling the requesting clinician immediately).

REFERENCES

1. Norman GR, Coblentz CL, Brooks LR, and Babcook CJ: Expertise in Visual Diagnosis: A Review of the Literature. Acad Med 1992; 67: S78-83.
2. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Invest Radiol 1978; 13: 175-81.
3. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. J Med Imaging Radiat Oncol 2012; 56: 173-8.
4. Kundel HL, Nodine CF, Conant EF, and Weinstein SP: Holistic Component of Image Perception in Mammogram Interpretation: Gaze-Tracking Study. Radiology 2007; 242: 396-402.
5. Nodine CF, Kundel HL, Lauver SC, and Toto LC: Nature of Expertise in Searching Mammograms for Breast Masses. Acad Radiol 1996; 3: 1000-6.
6. Christensen EE, Murry RC, Holland K, Reynolds J, Landay MJ, and Moore JG: The Effect of Search Time on Perception. Radiology 1981; 138: 361-5.
7. Norman G: Research in Clinical Reasoning: Past History and Current Trends. Med Educ 2005; 39: 418-27.
8. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Acad Radiol 1996; 3: 137-44.
9. Kundel HL, and La Follette PS, Jr.: Visual Search Patterns and Experience with Radiological Images. Radiology 1972; 103: 523-8.
10. Nodine CF, and Kundel HL: The Cognitive Side of Visual Search. North-Holland, Elsevier Science Publishers B.V., 1987.
11. Nodine CF, Mello-Thoms C, Kundel HL, and Weinstein SP: Time Course of Perception and Decision Making During Mammographic Interpretation. AJR Am J Roentgenol 2002; 179: 917-23.
12. Cooper L, Gale A, Darker I, Toms A, and Saada J: Radiology Image Perception and Observer Performance: How Does Expertise and Clinical Information Alter Interpretation? Stroke Detection Explored through Eye-Tracking. Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment 2009; 7263: 72630K1-12.
13. Beutel J: Handbook of Medical Imaging: Physics and Psychophysics. Washington, SPIE, 2000.
14. Azevedo R, Lajoie S, Desaulniers M, Fleiszer D, and Bret P: Radtutor: The Theoretical and Empirical Basis for the Design of a Mammography Interpretation Tutor. Artificial Intelligence in Education. Edited by du Boulay B and Mizoguchi R. Amsterdam, IOS Press, 1997.
15. Lesgold AM, Rubinson H, Feltovich P, Glaser R, Klopfer D, and Wang Y: Expertise in a Complex Skill: Diagnosing X-Ray Pictures. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1988.
16. Raufaste E, Eyrolle H, and Mariné C: Pertinence Generation in Radiological Diagnosis: Spreading Activation and the Nature of Expertise. Cognitive Science 1998; 22: 517-46.
17. Rogers E: A Study of Visual Reasoning in Medical Diagnosis. Mahwah, N.J., Lawrence Erlbaum, 1996.
18. Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interactions of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
19. Nodine CF, and Kundel HL: Using Eye Movements to Study Visual Search and to Improve Tumor Detection. Radiographics 1987; 7: 1241-50.
20. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, and Seltzer SE: Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section Ct Scan Can Ruin Your Whole Day. Radiology 2011; 259: 346-62.

21. Reiner BI, Siegel EL, and Siddiqui K: Evolution of the Digital Revolution: A Radiologist Perspective. J Digit Imaging 2003; 16: 324-30.
22. Wang XH, Durick JE, Lu A, Herbert DL, Golla SK, Foley K, Piracha CS, Shinde DD, Shindel BE, Fuhrman CR, Britton CA, Strollo DC, Shang SS, Lacomis JM, and Good WF: Characterization of Radiologists' Search Strategies for Lung Nodule Detection: Slice-Based Versus Volumetric Displays. Journal of Digital Imaging 2008; 21: S39-S49.
23. Cooper L, Gale A, Saada J, Gedela S, Scott H, and Toms A: The Assessment of Stroke Multidimensional Ct and Mr Imaging Using Eye Movement Analysis: Does Modality Preference Enhance Observer Performance? Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment 2010; 7627: 76270B.
24. Matsumoto H, Terao Y, Yugeta A, Fukuda H, Emoto M, Furubayashi T, Okano T, Hanajima R, and Ugawa Y: Where Do Neurologists Look When Viewing Brain Ct Images? An Eye-Tracking Study Involving Stroke Cases. PLoS One 2011; 6: e28929.
25. Venjakob A, Marnitz T, Mahler J, Sechelmann S, and Roetting M: Radiologists' Eye Gaze When Reading Cranial Ct Images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment 2012; 8318.
26. Hegarty M: Components of Spatial Intelligence, 2010.
27. Berbaum KS, Smoker WR, and Smith WL: Measurement and Prediction of Diagnostic Performance During Radiology Training. AJR Am J Roentgenol 1985; 145: 1305-11.
28. Cohen CA, and Hegarty M: Individual Differences in Use of External Visualisations to Perform an Internal Visualisation Task. Applied Cognitive Psychology 2007; 21: 701-11.
29. Kundel HL, and Nodine CF: Interpreting Chest Radiographs without Visual Search. Radiology 1975; 116: 527-32.
30. Kundel HL, and Wright DJ: The Influence of Prior Knowledge on Visual Search Strategies During the Viewing of Chest Radiographs. Radiology 1969; 93: 315-20.
31. Norman G, Young M, and Brooks L: Non-Analytical Models of Clinical Reasoning: The Role of Experience. Med Educ 2007; 41: 1140-5.
32. Marchiori DM, Adams TL, and Henderson CN: Developing a Clinical Competency Examination in Radiology: Part I--Test Structure. J Manipulative Physiol Ther 1999; 22: 57-62.
33. Marchiori DM, Henderson CN, and Adams TL: Developing a Clinical Competency Examination in Radiology: Part II--Test Results. J Manipulative Physiol Ther 1999; 22: 63-74.
34. Loy CT, and Irwig L: Accuracy of Diagnostic Tests Read with and without Clinical Information: A Systematic Review. JAMA 2004; 292: 1602-9.
35. Callaway MP, Boggis CR, Astley SA, and Hutt I: The Influence of Previous Films on Screening Mammographic Interpretation and Detection of Breast Carcinoma. Clin Radiol 1997; 52: 527-9.
36. Quekel LG, Goei R, Kessels AG, and van Engelshoven JM: Detection of Lung Cancer on the Chest Radiograph: Impact of Previous Films, Clinical Information, Double Reading, and Dual Reading. J Clin Epidemiol 2001; 54: 1146-50.
37. Aideyan UO, Berbaum K, and Smith WL: Influence of Prior Radiologic Information on the Interpretation of Radiographic Examinations. Acad Radiol 1995; 2: 205-8.
38. Weber EU, Böckenholt U, Hilton DJ, and Wallace B: Determinants of Diagnostic Hypothesis Generation: Effects of Information, Base Rates and Experience. Journal of Experimental Psychology: Learning, Memory, and Cognition 1993; 19: 1151-64.
39. Markonis D, Holzer M, Dungs S, Vargas A, Langs G, Kriewel S, and Müller H: A Survey on Visual Information Search Behavior and Requirements of Radiologists. Methods of Information in Medicine 2012; 51: 539-48.

40. Rikers RM, Schmidt HG, and Boshuizen HP: Knowledge Encapsulation and the Intermediate Effect. Contemporary Educational Psychology 2000; 25: 150-66.
41. Schmidt HG, and Boshuizen HPA: On the Origin of Intermediate Effects in Clinical Case Recall. Memory & Cognition 1993; 21: 338-51.
42. Ericsson KA, and Simon HA: Protocol Analysis: Verbal Reports as Data. London, Massachusetts Institute of Technology, 1993.
43. Chi MT: Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. Journal of the Learning Sciences 1997; 6: 271-315.
44. Morita W, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interaction of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
45. Heiberg Engel PJ: Tacit Knowledge and Visual Expertise in Medical Diagnostic Reasoning: Implications for Medical Education. Med Teach 2008; 30: e184-8.
46. Linn MC, and Petersen AC: Emergence and Characterisation of Sex Differences in Spatial Ability: A Meta-Analysis. Child Development 1985; 56: 1479-98.
47. Wolfe JM, Võ MLH, Evans KK, and Greene MR: Visual Search in Scenes Involves Selective and Nonselective Pathways. Trends Cogn Sci 2010; 15: 77-84.
48. Boutis K, Pecaric M, Seeto B, and Pusic M: Using Signal Detection Theory to Model Changes in Serial Learning of Radiological Image Interpretation. Adv Health Sci Educ Theory Pract 2010; 15: 647-58.

Chapter 3

Volumetric and Two-Dimensional Image Interpretation Show Different Cognitive Processes in Learners

This chapter is based on:
Van der Gijp, A., Ravesloot, C.J., Van der Schaaf, M. F., Van der Schaaf, I.C., Huige, J.C.B.M., Vincken, K.L., Ten Cate, Th. J., Van Schaik, J.P.J. Volumetric and two-dimensional image interpretation show different cognitive processes in learners. Academic Radiology 2015; 22(5): 632-9.

ABSTRACT

Objective In current practice radiologists interpret digital images, including a substantial amount of volumetric images. We hypothesized that interpretation of a stack of a volumetric dataset demands different skills than interpretation of two-dimensional (2D) cross-sectional images. This study aimed to investigate and compare knowledge and skills used for the interpretation of volumetric versus 2D images.

Methods Twenty radiology clerks were asked to think out loud while reading four or five volumetric CT images in stack mode and four or five 2D CT images. Cases were presented in a digital testing program allowing stack viewing of volumetric datasets and changing views and window settings. Thoughts verbalized by the participants were registered and coded by a framework of knowledge and skills concerning three components: perception, analysis and synthesis. The components were subdivided into sixteen discrete knowledge and skill elements. A within-subject analysis was performed to compare cognitive processes during volumetric image readings versus 2D cross-sectional image readings.

Results Most utterances contained knowledge and skills concerning *perception* (46%). A smaller part involved *synthesis* (31%) and *analysis* (23%). More utterances regarded *perception* in volumetric image interpretation than in 2D image interpretation (Mdn 48% versus 35%, $z = -3.9$, $p<.001$). *Synthesis* was less prominent in volumetric than in 2D image interpretation (Mdn 28% versus 42%, $z = -3.9$, $p<.001$). No differences were found in *analysis* utterances.

Conclusion Cognitive processes in volumetric and 2D cross-sectional image interpretation differ substantially. Volumetric image interpretation draws predominantly on perceptual processes while 2D image interpretation is mainly characterized by synthesis. The results encourage the use of volumetric images for teaching and testing perceptual skills.

INTRODUCTION

The daily practice of radiologists has changed since the introduction of cross-sectional imaging techniques (e.g., CT and MRI) and digital viewing systems[1, 2]. Digital volumetric datasets have been introduced, which can be scrolled through in different planes and window settings. Volumetric image sets are increasingly used, because this is advantageous for identification and analysis of radiologic abnormalities[3]. We expect that the interpretation of stacks of volumetric datasets demands different skills than interpretation of two-dimensional (2D) images[1, 4]. For example, visual search patterns in stack mode viewing of CT images differ from tiled mode viewing[5]. Drew et al. found that the pattern of errors made in volumetric CT image interpretation differs from error patterns in the interpretation of 2D images, which were chest x-rays in this case, as decision errors are less common in CT image interpretation[6, 7]. In volumetric image interpretation, radiologists need to navigate through and manipulate images in order to identify and analyze lesions. Although the multidimensional information enables a radiologist to observe the image features in detail, this requires the processing of much more information which could make the radiologists' search more complex and time-consuming[1].

As radiology practice has changed, and cognitive processes in image interpretation may have consequently altered, traditional 2D teaching methods may not align well with the knowledge and skills required for current practice[8]. To gain insight in image interpretation skills for educational purposes, it is useful to explore which cognitive processes occur in volumetric image interpretation and how these differ from 2D image interpretation.

Cognition is a generic term for processes that involve for example perceiving, recognizing, problem-solving, judging, reasoning and decision making[9]. Cognitive processes in radiological image interpretation encompass extracting image information and combining this with information acquired from patient history and external sources to understand and make inferences about the meaning of the image. The cognitive processes of interest in this paper are the use of knowledge and skills for image interpretation.

So far research has mainly focused on cognitive processes in the interpretation of 2D images such as chest X-rays[10, 11]. Differences in cognitive processes of radiologists and radiology trainees in volumetric image interpretation are only recently researched. Morita et al. focused on the interaction between perceptual and conceptual processes[12]. Perceptual processing pertains to retrieving visual information from a CT image, e.g. density or shape of an abnormality. Conceptual processing refers to relating perceived image features to existing knowledge in the observer's memory, for instance knowledge about radiological appearances of diseases or normal anatomy. This study showed that the interaction between the two processes was more prominent and occurred at an earlier stage among radiologists than among radiology trainees[12].

In a previous study, we developed a framework representing knowledge and skills required for radiological image interpretation based on an interview and survey study among experts[13]. The framework has three main components: perception, analysis and synthesis, and sixteen subcomponents. Six requisite knowledge and skill items are related to more than one main component and are placed separately. The framework is presented in Figure 1. The framework proved to be convenient for coding cognitive processes during image interpretation, with a high interrater reliability[13]. In the present study, the framework was used to characterize and compare cognitive processes used in volumetric and 2D image interpretation.

As current radiology practice largely involves volumetric image interpretation, the cognitive processes underlying volumetric image interpretation should be further explored to improve education in modern image interpretation. The aim of this study is to reveal cognitive processes during volumetric cross-sectional image interpretation and to compare these to cognitive processes used during 2D cross-sectional image interpretation. The research questions are: 1) Which cognitive processes occur during volumetric image interpretation? 2) Which cognitive processes during volumetric image interpretation differ from those in 2D image interpretation in radiology education? We hypothesized that perceptual processes are more important in volumetric image interpretation, because searching for abnormalities could be more complex and time-consuming than in 2D image interpretation.

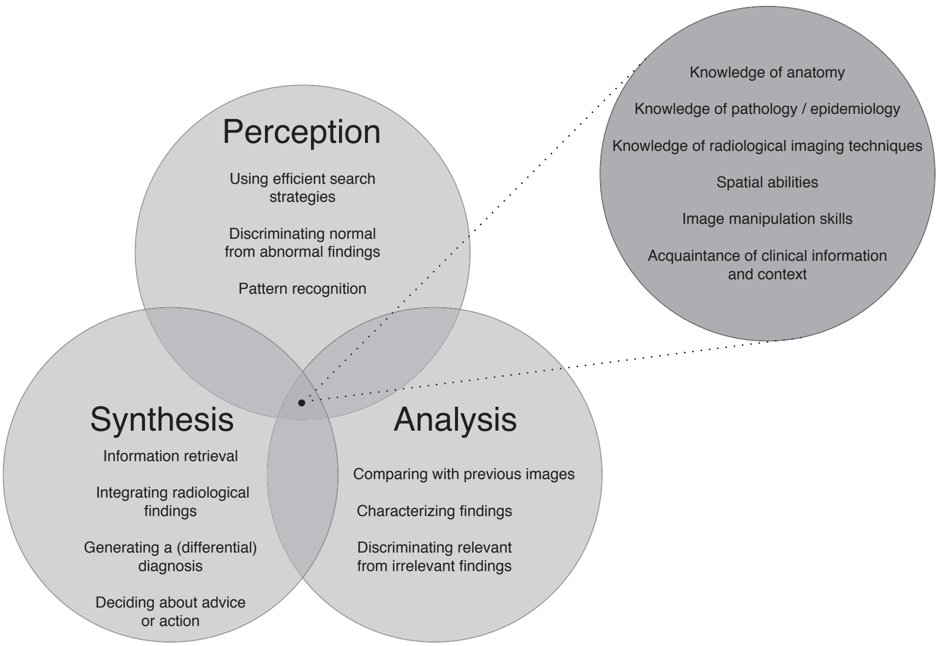

Figure 1. Framework representing important knowledge and skills for radiological image interpretation. Reprinted from Van der Gijp et al.[13]. With kind permission from Springer Science and Business Media.

METHODS

Study Design

A within-subjects design was used. Concurrent verbal protocols were used as a proxy of cognitive processes[14]. Verbalizations of participants during volumetric and 2D radiologic image interpretation were compared. All participants gave written informed consent for the study.

Participants And Setting

The study focused on an intermediate level study population, since intermediates are likely to verbalize more than both novices and experts do, known as the 'intermediate effect'[15]. Twenty clerks of the radiology department of University Medical Center Utrecht agreed to volunteer in the study. Participants were fourth to sixth-year medical students with 75% women. In the Netherlands, medical school takes six years and includes clerkships. Their mean age was 24.7 years. All participants had completed an elective radiology clerkship (six to twelve weeks) less than two weeks prior to the study. The objectives of the clerkship focused on acute or sub-acute pathology in four radiology subareas: neuroradiology, musculoskeletal radiology, abdominal radiology and chest radiology.

Instrumentation

Image Cases

Participants were asked to solve an even number of volumetric CT image and 2D CT image cases; fifteen participants received four of each and five participants received five of each. The volumetric CT images were read in stack mode. The 2D CT images consisted of one to four cross-sectional slices, selected from a volumetric dataset. All cases involved prevalent acute or sub-acute diseases in four radiology subareas: neuroradiology (e.g. brain infarction), musculoskeletal radiology (e.g. vertebral fracture), abdominal radiology (e.g. diverticulitis) and chest radiology (e.g. pneumonia). The cases were introduced with clinical background information, and corresponded with the teaching objectives of the clerkship. Two sets of volumetric and 2D images were used. All volumetric image cases in set A were presented as 2D image cases in set B and vice versa. Participants read eight or ten cases out of set A or B, half of the cases being volumetric and half 2D. Each case was read only once by each participant. Cases were presented in random order.

Digital Assessment Environment

Cases were shown in VQuest[16], a digital assessment environment which allows to scroll through a stack of images and change view or window setting. VQuest was found to be user-friendly and image display was found to be a fair representation of clinical practice, especially when images were displayed in stack mode[8, 17]. In this study setting, it was not possible to retrieve images from the patient archive and therefore images could not be compared with previous patient studies.

Image Display
Both volumetric and 2D CT images were presented in greyscale. Volumetric images were displayed in stack mode. Participants had the opportunity to scroll back and forth through the stack of images at desired speed in axial, sagittal and coronal view. Window setting could be adjusted in soft tissue, bone, lung and brain setting. 2D image cases contained a selection of one to four slices of a volumetric dataset, displayed in tiled mode. The selection of images showed the abnormalities needed to interpret the case. The window setting of each slice was fixed on the optimal setting for visualization of the abnormalities and could not be adjusted.

Coding Scheme
The coding scheme was based on the framework as shown in Figure 1, containing knowledge and skill elements related to image interpretation[13]. All subcomponents of perception, analysis and synthesis were labeled with a unique code. The framework also included six requisite knowledge and skill items, which were not component-specific. Requisite knowledge and skill items were all tagged with a code for perception, analysis and synthesis.

Procedure

Image Reading
Prior to the think aloud investigation, participants received a standard instruction and did an exercise in thinking out loud[14]. During the study, all participants were asked to think out loud while reading eight or ten digital radiologic CT cases. They were asked to report a diagnosis or differential diagnosis and to give an advice (e.g. for follow-up imaging) if necessary. They were also asked to indicate if they felt the need for additional information to solve the case. However, additional information sources were not available in the room. A supervisor stayed in the room during the entire think aloud investigation, to encourage participants to keep thinking aloud. The supervisor was the same person during the entire data collection.

Coding Process
Two independent raters coded the verbalizations of 25% of all participants. Because inter-rater reliability was found to be satisfactory (mean Cohen's $\kappa = .76$), one rater continued to code the remaining verbalizations. Utterances related to requisite knowledge and skills were all assigned to one of the three main components perception, analysis or synthesis, based on the content and context of the utterance. Although image manipulation was not allowed in 2D images, 'image manipulation skills' could be coded during 2D image interpretation in case knowledge about image manipulation was expressed, for example by stating how the image could be manipulated to support the interpretation process. Because participants

could not retrieve images from the patient archive, the variable 'comparing with previous images' could only be coded in case participants verbalized they would have liked to compare with previous patient images. As additional information sources, for example books or the internet were not available during case reading, utterances related to the intention to search for additional information were coded as 'information retrieval'.

Data Analysis

Utterances that could not be related to image interpretation knowledge or skills were excluded from the analysis. The total number of utterances for each code was registered per participant per case. For each participant, the number and percentage of utterances per code and per main component were calculated for volumetric and 2D image cases. Wilcoxon signed-rank tests were conducted to compare the codes of the main components in volumetric readings and 2D readings. If main components were significantly different, codes of subcomponents in volumetric and 2D readings were compared using Wilcoxon signed-rank tests with a Bonferroni correction.

Institutional Review Board Approval

The study was approved by the Ethical Review Board of the Netherlands Association for Medical Education.

RESULTS

Except for spatial abilities, all knowledge and skill items were found to be verbalized by participants. Examples of utterances of each item are presented in Table 1.

Fourteen percent of all utterances could not be related to image interpretation knowledge and skills and were excluded from the analysis. The total number of the remaining coded utterances was slightly more than twice as high in volumetric image cases compared to 2D image cases. This ratio was comparable in all four radiology subareas. Completion time was also more than twice as high in volumetric image cases. A wide range in the number of utterances was found among cases and among participants. Distribution of the number of utterances among volumetric and 2D image cases is given in Table 2.

Most utterances were categorized as perception (46%), followed by synthesis (31%) and analysis (23%). In volumetric image interpretation most utterances represented perception (Mdn 47.9%), in contrast to 2D image interpretation (Mdn 34.6%). The percentage of perceptual utterances was significantly higher in volumetric images than in 2D images ($z=-3.9$, $p<.001$, $r=.61$). The percentage of utterances related to synthesis was significantly larger in 2D images than in volumetric images ($z=-3.9$, $p<.001$, $r=.61$). The percentage of analysis was almost equal in both groups. Results of Wilcoxon Signed Rank tests are presented in Table 3.

Table 1. Examples of verbalizations of knowledge and skill items

Framework components	Code items	Examples
Requisites*	Knowledge of anatomy	"*This is the posterior edge of the maxillary sinus*." (perception, subdural hematoma and maxillary sinus fracture)
	Knowledge of pathology/ epidemiology	"The filling defect seems to be situated in the middle of the vessel. *Of course, this is a sign of* an acute pulmonary embolism." (synthesis, pulmonary embolism with lung infarction)
	Knowledge of radiological imaging techniques	"Let's see if there is contrast leakage present. *I am not sure if a contrast agent is used. I think there is."* (analysis, spleen laceration)
	Spatial abilities**	-
	Image manipulation skills (navigating through, changing views or contrast)	"*I always count vertebral bodies in the sagittal view.*" (perception, Jefferson fracture)
	Acquaintance of clinical information and context	"I see a large hypodense area in the right hemisphere. *This corresponds to the hemiparesis at the left.*" (analysis, brain infarction)
Perception	Using efficient search strategies	"*I try to divide the head into three parts. First, I examine the upper part, then the middle part and finally the lower part*." (subarachnoid hemorrhage)
	Discriminating normal from abnormal findings	"*I see abnormalities* in the lung. The heart *looks normal."* (pulmonary contusion)
	Pattern recognition	"Then we *directly see what this is*: a scapular fracture." (scapular fracture)
Analysis	Comparing with previous images	"A lesion in the left adrenal gland, probably an incidentaloma, though I can't exclude malignancy. Besides, *I don`t have anything to compare with.*" (diverticulitis)
	Characterizing findings	"There is an *increased density especially at the right side.*" (subarachnoid hemorrhage)
	Discriminating relevant from irrelevant findings	"This is probably a renal cyst. *I don't think this is causing any problems.*" (aneurysm of the abdominal aorta)
Synthesis	Information retrieval	"Fracture of (…) *I would have to look up* which bone has been broken." (subdural hematoma and maxillary sinus fracture)
	Integrating radiological findings	"*I see air in the brain which is probably coming from the maxillary sinus which is fractured* (…) *This means there is a connection between the maxillary sinus and the brain which causes intracranial air.*" (subdural hematoma and maxillary sinus fracture)
	Generating a (differential) diagnosis	"I believe this is *an aortic dissection*. Where does it start, before or behind the vessels? No behind, *so it is type B*." (aortic dissection)
	Deciding about advice or action	"Cerebral herniation…if I was a radiology resident, *I would have warned them that they should pay attention to that*." (epidural hematoma)

*Knowledge and skill items of the category 'requisites' could relate to perception, analysis or synthesis.

**The variable 'spatial abilities' was not identified.

Table 2. Distribution of utterances among volumetric and 2D image interpretation

	Volumetric	2D
N participants	20	20
Total number of utterances	5429	2563
Range among cases	11 - 253	7 - 60
Range among participants (average per case)	19 - 168	14 - 54
Mean completion time per case (s)	390	181

Table 3. Differences in image interpretation components in 2D and volumetric image interpretation

	Volumetric		2D			
Component	**Total**	**Median %**	**Total**	**Median %**	**p-value***	**Effect size r**
Perception	2751	47.9	945	34.6	<.001	.61
Analysis	1245	23.4	555	21.7	.31	.16
Synthesis	1433	27.8	1063	41.5	<.001	.61

* Wilcoxon Signed Rank Test

In volumetric image interpretation, 'knowledge of anatomy' was verbalized most (29.0% of the utterances), followed by 'discriminating normal from abnormal findings' (19.8%). 'Comparing with previous images' and 'discriminating relevant from irrelevant findings' were coded the least, with one and five coded utterances respectively. 'Characterizing findings' was most verbalized within analysis and 'generating a differential diagnosis' was most verbalized within synthesis. Figure 2 and 3 show the distributions of knowledge and skills in volumetric image interpretation and 2D image interpretation respectively. The horizontal scales of Figure 2 and 3 are not equal.

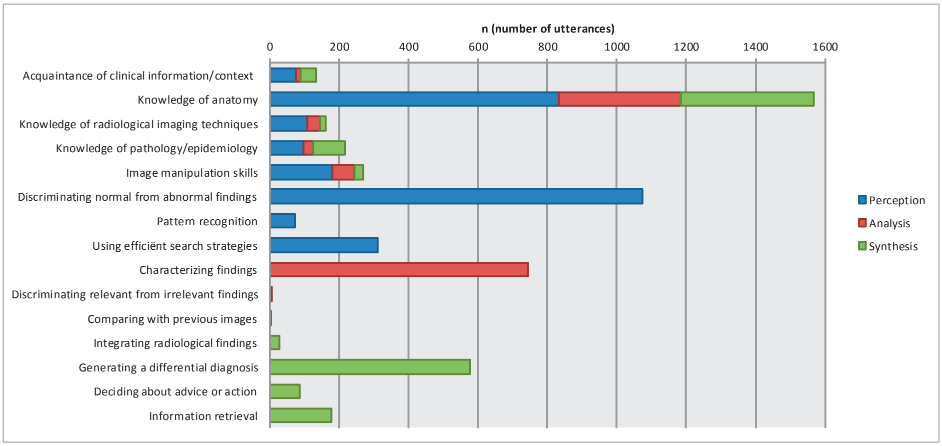

Figure 2. Distribution of knowledge and skill items used in volumetric image interpretation.

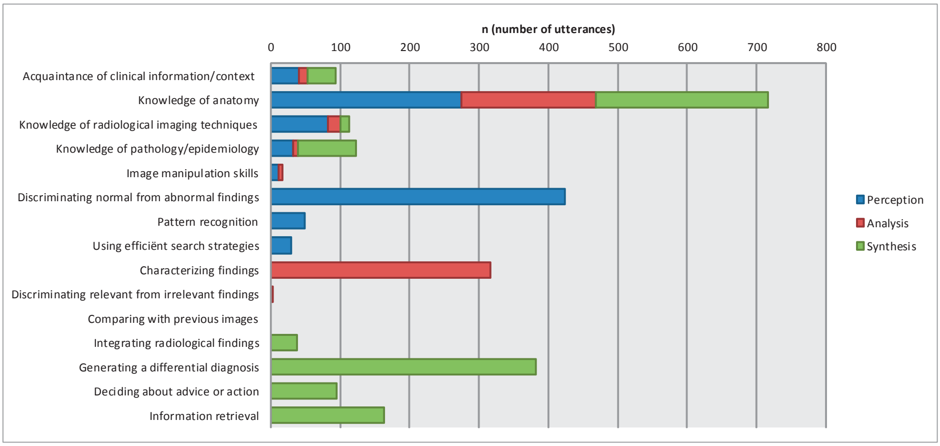

Figure 3. Distribution of knowledge and skill items used in 2D image interpretation.

The perception subcomponents 'knowledge of anatomy', 'using efficient search strategies' and 'image manipulation skills' represented a larger part of utterances in volumetric image interpretation than in 2D image interpretation, z=-3.1, p<.005, r=.49; z=-3.6, p<.001, r=.57 and z=-3.9, p<.001, r=.62 respectively. The synthesis component 'generating a differential diagnosis' represented a larger part of 2D image interpretation than volumetric image interpretation, z=-3.2, p<.001, r=.51.

DISCUSSION

Cognitive processes in volumetric image interpretation were found to differ from 2D cross-sectional image interpretation. Volumetric image interpretation involves more cognitive processes than 2D image interpretation, but completion time differed with the same factor. Perception is more prominent in volumetric image interpretation than in 2D image interpretation. Cognitive processes related to synthesis have a bigger share in 2D image interpretation (though in absolute numbers, synthesis is also more prevalent in volumetric image interpretation). A framework of knowledge and skills in radiological image interpretation was used to subcategorize these cognitive processes on a knowledge and skill level [13]. In the category perception, 'knowledge of anatomy', 'the use of efficient search strategies' and 'image manipulation skills' were more prominent in volumetric than in 2D image interpretation. In the category synthesis, 'generating a differential diagnosis' was less prominent in volumetric than in 2D image interpretation.

The prominence of perception in volumetric image interpretation could be due to a more complex and time-consuming search for abnormalities in volumetric image interpretation, which is supported by our results showing more verbalizations of using efficient search strategies and image manipulation skills in volumetric image interpretation. Knowledge of anatomy was also more frequently used in volumetric image interpretation, which is probably due to a larger amount of anatomical information that has to be processed. In absolute numbers, synthesis is found more in volumetric image interpretation, but less than expected based on completion time differences. This could be caused by a relatively less complex and time-consuming search, which leaves time and mental effort for a more thorough synthesis process.

Based on the number of verbalized cognitive processes and the completion time, the mental effort seems higher in volumetric image interpretation than in 2D image interpretation. This aligns with previous research showing a positive relation between 'multiplanar image information' and self-reported mental effort of medical students performing a radiological image interpretation task[18].

We did not find other studies reporting differences in cognitive processes in volumetric versus 2D image interpretation in the context of radiology education. Previously reported differences in visual search patterns and error patterns[5-7] only suggest a difference in cognitive processes, though the actual cognitive processes during image interpretation, in terms of the thinking process of the image interpreter, was not investigated. Most verbal protocol studies investigating cognitive processes in radiological image interpretation research are related to differences in expertise levels. For example, studies in X-ray image interpretation show that experts report more different and connecting findings than novices and generate longer chains of reasoning[10, 11]. Besides, experts appear to ignore irrelevant data, but take immediate account of relevant data[10, 11]. Lesgold et al. studied the cognitive processes of experts and concluded they 'see things differently', 'build mental representations of patient anatomy' and 'exhibit flexibility'. These studies describe important differences in verbalizations between expert levels, but do not translate these into knowledge and skills needed for the task. In our study verbal protocols during image interpretation were analyzed on a level of knowledge and skills, to make the results useful for educational purposes. One other study, conducted in CT image interpretation reported a detailed analysis of verbal protocols in numerous subcategories[19]. The authors reported the content of the information verbalized by the participants but did not translate the information into knowledge and skills. For example, it was registered how many times words like 'density', 'shape' and 'size' of the lesion was mentioned, which we translated into the skill: 'characterizing findings'.

Our results regarding examination time, being longer in volumetric image interpretation than in 2D image interpretation, are contradictory to the results of Ellis et al. [5] who found longer examination times in tiled reading versus stack reading. This difference can be explained by the difference in number of 2D images presented to participants. Tiled reading includes multiple images presented next to each other, while in our study only a few 2D CT slices were presented, as this is more common in radiology education and we tried to find differences in 2D and volumetric image interpretation relevant for education.

Our study has some limitations. Verbal protocols are used as a proxy of cognitive processes. Utterances of participants are used to capture knowledge and skills that might underlie those utterances[20]. Relying on verbal protocols could possibly underestimate the amount of cognitive processes and could blur the distribution. Cognitive processes can be unconscious and therefore might not to be verbalized. For example, 'discriminating relevant from irrelevant findings' is a process that probably frequently occurs during image interpretation as not all findings are being reported, though the decision to report a finding because it is judged relevant (or not to report a finding because it is judged irrelevant) is barely verbalized. This potential bias increases with growing automaticity of the processes and is therefore more prominent in experts[14]. To minimize discrepancies between the cognitive processes actually used during the performance and the cognitive processes reported, we collected concurrent instead of retrospective verbal reports[14]. Thinking aloud is expected to have little effect on perceptual performance in radiological image interpretation, except for prolonging the task[21]. Some cognitive processes may have been underestimated or overestimated due to the experimental conditions. Because previous patient studies were not available, the variable 'comparing with previous images' could only be coded when participants made explicit that they wanted to compare. As a consequence, 'comparing with previous images' was hardly verbalized in this study. One skill that may have been overestimated is 'information retrieval'. Because information sources were not available in the study setting, participants were asked whether they needed additional information to solve the case. This could have encouraged participants to verbalize thoughts about the intention to search for additional information, which was coded as 'information retrieval'. We do not suppose that these under- or overestimations influenced our results concerning the comparison of 2D and 3D image interpretation, because we used a within-subject design.

The study population consisted of radiology clerks, i.e. fourth to sixth-year medical students. The distribution of cognitive processes could be different in novices and experts. It is not sure if the differences in cognitive processes between volumetric and 2D image interpretation will persist in radiology experts, because experts have proven to be more proficient searchers as shown in numerous eye tracking studies[22-26] and perceptual processes might be less prominent in volumetric image interpretation. Characterizing differences in cognitive processes in volumetric and 2D image interpretation of novices and experts could be subject for further investigation. A suggestion for further research would be to validate our results in other expertise levels.

Our results encourage the use of volumetric images in radiology education, because its interpretation requires different and more extensive knowledge and skills than 2D image interpretation. Especially for teaching and testing perceptual skills, volumetric images are recommended.

REFERENCES

1. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, and Seltzer SE: Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section Ct Scan Can Ruin Your Whole Day. Radiology 2011; 259: 346-62.
2. Reiner BI, Siegel EL, and Siddiqui K: Evolution, of the Digital Revolution: A Radiologist Perspective. Journal of Digital Imaging 2003; 16: 324-30.
3. Horton KM, Johnson PT, and Fishman EK: Mdct of the Abdomen: Common Misdiagnoses at a Busy Academic Center. AJR Am J Roentgenol 2010; 194: 660-7.
4. Hegarty M: Components of Spatial Intelligence., 2010.
5. Ellis SM, Hu X, Dempere-Marco L, Yang GZ, Wells AU, and Hansell DM: Thin-Section Ct of the Lungs: Eye-Tracking Analysis of the Visual Approach to Reading Tiled and Stacked Display Formats. Eur J Radiol 2006; 59: 257-64.
6. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13: 3.
7. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Invest Radiol 1978; 13: 175-81.
8. Ravesloot CJ, Van der Schaaf MF, van Schaik JPJ, ten Cate TJ, van der Gijp A, Mol C, and Vincken KL: Volumetric Ct-Images Improve Testing of Radiological Image Interpretation Skills. Eur J Radiol 2015; 84: 856-61.
9. Snow RE, Jackson D, and Corno L: Individual Differences in Affective and Conative Functions. Handbook of Educational Psychology. Edited by Berliner DC and Calfee RC. New York, Simon & Schuster Macmillan, 1996, 243.
10. Lesgold AM, Rubinson H, Feltovich P, Glaser R, Klopfer D, and Wang Y: Expertise in a Complex Skill: Diagnosing X-Ray Pictures. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1988.
11. Raufaste E, Eyrolle H, and Mariné C: Pertinence Generation in Radiological Diagnosis: Spreading Activation and the Nature of Expertise. Cognitive Science 1998; 22: 517-46.
12. Morita W, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interaction of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
13. van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JC, Ravesloot CJ, van Schaik JP, and Ten Cate TJ: Interpretation of Radiological Images: Towards a Framework of Knowledge and Skills. Adv Health Sci Educ Theory Pract 2014.
14. Ericsson KA, and Simon HA: Protocol Analysis: Verbal Reports as Data. London, Massachusetts Institute of Technology, 1993.
15. Rikers RM, Schmidt HG, and Boshuizen HP: Knowledge Encapsulation and the Intermediate Effect. Contemporary Educational Psychology 2000; 25: 150-66.
16. Vincken KL, Ravesloot CJ, and Mol C: Vquest. <www.vquest.nl> [Accessed August 2014].
17. Ravesloot CJ, van der Gijp A, van der Schaaf MF, Huige JCBM, Vincken KL, Mol CP, Bleys RLAW, ten Cate TJ, and van Schaik JPJ: Support for External Validity of Radiological Anatomy Tests Using Volumetric Images. Acad Radiol 2015; 84: 856-61.
18. Stuijfzand BG, van der Schaaf MF, Kirschner FC, Ravesloot CJ, Van der Gijp A, and Vincken KL: Medical Students' Working Memory Load in Multiplanar Image Interpretation: Insights from Human-Computer Interactions and Eye Movements. Computers in Human Behavior 2014; 62: 394-403.

19. Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interactions of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
20. Chi MT: Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. Journal of the Learning Sciences 1997; 6: 271-315.
21. Littlefair S, Brennan P, Reed W, Williams M, and Pietrzyk MW: Does the Thinking Aloud Condition Affect the Search for Pulmonary Nodules? Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment. Edited by Abbey CK and MelloThoms CR. 2012.
22. Nodine CF, Mello-Thoms C, Kundel HL, and Weinstein SP: Time Course of Perception and Decision Making During Mammographic Interpretation. AJR Am J Roentgenol 2002; 179: 917-23.
23. Nodine CF, Kundel HL, Lauver SC, and Toto LC: Nature of Expertise in Searching Mammograms for Breast Masses. Academic radiology 1996; 3: 1000-06.
24. Christensen EE, Murry RC, Holland K, Reynolds J, Landay MJ, and Moore JG: The Effect of Search Time on Perception. Radiology 1981; 138: 361-5.
25. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Academic radiology 1996; 3: 137-44.
26. Kundel HL, Nodine CF, Conant EF, and Weinstein SP: Holistic Component of Image Perception in Mammogram Interpretation: Gaze-Tracking Study. Radiology 2007; 242: 396-402.

Chapter 4

How Visual Search Relates to Visual Diagnostic Performance: a Narrative Systematic Review of Eye Tracking Research in Radiology

This chapter is based on:
Van der Gijp, A., Ravesloot, C.J., Jarodzka, H., Van der Schaaf, M.F., Van der Schaaf, I.C., Van Schaik, J.P.J., Ten Cate, Th.J. How visual search relates to visual diagnostic performance: a narrative systematic review eye tracking research in radiology. Advances in Health Sciences Education 2017; 22(3): 765-787.

ABSTRACT

Objective Eye tracking research has been conducted for decades to gain understanding of visual diagnosis such as in radiology. For educational purposes, it is important to identify visual search patterns that are related to high perceptual performance and to identify effective teaching strategies. This review of eye tracking literature in the radiology domain aims to identify visual search patterns associated with high perceptual performance.

Methods Databases PubMed, EMBASE, ERIC, PsycINFO, Scopus and Web of Science were searched using 'visual perception' OR 'eye tracking' AND 'radiology' and synonyms. Two authors independently screened search results and included eye tracking studies concerning visual skills in radiology published between January 1, 1994 and July 31, 2015. Two authors independently assessed study quality with the Medical Education Research Study Quality Instrument (MERSQI), and extracted study data with respect to design, participant and task characteristics, and variables. A thematic analysis was conducted to extract and arrange study results, and a textual narrative synthesis was applied for data integration and interpretation.

Results The search resulted in 22 relevant full-text articles. Thematic analysis resulted in six themes that informed the relation between visual search and level of expertise: 1) time on task, 2) eye movement characteristics of experts, 3) differences in visual attention, 4) visual search patterns, 5) search patterns in cross-sectional stack imaging, and 6) teaching visual search strategies. Expert search was found to be characterized by a global-focal search pattern, which represents an initial global impression, followed by a detailed, focal search-to-find mode. Specific task-related search patterns, like drilling through CT scans and systematic search in chest X-rays, were found to be related to high expert levels. One study investigated teaching of visual search strategies, and did not find a significant effect on perceptual performance.

Conclusion Eye tracking literature in radiology indicates several search patterns are related to high levels of expertise, but teaching novices to search as an expert may not be effective. Experimental research is needed to find out which search strategies can improve image perception in learners.

INTRODUCTION

An important part of clinical reasoning is based on visual information. Physicians use visual information derived from direct observation of patients (e.g. dermatologic findings) and from other means such as electrocardiograms, histopathology and radiological images. Eye tracking can reflect physicians' attention with objective measures and provide insight into clinical reasoning processes[1]. Visual diagnosis plays a central role in diagnostic radiology and eye tracking procedures have proven to be a valuable tool for investigating the visual diagnostic process in radiology for decades[2].

Several investigators have proposed models that attempt to capture the complexity of radiologists' visual search. In 1978, Nodine and Kundel proposed a 3-phase theory of visual search and detection, distinguishing initial overall pattern recognition, focal attention for image detail and final decision making[3]. The first phase refers to the first glance at the image, in which observers compare an anatomic mind map with the perceived image and localize perturbations. Evidence of this holistic search derives from studies in which expert radiologists were found to be able to correctly identify abnormal images within approximately one fourth of a second[4-6]. In the following scanning phase, observers examine potential targets and perturbations through fixation (i.e., directing their gaze and visual attention to a specific location). In the third phase, a decision about the presence of a lesion is made, characterized by prolonged or multiple clustered fixations.

More recent research discerns only two components of visual search, embracing a similar concept of a global-focal search model: 1) a relatively fast global impression that signals possible abnormalities and 2) a slower, more detailed focal search for recognition and evaluation of abnormalities. These two components of visual search are defined either as a fast holistic and slow search-to-find mode[7]; a pre-attentive filter and cognitive evaluation stage[8]; or nonselective and selective pathways running in parallel[9]. The concept also aligns with the system 1 and system 2 thinking modes described by Kahneman[10]. The fast, holistic impression of a global search relates to system 1 thinking, which is an automatic and relatively quick thinking process, while the slower focal search mode can be associated with the more attentional and effortful mental activity of system 2 thinking.

These models were developed for static 2D images and may not apply to cross-sectional stack imaging where visual search patterns are even more complex. Cross-sectional images, such as Computed Tomography (CT) and Magnetic Resonance (MR) scans, are currently viewed in stack mode which involves scrolling through a large set of consecutive cross-sections of a body region. Searching a stack of cross-sectional images differs fundamentally from searching a single image, because the identification and interpretation of abnormalities requires scrolling through a set of images and the visual information changes continuously depending on the level of the cross-section.

Perceptual errors, i.e. errors in the detection of abnormalities, account for a major part of misdiagnoses in radiology[11], and can result from cognitive biases[12] or a faulty visual search[13]. It is important to identify visual search patterns that do or do not lead to an accurate perception of lesions on radiological images. From an educational perspective, the ultimate goal of eye tracking research in medical image perception is to improve image interpretation by avoiding errors in visual search. When we understand the perceptual process, we may identify search strategies that lead to improved performance of clinicians and integrate these in radiology training. The aim of this review is to identify visual search characteristics that may lead to higher perceptual accuracy. The central question in this review is: How do visual search characteristics relate to diagnostic performance in radiology? We explored eye tracking research in the radiology expertise literature to address this question.

To understand this research, some background information is needed about eye movement parameters and how they relate to the global-focal search models. Eye movement parameters that are frequently used in radiological image perception are: *time to first fixation of the abnormality, fixation durations on relevant or redundant areas, number of fixations on relevant or redundant areas, saccades and image coverage.* A *fixation* is a period of time wherein the eye remains still. It reflects attention to that particular area in the image and actual intake of information. *Saccades* are the rapid eye movements in between fixations to re-allocate the focus of attention from one area to another. During a saccade, no visual information can be processed[1].

According to the global-focal search models[7-9], experts exhibit an efficient search by selecting potential lesions with a global search, taking advantage of a larger functional visual field than novices in combination with a greater conceptual knowledge about where to best look for abnormalities. The global search guides the eyes to suspicious areas for a more detailed focal search. In an effective global search, perturbations are localized fast and *time to first fixation of the abnormality* is expected to be shorter in experts than in less proficient observers[14]. The efficient selection process enables experts to take in all relevant information through fewer fixations, so that *image coverage*, based on location of fixations, may be relatively less in experts' searches. Consequently, required search time also decreases with increasing expertise. *Saccades* are presumed to enlarge with experience due to an increased visual span[14, 15]. An effective global search could decrease the need to fixate intensively on a lesion and could decrease *fixation duration* and *number of fixation* parameters[14]. On the other hand, an ineffective global search can prevent an observer from recognizing the abnormality at all, also decreasing the duration and number of fixations on a lesion. These parameters are probably task-specific[14, 15], so we expect variable relationships to expert level.

In addition to the traditional eye movement parameters, various derivative parameters are increasingly being used to capture and categorize complex search patterns that result from a combination of subsequent eye movements. In this article, we will refer to these parameters as 'visual search patterns'.

METHODS

A narrative systematic review[16] was conducted. The search strategy was systematic, with a broad sensitive approach. The data of the included studies did not allow for a meta-analysis, due to large variations in methodology, variables and outcome measures. A thematic analysis approach[17] was used to extract and arrange the data. Next, results within each theme were integrated and interpreted with a textual narrative synthesis[18], visualizing the differences and similarities in results among varying study methodologies.

Literature Review Protocol

We searched the databases PubMed, EMBASE, ERIC, PsycINFO, Scopus and Web of Science on August 15, 2014. A librarian assisted with the electronic search. The search terms were synonyms of 'visual perception' and 'eye tracking', combined with the Boolean operator AND to synonyms of radiology. The search strategy consisted of keywords and controlled vocabulary (Figure 1). Because radiology imaging has tremendously changed in the last decades and only fairly recent literature can reflect current radiology practice, we limited the search to the last twenty years. The search was repeated on July 31, 2015.

Search syntax		Date of search: 15 aug 2014
Databases searched:	**Search terms entered into databases:**	**Relevant mesh terms/emtrees/subject headings used:**
Pubmed in title and abstract **Embase** in title and abstract **ERIC** in title and abstract **Psychinfo** in title and abstract **Scopus** in title, abstract and keywords **Web of science** in title or topic	visual perception OR visual search OR visual scanning OR visual tracking OR perceptual OR eye movement OR eye movements OR eye movement measurements OR ocular movement OR ocular movements OR ocular fixation OR eyetracking OR eye tracking OR eye-tracking OR eyetrack OR eyetracker OR eye-tracker OR eye tracker OR eye control OR eye controlled OR eye-controlled OR eye position OR eye-position OR eye positions OR eye-positions OR eye fixation OR eye fixations OR gaze tracking OR gaze-tracking OR gaze tracker OR gaze trackers OR gaze duration OR fixation analysis OR relevant mesh terms/emtrees/subject headings AND radiology OR radiography OR radiologic OR radiological OR radiologist OR radiologists OR radiographic OR radiograph OR radiographs OR Roentgenography OR relevant mesh terms/emtrees/subject headings	**Pubmed** *Mesh terms* Visual perception, eye movements, eye movement, measurements, radiology and radiography **Embase** *Emtree terms* pattern recognition, signal detection, gaze, eye fixation, eye movement, eye tracking, radiology, radiography and radiologist **PsycINFO** *Subject Headings (Exp)* Visual perception, eye movements, eye fixation, visual tracking, roentgenography, radiology **ERIC** *Subject Headings (Exp)* Visual perception, eye movements, radiology
Limit: Publication year 1994-2014		

Figure 1. Search syntax

Study Eligibility and Selection

We included empirical studies published between January 1, 1994 and July 13, 2015. Studies had to address at least one of the three following topics by means of eye tracking research: (1) the relationship between level of expertise and visual search characteristics, (2) the relationship between performance and visual search characteristics, or (3) the effect of teaching visual search strategies on performance. Reviews were only included for bibliography search. Inclusion and exclusion criteria are summarized in the search flowchart (Figure 2).

After removing duplicate articles, two authors (C.J.R and A.G.) independently screened titles and abstracts for relevance. Inclusion and exclusion criteria were applied to all potentially relevant full-text articles. In the case of discrepancy between the two researchers, the eligibility of the study was discussed to reach consensus. If two similar articles were published based on overlapping study data, only the one with the largest sample was included. The bibliographies of included articles were screened for additional relevant articles.

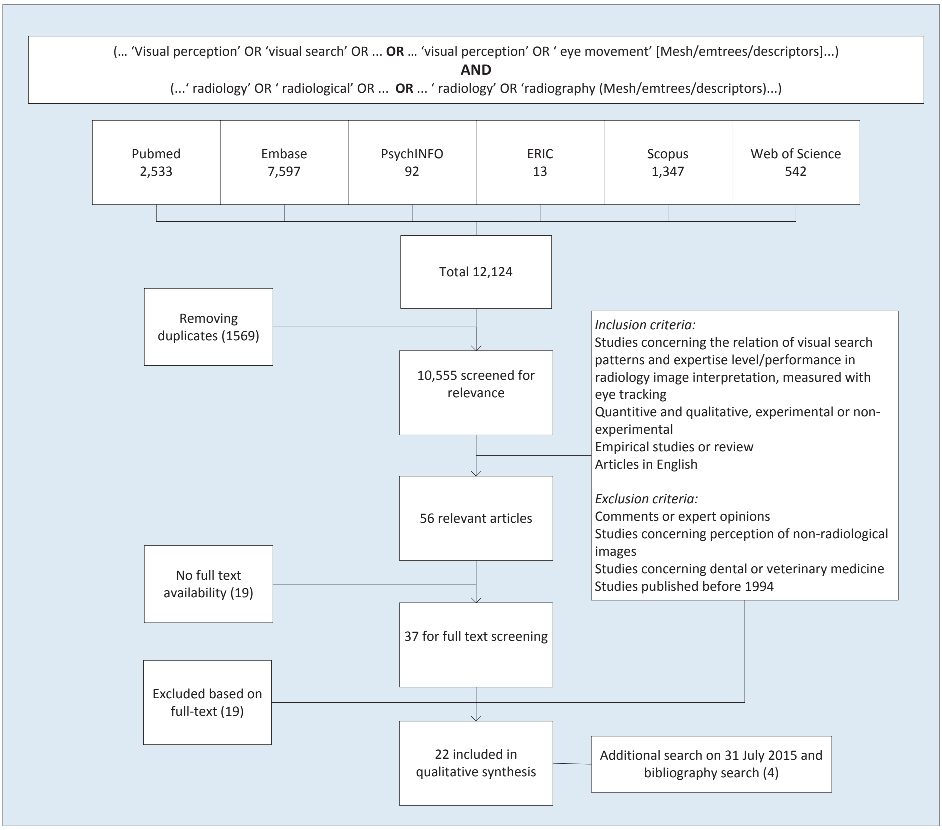

Figure 2. Search flow chart

Data Collection, Quality Assessment

Two reviewers independently extracted characteristics and outcomes of the studies that met the inclusion criteria. Discrepancy between the researchers was discussed until consensus was reached. The quality of the studies was assessed with the Medical Education Research Study Quality Instrument (MERSQI) score[19].

Data Analysis and Synthesis

The relationship between visual search characteristics and level of expertise was approached in several ways. We could identify six topics for thematic analysis: (1) time on task, (2) eye movement characteristics of experts, (3) differences in visual attention, (4) visual search patterns, (5) search patterns in cross-sectional stack imaging, and (6) teaching visual search strategies. All data were categorized in one of these themes. One single study could address multiple themes. For the textual narrative synthesis, the results within each theme were pooled in different subgroups to discover patterns of differences and similarities in results in between or within subgroups. Subgroup results were compared based on differences in task, target, image modality and lesion subtlety (Table 2); similarities and differences in subgroup results were reported.

RESULTS

The primary search on August 15, 2014 yielded 12,125 articles; after removing duplicates 10,555 remained. Narrowing the search strategy was considered, but key words broadening the search (e.g. visual perception, eye movements) were essential for the research questions and could not be omitted. Titles were screened for relevance. Applying inclusion and exclusion criteria on abstracts of potentially relevant studies resulted in 56 relevant abstracts, of which 37 articles could be retrieved in full-text. After applying inclusion and exclusion criteria on full-text articles, 18 relevant studies remained. Four additional articles were added from the second search on July 31, 2015. Screening reference lists of all included articles and six review articles[2, 9, 14, 20-22] did not provide additional relevant articles. Search results are summarized in Figure 2.

Characteristics of Included Studies

All included studies contained quantitative data. Four studies additionally analyzed qualitative data with respect to the research question[23-26]. In total, 526 observers participated in the studies. Quality assessment yielded a mean MERSQI score of 11.5 (out of 18), ranging from 8.5 to 14.5. A major limitation of most studies was the study design. Almost all studies were cross-sectional studies, investigating associations between level of expertise and visual search characteristics, which may have been influenced by many co-varying factors. Only one study investigated the effect of visual search strategies on performance with a randomized controlled trial[27]. Triangulation of eye tracking data with other measures is important for a sound interpretation of underlying cognitive processes[28]. Unfortunately, the vast majority of the studies only related the eye tracking data to performance data and did not triangulate results with verbal data. Sampling was generally done from one institution, except for one study that sampled two institutions[29]. The validity evidence for the evaluation instruments varied between studies. For example, several studies did not report any properties of the applied eye tracking system and whether or not image findings were validated by an expert panel. Overall, strengths of the studies were the objectivity of the data and outcome measures.

The included studies are listed in Table 1. Detailed characteristics of the studies are shown in the appendix. The results of the thematic analysis are listed in Table 2.

Time on Task

In general visual search time decreases with increasing levels of expertise[24, 26, 27, 30-36]. The average time experts took to view an image varied largely between studies, ranging from four seconds up to around 45 seconds, due to differences in task characteristics (detection, interpretation or both; lesion subtlety) and time limits. Novices or lay persons took approximately 1.5 to 2.5 times longer. In some studies viewing time did not significantly differ between expert levels[37-39].

Eye Movement Characteristics of Experts

Studies show a high variance in the type of tasks, relating to detection or interpretation, subtlety and number of lesions, image modality and time limits. Despite these differences, there are some consistent patterns in how experts move their eyes compared to less proficient observers: experts fixate faster on an abnormality[7, 24, 30, 34, 39, 40] and make more fixations[26, 32, 33, 35]. Experts or subspecialized experts tended to fixate on an abnormality within 0.5 to two seconds in mammography and CT studies, around three seconds in skeletal and chest X-rays, and up to five seconds in subtle fracture cases. Novices typically took around 1.5 to two times longer to fixate on a lesion, and up to 4.5 times longer in cases with subtle abnormalities or a second abnormality. Most studies found no significant differences in total fixation duration between expert levels[15, 31, 33].

Other findings were less consistent across studies. Saccade length was found to be larger in experts than novices in two studies[31, 32], but smaller in one study[15]. The short saccades of experts were found in CT scans with visceral abnormalities (e.g. cancer) and related enlarged lymph nodes, which are often grouped and in the vicinity of the visceral abnormality. Variable results were found concerning image coverage: two chest X-ray studies found a negative relation between image coverage and expertise level[27, 32], while two chest CT studies found a positive relationship between image coverage and performance[29, 36]. Eye movement characteristics of experts are summarized in Table 3.

Differences in Visual Attention

Apart from describing pure eye movement characteristics of experts, examining which areas in the image receive more or less attention can give additional insight into the nature of visual expertise. For example, do experts pay more attention to the abnormality (the area of interest, AOI), or do they spend more time on the rest of the image to check for other abnormalities?

Table 1. Characteristics and quality assessment of included studies

				Design				**Variables**				**Quality assessment (MERSQI)**						
Publication year	**First author**	**Type of article**	**Specialty**	Correlational	Causal-comparative	Randomized trial	Qualitative components	1. Level of expertise	2. eye tracking measure(s)	3. Retrospective reporting	4. performance	Study design	Sampling	Type of data	Validity of evaluation instrument	Data analysis	Outcomes	Total
2009	Alzubaidi, M.[a]	CONF	RAD	x			x	x	x	x		1	1	3	0	2	1.5	8.5
2010	Alzubaidi, M.[b]	CONF	RAD	x				x	x	x		1	1	3	1	2	1.5	9.5
2013	Bertram, R.	JOUR	RAD		x			x	x		x	1	1	3	2	3	1.5	11.5
2010	Cooper, L.	CONF	RAD		x			x	x		x	1	1	3	3	3	1.5	12.5
2015	Diaz, I.	JOUR	RAD		x			x	x		x	1	1	3	1	1	1.5	8.5
2013	Donovan, T.	JOUR	RAD		x			x	x		x	1	1	3	3	3	1.5	12.5
2013	Drew, T.	JOUR	RAD		x				x		x	1	2	3	3	3	1.5	13.5
2015	Giovinco N.A.	JOUR	ORT		x			x	x			1	1	3	2	2	1.5	10.5
1994	Hu, C. H.	JOUR	RAD		x		x	x	x		x	1	1	3	3	3	1.5	12.5
2012	Kok, E.[a]	JOUR	RAD		x			x	x		x	1	1	3	3	3	1.5	12.5
2016	Kok, E.[b] - exp.1	JOUR	RAD		x			x	x		x	1	1	3	3	3	1.5	12.5
2016	Kok, E.[b] - exp.2	JOUR	RAD			x			x		x	3	1	3	3	3	1.5	14.5
1996	Krupinski, E. A.	JOUR	RAD		x		x	x	x		x	1	1	3	3	3	1.5	12.5
2007	Kundel, H. L.	JOUR	RAD	x					x		x	1	1	3	3	3	1.5	12.5
2007	Leong, J. J. H.	JOUR	ORT/ RAD/ A&E		x			x	x		x	1	1	3	2	3	1.5	11.5
2014	Mallet, S.	JOUR	RAD		x			x	x		x	1	1	3	3	2	1.5	11.5
2006	Manning, D.[b]	JOUR	RAD		x			x	x		x	1	1	3	1	3	1.5	10.5
2011	Matsumoto, H.	JOUR	NEU		x		x	x	x			1	1	3	2	3	1.5	11.5
1996	Nodine, C. F.[a]	JOUR	RAD		x			x	x		x	1	1	3	0	3	1.5	9.5
2002	Nodine, C. F.[b]	JOUR	RAD		x			x	x		x	1	1	3	2	3	1.5	11.5
2015	Rubin, G.D.	JOUR	RAD	x					x		x	1	1	3	3	3	1.5	12.5
2013	Voisin, S.	CONF	RAD	x					x		x	1	1	3	2	2	1.5	10.5
2012	Wood, G.	JOUR	RAD	x	x			x	x		x	1	1	3	3	3	1.5	12.5

‡Research question categories: 1. Expert level related to visual search patterns; 2. Visual search patterns related to performance; 3. Visual search patterns or expert level related to error types

CONF = Conference paper, JOUR = Journal article, RAD = Radiology, ORT = Orthopedics, A&E = Accident and Emergency, NEU = Neurology

MERSQI = Medical Education Research Study Quality Instrument. Exp. = Experiment

Table 2. Results of thematic analysis

	Participants		**Type of task**			**Cases**		**Themes**					
First author	**Total participants**	**Expert level comparison**	**Task**	**Target**	**Image modality**	**Total**	**Lesion subtlety**	**1. Time on task**	**2. Eye movement characteristics**	**3. Differences in visual attention**	**4. Visual search patterns**	**5. Search patterns in stack imaging**	**6. Teaching visual search strategies**
Alzubaidi, M.[a]	5	S-E	D+I	chest lesions	XR	20	NS	X	X		X		
Alzubaidi, M.[b]	5	S-E	D+I	chest lesions	XR	20	NS				X		
Bertram, R.	38	E-N-L	D	abdominal lesions	sCT	9	M		X	X			
Cooper, L.	28	E-I-L	D	stroke	sCT / MR	48	M	X	X	X			
Diaz, I.	6	E-N	D	lung nodules	sCT	NS	M					X	
Donovan, T.	40	E-N-L	D	lung nodules	XR	30	S	X		X			
Drew, T.	25	E	D	lung nodules	sCT	5	S		X			X	
Giovinco, N.A.	16	E-I	I	bunion	XR	25	NA	X	X				
Hu, C. H.	15	S-E-I-N	D	lung nodules / fractures	XR	10	S				X		
Kok, E.[a]	30	E-I-N	D+I	chest lesions	XR	24	S	X	X		X		
Kok, E.[b] - exp.1	20	E-I-N	D+I	chest lesions	XR	5	NA	X	X		X		
Kok, E.[b] - exp.2	75	N		chest lesions	XR	22	M						X
Krupinski, E. A.	6	S-I	D	breast cancer	XR	20	M	X	X		X		
Kundel, H. L.	9	S-E-I	D+I	breast cancer	XR	40	S		X		X		
Leong, J. J. H.	25	E-I	D	fractures	XR	33	NS	X		X	X		
Mallet, S.	65	S-E	D	colon polyps	CTC	23	M	X	X				
Manning, D. [b]	21	E-N	D	lung nodules	XR	120	M	X	X				
Matsumoto, H.	24	E-N	D+I	stroke	tCT	6	M			X			
Nodine, C. F.[a]	15	S-I-N-L	D+I	breast cancer	XR	9	NS		X				
Nodine, C. F.[b]	9	S-E-I	D+I	breast cancer	XR	40	S		X	X			
Rubin, G.D.	13	S-E-I	D	lung nodules	sCT	40	S	X	X				
Voisin, S.	6	S-I	I	breast masses	XR	40	NA	X	X				
Wood, G.	30	E-I-N	D+I	fractures	XR	9	M	X	X	X			

S = subspecialized experts, E = experts, I = intermediates, N = novices, L = lay people
D = detection, I = interpretation
XR = X-rays, sCT = stack mode Computed Tomography, tCT = tiled mode Computed Tomography, MR = Magnetic Resonance
S = subtle lesions, M = mixed subtle and non-subtle, NA = not applicable, NS = not specified
exp. = experiment

Table 3. The relation between level of expertise and eye tracking parameters

Eye tracking parameters	Association with high levels of expertise (number of studies*)
Total time	Decrease (10)
Time to first fixation	Decrease (6)
Total fixation duration	-
Fixation duration on AOI	Increase (3) or decrease (3)
Dwell time ratio	Increase (1) or decrease (1)
Total number of fixations	Decrease (4)
Number of fixations on AOI	Increase (1) or decrease (1)
Saccade length	Increase (2) or decrease (1)
Image coverage	Decrease (2) or increase (2)

**Only significant results are included*
- = No significant difference reported

Results were found to differ across tasks: fixation duration on AOIs decreased in higher level experts in detection only tasks[30, 37, 38], but increased in tasks combining detection and interpretation[23, 34, 41]. The increase in fixation duration was most prominent in subtle lesions[23, 34]. The effect of expertise on the frequency of fixations on a relevant area also differed across tasks. During lung nodule detection, experts fixated less frequently on nodules than novices did [37]. In contrast to novices, experts did not increase their fixation frequency, when hydronephrosis (i.e. a dilated renal collecting system) was visible on CT[15]. However, when searching CT scans for enlarged lymph nodes, experts fixated more frequently than novices did on a region were enlarged lymph nodes prevail[15]. When lymph nodes were present, they further increased their fixation frequency in contrast to novices[15]. This difference may be explained by the nature of the tasks. The experts probably did not need as many fixations as the novices to identify a lung nodule or hydronephrosis. They instead deliberately sampled other areas where abnormalities frequently are expected, and increased their attention to assess lymph nodes for enlargement.

Compared to experts, novices are found to give more attention to salient structures, regardless of their relevance. For example, the heart is a salient though relatively unimportant structure on a chest X-ray, because there is generally not much to report about it apart from its size. Naïve observers spend much time on the heart, while experts spend a large percentage of their time on the lungs[37]. Similarly, Matsumoto et al. found that novices fixated more or equally often on salient features in brain CT scans (e.g. physiological calcifications), while experts fixated more on non-salient, but relevant areas, such as the brain parenchyma and inconspicuous lesions[23].

Visual Search Patterns

Several studies distinguished different visual search patterns among different types of tasks (Table 4). Hu et al.[25] described four types of search patterns (circular, radial, zigzag and complex) in wrist and hand X-ray search and found that experts predominantly follow a radial pattern, i.e. following the digits out and back from the carpus, while novices' search patterns were more variable. Krupinski et al.[24] reported that, in mammograms with more than one lesion, after detection of the first lesion, intermediates generally followed a circumferential search pattern, while experts directly proceeded to the second (and third) lesion. Also, experts showed more comparison scanning between the left and right breast. Kok et al.[31] found a higher global/local ratio in experts viewing normal chest X-ray images, meaning that experts' visual search patterns were more diffuse than novices. In abnormal images no significant global/local ratio differences were found between expert levels. This research group also found that experts searched normal chest X-rays more systematically than novices[27].

Table 4. Characteristics of visual search patterns associated with high levels of expertise

Image modality	Visual search patterns in high expert levels
Wrist and hand X-rays	Radial pattern; global-focal pattern
Mammography	Comparison scanning between left and right
	Global-focal pattern
Chest X-rays	Systematic search
	Global-focal pattern
	Diffuse pattern
Chest CT in stack mode	Drilling (scrolling up and down while focusing on one area)

Global-Focal Search Patterns

Some studies found quantitative or qualitative evidence that the visual search of experts is characterized by a global-focal search pattern. Quantitative evidence was based on distance to target measures and showed that experts' search was consistent with a two-stage search pattern: 1) spending time on or near the abnormality, interpreted as time for identification and decision making and 2) spending time relatively far from the abnormality, for cross-referencing or identification of other abnormalities[38]. The use of a two-stage search was corroborated by qualitative data from another study where radiologists were interviewed following an eye tracking experiment[26]. The most experienced radiologists described a two-step search strategy, starting with holistic perception, followed by more spatially focused visual perception analyzing areas of the X-ray image in detail. In contrast, the radiologist with the least experience described his search strategy as a sequential examination of small areas, a predominantly focal search pattern which was supported by his eye movements[26]. Qualitative evidence for a global-focal search pattern was also found in mammography: the best observers typically jumped directly to the malignancy at the beginning of their search, followed by a circumferential scan of the view, and consecutively a long saccade to the malignancy at the second view, again followed by a circumferential scan[7]. Alzubaidi et al. found that less experienced observers fixate more often than experienced observers within one region repetitively[42], again reflecting a focal search pattern in the less experienced.

Search Patterns in Cross-Sectional Stack Imaging

Two studies combined eye tracking and scroll behavior data and distilled different search patterns used in cross-sectional stack image viewing. Drew et al.[29] distinguished two visual search patterns in lung nodule detection in a stack of CT images: scanners and drillers. Scanners tend to visually search a single slide, before scrolling further through the stack, while drillers focus their eyes on one quadrant of the lung field and quickly scroll through the stack in depth before moving to another quadrant. Scanners' search patterns are characterized by longer saccades and more quadrant fixation clusters compared to drillers. Drillers reached larger lung coverage and identified more pulmonary nodules than scanners. As a co-varying factor, drillers tended to have more experience in searching through chest CT scans than scanners. A second, descriptive study of radiologists' approach to viewing CT scans[43] found a search approach similar to drilling: an organized way of scrolling and viewing through CT scans by typically scrolling up and down through a quadrant once before moving to a next quadrant. Lay people's search patterns showed more frequent changes in scrolling and viewing direction and varied more than expert patterns.

Teaching Visual Search Strategies

Only one study addressed the effect of teaching specific search strategies and compared three visual search trainings in chest X-ray viewing: systematic viewing, non-systematic viewing and full coverage viewing. The systematic viewing training led to a more systematic search and the full coverage training to a larger image coverage, but neither resulted in higher diagnostic performance compared to the non-systematic viewing group. The full coverage group even performed worse than the non-systematic viewing group[27].

DISCUSSION

The eye tracking literature of the past twenty years details the relationship between visual search characteristics and level of expertise within six main themes: time on task, eye movement characteristics of experts, differences in visual attention, visual search patterns, search patterns in cross-sectional stack imaging, and the effect of teaching visual search strategies. Studies varied considerably with respect to multiple study characteristics, such as expertise levels, type of tasks, eye tracking parameters and performance measures. However, some consistent results were found across studies and in general supported the global-focal search theory. Experts and high performers consistently need shorter viewing times and fixate on abnormalities faster than less experienced readers or low performers. Experts needed up to five seconds to first fixate on subtle abnormalities. From previous non-eye tracking studies we know that experts can make correct diagnostic decisions based on much shorter image exposures, even in exposures as short as 200 milliseconds[4]. Such short time spans do not allow for searching the image and fixating on abnormalities; rather abnormalities are probably perceived through taking advantage of their large visual span. Different from most of the cases reported in our review, the cases used in the 200-millisecond study of Kundel et al. were large or diffuse abnormalities that can be assessed more globally than subtle findings.

Findings regarding number of fixations and saccade length also supported the global-focal search theory. The number of fixations decreased with higher expertise levels, probably because experts need fewer fixations to collect relevant image information[14]. This also aligns with shorter viewing times needed by experts. Saccade length was found to be generally larger in experts than novices, which aligns with an effective global search, because large saccades towards a disturbed signal are likely to follow a successful global search[7]. However, experts were found to shorten their saccades in CT scans with grouped enlarged lymph nodes that were related to and in the proximity of visceral abnormalities, probably because they *suspected* the enlarged lymph nodes to be in proximity of the detected visceral abnormality and near each other, based on knowledge and experience. This underscores that search patterns of experts are driven by knowledge and that they anticipate their search based on what they have already found.

Differences in visual attention across levels of expertise revealed that experts tend to focus on relevant though not necessarily salient structures, whereas novices tend to look at salient structures regardless of their relevance. We found some inconsistent results across studies with respect to differences in visual attention to abnormalities, though this may be explained by differences in the type of task. Experts spend less time fixating on lesions than novices in detection only tasks, probably because they make fast normal – abnormal decisions and quickly proceed searching the rest of the image for other findings. When diagnostic reasoning is added to the task, experts spend more time on the lesion than novices. Because this difference was most pronounced in subtle or non-salient findings, it may be explained by novices' failures to recognize the findings as abnormal hindering them from analyzing its features.

Both quantitative and qualitative analyses yielded evidence for a global-focal search pattern that is more prominent in those with higher expertise levels. Quantitative evidence included a fast initial fixation on the abnormality by experts due to developed global searches, and distance to target measures showing that experts adhere to a two-stage search pattern distinguishing searches near and far from the abnormality. Qualitatively, the global-focal search pattern was confirmed by experts in retrospective reports and by sequenced eye tracking data showing a fast jump towards the abnormality followed by a circumferential scan of the rest of the image. In contrast to experts, novices seem to exhibit a predominantly focal search pattern. Beyond the global-focal search patterns, a wide variety of other search patterns has been explored in the included studies. Some patterns are potentially relevant for teaching purposes, such as the radial search in hand and wrist X-rays, or drilling through chest CT`s. These patterns suggest that expert observers have deliberately adopted such strategies to accomplish a complete and efficient search. However, this is not confirmed by verbal data in the vast majority of studies and the observer's experience could be tacitly driving the search pattern. Therefore, it is doubtful learners' perceptual performance will improve upon incorporating the search strategy of experts if they lack experience and knowledge.

Some visually dependent specialties, such as radiology, have witnessed huge technological developments. Only a minority of the studies investigated visual search in volumetric image perception in radiology, possibly because of challenges that are related to eye tracking in such imaging[44]. The large amount of information in volumetric images increases the risk of perceptual errors and introduces particular challenges such as abnormalities that can only be visualized in certain contrast settings[45]. Because visual search in cross-sectional image stacks requires image manipulation[46] and knowledge and skills needed to interpret stacks of images differ substantially from 2D image interpretation[47], evidence for effective search patterns in both visualization modes is not indisputably exchangeable. For example this review shows contrary results with respect to image coverage which can be possibly explained by differences in image modality. In X-ray viewing, experts tended to cover a smaller area of the image than novices did, which can be explained by a larger visual span of experts, but contrary results were found in CT scan reading. In the study of Drew et al.[29] high performers who tended to scroll back and forth through the image while fixating on one area (i.e. drillers) showed larger image coverage than lower performers that scanned the complete image slide by slide (i.e. scanners). Searching through cross-sectional image datasets in multiple planes and contrast settings[48] demands a joint effort of the observer's hands and eyes to visualize the image information. When scanning the complete area of one slide before proceeding to the next slide, large distances have to be bridged with eye movements which is inefficient and may cause skipping of areas. Drilling through images, fixating on one area and scrolling through, requires much fewer eye movements and more systematically addresses all areas of the image. Because cross-sectional imaging is increasing and contains much visual information[48], finding efficient strategies in different types of cross-sectional search tasks is gaining importance. A recent study investigated zooming behavior in microscope navigation in clinical pathology[49]. Expertise was characterized by fewer magnifications. This aligns with the global-focal search pattern; magnification allows for a more detailed search consistent with the focal search-to-find search patterns of novices, versus a more global (zoomed out) view of experts.

Only one eye tracking study investigated whether training learners to apply expert visual search strategies, in this case systematic search, can improve perceptual performance. Even though the systematic training had a positive effect on the use of systematic viewing, diagnostic performance did not improve. This underscores that we should not assume that teaching expert skills directly to learners is likely to be effective. Much of the evidence in this review supports the global-focal search pattern. But should we teach novices to start with a proficient global search, followed by a focal search? The efficiency of this approach is questionable. The efficient, quick approach of experts is probably driven by hypotheses based on knowledge and experience, in which case it would not make sense to instruct a novice to look at an abnormality more quickly. It may be more beneficial to teach novices where to look in certain clinical circumstances, guided by clinical information or other related findings. Another approach to direct learners` attention to specific areas are Eye Movement Modelling Examples. These are scan paths of experts that can guide learners` visual attention to relevant areas[50, 51]. Other, more defined search strategies mentioned in this review may be relevant to teach to novices, for example drilling though CT scans or radial search in wrist X-rays. But even if we can teach students to search accordingly, will this help them perform better? Randomized trials comparing these search strategies with each other or with free searching are needed to find out which visual search strategies can and should be taught.

Eye tracking studies in other visual diagnostic professions such as pathology[49] and in patient observations for diagnosing neurological disorders[52], show similar evidence for holistic and efficient visual search skills of experts. The holistic search characterized by a fast identification of the abnormality and efficient sampling of the rest of the image aligns with the rapid hypothesis generation and efficient and limited data gathering seen in expert clinical reasoning, attributed to the efficient data processing of System 1 thinking processes[53]. This is in contrast to the more extensive data gathering in System 2 thinking processes, most apparent in novices who sample the entire image with a focal search pattern[26, 27, 32, 42]. It is obvious that novices' and experts' search and thinking processes differ, but for both clinical reasoning and visual search it is unclear which approach is most effective for novices[53], or how either approach should be taught.

Some limitations of our study must be reported. Although we did an exhaustive literature search, we may not have captured all literature concerning the relationship between visual search characteristics and perceptual performance. We only included studies from the last twenty years, because technical improvements have had a large influence on image quality and representation in the last decades[48]. We also confined our search to radiology, and although the results might apply to other visual search tasks, we cannot assure that the results are transferable. Finally, we focused on eye tracking studies and did not include, for example, think-aloud studies that give more insight on the cognitive aspects of visual search. We reduced a large number of titles to a fairly small number of relevant articles, and relevant articles could have been missed. To reduce this risk, two raters applied clear, objective criteria to exclude articles. Cross referencing did not reveal any missed relevant articles.

In conclusion, visual diagnosis research of the past 20 years has predominantly focused on finding differences between experts and novices, providing clues for improving image perception. Heterogeneity of tasks and outcome measures complicates data synthesis and drawing conclusions. In addition, hardly any studies looked at how to improve image perception by teaching visual search strategies based on these theories. For educational purposes, we should shift from describing differences between experts and novices to defining search strategies and investigating methods to teach novices to search efficiently.

Appendix 1. Detailed study characteristics, sorted by target

	Participants									Type of task			
First author	Total participants	Subspecialized experts	Radiologists or other medical specialists	Interns/residents/SHO/registrars	Medical students	Radiographers (technicians)	Student radiographers	Naïve observers	Novices (not specified per group)	Detection task	Detection > 1 lesion	Interpretation task	Time limit
Alzubaidi, M.[a]	5	3	2	-	-	-	-	-	-	x	NS	x	-
Alzubaidi, M.[b]	5	3	2	-	-	-	-	-	-	x	NS	x	-
Kok, E.[a]	30	-	9	10	11	-	-	-	-	x	x	x	-
Kok, E.[b] - exp.1	20	-	9	10	11	-	-	-	-	x	-	x	-
Kok, E.[b] - exp.2	75	-	-	-	75	-	-	-	-	x	x	x	-
Diaz	6	-	3	-	-	-	-	3	-	x	-	-	-
Donovan, T.	40	-	10	-	-	20	-	10	-	x	x	-	-
Drew, T.	25	-	25	-	-	-	-	-	-	x	x	-	x
Manning, D. [b]	21	-	8	-	-	5	8	-	-	x	x	-	x
Rubin, G.D.	13	5	2	6	-	-	-	-	-	x	x	-	-
Krupinski, E. A.	6	3	-	3	-	-	-	-	-	x	x	-	-
Kundel, H. L.	9	3	3	3	-	-	-	-	-	x	-	x	-
Nodine, C. F.[a]	15	3	-	5	-	2	-	5	-	x	NS	x	x
Nodine, C. F.[b]	8	3	3	3	-	-	-	-	-	x	x	x	-
Voisin, S.	6	2	-	4	-	-	-	-	-	-	-	x	NS
Cooper, L.	28	-	8	10	-	-	-	10	-	x	-	-	-
Matsumoto, H.	24	-	12	-	-	-	-	-	12	x	-	x	x
Leong, J. J. H.	25	-	11	14	-	-	-	-	-	x	x	-	-
Wood, G.	30	-	10	10	-	-	10	-	-	x	-	x	-
Hu, C. H.	15	3	1	8	3	-	-	-	-	x	-	-	x
Giovinco, N.A.	16	-	7	-	-	-	-	-	9	-	-	x	-
Bertram, R.	38	-	7	-	-	9	-	22	-	x	x	-	NS
Mallet, S.	65	27	38	-	-	-	-	-	-	x	x	-	x

SHO = Senior House Officer

XR = X-rays, sCT = stack mode Computed Tomography, tCT = tiled mode Computed Tomography, MR = Magnetic Resonance

S = subtle lesions, M = mixed subtle and non-subtle, NA = not applicable, NS = not specified

TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, acc = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, d = detectability index

Type of task		Cases			Performance measures	
Target	Image modality	Total	Abnormal (%)	Lesion subtlety	TP/FP/TN/FN	Other
chest lesions	XR	20	NS	NS	-	
chest lesions	XR	20	NS	NS	-	
chest lesions	XR	24	67	S	-	acc
chest lesions	XR	5	0	NA	TN	
chest lesions	XR	22	91	M	TP/TN	
lung nodules	sCT	NS	NS	M	TP	
lung nodules	XR	30	50	S	TP/FP	acc
lung nodules	sCT	5	100	S	TP/FN	
lung nodules	XR	120	NS	M	-	AUC
lung nodules	sCT	40	100	S	TP	
breast cancer	XR	20	75	M	TP/TN/FP/FN	
breast cancer	XR	40	50	S	TP/FN/FP/TN	AUC
breast cancer	XR	9	44	NS	TP/FP	d
breast cancer	XR	40	50	S	TP/FP/FN/TN	AUC
breast masses	XR	40	100	NA	-	
stroke	sCT / MR	48	75	M	TP/FP	
stroke	tCT	6	83	M	-	AUC
fractures	XR	33	85	NS	TP/FN	
fractures	XR	9	67	M	-	acc
lung nodules/ fractures	XR	10	70	S	TP/FN/FP	
bunion	XR	25	NS	NA	-	
abdominal lesions	sCT	9	67	M	TP/TN	acc
colon polyps	CTC	23	100	M	TP	

REFERENCES

1. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, and van de Weijer J: Eye Tracking: A Comprehensive Guide to Methods and Measures. New York, Oxford University Press, 2011.
2. Krupinski EA: Current Perspectives in Medical Image Perception. Attention, Perception, and Psychophysics 2010; 72: 1205-17.
3. Nodine CF, and Kundel HL: Using Eye Movements to Study Visual Search and to Improve Tumor Detection. Radiographics : a review publication of the Radiological Society of North America, Inc 1987; 7: 1241-50.
4. Kundel HL, and Nodine CF: Interpreting Chest Radiographs without Visual Search. Radiology 1975; 116: 527-32.
5. Carmody DP, Nodine CF, and Kundel HL: Finding Lung Nodules with and without Comparative Visual Scanning. Perception & Psychophysics 1981; 29: 594-98.
6. Oestmann JW, Greene R, Kushner DC, Bourgouin PM, Linetsky L, and Llewellyn HJ: Lung Lesions: Correlation between Viewing Time and Detection. Radiology 1988; 166: 451-3.
7. Kundel HL, Nodine CF, Conant EF, and Weinstein SP: Holistic Component of Image Perception in Mammogram Interpretation: Gaze-Tracking Study. Radiology 2007; 242: 396-402.
8. Swensson RG: A Two-Stage Detection Model Applied to Skilled Visual Search by Radiologists. Perception and Psychophysics 1980; 27: 11-16.
9. Drew T, Evans K, Võ MLH, Jacobson FL, and Wolfe JM: Informatics in Radiology: What Can You See in a Single Glance and How Might This Guide Visual Search in Medical Images? Radiographics 2013; 33: 263-74.
10. Kahneman D: A Perspective on Judgment and Choice: Mapping Bounded Rationality. American Psychologist 2003; 58: 697-720.
11. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. J Med Imaging Radiat Oncol 2012; 56: 173-8.
12. Jager G, Futterer JJ, and Rutten M: Cognitive Errors in Radiology: “Thinking Fast and Slow”. In European Congress of Radiology, Vienna, Austria, 2014.
13. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Invest Radiol 1978; 13: 175-81.
14. Gegenfurtner A, Lehtinen E, and Säljö R: Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review 2011; 23: 523-52.
15. Bertram R, Helle L, Kaakinen JK, and Svedstrom E: The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception. PLoS One 2013; 8.
16. Kastner M, Tricco AC, Soobiah C, Lillie E, Perrier L, Horsley T, Welch V, Cogo E, Antony J, and Straus SE: What Is the Most Appropriate Knowledge Synthesis Method to Conduct a Review? Protocol for a Scoping Review. BMC Med Res Methodol 2012; 12: 114.
17. Mays N, Pope C, and Popay J: Systematically Reviewing Qualitative and Quantitative Evidence to Inform Management and Policy-Making in the Health Field. J Health Serv Res Policy 2005; 10 Suppl 1: 6-20.
18. Lucas PJ, Baird J, Arai L, Law C, and Roberts HM: Worked Examples of Alternative Methods for the Synthesis of Qualitative and Quantitative Research in Systematic Reviews. BMC Med Res Methodol 2007; 7: 4.
19. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, and Wright SM: Association between Funding and Quality of Published Medical Education Research. JAMA 2007; 298: 1002-9.

20. Taylor PM: A Review of Research into the Development of Radiologic Expertise: Implications for Computer-Based Training. Academic Radiology 2007; 14: 1252-63.
21. Norman GR, Coblentz CL, Brooks LR, and Babcook CJ: Expertise in Visual Diagnosis: A Review of the Literature. Acad Med 1992; 67: S78-83.
22. Blondon K, Wipfli R, and Lovis C: Use of Eye-Tracking Technology in Clinical Reasoning: A Systematic Review. Stud Health Technol Inform 2015; 210: 90-4.
23. Matsumoto H, Terao Y, Yugeta A, Fukuda H, Emoto M, Furubayashi T, Okano T, Hanajima R, and Ugawa Y: Where Do Neurologists Look When Viewing Brain Ct Images? An Eye-Tracking Study Involving Stroke Cases. PLoS One 2011; 6: e28929.
24. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Academic radiology 1996; 3: 137-44.
25. Hu CH, Kundel HL, Nodine CF, Krupinski EA, and Toto LC: Searching for Bone Fractures: A Comparison with Pulmonary Nodule Search. Academic Radiology 1994; 1: 25-32.
26. Alzubaidi M, Black Jr JA, Patel A, and Panchanathan S: Conscious Vs. Subconscious Perception, as a Function of Radiological Expertise. Albuquerque, NM, 2009.
27. Kok EM, Jarodzka H, de Bruin AB, BinAmir HA, Robben SG, and van Merrienboer JJ: Systematic Viewing in Radiology: Seeing More, Missing Less? Adv Health Sci Educ Theory Pract 2016; 21: 189-205.
28. Kok EM, and Jarodzka H: Before Your Very Eyes: The Value and Limitations of Eye Tracking in Medical Education. Med Educ 2016.
29. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13.
30. Cooper L, Gale A, Saada J, Gedela S, Scott H, and Toms A: The Assessment of Stroke Multidimensional Ct and Mr Imaging Using Eye Movement Analysis: Does Modality Preference Enhance Observer Performance? Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment 2010; 7627: 76270B.
31. Kok EM, de Bruin ABH, Robben SGF, and van Merriënboer JJG: Looking in the Same Manner but Seeing It Differently: Bottom-up and Expertise Effects in Radiology. Applied Cognitive Psychology 2012; 26: 854-62.
32. Manning D, Ethell S, Donovan T, and Crawford T: How Do Radiologists Do It? The Influence of Experience and Training on Searching for Chest Nodules. Radiography 2006; 12: 134-42.
33. Giovinco NA, Sutton SM, Miller JD, Rankin TM, Gonzalez GW, Najafi B, and Armstrong DG: A Passing Glance? Differences in Eye Tracking and Gaze Patterns between Trainees and Experts Reading Plain Film Bunion Radiographs. J Foot Ankle Surg 2014; 54: 382-91.
34. Wood G, Knapp KM, Rock B, Cousens C, Roobottom C, and Wilson MR: Visual Expertise in Detecting and Diagnosing Skeletal Fractures. Skeletal Radiology 2013; 42: 165-72.
35. Voisin S, Pinto F, Xu S, Morin-Ducote G, Hudson K, and Tourassi GD: Investigating the Association of Eye Gaze Pattern and Diagnostic Error in Mammography. In Proceedings of SPIE - The International Society for Optical Engineering, 2013.
36. Rubin GD, Roos JE, Tall M, Harrawood B, Bag S, Ly DL, Seaman DM, Hurwitz LM, Napel S, and Roy Choudhury K: Characterizing Search, Recognition, and Decision in the Detection of Lung Nodules on Ct Scans: Elucidation with Eye Tracking. Radiology 2015; 274: 276-86.
37. Donovan T, and Litchfield D: Looking for Cancer: Expertise Related Differences in Searching and Decision Making. Applied Cognitive Psychology 2013; 27: 43-49.
38. Leong JJH, Nicolaou M, Emery RJ, Darzi AW, and Yang GZ: Visual Search Behaviour in Skeletal Radiographs: A Cross-Speciality Study. Clinical Radiology 2007; 62: 1069-77.

39. Mallett S, Phillips P, Fanshawe TR, Helbren E, Boone D, Gale A, Taylor SA, Manning D, Altman DG, and Halligan S: Tracking Eye Gaze During Interpretation of Endoluminal Three-Dimensional Ct Colonography: Visual Perception of Experienced and Inexperienced Readers. Radiology 2014; 273: 783-92.
40. Nodine CF, Kundel HL, Lauver SC, and Toto LC: Nature of Expertise in Searching Mammograms for Breast Masses. Academic radiology 1996; 3: 1000-06.
41. Nodine CF, Mello-Thoms C, Kundel HL, and Weinstein SP: Time Course of Perception and Decision Making During Mammographic Interpretation. American Journal of Roentgenology 2002; 179: 917-23.
42. Alzubaidi M, Patel A, Panchanathan S, and Black Jr JA: Reading a Radiologist's Mind: Monitoring Rising and Falling Interest Levels While Scanning Chest X-Rays. In Progress in Biomedical Optics and Imaging - Proceedings of SPIE, San Diego, CA, 2010.
43. Diaz I, Schmidt S, Verdun FR, and Bochud FO: Eye-Tracking of Nodule Detection in Lung Ct Volumetric Data. Med Phys 2015; 42: 2925.
44. Venjakob A, and Mello-Thoms C: Review of Prospects and Challenges of Eye Tracking in Volumetric Imaging. Journal of Medical Imaging 2015; 3.
45. McCreadie G, and Oliver TB: Eight Ct Lessons That We Learned the Hard Way: An Analysis of Current Patterns of Radiological Error and Discrepancy with Particular Emphasis on Ct. Clin Radiol 2009; 64: 491-9; discussion 500-1.
46. van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JC, Ravesloot CJ, van Schaik JP, and Ten Cate TJ: Interpretation of Radiological Images: Towards a Framework of Knowledge and Skills. Adv Health Sci Educ Theory Pract 2014; 19: 565-80.
47. van der Gijp A, Ravesloot CJ, van der Schaaf MF, van der Schaaf IC, Huige JC, Vincken KL, Ten Cate OT, and van Schaik JP: Volumetric and Two-Dimensional Image Interpretation Show Different Cognitive Processes in Learners. Acad Radiol 2015; 22: 632-9.
48. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, and Seltzer SE: Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section Ct Scan Can Ruin Your Whole Day. Radiology 2011; 259: 346-62.
49. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, and Boshuizen HP: Expertise in Clinical Pathology: Combining the Visual and Cognitive Perspective. Adv Health Sci Educ Theory Pract 2015; 20: 1089-106.
50. Jarodzka H, Balslev T, Holmqvist K, Nystrom M, Scheiter K, Gerjets P, and Eika B: Conveying Clinical Reasoning Based on Visual Observation Via Eye-Movement Modelling Examples. Instructional Science 2012; 40: 813-27.
51. Vitak SA, Ingram JE, Duchowski AT, Ellis S, and Gramaopadhye AK: Gaze-Augmented Think-Aloud as an Aid to Learning. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2012; 2991-3000.
52. Balslev T, Jarodzka H, Holmqvist K, de Grave W, Muijtjens AM, Eika B, van Merrienboer J, and Scherpbier AJ: Visual Expertise in Paediatric Neurology. Eur J Paediatr Neurol 2012; 16: 161-6.
53. Norman G: Dual Processing and Diagnostic Errors. Adv Health Sci Educ Theory Pract 2009; 14 Suppl 1: 37-49.

Chapter 5

The Effect of Teaching Search Strategies on Perceptual Performance

This chapter is based on:
Van der Gijp, A., Vincken, K.L., Boscardin, C., Webb, E.M., Ten Cate, Th.J., Naeger, D.M. The effect of teaching search strategies on perceptual performance. Academic Radiology 2017; 24(6):762-767.

ABSTRACT

Objective Radiology expertise is dependent on the use of efficient search strategies. The aim of this study is to investigate the effect of teaching search strategies on trainee's accuracy in detecting lung nodules at CT.

Methods Two search strategies, 'scanning' and 'drilling', were tested with a randomized crossover design. Nineteen junior radiology residents were randomized into two groups. Both groups first completed a baseline lung nodule detection test allowing a free search strategy, followed by a test after scanning instruction and drilling instruction or vice versa. True and false positive scores and scroll behavior were registered. A mixed design ANOVA was applied to compare the three search conditions.

Results Search strategy instruction had a significant effect on scroll behavior, $F(1.3)=54.2$, $p<0.001$, TP score, $F(2)=16.1$, $p<0.001$, and FP score, $F(1.3)=15.3$, $p<0.001$. Scanning instruction resulted in significantly lower TP scores than drilling instruction (M10.7, SD5.0 versus M16.3, SD5.3); $t(18)=4.78$, $p<0.001$ or free search (M15.3, SD4.6); $t(18)=4.44$, $p<0.001$. TP scores for drilling did not significantly differ from free search. FP scores for drilling (M7.3, SD5.6) were significantly lower than for free search (M12.5, SD7.8); $t(18)=4.86$, $p<0.001$.

Conclusion Teaching a drilling strategy is preferable to teaching a scanning strategy for finding lung nodules.

INTRODUCTION

Perceptual errors account for a substantial part of misdiagnoses in radiology[1] and can be related to the search behavior of the observer[2]. For educational purposes, it is important to identify which visual search patterns are most effective and to investigate if teaching search strategies improves perception.

Visual search characteristics that are related to expertise and high performance have been identified in various radiology perception tasks[3]. For example experts tend to fixate on abnormalities faster[4-6] and need less time and a smaller number of eye fixations to inspect the image[7, 8]. These characteristics derive from experience, and they lack an underlying structure that can be taught to novices.

Some specific visual search patterns are found to be related to high performance[9-12]. Most patterns apply to visual search in X-rays, such as chest X-rays or mammography. Two visual search types are distinguished for searching chest CT images: "scanners" and "drillers"[12]: Scanners tend to visually search a single slice, before scrolling further through the stack, whereas drillers focus their eyes on one quadrant of the lung fields and quickly scroll through the stack in depth before moving to another quadrant. Drillers outperformed scanners with respect to higher true positive rates and a larger lung coverage[12]. One interesting finding was that, when given the option to search freely, more experienced readers tend to select "drilling" as a search pattern (the more effective pattern), suggesting it might be a pattern that has, consciously or unconsciously, evolved through instruction or practice. The relationship between search patterns and experience has been noted in several other studies[9-12], although it is unknown if experts unconsciously adopt these patterns or deliberately chose or had acquired one, as a *strategy*.

Teaching junior trainees to use expert search strategies may not necessarily be effective. First, learning the strategy may not be easy, particularly given that most experts acquire their behaviors after years of practice. Second, the improvement in perceptual performance that comes with experience is probably due to multiple factors. Knowledge gained and feedback received are known to be critical factors in developing visual expertise[13-16]. Therefore, it is not evident that learners' perceptual performance will improve simply by using the search strategies of experts. However, some perceptual tasks, such as finding lung nodules on chest CT scans, do not depend on a large knowledge base, and therefore teaching a search strategy may improve detection. Experimental studies may be beneficial to determine if search patterns can be taught to junior observers, and if this can improve perceptual performance.

The aim of this research study is twofold: 1) to investigate if drilling and scanning search strategies can be taught to junior radiology trainees, and 2) to compare the effect of teaching each search strategy on trainee's perception accuracy of lung nodule detection. We hypothesized that junior radiology trainees could adopt a new search strategy after instruction and that the use of a drilling strategy would improve the trainees' perceptual performance compared to a scanning strategy.

METHODS

Design

An experimental study was conducted to compare the effect of two teaching methods on perceptual performance. A randomized cross over design was chosen to adjust for individual variation in performance, differences in search strategies prior to any search strategy instruction, and possible differences in search behavior due to the sequence of the search strategy instructions. The design is illustrated in Figure 1.

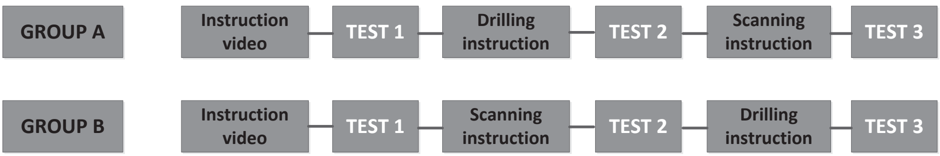

Figure 1. Study design

Study Population and Procedure

Over a 3-month period, nineteen (70%) first and second-year radiology trainees of a US academic medical center's radiology residency program enrolled in the study. Participants were randomly divided into two groups: both groups first watched an instructional video that provided lung nodule definitions. For the purpose of the study, a pulmonary nodule was defined as a solid opacity with a diameter greater than or equal to 3.0 mm. Ground glass or calcified nodules were not included. The instructional video showed examples of true nodules and also addressed other pulmonary abnormalities that were not considered nodules, such as consolidations, linear densities, pleural irregularities and apical scarring, all illustrated by examples. Study participants then completed a pretest using free search (Test 1).

After the free search test (Test 1), group A started with the drilling instruction video, followed by Test 2, the scanning instruction video, and Test 3. Group B received the scanning and drilling instruction in opposite order. The drilling instructions explained the drilling search strategy: mentally dividing each lung into three regions (anterior, middle and posterior) and scrolling through each region individually, while keeping the eyes fixated in that region. The scanning instruction explained the scanning search strategy: reviewing all visible lung parenchyma at once (both sides), while slowly scrolling down, image by image.

In all three tests, participants were asked to mark as many true pulmonary nodules as possible, while avoiding marking any foci not meeting the study's definition of a true nodule. There was a time limit of four minutes per case.

The digital assessment program VQuest (www.vquest.eu) was used for the viewing and marking of lung nodules. This program is designed to deliver tests containing volumetric images, and allows for registering all scroll movements and mouse clicks. During the tests, participants could scroll through the stack of images, zoom in or out, adjust contrast settings and measure findings. All stacks were viewed in axial plane. To select a lung nodule, participants were instructed to place a marker by clicking in the center of the nodule.

Tests

Test 1, 2 and 3 were unique tests, each containing seven volumetric pulmonary CT-scans. In total, each test consisted of 31 true nodules spread out over the seven CT scans. Nodules were 3 to 6 mm, with an average of 4 mm. The scans were retrieved from the Picture Archiving and Communication System (PACS) of the institution and were reviewed by two experienced radiologists (10 and 5 years of experience). Disagreement was resolved in consensus format. The selected chest CT-scans had on average 349 slices, and slice thickness was 1.25 mm in all cases. The tests were made as equivalent as possible, by means of a test blue print (Table 1). Each test was similar with regards to total number of nodules, the size of the nodules, the distribution of cases with fewer and more nodules, and the distribution of easy versus difficult cases. Nodules attached to vessels, bronchi, mediastinal structures or diaphragm were considered difficult, all other locations were considered easy.

Table 1. Tests blue prints with number, size and location of nodules

	Number per case			**Size**			**Location***	
	0-2	3-6	>6	3-4 mm	5-6 mm	Average size (mm)	Easy	Difficult
Test 1	2	3	2	23	8	4.0	25	6
Test 2	2	3	2	23	8	4.1	25	6
Test 3	2	3	2	23	8	4.0	25	6

* Difficult: attached to vessels, bronchi, mediastinal structures or diaphragm; Easy: anywhere else

Questionnaire

Participants completed a short questionnaire to evaluate the study set up and to determine how participants searched for nodules in the case of free search and how they perceived the two instructed search strategies. Response formats were: a five-point Likert scale (5 = strongly agree, 4 = agree, 3 = undecided, 2 = disagree, 1 = strongly disagree), dichotomous (yes or no) or open. The survey questions are listed in the appendix.

Analysis

Reliability of the true positive scores was estimated with Cronbach`s alpha, using the true nodules as items. Average p-values of the items were calculated to estimate test difficulty.

Two main outcomes were measured, related to the two research questions: (i) Scroll behavior and (ii) perceptual performance.

(i) Scroll behavior: for each case the number of runs was determined. A run is defined by Venjakob et al. as a scroll movement in one direction that covers at least 50% of the total scan length, ignoring interruptions by any smaller scroll movements in the other direction[17].

Figure 2 presents an example of a scroll pattern with five runs, despite some interrupting smaller movements. The total scrolling time was also calculated for each case.

(ii) Perceptual performance: the average numbers of true and false positives of Test 1, 2 and 3 were calculated for both groups. We were primarily interested in differences in true positives. False positive numbers were also measured, given an increase in true positives is sometimes achieved at the cost of an increase in false positives.

For both outcome measures, a mixed design ANOVA was used for the statistical analysis. The instructed search strategy (no scroll instruction, drilling instruction or scanning instruction) was the within-subject factor. Between subject factors were study group (A or B) and year of training (first or second year of residency). Paired T-tests were used for post hoc testing in normally distributed data and Wilcoxon signed-rank tests in non-normal distributed data.

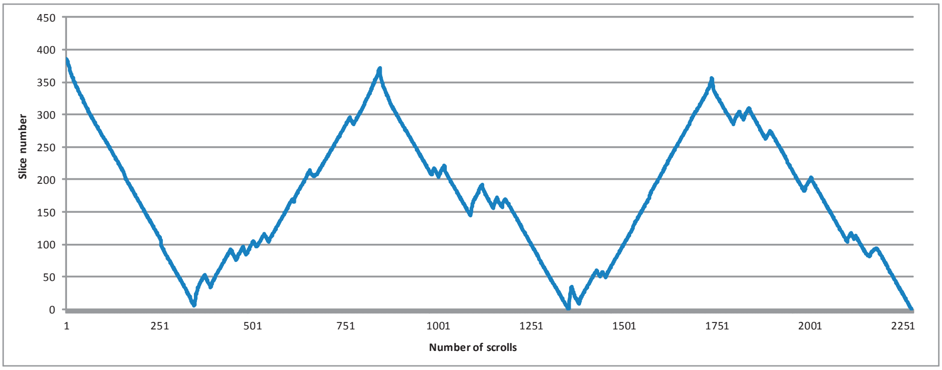

Figure 2. Example of a scroll pattern with five runs

IRB Approval

The study was approved by the Institutional Review Board of the institution. Participation was voluntary and all participants gave written informed consent. This study did not involve protected health information.

RESULTS

Participants

After randomization, nine participants were assigned to group A and ten to group B, of which four in each group were second-year residents.

Test Performance

Reliabilities of the tests were acceptable: Cronbach`s α was .74, .82, and .84 for Test 1, 2 and 3 respectively. Average p-values were: .49, .39 and .48 respectively.

Scroll Behavior

The median number of runs and median scrolling time for each search strategy are provided in Table 2.

There was no significant effect of study group or year of residency on the number of runs or time. There was a significant within-subjects effect of search strategy on number of runs, $F (1.3) = 54.2$, $p < 0.001$, and on time, $F (2) = 10.5$, $p < 0.001$, indicating that the instructed search strategy significantly affected the amount of drilling and the time required to complete the task. There were no significant interaction effects for both outcomes (Table 3).

Table 2. Scroll behavior and perceptual performance measures per search strategy condition

	Free search	**With drilling instruction**	**With scanning instruction**
Scroll behavior			
Number of runs per case, Mdn (IQR)	10 (4)	10 (4)	2 (3)
Scrolling time per case in seconds, Mdn (IQR)	219.6 (35.9)	208.0 (39.3)	167.0 (96.6)
Perceptual performance			
True Positives, M (SD)	15.3 (4.6)	16.3 (5.3)	10.7 (5.0)
False Positives, M (SD)	12.5 (7.8)	7.3 (5.6)	5.6 (4.8)

M = mean, SD = Standard Deviation, Mdn = Median, IQR = Interquartile Range

Table 3. Mixed design ANOVA for scroll behavior outcomes

	Number of runs		**Time**	
	F	p	F	p
Between subjects				
year of residency	1.0	0.33	0.4	0.55
study group	3.5	0.08	0.7	0.41
Within subjects				
search strategy	54.2	<0.001	10.5	<0.001
search strategy * year of residency	0.1	0.88	0.7	0.49
search strategy * study group	28.3	0.74	0.9	0.43
search strategy * year of residency * study group	0.4	0.37	1.8	0.18

study group: intervention group A or B, search strategy: free search, drilling instruction or scanning instruction, year of residency: first or second year

The number of runs was significantly lower after participants were instructed to scan (Mdn 2, IQR 3), than after drilling instruction (Mdn 10, IQR 4), $z = -3.8$, $p < 0.0$, or when allowed to search as desired (Mdn 10, IQR 4), $z = -3.8$, $p < 0.001$. The participants' free search and drilling instructed search had similar amounts of drilling.

The scanning search required significantly less time compared to the free search (Mdn 167.0, IQR 96.6 versus Mdn 219.6, IQR 35.9), $z = -3.6$, $p < 0.001$, although the drilling search (Mdn 208.0, IQR 39.3) was not found to be significantly different from scanning, $z = -1.4$, $p = 0.147$, or free search, $z = -1.9$, $p = 0.053$.

Perceptual Performance

The mean number of true positives and false positives for each search strategy condition are given in Table 2. There was a significant effect of year of residency on the true positive score, $F(1) = 9.0$, $p < 0.01$, although no significant effect on the false positive score (Table 4).

Within-subject analysis shows a significant effect of instructed search strategy on the number of true positives, $F(2) = 16.1$, $p < 0.001$, and false positives, $F(1.3) = 15.3$, $p < 0.001$. Search strategy accounted for 51.8% of the variability in the number of true positives and 50.1 % of the variability in false positives. There were no significant interaction effects with the two between-subject factors, study group and year of residency, for both performance outcomes.

Pair-wise analysis showed that drilling (M 16.3, SD 5.3) and free search (M 15.3, SD 4.6) both resulted in significantly higher true positive scores than scanning (M 10.7, SD 5.0); $t(18)=4.78$, $p < 0.001$ and $t(18)=4.44$, $p<0.001$ respectively. True positive scores of free search and drilling did not significantly differ.

For the false positive scores, post hoc analysis revealed that free search (M 12.5, SD 7.8) resulted in significantly higher FP rates than drilling (M 7.3, SD 5.6); $t(18)=4.86$, $p < 0.001$, and scanning (M 5.6, SD 4.8); $t(18)=4.44$, $p < 0.001$.

Table 4. Mixed design ANOVA for perceptual performance outcomes

	TP		FP	
	F	p	F	p
Between subjects				
year of residency	9.0	<0.01	1.4	0.26
study group	.12	.74	.04	.85
Within subjects				
search strategy	16.1	<0.001	15.3	<0.001
search strategy * year of residency	1.5	0.85	0.3	0.66
search strategy * study group	28.3	0.07	0.9	0.84
search strategy * year of residency * study group	0.4	0.96	0.9	0.38

Questionnaire

The vast majority of participants (89%) agreed or totally agreed that reviewing chest CT scans in this research study using VQuest was substantially similar to clinical practice. Both instructional videos were found to provide sufficient instruction on using the search strategies by 95% of the participants.

After completing the free search test, all participants reported that they were aware of their search strategy when reading chest CT scans. Seventeen participants specified their search strategy and all were variants of drilling: 53% reported dividing the lungs into two, three or four regions and searched each region individually, the remainder searched lobe by lobe. In addition, seven participants (37%) built in an extra check of the edges of the lungs at the pleura, fissures, mediastinum and hilar structures.

The drilling search strategy was not new for most participants, 74% had heard about it, whereas only 11% was familiar with the scanning search strategy. Participants felt drilling led to a higher detection rate than scanning (M 3.7 vs. 1.8, $z = 3.8$, $p < 0.001$), and found drilling to be more time efficient (M 4.2 vs. 1.8, $z = 3.6$, $p < 0.001$), and easier to use continuously (M 4.3 vs. 1.9, $z = 3.9$, $p < 0.001$). Of all participants, 58% planned to use drilling as their new search strategy, which reportedly was different or somewhat different from the strategy they used before the study. The remainder preferred to keep using their own search strategy.

DISCUSSION

Search strategy instruction had a significant effect on both scroll behavior and perceptual performance. The scanning instruction decreased the number of long scroll movements and scrolling time. The drilling instruction did not alter scroll behavior significantly. However, the majority of participants reported already using some kind of drilling strategy at their free search. Perceptual performance following drilling search instructions outperformed performance following scanning search instructions in terms of true positives. Compared to a free search, the use of a drilling search strategy did not result in more true positive findings. The study confirms our hypothesis that drilling outperforms scanning for detecting lung nodules, although drilling did not improve the baseline performance, probably explained by an a priori search strategy similar to drilling.

The benefit after drilling instruction may have been caused by initial habits of drilling and not by the instruction. Still, there was an improvement after instruction: it did reduce the number of false positive findings, on top of an initial habit of a more productive drilling strategy at baseline. This improvement did not coincide with a decline in true positive rate, while the scanning instruction reduced the false positive rate at the cost of true positives.

The decline in false positive findings after instruction was an unexpected result, as the strategies were anticipated to improve actual detection rather than avoiding overcalls of non-nodule structures. The search strategy instruction may have induced a more focused attention and kept participants from being distracted by non-nodule structures. Another explanation may be that participants were more vigilant to mark any possible abnormality they found at the first test and became less attentive and less willing to spend extra time as the experiment proceeded, resulting in a drop of false positives in the second and third test. False positive findings have been associated with an increase in reading time, as previously found in breast cancer detection[18].

None of the participants reported the use of scanning in their free search, nor did anyone show a scroll pattern that indicated scanning behavior. This differs from the study of Drew et al., reporting both ‘scanners’ and ‘drillers’ among radiologists in a lung nodule detection task, although scanners were in the minority[12]. Interestingly, in that study the scanners were less experienced readers than the drillers, which would raise the expectation of finding even more scanners among the junior residents in our study. Possibly the use of thin-section scans intrinsically evokes a drilling search, as scanning simply takes too much time. Another way to deal with large numbers of thin slices is the use of Maximum Intensity Projections (MIP), which have been shown to improve lung nodule detection accuracy[19].

Our study has several limitations. First, the sample size of nineteen participants is relatively small. However, each participant was asked to review multiple scans with multiple nodules, and indeed the study was powered to detect a difference as evidenced by our results. We do not expect that the main results would substantially change with a larger sample size although subtle differences might have emerged in comparisons that were not statistically significantly different. Second, although we tried to equalize difficulty across the three tests by means of a test blue print, test 2 proved to be more difficult than the other two tests. The nonsignificant effect of the search strategy and study group interaction is probably due to this inequality in test difficulty that may have diminished the degree to which a participant could profit from the search strategy that was applied in the second test. This did not jeopardize our scanning versus drilling comparison, because the scanning and the drilling conditions were equally affected. However, it may have caused an underestimation of the effect of instruction on perceptual performance. Third, our study was specifically targeted to the detection of lung nodules (selected for a number of reasons); this means that we cannot generalize the results to other detection tasks. Finally, we only used scroll behavior and self-reports to estimate participants` search strategies, and did not apply eye tracking to further characterize visual attention. A higher drilling fraction does indicate more drilling, although we cannot further specify how participants divided the lung into regions and in which order.

Future research can be directed toward comparing scanning and drilling search strategies in other detection tasks, and also to exploring which *type* of drilling strategy would be optimal. For example, it would be helpful to compare drilling strategies dividing the lung into different numbers of regions, or to investigate if building in checks of the lung edges is beneficial.

Search strategy instruction can influence scroll behavior and perceptual performance of junior radiology residents completing a lung nodule detection task. In junior trainees, a drilling strategy yields a better perceptual performance than a scanning strategy. Teaching a scanning strategy further decreases the perceptual performance of junior radiology residents below their baseline performance. Teaching a drilling strategy for detecting lung nodules in chest CT scans is therefore preferable.

Appendix 1. List of survey questions with response formats

After Test 1

1) When I am reading chest CT scans, I am aware of the search strategy that I use. (strongly agree/agree/undecided/disagree/strongly disagree)

If you indicated agree or strongly agree, please describe your search strategy:

2) Do you know any (other) search strategies that can be used to read scans or X-rays? (yes/no)

If yes, please list:

After both Test 2 and 3

1) Did you hear about this search strategy before it was described as part of this study? (yes/no)
2) The instructional video provided as part of this study provided sufficient instruction on using this search strategy. (strongly agree/agree/undecided/disagree/strongly disagree)
3) I found this search strategy to be easy to use continuously on a group of chest CT cases. (strongly agree/agree/undecided/disagree/strongly disagree)
4) I found this search strategy to be a time efficient way to look for pulmonary nodules. (strongly agree/agree/undecided/disagree/strongly disagree)
5) I found this search strategy leads to a high rate of detection of pulmonary nodules. (strongly agree/agree/undecided/disagree/strongly disagree)

After completion of all tests

1) In clinical practice I plan to use: (drilling/scanning/undecided/other)

If other, specify:

2) Does the strategy you plan to use differ from the strategy you were using before this study? (exactly the same/somewhat similar/different/not sure)
3) I found reviewing chest CT scans for pulmonary nodules in this research study was substantially similar to the clinical practice of reviewing chest CT scans for pulmonary nodules (i.e. Similar computer interface, similar reading environment etc.). (strongly agree/agree/undecided/disagree/strongly disagree)

If not, why not?

REFERENCES

1. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. J Med Imaging Radiat Oncol 2012; 56: 173-8.
2. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Invest Radiol 1978; 13: 175-81.
3. van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JP, and Ten Cate TJ: How Visual Search Relates to Visual Diagnostic Performance: A Narrative Systematic Review of Eye-Tracking Research in Radiology. Adv Health Sci Educ Theory Pract 2016.
4. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Acad Radiol 1996; 3: 137-44.
5. Cooper L, Gale A, Saada J, Gedela S, Scott H, and Toms A: The Assessment of Stroke Multidimensional Ct and Mr Imaging Using Eye Movement Analysis: Does Modality Preference Enhance Observer Performance? Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment. Edited by Manning DJ and Abbey CK. 2010.
6. Wood G, Knapp KM, Rock B, Cousens C, Roobottom C, and Wilson MR: Visual Expertise in Detecting and Diagnosing Skeletal Fractures. Skeletal Radiol 2013; 42: 165-72.
7. Alzubaidi M, Black Jr JA, Patel A, and Panchanathan S: Conscious Vs. Subconscious Perception, as a Function of Radiological Expertise. In Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2009.
8. Manning D, Ethell S, Donovan T, and Crawford T: How Do Radiologists Do It? The Influence of Experience and Training on Searching for Chest Nodules. Radiography 2006; 12: 134-42.
9. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Academic radiology 1996; 3: 137-44.
10. Kok EM, de Bruin ABH, Robben SGF, and van Merriënboer JJG: Looking in the Same Manner but Seeing It Differently: Bottom-up and Expertise Effects in Radiology. Applied Cognitive Psychology 2012; 26: 854-62.
11. Hu CH, Kundel HL, Nodine CF, Krupinski EA, and Toto LC: Searching for Bone Fractures: A Comparison with Pulmonary Nodule Search. Academic Radiology 1994; 1: 25-32.
12. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13.
13. Baghdady MT, Pharoah MJ, Regehr G, Lam EW, and Woods NN: The Role of Basic Sciences in Diagnostic Oral Radiology. J Dent Educ 2009; 73: 1187-93.
14. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, and Conant EF: How Experience and Training Influence Mammography Expertise. Acad Radiol 1999; 6: 575-85.
15. Nodine CF, Kundel HL, Lauver SC, and Toto LC: Nature of Expertise in Searching Mammograms for Breast Masses. Acad Radiol 1996; 3: 1000-6.
16. Elmore JG, Wells CK, and Howard DH: Does Diagnostic Accuracy in Mammography Depend on Radiologists' Experience? J Womens Health 1998; 7: 443-9.
17. Venjakob A, Marnitz T, Mahler J, Sechelmann S, and Roetting M: Radiologists' Eye Gaze When Reading Cranial Ct Images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment 2012; 8318.
18. Nodine CF, Mello-Thoms C, Kundel HL, and Weinstein SP: Time Course of Perception and Decision Making During Mammographic Interpretation. AJR Am J Roentgenol 2002; 179: 917-23.
19. Valencia R, Denecke T, Lehmkuhl L, Fischbach F, Felix R, and Knollmann F: Value of Axial and Coronal Maximum Intensity Projection (Mip) Images in the Detection of Pulmonary Nodules by Multislice Spiral Ct: Comparison with Axial 1-Mm and 5-Mm Slices. Eur Radiol 2006; 16: 325-32.

Chapter 6

Identifying Error Types in Visual Diagnostic Skill Assessment

This chapter is based on:
Ravesloot, C.J.*, Van der Gijp, A.*, Van der Schaaf, M.F., Huige, J.C.B.M., Ten Cate, Th. J., Vincken, K.L., Mol, C.P., Van Schaik J.P.J. Identifying error types in visual diagnostic skill assessment. *Diagnosis 2017; 4(2).*

*joint first authors

ABSTRACT

Introduction Misinterpretation of medical images is an important source of diagnostic error. Errors can occur in different phases of the diagnostic process. Insight in the error types made by learners is crucial for training and giving effective feedback. Most diagnostic skill tests however penalize diagnostic mistakes without an eye for the diagnostic process and the type of error. A radiology test with stepwise reasoning questions was used to distinguish error types in the visual diagnostic process. We evaluated the additional value of a stepwise question-format, in comparison with only diagnostic questions in radiology tests.

Methods Medical students in a radiology elective (n = 109) took a radiology test including 11 to 13 cases in stepwise question-format: marking an abnormality, describing the abnormality and giving a diagnosis. Errors were coded by two independent researchers as perception, analysis, diagnosis, or undefined. Erroneous cases were further evaluated for the presence of latent errors or partial knowledge. Inter-rater reliabilities and percentages of cases with latent errors and partial knowledge were calculated.

Results The stepwise question-format procedure applied to 1351 cases completed by 109 medical students revealed 828 errors. Mean inter-rater reliability of error type coding was Cohen's $\kappa = .79$. Six-hundred-fifty errors (79%) could be coded as a perception, analysis or diagnosis error. The stepwise question-format revealed latent errors in 9% and partial knowledge in 18% of cases.

Conclusion A stepwise question-format can reliably distinguish error types in the visual diagnostic process, and reveals latent errors and partial knowledge.

INTRODUCTION

Due to increased use of diagnostic imaging[1, 2] and improved accessibility of medical images throughout the hospital, many medical doctors now interpret radiology images on a daily basis. Radiological image interpretation is a complex skill, requiring application and integration of different sorts of knowledge and skills[3], and substantial training and experience to develop[4-6]. Many medical doctors start their careers in the emergency room or on ward, interpreting radiology images of critically ill patients, while their experience is generally limited. Diagnostic errors can have significant consequences in this acute context[7,8]. A major part is due to incorrect image interpretation skill of junior doctors[8-11]. Appropriate radiology training and assessment is imperative to improve radiological image interpretation of learners.

Three main components of radiological image interpretation can be distinguished: perception (detection of a lesion), analysis (characterization of a lesion) and synthesis or diagnosis (synthesizing all information into a conclusion or diagnosis)[3]. Specific knowledge and skills are important in each component[3]. For example, accurate perception requires efficient search strategies[12-14], and for correct analysis the ability to characterize findings of abnormalities is essential[3]. Image interpretation errors can occur in all three components and may be caused by different knowledge and skill gaps related to this component[15-17]. Insight into the nature of errors made by students may help to improve training. Research into diagnostic errors, however, mostly involves eye tracking studies with a focus on perception errors[12, 18-20]. Little is known about errors beyond perception, like analyzing features of the abnormality and generating a differential diagnosis[3].

Radiology performance tests that are able to trace incorrect image interpretation back to one of the components could be useful to direct feedback, tailor training and stimulate a specific learning process[21]. Radiology tests asking trainees to interpret an image by only asking for a diagnosis might leave perception or analysis errors unattended (latent error). Collateral information or knowledge, such as clinical information or prevalence of the disease, may lead to the correct answer. On the other hand, when the diagnosis is incorrect, perception and analysis might still be correct (partial knowledge).

In this study we evaluated the value of a stepwise reasoning question format in image interpretation assessment. Clinical cases were provided followed by questions regarding perception, analysis and diagnostic skills[3] in a stepwise format. Hypothetically, a set of questions in a stepwise format could be able to unravel the process and provide meaningful performance information such as partial knowledge, latent errors or insights in specific component error types.

Our research questions are 1) Can errors during one of the components of image interpretation (perception, analysis, synthesis) reliably be identified using the stepwise question-format in a radiology performance test? 2) Does the stepwise question approach provide additional performance information compared to single diagnostic questions?

METHODS

Study Design

We developed a radiology test with stepwise questions, assessing perception, analysis and synthesis skills in image interpretation. The test was administered to one hundred-twelve medical students in a radiology elective at the University Medical Center Utrecht in The Netherlands from March 2012 to February 2014. The reliability of the stepwise question approach for identifying component errors was estimated with an interrater comparison of scores by two raters. Additional performance information of the stepwise question-format was evaluated by comparing answers on the stepwise questions with the single diagnostic question. The study was approved by the Ethical Review Board of the Netherlands Association of Medical Education (NVMO) and all participants gave informed consent.

Participants

One hundred-twelve medical students took a digital radiology test as a mandatory exam following their radiology clerkship. One hundred-nine of them gave informed consent to analyze their test results and were included in the study. Participants were all fourth to sixth-year medical students. All participants had followed an intensive longitudinal radiology course on radiology anatomy and basic radiology knowledge and image interpretation skill on prevalent illnesses as a mandatory part of the undergraduate program. Further, they followed a six-week radiology elective clinical rotation on knowledge and image interpretation skill of acute and sub-acute diseases in the subareas musculoskeletal, abdominal, chest, and neuroradiology. In addition to the clinical rotation they attended three case-based lessons addressing the radiology subareas.

Instruments

Radiology Test

A panel of five radiology education experts created a question bank with image-based cases. Images of acute and subacute diseases were collected and test questions were constructed in the stepwise question-format. The panel also constructed an answer sheet in consensus. A case included one clinical scenario with concomitant imaging and was composed of multiple questions testing perceptual, analytical and diagnostic skills, or knowledge related to the pathology visible on the image (see for an example figure 1). To assess perception participants were asked to mark the abnormality in the image. Analytical and diagnostic skills were assessed by questions in multiple choice or long menu format. A case did not always include all three components. For some obvious abnormalities, for example a large brain hemorrhage, it was not asked to mark them first. The distribution of the component questions among all completed cases was 40% perception, 22% analysis and 37% diagnosis.

Three versions were composed based on a test matrix: 1A, 2A and 3A. Each test contained eleven to thirteen cases in stepwise question-format, including five volumetric CT image cases, five 2D CT image cases and three X-ray image cases, except for test 1A , that contained only one x-ray case in the stepwise question-format. All images were abnormal. Each CT image case was constructed in a volumetric (complete CT-scan with stack viewing) and a 2D format (presenting only selected CT-slices). Of each test, a parallel version was constructed in which all volumetric questions of version A were in 2D format and vice versa: version 1B, 2B and 3B. A scoring model was developed by the expert panel before the first test was administered. The different versions 1A, 2A and 3A were pilot tested by four, one and two medical students respectively and, if necessary, test questions were rephrased and the scoring model was adjusted.

The tests were administered with VQuest, a digital testing program that allows for radiology image viewing and manipulation, including stack viewing of volumetric CT images in a way that is representative of clinical practice[22, 23]. Participants were allowed to change image contrast and scroll through volumetric images in three viewing directions.

Coding Scheme

1. Error types: all errors were coded as related to one of the three components of image interpretation, perception, analysis, and synthesis. If an error could not be identified as a component error it was coded as undefined.
2. Additional performance information: at case level we compared the answers on the stepwise question-format questions with the answer on the final diagnostic question (the single diagnostic question, figure 1, sub question c) and classified if the stepwise questions provided additional performance information compared to the final diagnostic question alone. Additional performance information was considered to be present when the performance information could only be derived from the answers given on the stepwise questions, and could not be derived from the answers given on the final diagnostic question alone. Each case including at least one error was classified into three categories: 1) Cases with latent errors, being cases in which the student gave the correct diagnosis, while there was either a perception error or analysis error. 2) Cases in which partial knowledge could be valued due to the stepwise question-format. For example, a student marked the correct abnormality, but failed to give the correct diagnosis. 3) Cases in which the stepwise question approach did not reveal latent errors or partial knowledge.

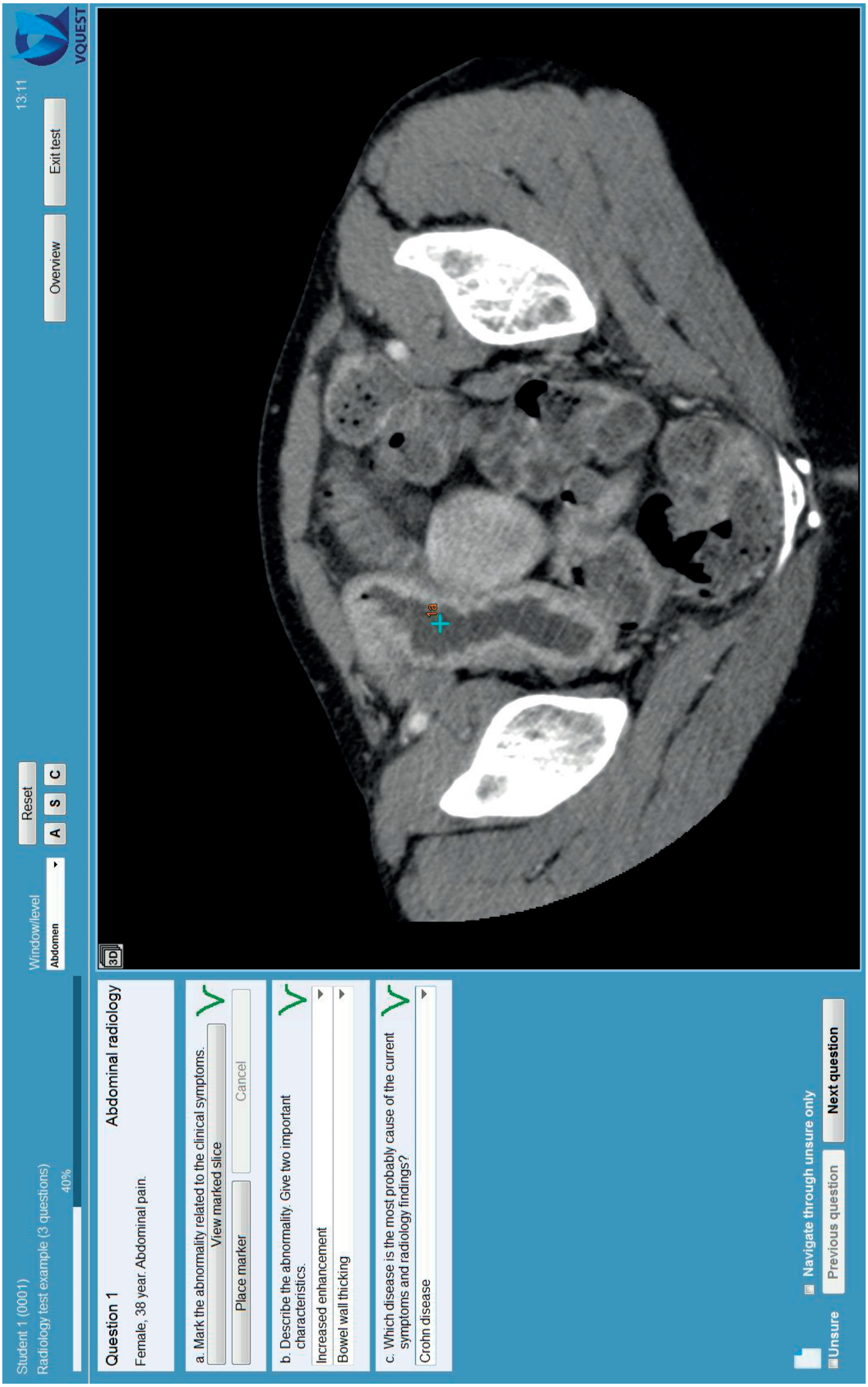

Figure 1. An example of the stepwise question-format with questions that aim to test (a) perception, (b) analysis, and (c) diagnosis skill.

Procedure

The radiology tests were administered in groups of one to seven students. Answers of all participants (on question level) were scored by two researchers (A.G. and C.R., both radiology residents at the time of the study) and discrepancies were solved after reaching consensus. More than one error could be identified within a case. However, if the perception question was incorrectly answered, all other related analytical and diagnostic questions were not marked as separate errors. All identified errors were independently coded and categorized by two researchers (A.G and C.R.). Inter-rater agreement was calculated (with Cohen's κ) per test version after coding all responses. Discrepancies were discussed by the raters until consensus was reached.

Data Analysis

Inter-rater agreement for coding errors by two raters was calculated by Cohen`s κ per test version to estimate the reliability of error identification.

The percentage of cases with latent errors and partial knowledge were calculated (see Figure 3).

RESULTS

Baseline Characteristics

In total 16 to 21 students participated per test version which consisted of eleven to thirteen cases. Estimated reliabilities of the tests ranged from low to sufficient on case level (Cronbach's alpha .28 - .69), and from weak to good (Cronbach's alpha .57 to .84) on question level. One hundred-nine participants completed in total 1351 cases. In 715 completed cases no errors were identified. In the remaining 636 cases 828 errors were found. The average number of errors per case was 0.6 (SD 0.7). The median test score on the stepwise questions was 77.8% with an interquartile range of 15.4%.

Inter-Rater Agreement

Cohen`s κ for classification of error types and additional performance information are given in Table 1.

Table 1. Inter-rater agreement for coding error types and additional performance information by two raters, estimated by Cohen`s κ per test version

	Error types		Additional performance information	
	Version A	**Version B**	**Version A**	**Version B**
Test 1	0.73	0.86	0.66	0.81
Test 2	0.77	0.88	0.84	0.95
Test 3	0.68	0.90	0.63	0.81

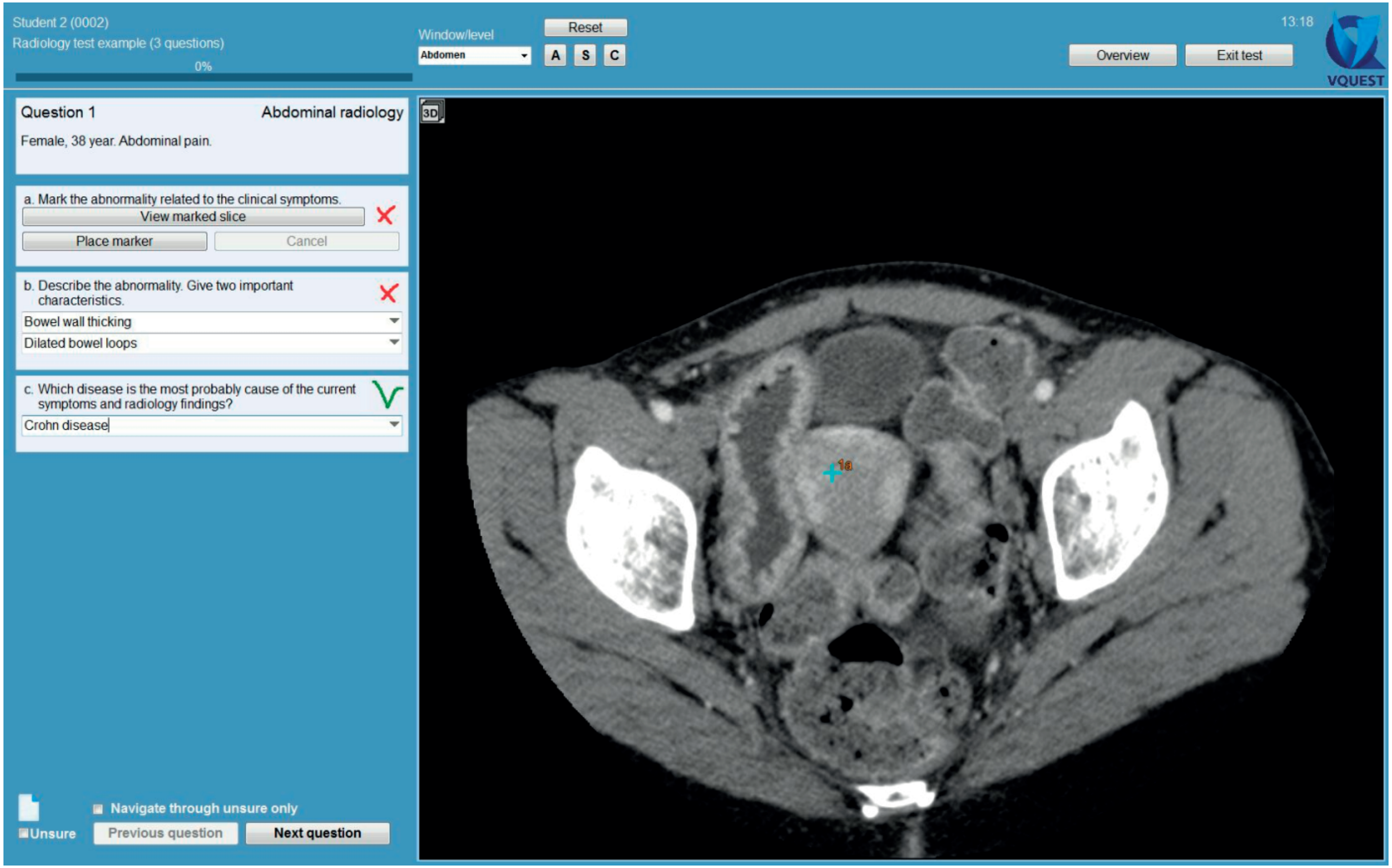

Figure 2. Examples of additional performance information of the stepwise question-format by (a) showing latent errors or (b) partial knowledge

a. Latent error. This student gave the correct diagnosis "Crohn disease", but marked a normal structure (uterus) as the abnormality (perception error).

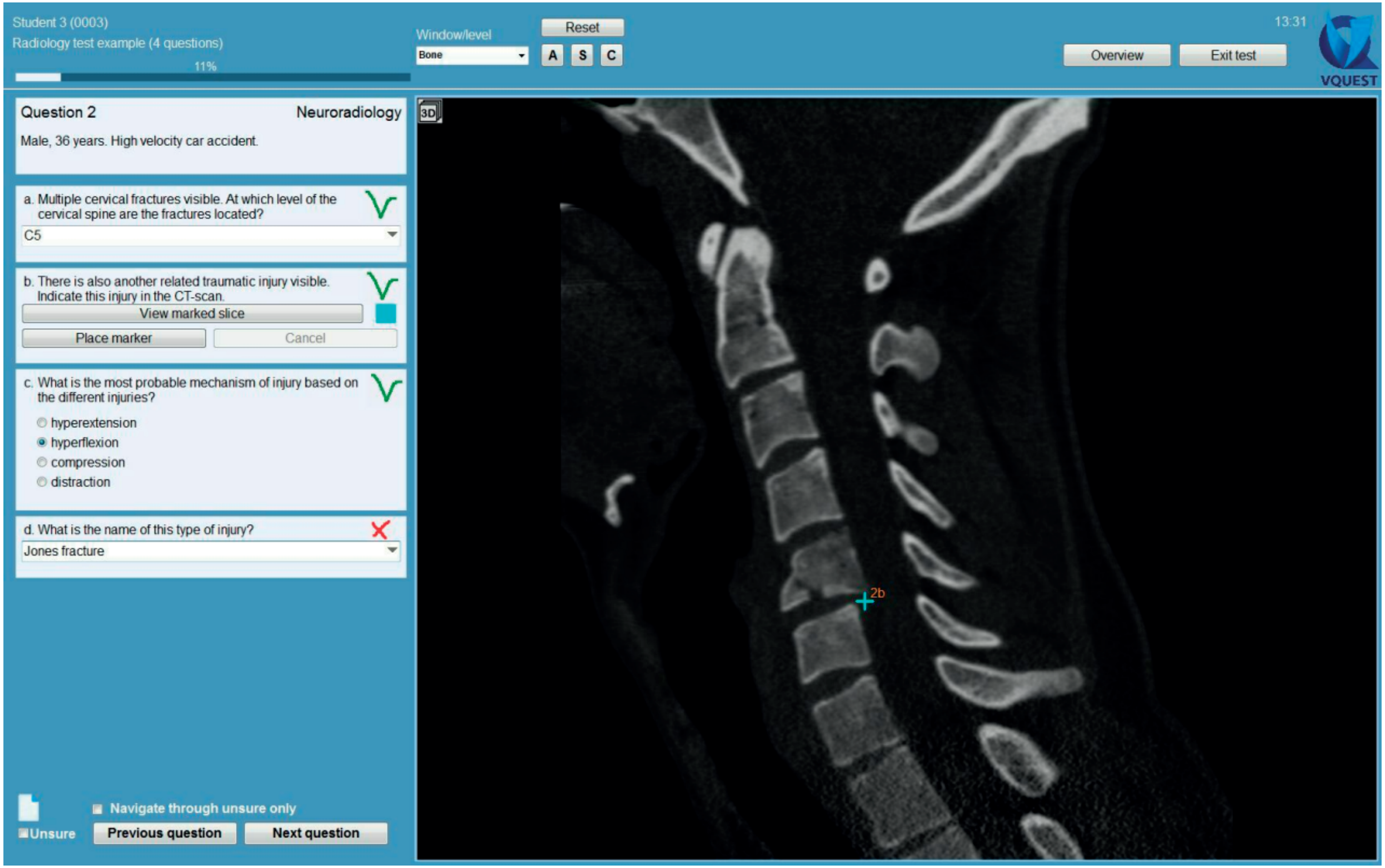

Figure 2. Examples of additional performance information of the stepwise question-format by (a) showing latent errors or (b) partial knowledge

b. Partial knowledge. This student gave the incorrect diagnosis: "Jones fracture" instead of "teardrop fracture" (diagnosis error), however the answers to the other questions, addressing perception and analysis, revealed that the student interpreted the image otherwise correctly.

Error Types

In total 651 errors could be identified as a component error (79%). The remaining 21% was left undefined. The distribution of the component errors was 38%-20%-41% and is comparable with the distribution of the questions per component in the test.

Additional Performance Information

In 27% of the cases (367 of 1351 cases) the stepwise question-format revealed latent errors or partial knowledge, that could not be derived from the final diagnostic question alone. In 18% of the cases (243 cases) participants were evaluated as being a more competent interpreter based on the stepwise approach, because the stepwise questions revealed partial knowledge, even though they provided a wrong diagnosis. For example, a participant gave an incorrect diagnosis (a Jones fracture instead of a tear drop fracture), but correctly indicated and analyzed the lesion, as shown by the answers to the preceding questions (see figure 2b). In 9% of the cases (125 cases) the diagnosis was correct, but the stepwise questions showed an error in either perception (60 cases) or analysis (65 cases) of the abnormality, signifying a latent error. For example, a participant gave the correct diagnosis Crohn disease, but failed to correctly mark the abnormal bowel loop (see figure 2a). Another participant correctly diagnosed a patient with pneumonia, however did not mark the pneumonia, but indicated normal lung tissue as abnormal. The case text "dyspnea and fever" and the participant's knowledge about symptoms of pneumonia possibly resulted in the correct answer. These participants' image interpretation performances were valued as being lower, due to the latent errors revealed by the stepwise question-format. As shown in figure 3 in 20% of the cases (268 cases) the stepwise question-format revealed no partial knowledge or latent errors.

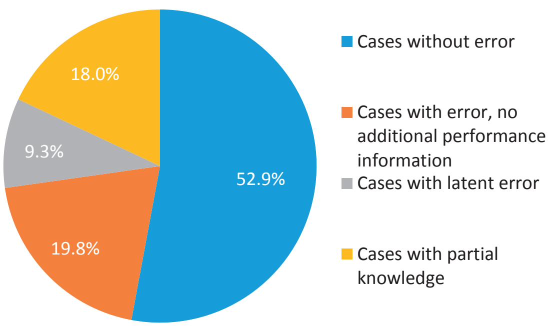

Figure 3. Percentage of cases with additional performance information derived from the stepwise question-format

DISCUSSION

The stepwise question-format reliably distinguished errors related to the image interpretation process of medical students in a radiology elective. Almost 80% of the errors could be related to an error in one of the components of image interpretation.

The percentages of error types should be interpreted with caution and should not be interpreted as actual prevalence of errors, because they are partly test-driven due to variation in the stepwise questions. The distribution of the different component errors (39%-20%-41%) was similar to the distribution of the stepwise questions on perception, analysis, and diagnosis (40%-22%-38%), so all errors seemed to occur almost with equal frequency in proportion to the questions that provoke the errors. However, this result cannot be compared with the relatively high proportion of perceptual errors in clinical practice reported in the literature[15, 17], because the current test only included abnormal cases. Therefore students were biased in the sense that they knew that an abnormality was present, which differs from clinical practice in which many cases are normal.

In more than a quarter of all cases the stepwise question-format revealed additional performance information of the learner that could not be derived from the final diagnostic question alone. In almost two-thirds of these cases partial knowledge was uncovered, while in one-third of the cases flaws in perception or analysis appeared which would have stayed unnoticed if only a diagnosis was asked.

These results indicate that the stepwise question-format provides the teacher with additional performance information compared to only diagnostic questions formats, because this method can reliably identify in which component of the image interpretation process the student succeeded or failed. Stepwise questions give a more detailed and probably more accurate representation of the performance of students and uncover latent errors in the image interpretation process.

The additional information can be useful for both teachers and students. Teachers will get more insight into the errors of students and their actual performance level which may be used to inform individual feedback to learners. Students may use the information to detect their own strengths and weaknesses, and to monitor their development in specific image interpretation skills.

Some limitations of the study should be addressed. Image interpretation is a complex process and the distinction between error types is rather artificial. Perception, analysis and synthesis processes are highly interrelated and constantly alternate within the image interpretation process[24]. The stepwise question approach directed the diagnostic reasoning process into a stepwise reasoning process, broken down into three components, and therefore allowed for distinguishing three components of the image interpretation process. Breaking down errors in these three components should only be recognized as one of many ways of specifying errors in diagnostic radiology for an improved understanding of decision-making.

A drawback of the stepwise question method is the increased time needed to complete the tests. Single questions would allow for a larger number of cases in the same testing time, which could increase reliability. The reliability of the current test ranged from low to sufficient on case level, probably because of the low number of cases (11 -13) in each test version. It is important to take the purpose of the test into account, when deciding for either the stepwise question-format or only diagnostic questions. The information revealed by the stepwise question-format approach can be valuable for feedback in formative testing, while a large number of questions can be necessary to increase reliability in summative tests.

In this study, we only applied the stepwise question method to medical students and we cannot generalize our results to other levels of expertise. For example, the error rate and distribution of error types are probably different due to other knowledge and experience levels. Testing the method in residents or radiologists is subject to further research.

We conclude that the stepwise question approach succeeded in identifying specific error types in the image interpretation process. Besides, it revealed partial knowledge and latent errors. Both can be used to tailor image interpretation training. Specifying error types may be advantageous for giving effective feedback, because a specific task evaluation is one of the characteristics of high-quality feedback[25, 26].

REFERENCES

1. Bhargavan M, and Sunshine JH: Workload of Radiologists in the United States in 2002-2003 and Trends since 1991-1992. Radiology 2005; 236: 920-31.
2. Bhargavan M, Kaye AH, Forman HP, and Sunshine JH: Workload of Radiologists in United States in 2006-2007 and Trends since 1991-1992. Radiology 2009; 252: 458-67.
3. van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JC, Ravesloot CJ, van Schaik JP, and Ten Cate TJ: Interpretation of Radiological Images: Towards a Framework of Knowledge and Skills. Adv Health Sci Educ Theory Pract 2014.
4. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, and Conant EF: How Experience and Training Influence Mammography Expertise. Acad Radiol 1999; 6: 575-85.
5. Norman GR, Coblentz CL, Brooks LR, and Babcook CJ: Expertise in Visual Diagnosis: A Review of the Literature. Acad Med 1992; 67: S78-83.
6. Taylor PM: A Review of Research into the Development of Radiologic Expertise: Implications for Computer-Based Training. Acad Radiol 2007; 14: 1252-63.
7. Gruen RL, Jurkovich GJ, McIntyre LK, Foy HM, and Maier RV: Patterns of Errors Contributing to Trauma Mortality: Lessons Learned from 2,594 Deaths. Ann Surg 2006; 244: 371-80.
8. Wechsler RJ, Spettell CM, Kurtz AB, Lev-Toaff AS, Halpern EJ, Nazarian LN, Feld RI, Needleman L, and Alexander AA: Effects of Training and Experience in Interpretation of Emergency Body Ct Scans. Radiology 1996; 199: 717-20.
9. Guly HR: Diagnostic Errors in an Accident and Emergency Department. Emerg Med J 2001; 18: 263-9.
10. Rhea JT, Potsaid MS, and DeLuca SA: Errors of Interpretation as Elicited by a Quality Audit of an Emergency Radiology Facility. Radiology 1979; 132: 277-80.
11. Gwynne A, Barber P, and Tavener F: A Review of 105 Negligence Claims against Accident and Emergency Departments. J Accid Emerg Med 1997; 14: 243-5.
12. Hu CH, Kundel HL, Nodine CF, Krupinski EA, and Toto LC: Searching for Bone Fractures: A Comparison with Pulmonary Nodule Search. Acad Radiol 1994; 1: 25-32.
13. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Acad Radiol 1996; 3: 137-44.
14. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13.
15. Renfrew DL, Franken Jr EA, Berbaum KS, Weigelt FH, and Abu-Yousef MM: Error in Radiology: Classification and Lessons in 182 Cases Presented at a Problem Case Conference. Radiology 1992; 183: 145-50.
16. Pinto A, Acampora C, Pinto F, Kourdioukova E, Romano L, and Verstraete K: Learning from Diagnostic Errors: A Good Way to Improve Education in Radiology. Eur J Radiol 2011; 78: 372-6.
17. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. Journal of Medical Imaging and Radiation Oncology 2012; 56: 173-78.
18. Donovan T, and Litchfield D: Looking for Cancer: Expertise Related Differences in Searching and Decision Making. Applied Cognitive Psychology 2013; 27: 43-49.
19. Rubin GD, Roos JE, Tall M, Harrawood B, Bag S, Ly DL, Seaman DM, Hurwitz LM, Napel S, and Roy Choudhury K: Characterizing Search, Recognition, and Decision in the Detection of Lung Nodules on Ct Scans: Elucidation with Eye Tracking. Radiology 2015; 274: 276-86.
20. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Investigative Radiology 1978; 13: 175-81.
21. Pecaric M, Boutis K, Beckstead J, and Pusic M: A Big Data and Learning Analytics Approach to Process-Level Feedback in Cognitive Simulations. Acad Med 2017; 92: 175-84.

22. Ravesloot CJ, van der Schaaf MF, van Schaik JP, ten Cate OT, van der Gijp A, Mol CP, and Vincken KL: Volumetric Ct-Images Improve Testing of Radiological Image Interpretation Skills. Eur J Radiol 2015; 84: 856-61.
23. Ravesloot CJ, van der Gijp A, van der Schaaf MF, Huige JC, Vincken KL, Mol CP, Bleys RL, ten Cate OT, and van Schaik JP: Support for External Validity of Radiological Anatomy Tests Using Volumetric Images. Acad Radiol 2015; 22: 640-5.
24. Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interactions of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
25. Sadler DR: Beyond Feedback: Developing Student Capability in Complex Appraisal. Assessment and Evaluation in Higher Education 2010; 35: 535-50.
26. Black P, and William D: Assessment and Classroom Learning. Assessment in education: principles, policy and practice 1998; 5.

Chapter 7

Increasing Authenticity of Simulation-Based Assessment in Diagnostic Radiology

This chapter is based on:
Van der Gijp, A., Ravesloot, C.J., Tipker, C.A., De Crom, K., Rutgers, D.R., Van der Schaaf, M.F., Van der Schaaf, I.C., Mol, C.P., Vincken, K.L., Ten Cate, Th.J., Maas, M., Van Schaik, J.P.J. Increasing authenticity of simulation-based assessment in diagnostic radiology. *Revisions submitted.*

ABSTRACT

Objective Clinical reasoning in diagnostic imaging professions is a complex skill that requires processing of visual information and image manipulation skills. We developed a digital simulation-based test method to increase the authenticity of image interpretation skill assessment.

Methods A digital application, allowing volumetric image viewing and manipulation, was used for three test administrations of the national Dutch Radiology Progress Test (DRPT) for residents. This study describes the implementation and development process in three phases. To assess the authenticity of the digital tests, perceived image quality and correspondence to clinical practice was evaluated and compared with previous paper-based tests. Quantitative and qualitative evaluation results were used to improve subsequent tests.

Results The authenticity of the first digital test was not rated higher than the paper-based tests. Test characteristics and environmental conditions, such as image manipulation options and ambient lighting, were optimized based on participants' comments. After adjustments in the third digital test, participants favored the image quality and clinical correspondence of the digital image questions over paper-based image questions.

Conclusion Digital simulations can increase authenticity of diagnostic radiology assessments compared to paper-based testing. However, authenticity does not necessarily increase with higher fidelity. It can be challenging to simulate the image interpretation task of clinical practice in a large-scale assessment setting, due to technological limitations. Optimizing image manipulation options, the level of ambient light, time limits and question types can help improving the authenticity of simulation-based radiology assessments.

INTRODUCTION

Complex clinical skills are challenging to assess. The closest reflection of actual clinical behavior is obtained by workplace based assessments in which trainees are judged for their performance at clinical tasks, without standardization of patients or settings[1]. However, the inherent lack of standardization and the difficulty to attain high levels of inter-rater agreement can diminish its reliability[2]. Besides, workplace based assessments generally require direct observation and much time and effort of faculty members. Simulation-based assessment aims to test participants' clinical performance in a standardized setting. Most medical simulations use virtual patients[3, 4], such as mannequins, standardized patients or laparoscopy simulators. These patient models attempt to approach the experience with real patients. In visual diagnostic domains, such as radiology, the imaging data of real patients can be used to simulate the clinical task of image interpretation.

Although the use of real patient data adds to the fidelity of the simulation, it does not guarantee a high level of authenticity[5]. Authenticity refers to the degree to which an assessment resembles the task in professional practice[6]. To simulate a clinical task in a credible way, a certain level of authenticity is needed. Current radiological practice -for a great part- involves reading volumetric images. Volumetric images are sets of successive cross-sections of a human body part. These cross-sections are usually MR (Magnetic Resonance) or CT(Computed Tomography) images. Radiologists scroll through the set of images to detect lesions and diagnose diseases. Advanced image interaction or manipulation tools can be used, such as scrolling through images in any direction and adjusting contrast settings[7-9]. Interpreting volumetric images requires different cognitive skills than interpreting 2D images[10] and requires the processing of large amounts of visual data. These skills cannot be captured in a paper-based test. A computer-based test is needed to include human-computer interactions and to reach an acceptable level of authenticity. However, a higher resemblance of clinical practice does not necessarily mean that examinees experience this as such, because the authenticity of a test is partly in the eyes of the beholder[11].

To evoke true radiological diagnostic reasoning, not only the viewing mode but also the cognitive characteristics of the test should align with clinical practice. In clinical reasoning literature, many question types have been developed and investigated in an attempt to test clinical reasoning[12]. The diagnostic process in visual domains primarily focuses on hypothesis generation, based on image characteristics and available clinical information. Image-based questions, accompanied with limited clinical information, and a response format that requires the active generation of a diagnosis aligns best with this clinical task. The perceptual component may be captured with question types that test detection skills, e.g. by asking for marking abnormalities or anatomical structures.

Many initiatives to simulate the radiological image interpretation task have been reported in radiology education literature. Most simulations are used for e-learning purposes[13, 14]. The image interaction possibilities in these simulations are usually absent or limited[14], and learners tend to encourage possibilities for image interaction[15-17]. Some studies even suggest that introducing teaching material with image interaction possibilities has a positive effect on learning outcomes[18, 19], but the level of evidence is low. The use of simulations for radiology assessments is less widely reported, and image interaction possibilities are lacking or limited[20, 21].

In a previous study, we found that the introduction of volumetric images has the potential to improve the validity of radiology anatomy tests in medical students[22]. According to the medical students, testing with volumetric images reflected clinical practice better than testing with 2D images[22]. This population had no experience in clinical practice and the questions only involved normal anatomy, whereas clinical radiology involves recognition and interpretation of pathological images. These results should therefore be verified in a population that has a good understanding of radiology practice and with questions that aim to test image interpretation in pathological cases as well.

The purpose of this study was threefold:

1. To develop and implement a digital simulation-based assessment method for monitoring image interpretation skills of radiology trainees.
2. To describe the methods, challenges and lessons learned from this development and implementation process.
3. To evaluate the authenticity of the digital test in comparison with former paper-based assessments.

METHODS

Setting

In the Netherlands, radiology residency involves five years of training in academic and non-academic hospitals. The DRPT is a semi-annual mandatory test for residents. The DRPT aims to test development of radiological knowledge: all residents, regardless of their level of experience, take the same end-of-training level test. The DRPT has a formative purpose, i.e. to provide feedback to trainees, to reflect their progress and guide self-directed learning, rather than to yield a summative score[23]. The questions are constructed by expert radiologists of the examination committee of the Radiological Society of the Netherlands. The test has been described in more detail in a previous study[24]. After ten years of paper-based testing, the DRPT was transformed into a digital format in 2013. In April 2013 a pilot test was conducted among 383 participants. The aim of this pilot was to check the feasibility of the test procedure and its technical performance. The implementation and development process of the following three digital radiology progress tests (signified below as Digital Test 1, 2 and 3), administered in 2013 and 2014, is described and evaluated in the current study. To compare the digital test with previous paper-based testing, the three most recent paper-based tests before the transition (Paper-based Test 1, 2 and 3, administered in 2011 and 2012), were used. The current study focuses on the image-based questions assessing image interpretation skills.

Study Design: Three-Phase Developmental Process Evaluation

We longitudinally describe the implementation and development process of the three digital tests. In the method section of phase I, the initial implementation of DT 1 (Digital Test 1) will be outlined. The method section of phase II and III will focus on the changes that were made in DT 2 and DT 3 based on the results of the previous digital tests. Both quantitative and qualitative results informed decisions for further improvement of the following digital tests. In the development process we focused on improving the authenticity of the image interpretation task. Radiology expertise literature distinguishes visual and cognitive components of image interpretation[25-28]. Therefore, we distinguish viewing task and cognitive task characteristics to evaluate task authenticity. Viewing task characteristics include aspects of the task that are related to the images, i.e. how images are displayed and to what extent they can be manipulated. Cognitive task characteristics involve aspects of the task that are related to the thinking process of the trainee, i.e. processing the visual information, and diagnostic reasoning. After each digital test, we evaluated the participants' perceptions about the authenticity of the digital image questions compared to former paper-based questions, derived from quantitative and qualitative data from questionnaires. Besides, we compared the reliability of the digital image question subtests with former paper-based versions.

Participants

The three digital progress tests were taken by 356 (Digital Test 1, DT 1), 367 (DT 2) and 349 (DT 3) Dutch radiology residents. The mean duration of radiology training of the participants at the moment of testing was 2.4, 2.5 and 2.4 years respectively. The paper-based tests were taken by 357 (Paper-based Test 1, PT 1), 367 (PT 2) and 354 (PT 3) residents, who had completed an average duration of radiology training of 2.3 years across all three tests.

Questionnaire

Within one week after the digital tests, participants received an invitation to answer an online questionnaire with a reminder after one week. Questions concerned the perceived correspondence to clinical practice and image quality of the digital and paper-based image questions. Questions concerning the paper-based image questions were not filled out by respondents who never completed a paper-based version of the test. Response format was on a five-point Likert scale, ranging from "insufficient" to "good". Participants were asked to provide suggestions for improvement in four open comment sections of each questionnaire. Open comments regarding the image questions were used for qualitative analysis to explain or complement the quantitative findings. The questionnaire also included questions about the logistics of the test administration, the user-friendliness of the test application and the test environment, the answers of which were to be used to guide the improvement of subsequent tests.

Response rates were 52%, 46% and 43% after DT 1, 2 and 3 respectively. The distribution of participants over the training years across all three questionnaires together was: year 1: 20-24%; year 2: 20-21%; year 3: 15-25%; year 4: 20-27% and year 5: 10-16%.

Statistical Analysis

For comparing the reliability of the digital image questions with paper-based image questions, Cronbach`s alpha was calculated for each subtest level, after removal of flawed questions (3-5%) determined so by the examination committee and guided by item analysis. To compare reliabilities of the different subsets of image questions, Spearman Brown formula was applied to correct for test length differences[29]. Item-total correlations of the image questions were calculated.

After assumption checks, paired t-tests and Wilcoxon signed-rank tests were conducted to compare survey ratings concerning digital testing with those concerning paper-based testing. For qualitative analysis for the purpose of this study, comments concerning the image-based questions were categorized in perceptions regarding viewing task characteristics and regarding cognitive task characteristics. Themes concerning these characteristics were identified and reported when three or more comments were related to a theme. Comments given twice or more by one participant were counted only once. Comments concerning the non-image based questions and the organization of the test were analyzed separately, and used for the improvement of subsequent tests.

Ethical Approval

The Ethical Review Board of the Netherlands Association for Medical Education approved the study (ERB number 206).

Phase 1: Development and evaluation of Digital Test 1

Test Format

The first digital test contained 200 questions, equal to the paper-based test versions. The number of image-based questions was 36 and comparable to previous paper-based tests. In the first digital test, 56% of the image-based questions contained volumetric images. In the scoring model of the digital tests, each correct answer to a question yielded one point. The scoring model of the paper-based tests was based on formula scoring, because the questions included a don't know option. In formula scoring, scores are calculated by subtracting the number of wrong answers from the number of correct answers to correct for guessing. The don`t know option was removed prior to the digital test implementation, based on a previous experiment about the effect of the don`t know option[30].

Test Environment

A digital test environment, VQuest (http://www.vquest.eu or http://vquest.bluefountain.nl/en/), was used to administer the digital DRPT. The test application was developed at University Medical Center Utrecht and was used in previous studies to improve radiology tests for medical students[22, 31]. It allows for volume dataset viewing and image manipulation. Participants can navigate through volumetric images in different viewing directions. The program also allows for zooming in and out and adjusting image contrast. Examples of a volumetric image question are shown in Figure 1 and 2. A video of the test application is available in the supplementary of previously published work[22].

Simulating Viewing Task Characteristics

We listed the viewing characteristics of the task in clinical practice and the paper-based test. Within the possibilities and restrictions of the available hardware and software, we implemented viewing characteristics in the digital test that were closest to clinical practice.

Because of screen size and resolution restraints, the maximum number of images displayed at once was restricted. Even though ample image manipulation options were available in the test environment, not all options were included to their full extent. We anticipated that too many options could overwhelm participants and could increase reading time. Many manipulation options were therefore only made available if considered to have added value to a particular question. The viewing task characteristics of the first digital test, the paper-based test and clinical practice are described in Table 1.

Simulating Cognitive Task Characteristics

To improve simulation of the diagnostic reasoning task, hotspot and multiple choice questions were introduced, in addition to the true/false questions. The hotspot question aimed to test perception skills [9], by asking participants to place a marker in an abnormality or an anatomical structure. The multiple choice question could be used to test analysis or synthesis skills[9], for example testing the ability to diagnose, by listing a number of possible diseases.

RESULTS

Reliability of the digital image question subtest of DT 1 was comparable to the reliability of the paper-based image questions when corrected to a 60 image-based questions test with Spearman Brown formula (Table 2). Average item-total correlation (rit) values for digital image questions per question type are given in Table 3.

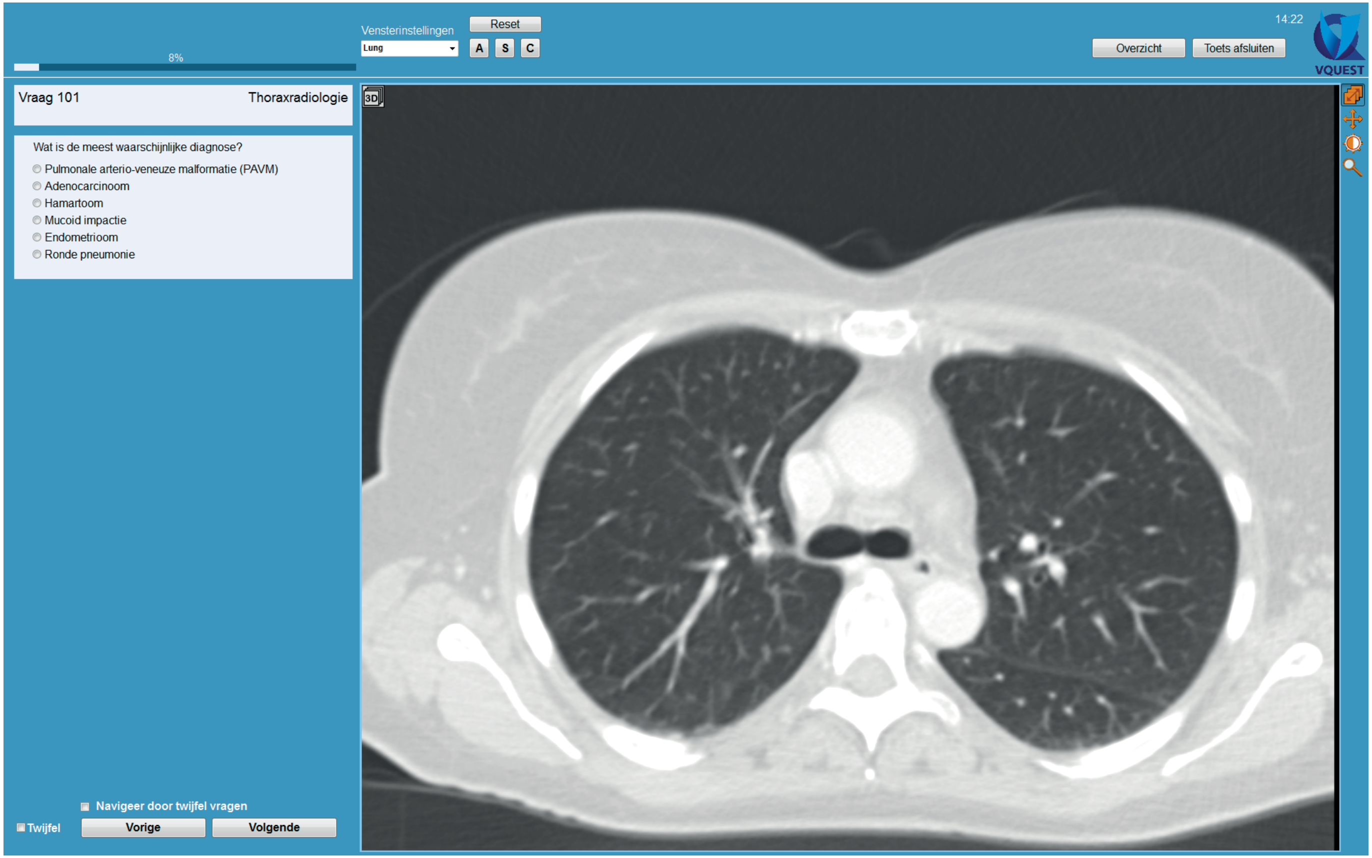

Figure 1. Example of a volumetric image question in VQuest

a. This is a chest CT scan with an abnormality. Participants are asked to choose the correct diagnosis. The abnormality is not visible on this cross-section, so participants have to scroll through the set of images to detect and analyze the abnormality.

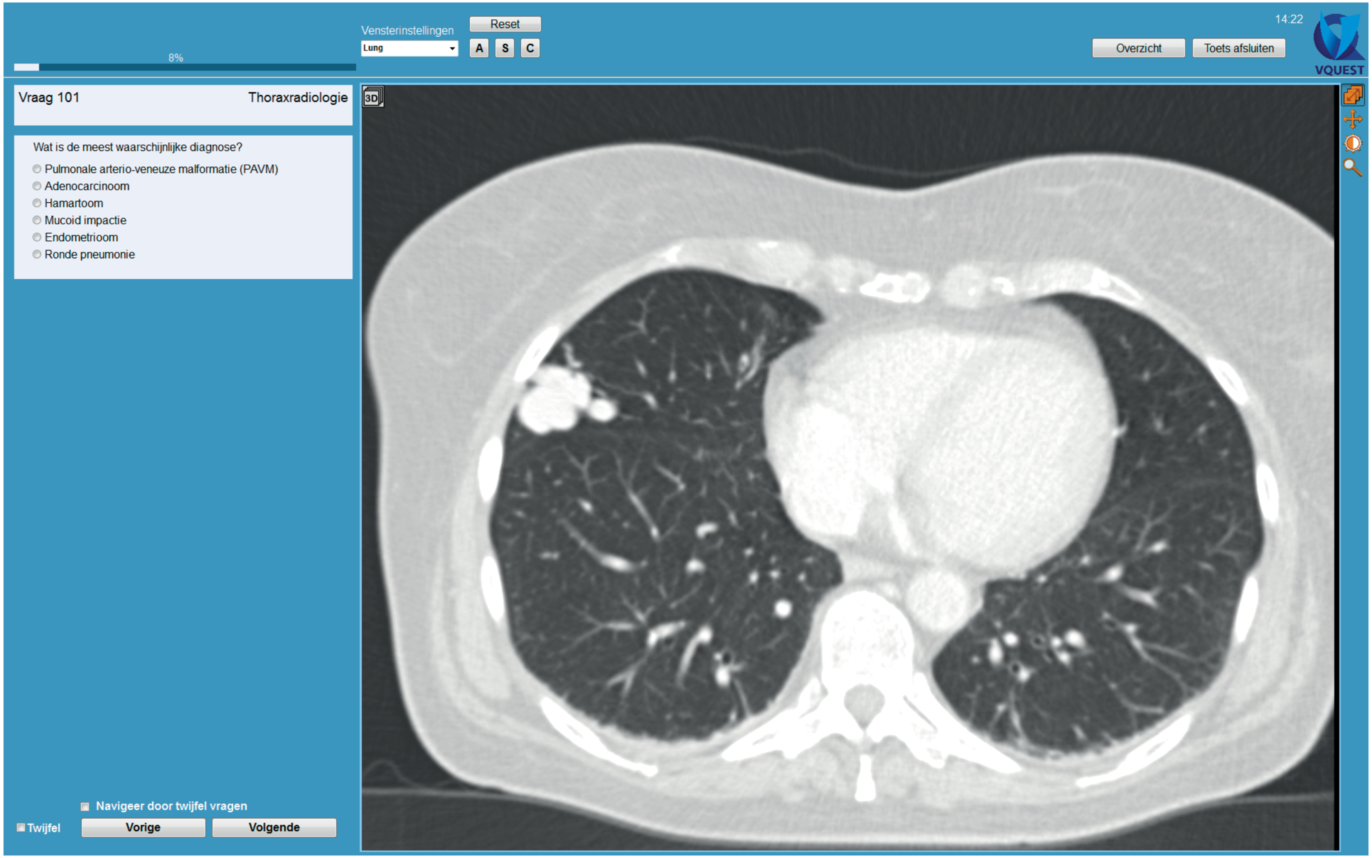

Figure 1. Example of a volumetric image question in VQuest

b. The same question as in Figure 1a, showing a different cross-section of the CT scan. On this cross-section, the abnormality is visible.

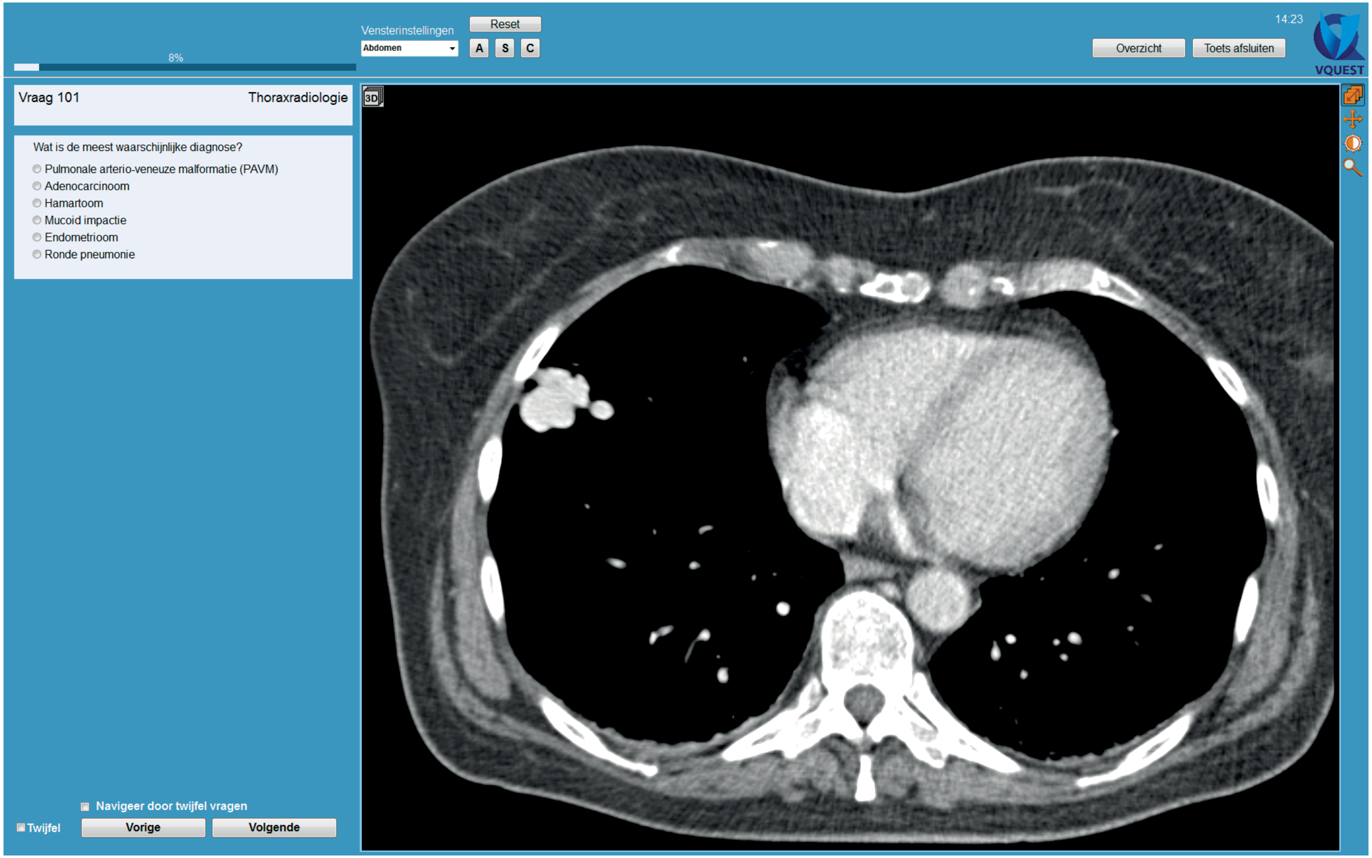

Figure 1. Example of a volumetric image question in VQuest

c. The same question as in Figure 1a and b, showing the same cross-section as in Figure 1b, but with a different contrast setting. Changing contrast can facilitate image interpretation, for example by showing the tissue characteristics of the abnormality.

Table 1. Viewing task characteristics in clinical practice, the digital image questions of DT 1 and the paper-based image questions

Viewing task characteristics	Clinical practice	DT 1	Paper-based tests
Image display			
Screen size	Typically 20-30 inch	15.6 inch	NA
Screen resolution	Typically 1600 x 1200 to 3280 x 2048	1366 x 768	NA
Display more than one image at once	Yes, multiple	Yes, maximum of four images	Yes, multiple
Comparing with previous images	On demand	Only when provided in the question	Only when provided in the question
Image manipulation			
Scrolling back and forth	Yes	Yes	No
Changing contrast setting	Yes, presets or free in any setting	Yes, presets or free in any setting	No
Changing viewing direction	Yes, presets or free in any direction	Presets, when considered to have added value	No
Zooming in and out	Yes	No	No
Making advanced reconstructions	Yes	Yes, when considered to have added value	No

DT = Digital Test, NA = Not Applicable

Table 2. Reliabilities of digital and paper-based image questions

	PT 1	PT 2	PT 3	DT 1	DT 2	DT 3
k (image items)	37	37	36	36	40	60
Reliability (Cronbach`s α)	.67	.78	.76	.74	.72	.80
Spearman Brown corrected α (k = 60)	.77	.85	.84	.83	.79	.80

PT = Paper-based Test, DT = Digital Test

Viewing Task Characteristics

Image quality was rated significantly lower than the former paper-based image quality, $t(142) = -2.84$, $p < 0.05$, $r = 0.32$. Qualitative data analysis showed five themes with respect to the suggestions for improvement of the viewing task. The most prevalent theme was 'image manipulation options', with the majority of the comments being related to the desire to zoom in and out. The second most important theme was 'image resolution/size' due to complaints about the small screen size and image resolution. In addition, the ambient lighting was criticized for being too high, leading to hindering reflections on the computer screens. The number of comments for each theme is listed in Table 4.

Table 3. Average rit-values of digital image questions per question type

Rit values image questions	LMQ (k)	HSQ (k)	MCQ (k)	TFQ (k)
DT 1	-	.31 (4)	.26 (24)	.15 (8)
DT 2	-	.34 (4)	.32 (24)	.18 (12)
DT 3	.35 (8)	.27 (4)	.30 (32)	.22 (16)

LMQ = long menu question; HSQ = hotspot question; MCQ = multiple choice question; TFQ = true/false question; k = number of questions

Table 4. Number of comments related to the viewing task, categorized into five themes

Theme	DT 1	DT 2	DT 3
Number of responders	186	169	152
Scrolling speed	4	32	30
Loading speed of images	3	4	19
Image resolution/size	78	31	11
Image manipulation options	112	37	10
Ambient light reflection on screen	34	6	1

DT = Digital Test

Cognitive Task Characteristics

Seventy-four % of the participants agreed that digital radiology progress testing corresponds better to clinical practice than paper-based testing. The participants' perceptions of correspondence with clinical practice between the image questions in DT 1 and the former paper-based image questions did not differ. Qualitative data analysis revealed that reflections upon the cognitive task characteristics were very scarce, only seven comments, and the predominant view was that the digital 2D and volumetric image questions reflected clinical practice better than paper-based image questions. For example, one of the participants wrote: "Image questions with full datasets fit the reality of practice better". Next, the participant added: "However, the quality of the images especially the X-rays, should be taken care of".

Phase 2: Improvement and Evaluation of Digital Test 2

Improving actual image quality was not possible due to hardware limitations. In an attempt to improve the perceived image quality, several adjustments were made in the viewing characteristics of the second test: 1) the zooming option was made available, 2) a full-screen option was implemented that enabled double clicking on of the images for a full-screen view of the image and 3) ambient light was reduced.

To compensate for the increase in time investment involved in answering volumetric image questions, the total number of questions was decreased from 200 to 180. Besides, the proportion of volumetric image questions was reduced to 40% of 40 image-based questions.

RESULTS

Reliability of the digital image question subtest of DT 2 was comparable to the reliability of the paper-based image questions when corrected to a 60 image-based questions test with Spearman Brown formula (Table 2). Average item-total correlation (rit) values for digital image questions per question type are given in Table 3.

Viewing Task Characteristics

Despite the attempts to improve perceived image quality, digital image quality was still rated significantly lower than the former paper-based image quality, $t(112) = -2.99$, $p < 0.05$, $r = 0.27$, respectively. In the qualitative data analysis, the emphasis was again on the image manipulation options and the image size, though also scrolling speed was subject of discussion (Table 4). Although the number of comments had decreased, there was still criticism on the screen size and image resolution. This was specifically related to the questions with mammography images, X-rays of the breasts, that can contain very small calcifications. One of the comments: "The details on the X-ray images, such as micro calcifications, are very hard to distinguish on these computer screens". The comments on the image manipulation options had decreased and changed direction. The zooming option was now available, but it was not always functioning well. The number of comments about ambient lighting had decreased from 34 to 6. In addition, comments about the scrolling speed had increased: slow and faltering scrolling was experienced.

Cognitive Task Characteristics

Seventy-two % of the participants agreed that digital radiology progress testing corresponds better to clinical practice than paper-based testing. According to the participants, digital image questions of the second test corresponded significantly better to clinical practice than paper-based image questions, $t(131) = 1.99$, $p < 0.05$, $r = 0.17$.

Again, the qualitative data again reflected the dominant view that digital 2D and volumetric image questions reflect clinical practice better than paper-based image questions. In addition, some participants shared the opinion that the value of testing with volumetric images is a trade-off between improved reflection of clinical practice and increased time needed to complete the questions. For example, one of their comments was: "changes in viewing direction and contrast settings enable better interpretation of the abnormalities on the image, but it takes a lot of time".

Figure 2. Another example of a volumetric image question in VQuest

a. This cross-section of a chest CT scan is shown in a different viewing direction than the cross-section in Figure 1. In addition, an advanced image reconstruction is applied: a Maximum Intensity Projection (MIP). This reconstruction method can be useful for detection of lung nodules or for showing patterns of lung nodules.

Figure 2. Another example of a volumetric image question in VQuest

b. The same question as Figure 2A, but with different contrast settings. This example illustrates that some structures become only visible after changing contrast settings. For example, the ribs become visible and multiple calcifications can be observed (bright white structures in the center of the image).

Phase 3: Improvement and Evaluation of Digital Test 3

To improve the viewing task characteristics, the software was optimized to improve the speed of image manipulation, especially the zooming option and the scroll function. The examination committee received extra instructions about what type of images and cases were preferable or not. For example, it was recommended not to include too large volumetric datasets and only include the relevant parts of the image, because too large datasets could negatively affect the performance of the software, such as loading time and scrolling speed.

To further enhance the test goal of image interpretation skills, the number of image-based questions was increased to 60 (of which 38% volumetric images). Besides, the options of question types were expanded with a long-menu question. This question type requires selecting an answer from a long list of answer options, which only appears after typing two or more letters corresponding to the available options. The long menu question was introduced to improve the authenticity of the simulation, since there are no multiple choice options in clinical practice, and diagnoses have to be actively generated.

RESULTS

Reliability of the digital image question subtest of DT 3 was comparable to the reliability of the paper-based image questions when corrected to a 60 image-based questions test with Spearman Brown formula (Table 2). Average item-total correlation (rit) values for digital image questions per question type are given in Table 3.

Viewing Task Characteristics

After these improvements, image quality of the third digital test was rated significantly higher than the former paper-based image quality, $t(108) = 2.91$, $p < 0.01$, $r = 0.27$. The number of comments about the image manipulation options and image size had decreased significantly and there was only one complaint left about ambient lighting. Scrolling speed and this time also loading speed were still prominent topics. The comments specified that especially loading of the volumetric datasets (CT and MR images) was perceived as slow.

Cognitive Task Characteristics

Eighty % of the participants agreed that digital radiology progress testing corresponds better to clinical practice than paper-based testing. Again, participants found that digital image questions corresponded significantly better to clinical practice than paper-based image questions, $t(109) = 3.28$, $p < 0.01$, $r = 0.30$ respectively. Apart from the positive comments towards digital testing with volumetric images, some participants feel that image reading should rather be tested in clinical practice, while the progress test should focus on knowledge and interpretation of static images. One of the comments: "Either you test knowledge through recognition of key images (that is what radiology is at least, and I think this is the goal of the test), or you let us read 20 full CT cases (however, you don`t need to test this because this is what we do all day)". And another comment: "I think the best clinical practice test is real clinical practice. The progress test is primarily a knowledge test and navigating through images does not add much".

DISCUSSION

We described the implementation and development process of a digital simulation-based assessment in radiology residency. We aimed to improve the authenticity of image interpretation assessment. Even though high fidelity simulations were used, the authenticity of the first digital test was not rated higher than the paper-based tests. After optimizing test characteristics, image manipulation options and environmental conditions, participants favored the authenticity of the digital image questions over paper-based image questions.

An improved representation of clinical practice with volumetric images was reported previously among medical students[22], who have very limited experience with interpreting images in clinical practice. Standards for the viewing task are probably higher for residents, who are used to working with high-quality images and advanced, high-speed image manipulation equipment. Unfortunately, computers available for large-scale tests usually do not meet the high criteria of screen size, screen resolution and processor speed utilized in radiology practice. In our study, this discrepancy was reflected by the initial high number of suggestions for improvement of the viewing task, expressed by the residents.

One of the most important challenges was how to improve perceived image quality. Digital image quality was initially rated lower than paper-based image quality. Based on the comments this was probably due to the small displays with a limited screen resolution that were available in this large-scale assessment setting. The screen resolution was below the 1280 x 1024 displays that were previously recommended for radiology assessments[20]. Small screen size had to be compensated by optimizing image manipulation options. We therefore introduced and optimized the zooming function and introduced a full-screen option for cases with multiple images. Besides, speed and smoothness of scrolling and zooming were criticized by the participants and had to be optimized. Our results underscore that only implementing image manipulation options is not enough, and an optimal functionality of the options is crucial to reach a satisfactory simulation. This relates to the human-computer interaction literature, showing that intuitive and direct interactions are crucial for user acceptance of a computerized system[32].

Optimizing the ambient light was another challenge. A high level of ambient light can have a negative effect on observer performance[33-35]. Some studies report this can (partly) be compensated for by means of interactive contrast adjustments[33, 36]. Even though participants could adjust contrast settings, a substantial part of their comments was related to hinder of ambient light. Since there was no possibility to dim lights, we ultimately turned the lights off, which was appreciated by the participants. On the other hand, there is evidence that a dark reading room may increase visual fatigue due to pupil contraction and dilation, which negatively affects reader performance[37]. It is therefore recommendable to include possibilities for light control in assessment rooms for image interpretation tests.

Screen size, image manipulation options and the level of ambient light appear to be crucial factors for establishing an optimal reflection of clinical practice in radiology assessment. The importance of these factors probably varies in other visual domains, such as dermatology and clinical pathology. For example, image manipulation options are probably more important in clinical pathology[38] than in clinical dermatology, whereas ambient light reflection on computer screens may affect any image interpretation task.

Another important consideration is the increased time needed to review volumetric images. Loading of and scrolling through volumetric images takes time. The improved reflection of clinical practice with volumetric images has a trade-off with time constraints. Increased time and effort of participants, faculty and staff support are common drawbacks of simulations[39]. It requires careful consideration whether the desired test validity justifies the investments[40]. We therefore recommend that volumetric images should only be used if they are considered to better match the goal of the question.

Throughout the development process, we introduced multiple choice, hotspot and long menu questions to better reflect the cognitive task of image interpretation. In most of the tests, the multiple choice, hotspot and long menu questions reached an average rit-value of .30 or larger, as is recommended for high-stake tests[41], in contrast to the true/false questions that showed average rit-values from .15 to .22. However, the long menu questions have some disadvantages for teachers, because they require extensive lists of answer options, including synonyms, and are time-consuming to construct[12].

The development process with changes in image manipulation options, ambient lighting, time limits and question types went along with an improved authenticity of the digital simulation-based assessment compared to its paper-based counterpart. Since we implemented and developed the test in three phases and changed multiple factors in each phase, we cannot determine the effect of each separate factor. We can only conclude that the set of changes was advantageous and recommend to carefully consider these factors when developing an image interpretation test.

Some limitations should be addressed. Although the whole cohort of Dutch radiology residents participated in the three digital tests, we cannot generalize these results to radiology programs in other countries. Also, response rates of the questionnaires were moderate, and slightly decreased throughout the study. However, only the fifth year residents were slightly underrepresented in the questionnaire responses, possibly because they felt they would not benefit from future test improvements.

When interpreting the test results over time, we should acknowledge differences in group composition across the three tests, primarily due to residents continuously phasing in and out of the program throughout the year. However, the average training year remained virtually constant across the tests.

This study underscores that authenticity of a simulation does not necessarily increase with higher fidelity. Simulating the image interpretation task of radiology practice in a large-scale assessment setting is challenging, due to technological limitations. Optimizing image manipulation options, the level of ambient light, time limits and question types can help improving the authenticity of simulation-based radiology assessments.

REFERENCES

1. Miller GE: The Assessment of Clinical Skills/Competence/Performance. Acad Med 1990; 65: S63-7.
2. Kogan JR, Holmboe ES, and Hauer KE: Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review. JAMA 2009; 302: 1316-26.
3. Cook DA, and Triola MM: Virtual Patients: A Critical Literature Review and Proposed Next Steps. Med Educ 2009; 43: 303-11.
4. Okuda Y, Bryson EO, DeMaria S, Jr., Jacobson L, Quinones J, Shen B, and Levine AI: The Utility of Simulation in Medical Education: What Is the Evidence? Mt Sinai J Med 2009; 76: 330-43.
5. Bland AJ, Topping A, and Tobbell J: Time to Unravel the Conceptual Confusion of Authenticity and Fidelity and Their Contribution to Learning within Simulation-Based Nurse Education. A Discussion Paper. Nurse Educ Today 2014; 34: 1112-8.
6. Gulikers JTM, Bastiaens TJ, and Kirschner PA: A Five-Dimensional Framework for Authentic Assessment. Educational Technology Research and Development 2004; 52: 67-86.
7. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, and Seltzer SE: Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section Ct Scan Can Ruin Your Whole Day. Radiology 2011; 259: 346-62.
8. Reiner BI, Siegel EL, and Siddiqui K: Evolution, of the Digital Revolution: A Radiologist Perspective. Journal of Digital Imaging 2003; 16: 324-30.
9. van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JC, Ravesloot CJ, van Schaik JP, and Ten Cate TJ: Interpretation of Radiological Images: Towards a Framework of Knowledge and Skills. Adv Health Sci Educ Theory Pract 2014; 19: 565-80.
10. van der Gijp A, Ravesloot CJ, van der Schaaf MF, van der Schaaf IC, Huige JC, Vincken KL, Ten Cate OT, and van Schaik JP: Volumetric and Two-Dimensional Image Interpretation Show Different Cognitive Processes in Learners. Acad Radiol 2015.
11. Gulikers JTM, Bastiaens TJ, Kirschner PA, and Kester L: Authenticity Is in the Eye of the Beholder: Student and Teacher Perceptions of Assessment Authenticity. Journal of Vocational Education and Training 2008; 60.
12. van Bruggen L, Manrique-van Woudenbergh M, Spierenburg E, and Vos J: Preferred Question Types for Computer-Based Assessment of Clinical Reasoning: A Literature Study. Perspect Med Educ 2012; 1: 162-71.
13. Zafar S, Safdar S, and Zafar AN: Evaluation of Use of E-Learning in Undergraduate Radiology Education: A Review. Eur J Radiol 2014; 83: 2277-87.
14. den Harder AM, Frijlingh M, Ravesloot CJ, Oosterbaan AE, and van der Gijp A: The Importance of Human-Computer Interaction in Radiology E-Learning. J Digit Imaging 2016; 29: 195-205.
15. Ernst RD, Sarai P, Nishino T, Collins T, Oto A, Hernandez A, Walser EM, and Chaljub G: Transition from Film to Electronic Media in the First-Year Medical School Gross Anatomy Lab. J Digit Imaging 2003; 16: 337-40.
16. Howlett D, Vincent T, Watson G, Owens E, Webb R, Gainsborough N, Fairclough J, Taylor N, Miles K, Cohen J, and Vincent R: Blending Online Techniques with Traditional Face to Face Teaching Methods to Deliver Final Year Undergraduate Radiology Learning Content. Eur J Radiol 2011; 78: 334-41.
17. Turmezei TD, Tam MD, and Loughna S: A Survey of Medical Students on the Impact of a New Digital Imaging Library in the Dissection Room. Clin Anat 2009; 22: 761-9.
18. Petersson H, Sinkvist D, Wang C, and Smedby O: Web-Based Interactive 3d Visualization as a Tool for Improved Anatomy Learning. Anat Sci Educ 2009; 2: 61-8.

19. Rengier F, Hafner MF, Unterhinninghofen R, Nawrotzki R, Kirsch J, Kauczor HU, and Giesel FL: Integration of Interactive Three-Dimensional Image Post-Processing Software into Undergraduate Radiology Education Effectively Improves Diagnostic Skills and Visual-Spatial Ability. Eur J Radiol 2013; 82: 1366-71.
20. Krupinski EA, Becker GJ, Laszakovits D, Gerdeman AM, and Evanoff MG: Evaluation of Off-the-Shelf Displays for Use in the American Board of Radiology Maintenance of Certification Examination. Radiology 2009; 250: 658-64.
21. Mullins ME, Will M, Mehta A, and Novelline RA: Evaluating Medical Students on Radiology Clerkships in a Filmless Environment: Use of an Electronic Test Prepared from Pacs and Digital Teaching Collection Images. Acad Radiol 2001; 8: 514-9.
22. Ravesloot CJ, van der Schaaf MF, van Schaik JP, ten Cate OT, van der Gijp A, Mol CP, and Vincken KL: Volumetric Ct-Images Improve Testing of Radiological Image Interpretation Skills. Eur J Radiol 2015; 84: 856-61.
23. Albanese M, and Case SM: Progress Testing: Critical Analysis and Suggested Practices. Adv Health Sci Educ Theory Pract 2015.
24. Ravesloot C, van der Schaaf M, Haaring C, Kruitwagen C, Beek E, Ten Cate O, and van Schaik J: Construct Validation of Progress Testing to Measure Knowledge and Visual Skills in Radiology. Med Teach 2012; 34: 1047-55.
25. Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interactions of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
26. Norman GR, Coblentz CL, Brooks LR, and Babcook CJ: Expertise in Visual Diagnosis: A Review of the Literature. Acad Med 1992; 67: S78-83.
27. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Investigative Radiology 1978; 13: 175-81.
28. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. Journal of Medical Imaging and Radiation Oncology 2012; 56: 173-78.
29. Ebel RL: Estimation of the Reliability of Ratings. Psychometrika 1951; 16: 407-24.
30. Ravesloot CJ, Van der Schaaf MF, Muijtjens AM, Haaring C, Kruitwagen CL, Beek FJ, Bakker J, Van Schaik JP, and Ten Cate TJ: The Don't Know Option in Progress Testing. Adv Health Sci Educ Theory Pract 2015; 20: 1325-38.
31. Ravesloot CJ, van der Gijp A, van der Schaaf MF, Huige JC, Vincken KL, Mol CP, Bleys RL, ten Cate OT, and van Schaik JP: Support for External Validity of Radiological Anatomy Tests Using Volumetric Images. Acad Radiol 2015; 22: 640-5.
32. O'Brien MA, Rogers WA, and Fisk AD: Developing a Framework for Intuitive Human-Computer Interaction. Proc Hum Factors Ergon Soc Annu Meet 2008; 52: 1645-49.
33. Fuchsjager MH, Schaefer-Prokop CM, Eisenhuber E, Homolka P, Weber M, Funovics MA, and Prokop M: Impact of Ambient Light and Window Settings on the Detectability of Catheters on Soft-Copy Display of Chest Radiographs at Bedside. AJR Am J Roentgenol 2003; 181: 1415-21.
34. Hellen-Halme K, and Lith A: Effect of Ambient Light Level at the Monitor Surface on Digital Radiographic Evaluation of Approximal Carious Lesions: An in Vitro Study. Dentomaxillofac Radiol 2012; 41: 192-6.
35. Uffmann M, Prokop M, Kupper W, Mang T, Fiedler V, and Schaefer-Prokop C: Soft-Copy Reading of Digital Chest Radiographs: Effect of Ambient Light and Automatic Optimization of Monitor Luminance. Invest Radiol 2005; 40: 180-5.

36. Goo JM, Choi JY, Im JG, Lee HJ, Chung MJ, Han D, Park SH, Kim JH, and Nam SH: Effect of Monitor Luminance and Ambient Light on Observer Performance in Soft-Copy Reading of Digital Chest Radiographs. Radiology 2004; 232: 762-6.
37. Pollard BJ, Chawla AS, Delong DM, Hashimoto N, and Samei E: Object Detectability at Increased Ambient Lighting Conditions. Med Phys 2008; 35: 2204-13.
38. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, and Boshuizen HP: Expertise in Clinical Pathology: Combining the Visual and Cognitive Perspective. Adv Health Sci Educ Theory Pract 2015; 20: 1089-106.
39. Srinivasan M, Hwang JC, West D, and Yellowlees PM: Assessment of Clinical Skills Using Simulator Technologies. Acad Psychiatry 2006; 30: 505-15.
40. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, Waugh RA, Brown DD, Safford RR, Gessner IH, Gordon DL, and Ewy GA: Simulation Technology for Health Care Professional Skills Training and Assessment. JAMA 1999; 282: 861-6.
41. Downing SM: Statistics of Testing. Assessment in Health Professions Education. Edited by Downing SM and Yudkowsky R. New York, Routlegde, 2009.

Chapter 8

General Discussion

To improve education in radiology, it is key to understand expertise of visual diagnostic reasoning. As elaborated in the introduction chapter of this thesis, existing theories such as the Visual Search and Detection Model of Kundel and Nodine[1] and global-focal search theories[2-4], describe the visual diagnostic reasoning process of radiologists. Other expertise theories, the Dual Process Theory[5, 6] and the Expertise Theory of Anders Ericsson[7], are helpful for understanding expertise development in visual diagnostic reasoning. The ability of expert radiologists to recognize a disease in a split second may be explained by the fast non-analytical way of thinking described in the Dual Process Theory. The Expertise Theory provides recommendations for training to reach expert performance.

However, the theories do not give answers to the domain-specific issues of radiology expertise development: which knowledge and skills are needed to support the development of radiology expertise, and more specifically image interpretation expertise in the contemporary era of advanced medical images? And how should we adapt our training and assessment methods to those advancements? This thesis aimed to 1) investigate which knowledge and skills are important for the development of expertise in radiological 2D and volumetric image interpretation and 2) contribute to the development and evaluation of training and assessment methods to stimulate expertise development in radiological 2D and volumetric image interpretation. The studies in this thesis focus on diagnostic radiology, with a specific interest in volumetric image interpretation. In relation to the first aim: in chapter 2 we investigated which domain-specific knowledge and skills are needed for radiological image interpretation, in chapter 3 we compared the use of these knowledge and skills in 2D and volumetric image interpretation and in chapter 4 we described the relation between visual search and diagnostic performance. In relation to the second aim: in chapter 5 we investigated the effect of teaching search strategies on perceptual performance, in chapter 6 we evaluated how we can identify different types of errors that occur in the visual diagnostic process and in chapter 7 we described how the authenticity of volumetric image interpretation skill testing can be increased with a digital simulation-based assessment.

DISCUSSION OF MAIN FINDINGS

Expertise Development in Image Interpretation

Which domain-specific knowledge and skills are needed for radiological image interpretation?

We developed a conceptual framework to capture knowledge and skills required for 2D and volumetric image interpretation. We identified three main components of image interpretation: perception, analysis and synthesis. This was further specified into sixteen knowledge or skill elements, described in detail in chapter 2. The framework adds to previous visual expertise research which primarily describes processes or strategies of image interpretation[8-11] and does not specify knowledge and skills that are necessary. Image interpretation is a high-level cognitive process, involving data representation, visual information processing, and reasoning. The specific knowledge and skills that are used differ from case to case. For example, instant pattern recognition takes place when visual information can be compared with similar previous examples. In these cases, image interpretation can be considered as trivial as the recognition of an object, such as a chair. Previous examples of cases must be encountered to enable the instant recognition of a new case, but knowledge about anatomy or pathology may not be a prerequisite in all cases. A striking recent research report showed that even pigeons could be trained to detect subtle calcifications on mammography[12]. However, an attempt to train them in distinguishing benign from malignant breast masses failed, probably because this distinction requires knowledge about the features of the mass that differentiates between those conditions. In case an observer has to diagnose a disease or a manifestation of a disease one had never encountered before, one cannot rely on pattern recognition. Analytical reasoning processes are then required to interpret the findings. The complexity of uncovering the image interpretation process is not only caused by the internal complexity of the cognitive processes but also due to the variation wherein these cognitive processes occur. Still, we pursued in creating a framework of required knowledge and skills, to assist in further research and advances in radiology education. We tried to cover the most important knowledge and skills that are required to build expertise in diagnostic radiology. We emphasize that the framework should be considered a simplified model of the highly complex and variable nature of the image interpretation process. Another consideration is the sustainability of the framework. Radiology is developing rapidly, embracing new techniques and combining modalities to improve the diagnostic and prognostic value of imaging. For example, improved MRI techniques allow for visualizing movement of organs and the combination of radiology with nuclear imaging techniques provides functional information. Interpreting these images requires additional knowledge about physiology and function. Other knowledge and skills may become less important in the future. Perception may be partly taken over by computers, due to the development of Computer-Aided Detection (CAD) systems[13]. This would, in turn, introduce other skills such as how to deal with the results of CAD systems and how to report and communicate those. This means our framework should be interpreted in the context of the current practice of diagnostic radiology. The knowledge and skills required for image interpretation may evolve through time based on new insights or developments in clinical practice.

How do these skills differ between 2D and volumetric image interpretation?

The framework of the first study was used to investigate differences in cognitive processes between 2D and volumetric image interpretation in the second study (chapter 3). Cognitive processes in volumetric image interpretation were found to differ from 2D image interpretation. Perception was most prominent in volumetric image interpretation, while synthesis was most prominent in 2D image interpretation. This is probably related to different amounts of image information in 2D and volumetric image interpretation. Volumetric images contain much more visual information, requiring more visual processing, hence more perception. Previous research also shows that the amount of image information is positively related to self-reported mental effort spent in image interpretation[14]. In 2D images, the amount of visual input is limited and more time and working memory can be dedicated to other cognitive processes, such as diagnostic considerations.

These results should be interpreted with caution, especially when it comes to generalizability to other multidimensional tasks. Volumetric images should not be confused with true three-dimensional (3D) images. Whereas a volumetric image is a set of two-dimensional cross-sections which can be scrolled through, 3D images are reconstructions of the surface of structures the body that can be rotated. Still, even though volumetric images are not true 3D representations of the human body, performance in recognizing and describing anatomical structures in volumetric images and in cadavers was found to correlate heavily[15]. This correlation was not found for the same task in 2D images and cadavers[15]. However, using 3D models for teaching anatomy does not necessarily improve the understanding of volumetric images[16]. These differences in representations of the human body should be taken into account when interpreting and comparing studies about expertise development in image interpretation skills.

How does visual search relate to diagnostic performance?

These first two studies in this thesis were based on verbal protocols, whereas image interpretation involves a large visual component. Eye tracking can give information about the visual attention of observers, and provide insight into their visual search patterns. We reviewed the eye tracking literature of radiology expertise research to explore which visual search patterns are related to high diagnostic performance. The review study described in chapter 4 demonstrates that experts and high performers need shorter viewing times and fewer fixations with longer saccades in between, and they fixate on abnormalities faster than inexperienced readers or low performers. Experts need less time and fewer fixations to collect the relevant image information[17] and recognize disturbances in a field of view more quickly by taking advantage of a large visual span[17, 18]. This relates to global-focal search theories[2-4], advocating that experts use a fast holistic approach to images, while less proficient observers apply a slow search-to-find approach. In a broader context of expertise development, this also reflects the principles of Dual Process Theory: the holistic approach being intuitive and fast, and the search-to-find approach being a conscious and relatively slow process. In addition to these general conclusions about differences in eye movement parameters across expert levels, experts use specific search patterns in specific tasks. Examples are the radial search in hand and wrist X-rays[19] or the drilling search pattern through chest CTs[20]. However, the effect of teaching these expert search strategies to trainees has hardly been investigated. Besides, only a minority of the studies investigated visual search in volumetric images, while interpreting these images currently involves an increasing part of radiology practice[21, 22].

Radiology Training and Assessment Methods

What is the effect of teaching search strategies on perceptual performance?

Volumetric image interpretation requires navigating through the images. We developed a scroll strategy training derived from a study by Drew et al., who discovered that expert radiologists exhibit two main types of search strategies when searching for lung nodules in chest CT scans: 'scanning' and 'drilling'[20]. We investigated if teaching these expert search strategies to junior residents could improve perceptual performance. Search strategy instruction had a significant effect on both scroll behavior and perceptual performance. The scanning instruction decreased the true positive rate, while the drilling instruction did not have a significant effect. Probably this was due to the finding that the majority of participants already used a certain type of drilling for their free search. Still, the drilling instruction significantly decreased the number of false positive findings compared to the free search condition. These results partly deviate from a previous study of Kok et al.[23]. Although both studies showed that search strategy instruction can affect the search behavior of trainees, Kok et al. did not find a beneficial effect on performance[23]. This may be due to some important differences in study population and in the task at hand. In our study, the task was strictly dedicated to the detection of lung nodules, whereas the participants in the study of Kok et al. had to detect different types of abnormalities. Their explanation for the absence of a performance effect was that the knowledge necessary for recognizing the lesions was probably lacking. The contribution of knowledge in recognizing these diverse diseases is probably larger than for recognizing lung nodules, where the efficiency of the search may be a more important factor than specific radiological knowledge. Moreover, we studied residents who already had acquired more domain-specific radiology knowledge than medical students. Another explanation could be the difference in imaging modality. The efficiency of search is probably more important in volumetric images used in our study than in X-rays used in the study of Kok et al.. Volumetric images contain much more visual information and require navigation to visualize the abnormalities. An X-ray is a 2D projection of a 3D structure, with many structures projecting over each other, and it can be very challenging to distinguish normal from abnormal, even for expert observers[24]. So even with a very efficient search strategy, observers will not detect findings if they cannot discriminate between normal and abnormal.

The preference of drilling over scanning cannot be transferred directly to other detection tasks and expert levels. Search behavior differs between different types of tasks [19, 23, 25, 26] and types of abnormalities encountered[26]. Lung nodules are focal abnormalities and lung nodule detection may benefit more from a drilling strategy than the detection of diseases with more diffuse abnormalities. The training may also have a different effect in other expertise levels, due to expected differences in knowledge level and prior search behavior. Expert radiologists who perform similarly use completely different search strategies[27, 28], and have a different starting point prior to instruction of new strategies. Besides, it may be challenging to alter the habits of experienced observers with a search strategy training.

How can we identify different types of errors that occur in the visual diagnostic process?
We used an assessment method with a stepwise-reasoning approach to identify different types of errors in the image interpretation process of medical students. In this question format, the visual diagnostic reasoning process was assessed in steps, using subquestions. With this method, more than three-quarters of all errors could be related to one of the image interpretation components, of which 39% perception, 20% analysis, and 41% synthesis errors. The distribution of error types we found should not be compared with the error types in clinical practice[29, 30]. The purpose of the study was to find a method to distinguish error types in an educational setting, and not to investigate error type rates. Participants were aware of the presence of an abnormality because a radiology test with only pathological cases was used for the study, which does not compare to clinical practice.

In addition to performance information, the stepwise-reasoning approach provides process information by disentangling the image interpretation process in three components. Although many radiology exams are composed of multiple choice questions[31], other examples of assessments providing process information can be found in the literature[32, 33]. For example, the study group of Pecaric et al. reported a learning analytics approach to uncovering the process of image interpretation[32]. Heat maps of hotspot questions are used to provide insight in perceptual errors, for example, growth plates being mistaken for a fracture. Mouse click information is used to show whether all images are reviewed. These different types of process information can be used for feedback on the level of an individual learner[33], but also on a group level, for example, to guide teachers in their instructions in a classroom setting.

Besides error type characterization, the stepwise-reasoning question format has the potential to provide additional performance information of the learner, that stays unnoticed if only a diagnosis is asked. Latent errors are perception and analysis errors if followed by a correct diagnosis. Students may guess the correct diagnosis for example based on clinical information, without detecting the abnormality or without a valid analysis of the abnormality. Partial knowledge can be identified when students correctly detect and/or analyze an abnormality, while the diagnosis is incorrect. Partial knowledge cannot be rewarded in multiple choice questions with only a list of diagnoses. With the stepwise-reasoning approach in chapter 6 we were able to identify latent errors or partial knowledge in over 25% of patient cases.

How can digital simulation-based assessment increase the authenticity of volumetric image interpretation skill testing?

A digital simulation-based assessment method was developed for testing image interpretation skills of radiology residents. This assessment method was designed to test volumetric image interpretation skills and included advanced image manipulation options and question types to simulate this task. The method was previously tested in radiological anatomy tests in medical students and was found to reflect clinical practice better than testing with 2D images[34]. However, it took several developmental steps to increase the perceived authenticity of the residents' test. Due to technological limitations, it was challenging to simulate the image interpretation task in a way that meets the standards of residents who are used to work with high-quality images and advanced, high-speed image manipulation equipment. After optimizing image manipulation options, ambient lighting, time limits and question types, the residents eventually favored the authenticity of the digital image questions over paper-based image questions.

Taking into account these challenges and technical and environmental requirements, we may question whether it is worth the effort to test with volumetric images and advanced image manipulation options. Does the possibility of scrolling through and manipulating images really add to the quality of radiological skills testing? Findings in this thesis and other previous studies affirm the added value of volumetric images in radiology tests. There are indications that 2D image interpretation and volumetric image interpretation differ substantially. In chapter 3 we found that 2D and volumetric image interpretation require different knowledge and skills. Besides, testing with volumetric images improves several quality aspects of assessment, such as reliability and correlation with external measures[15, 34]. Only testing with 2D images would be a potential risk for underrepresentation of the intended construct in the measurement[35]. An alternative could be to test the skills that are needed for volumetric image interpretation in the workplace. However, workplace based assessments have other challenges, such as utility and standardization issues[36].

STRENGTHS AND LIMITATIONS

One of the main strengths of our research is the focus on volumetric image interpretation. Volumetric images account for a major part of radiology practice but are not included in most previous studies investigating radiology expertise[37-42]. Image manipulation skills are hardly addressed in the literature. Only a small number of studies investigated differences in scroll behavior[20, 28], and the effect of teaching scroll behavior has been unknown. Besides, we investigated *both* the process of visual diagnostic reasoning *and* the required knowledge and skills. Most studies focus only on the process of visual diagnostic reasoning, for example with eye tracking studies, while according to literature in diagnostic reasoning, domain-specific knowledge is crucial for a successful reasoning process[43]. Also in the visual diagnostic reasoning literature, there are indications that teaching a strategy does not improve performance when knowledge is lacking[23]. The explicit identification of required knowledge and skills makes our results applicable for educational purposes.

Visual diagnostic reasoning has a cognitive and a perceptual component, and in the case of volumetric images, image manipulation components. These components are integrated and cannot be strictly separated: perception requires cognition, image manipulation is guided by cognitive processes and cognitive processes are guided by perceptual input which in turn depends on the way the image is presented. Different study methods are used to study the image interpretation process. The perceptual component is primarily studied with eye tracking, the cognitive component with think-aloud studies and the image manipulation skills with analyzing scroll and mouse click behavior. However, since all components are integrated none of these methods captures the complete process. For example, verbal protocols cannot capture unconscious cognitive processes and visual attention, whereas eye tracking does not reveal the reasons and intention behind the eye movements. Therefore we used a combination of methods to study visual diagnostic reasoning, revealing cognitive components (knowledge and skills), visual components (visual search patterns) and image manipulation components (scroll behavior).

Finally, a substantial part of our research was dedicated to the implementation of training and assessment methods in radiology education. The existing literature around radiology expertise development and the theoretical advances in part one were used to develop and improve training and assessment methods in part two. The second part of the thesis does not only focus on the results but also on the development of these training and assessment methods, to provide useful practical information for educators who are developing or implementing radiology education programs.

There are also limitations. Although we used several research methods to investigate different aspects of visual diagnostic reasoning, we did not triangulate the methods within one study population. Connecting different types of data could further improve the understanding of the complete process of image interpretation and the interaction between cognitive, visual and image interaction processes. At the same time, the interaction of these processes[9] complicates the analysis of simultaneous research methods. For example, thinking aloud could affect the (speed of) image interpretation and eye movements[44]. Besides, some of our study settings, e.g. large-scale assessments, did not allow for collecting concurrent think-aloud or eye tracking data. An alternative possibility for connecting eye tracking and think-aloud data would be cued retrospective reporting[45]. This is a retrospective reporting method using the eye movements of participants superimposed on the image as a cue.

We investigated visual diagnostic reasoning skills in medical students and radiology residents. Some authors argue that we should not teach medical students to interpret images since most of them will not become radiologists[46]. They advocate that medical students are better served by a curriculum focussing on knowledge about indications for and risks and benefits of medical imaging modalities. However, many non-radiologist physicians interpret images of, often critically ill, patients[47] and the availability of images throughout the hospital and possibly even outside the hospital may even further improve, due to increasing possibilities of PACS (Picture Archiving and Communication Systems) technology. It is a challenge to control image interpretation activities of all physicians and to provide all physicians with sufficient knowledge and skills. However, this does, even more, justify an image interpretation curriculum in undergraduate medical education to reach all medical doctors.

In this thesis, we only address visual diagnostic reasoning in radiology. Obviously, many other skills are needed to become a radiologist[48, 49]. For example, radiologists develop motor skills for acquiring medical images such as ultrasound and angiography, and for performing image-guided interventions, such as biopsies. Non-medical professional skills, such as teaching skills and communication skills are also crucial for the radiologist[48, 49], especially in the current system with an array of multidisciplinary clinical meetings. We chose not to incorporate all these skills that relate to, but not primarily qualify as, visual diagnostic reasoning because these activities require specific knowledge, skills and attitudes and require specific research methods.

IMPLICATIONS & FUTURE RESEARCH

Expertise Development in Image Interpretation

Our studies contribute to the understanding of the development of visual diagnostic expertise. We did not only investigate this field of expertise out of fascination, we also searched for knowledge advances that could be useful for educators. The best example is the framework of knowledge and skills. This framework can be used to support educational programs. For example, it could be used as a template for learning objectives or test blueprints. This is not necessarily limited to radiology education programs. After validation and some adaptations, it may also apply to image interpretation in other medical and non-medical domains. Depending on the nature of the image interpretation task, including volumetric image exercises could be desirable, because volumetric image interpretation provokes different cognitive processes than 2D image interpretation. However, the learning objectives should be weighed against time considerations to avoid unnecessary time investments from learners and teachers. In addition to educational purposes, the framework can also serve as a theoretical framework for expertise research, for example for coding qualitative data.

The understanding of visual diagnostic expertise development could be further increased by triangulating data within study populations[50]. Although we now have a sense of the components and the knowledge and skill elements of image interpretation, we do not know how these components interact. It remains unclear what the rationale or drive is behind the search strategies of observers. The use of a search strategy may be a conscious choice or an unconscious process, driven by visual input in combination with diagnostic considerations. Combining information from cognitive processes and search behavior may further explain differences in search behavior of observers of different expert levels.

Radiology Training and Assessment Methods

The ultimate goal of studying expertise in radiology should be to optimize educational programs to improve the training of learners. Although there is an extensive body of literature about visual diagnostic expertise[51], the number of studies investigating the effect of using these expert characteristics in educational programs is limited[52]. For educational purposes, we should not only describe differences between experts and novices but also investigate methods to improve the performance of trainees. There are some initial indications that teaching expert strategies to novices may affect their search pattern[23, 52]. Whether this improves performance is not entirely clear and results differ between studies[23, 52]. In the presence of sufficient knowledge to recognize the abnormalities, search strategy instruction can have a positive effect on performance. Whether we should include scroll strategy training in radiology education remains debatable. At least our results suggest that 'scanning' should not be taught as a search strategy for lung nodule detection on CT scans, because it decreases diagnostic performance of radiology trainees. Trainees seem to benefit from the drilling strategy, as it improves their diagnostic performance, and drilling instruction may be effective for trainees who start with reading chest CT scans. It is not desirable to start implementing 'drilling' as a general scroll strategy for all volumetric image interpretation tasks and all expertise levels. The benefits of a specific search strategy for diagnostic performance may differ depending on the task and the level of expertise. Which search strategies are most effective in other detection tasks, such as the detection of more diffuse abnormalities, is subject to further research.

The evaluation of the two assessment methods in chapter 6 and 7 informs educators and can guide them whether or not to use these approaches of assessment. The goal of the education program and assessment plays an important role in choosing the best format. The stepwise reasoning approach informs learners and teachers about error types and performance and may be especially useful for formative assessments. Teachers may use the information to tailor their instruction towards individual learners. In a classroom setting, the information can be used to focus instruction on errors types that are common in a particular group of learners. The framework of knowledge and skills in chapter 2 can be used to select the required knowledge and skills for each component. In case of frequent perception errors, search strategy instruction can be used to improve the detection of abnormalities.

The simulation-based assessment approach can be used as an addition to workplace based assessment. For image interpretation tests, it is important to optimize image quality within the possibilities of the facilities. When screen size is limited, image manipulation options such as zooming and full screen options can be used to compensate. Possibilities for ambient light control are highly recommended.

CONCLUSIONS

This thesis contributes to the understanding of image interpretation expertise development. Image interpretation expertise requires many knowledge and skill elements that are integrated in a complex process and can differ depending on the type of image information. Whereas perception is dominant in volumetric image interpretation, synthesis is dominant in 2D image interpretation. The thesis provides some initial indications that search strategy instructions can significantly affect performance of trainees in a volumetric image interpretation task.

The thesis provides several tools that can be used to improve radiology education. A framework of knowledge and skills was created that can serve as a guideline for the development of radiology education. Two methods for assessment of image interpretation skills were introduced: a formative assessment method to increase insight into error types and performance of learners and a digital simulation-based assessment method that can increase the authenticity of image interpretation assessment. These methods can be used in radiology education to stimulate image interpretation expertise development of learners.

REFERENCES

1. Nodine CF, and Kundel HL: Using Eye Movements to Study Visual Search and to Improve Tumor Detection. Radiographics : a review publication of the Radiological Society of North America, Inc 1987; 7: 1241-50.
2. Kundel HL, Nodine CF, Conant EF, and Weinstein SP: Holistic Component of Image Perception in Mammogram Interpretation: Gaze-Tracking Study. Radiology 2007; 242: 396-402.
3. Swensson RG: A Two-Stage Detection Model Applied to Skilled Visual Search by Radiologists. Perception and Psychophysics 1980; 27: 11-16.
4. Drew T, Evans K, Võ MLH, Jacobson FL, and Wolfe JM: Informatics in Radiology: What Can You See in a Single Glance and How Might This Guide Visual Search in Medical Images? Radiographics 2013; 33: 263-74.
5. Kahneman D: Thinking, Fast and Slow. New York, Farra, Straus and Giroux, 2011.
6. Kahneman D: A Perspective on Judgment and Choice: Mapping Bounded Rationality. American Psychologist 2003; 58: 697-720.
7. Ericsson KA, and Lehmann AC: Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints Annual Review of Psychology 1996; 273-305.
8. Kundel HL, Nodine CF, and Carmody D: Visual Scanning, Pattern Recognition and Decision-Making in Pulmonary Nodule Detection. Invest Radiol 1978; 13: 175-81.
9. Morita W, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interaction of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.
10. Rogers E: A Study of Visual Reasoning in Medical Diagnosis. Mahwah, N.J., Lawrence Erlbaum, 1996.
11. Nodine CF, and Kundel HL: Using Eye Movements to Study Visual Search and to Improve Tumor Detection. Radiographics 1987; 7: 1241-50.
12. Levenson RM, Krupinski EA, Navarro VM, and Wasserman EA: Pigeons (Columba Livia) as Trainable Observers of Pathology and Radiology Breast Cancer Images. PLoS One 2015; 10: e0141357.
13. Krupinski EA: Computer-Aided Detection in Clinical Environment: Benefits and Challenges for Radiologists. Radiology 2004; 231: 7-9.
14. Stuijfzand BG, van der Schaaf MF, Kirschner FC, Ravesloot CJ, Van der Gijp A, and Vincken KL: Medical Students’ Working Memory Load in Multiplanar Image Interpretation: Insights from Human-Computer Interactions and Eye Movements. Computers in Human Behavior 2014; 62: 394-403.
15. Ravesloot CJ, van der Gijp A, van der Schaaf MF, Huige JC, Vincken KL, Mol CP, Bleys RL, ten Cate OT, and van Schaik JP: Support for External Validity of Radiological Anatomy Tests Using Volumetric Images. Acad Radiol 2015; 22: 640-5.
16. Metzler R, Stein D, Tetzlaff R, Bruckner T, Meinzer HP, Buchler MW, Kadmon M, Muller-Stich BP, and Fischer L: Teaching on Three-Dimensional Presentation Does Not Improve the Understanding of According Ct Images: A Randomized Controlled Study. Teach Learn Med 2012; 24: 140-8.
17. Gegenfurtner A, Lehtinen E, and Säljö R: Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review 2011; 23: 523-52.
18. Bertram R, Helle L, Kaakinen JK, and Svedstrom E: The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception. PLoS One 2013; 8.
19. Hu CH, Kundel HL, Nodine CF, Krupinski EA, and Toto LC: Searching for Bone Fractures: A Comparison with Pulmonary Nodule Search. Academic Radiology 1994; 1: 25-32.

20. Drew T, Le-Hoa Vo M, Olwal A, Jacobson F, Seltzer SE, and Wolfe JM: Scanners and Drillers: Characterizing Expert Visual Search through Volumetric Images. Journal of Vision 2013; 13.
21. Bhargavan M, and Sunshine JH: Workload of Radiologists in the United States in 2002-2003 and Trends since 1991-1992. Radiology 2005; 236: 920-31.
22. Bhargavan M, Kaye AH, Forman HP, and Sunshine JH: Workload of Radiologists in United States in 2006-2007 and Trends since 1991-1992. Radiology 2009; 252: 458-67.
23. Kok EM, Jarodzka H, de Bruin AB, BinAmir HA, Robben SG, and van Merrienboer JJ: Systematic Viewing in Radiology: Seeing More, Missing Less? Adv Health Sci Educ Theory Pract 2016; 21: 189-205.
24. Berlin L: Radiologic Errors, Past, Present and Future. Diagnosis 2014; 1: 79–84.
25. Krupinski EA: Visual Scanning Patterns of Radiologists Searching Mammograms. Academic radiology 1996; 3: 137-44.
26. Kok EM, de Bruin ABH, Robben SGF, and van Merriënboer JJG: Looking in the Same Manner but Seeing It Differently: Bottom-up and Expertise Effects in Radiology. Applied Cognitive Psychology 2012; 26: 854-62.
27. Mello-Thoms C, Ganott M, Sumkin J, Hakim C, Britton C, Wallace L, and Hardesty L: Different Search Patterns and Similar Decision Outcomes: How Can Experts Agree in the Decisions They Make When Reading Digital Mammograms? Digital Mammography, Proceedings. Edited by Krupinski EA. 2008, 212-19.
28. Venjakob A, Marnitz T, Mahler J, Sechelmann S, and Roetting M: Radiologists' Eye Gaze When Reading Cranial Ct Images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment. Edited by Abbey CK and MelloThoms CR. 2012.
29. Donald JJ, and Barnard SA: Common Patterns in 558 Diagnostic Radiology Errors. Journal of Medical Imaging and Radiation Oncology 2012; 56: 173-78.
30. Renfrew DL, Franken Jr EA, Berbaum KS, Weigelt FH, and Abu-Yousef MM: Error in Radiology: Classification and Lessons in 182 Cases Presented at a Problem Case Conference. Radiology 1992; 183: 145-50.
31. Lewis PJ, Chen JY, Lin DJ, and McNulty NJ: Radiology Examweb: Development and Implementation of a National Web-Based Examination System for Medical Students in Radiology. Acad Radiol 2013; 20: 290-6.
32. Pecaric M, Boutis K, Beckstead J, and Pusic M: A Big Data and Learning Analytics Approach to Process-Level Feedback in Cognitive Simulations. Acad Med 2017; 92: 175-84.
33. Boutis K, Pecaric M, Shiau M, Ridley J, S PG, J SA, and M VP: A Hinting Strategy for Online Learning of Radiograph Interpretation by Medical Students. Med Educ 2013; 47: 877-87.
34. Ravesloot CJ, van der Schaaf MF, van Schaik JP, ten Cate OT, van der Gijp A, Mol CP, and Vincken KL: Volumetric Ct-Images Improve Testing of Radiological Image Interpretation Skills. Eur J Radiol 2015; 84: 856-61.
35. Messick S: Validity of Psychological Assessment. Validation of Inferences from Persons` Responses and Performances as Scientific Inquiry into Score Meaning. American Psychologist 1995; 50: 741-49.
36. Kogan JR, Holmboe ES, and Hauer KE: Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review. JAMA 2009; 302: 1316-26.
37. Cooper L, Gale A, Darker I, Toms A, and Saada J: Radiology Image Perception and Observer Performance: How Does Expertise and Clinical Information Alter Interpretation? Stroke Detection Explored through Eye-Tracking. Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment 2009; 7263: 72630K1-12.
38. Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y, Iwano S, Ikeda M, and Ishigaki T: Interactions of Perceptual and Conceptual Processing: Expertise in Medical Image Diagnosis. International journal of human-computer studies 2008; 66: 370-90.

39. Wang XH, Durick JE, Lu A, Herbert DL, Golla SK, Foley K, Piracha CS, Shinde DD, Shindel BE, Fuhrman CR, Britton CA, Strollo DC, Shang SS, Lacomis JM, and Good WF: Characterization of Radiologists' Search Strategies for Lung Nodule Detection: Slice-Based Versus Volumetric Displays. Journal of Digital Imaging 2008; 21: S39-S49.
40. Cooper L, Gale A, Saada J, Gedela S, Scott H, and Toms A: The Assessment of Stroke Multidimensional Ct and Mr Imaging Using Eye Movement Analysis: Does Modality Preference Enhance Observer Performance? Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment 2010; 7627: 76270B.
41. Matsumoto H, Terao Y, Yugeta A, Fukuda H, Emoto M, Furubayashi T, Okano T, Hanajima R, and Ugawa Y: Where Do Neurologists Look When Viewing Brain Ct Images? An Eye-Tracking Study Involving Stroke Cases. PLoS One 2011; 6: e28929.
42. Venjakob A, Marnitz T, Mahler J, Sechelmann S, and Roetting M: Radiologists' Eye Gaze When Reading Cranial Ct Images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment 2012; 8318.
43. Elstein AS, Shulman LS, and Sprafka SA: Medical Problem Solving: An Analysis of Clinical Reasoning. . Cambridge, MA, Harvard University Press, 1978.
44. Ericsson KA, and Simon HA: Protocol Analysis: Verbal Reports as Data. London, Massachusetts Institute of Technology, 1993.
45. van Gog T, Paas F, van Merriënboer JJG, and Witte P: Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. Journal of Experimental Psychology: Applied 2005; 11: 237-44.
46. Naeger DM, Webb EM, Zimmerman L, and Elicker BM: Strategies for Incorporating Radiology into Early Medical School Curricula. J Am Coll Radiol 2014; 11: 74-9.
47. Saha A, Roland RA, Hartman MS, and Daffner RH: Radiology Medical Student Education: An Outcome-Based Survey of Pgy-1 Residents. Acad Radiol 2013; 20: 284-9.
48. Chhem RK, Hibbert KM, and van Deven T, eds., Radiology Education Berlin Heidelberg, Springer, 2009.
49. Tielbeek AV, Aarts JCNM, Griever GER, van Hoogstraten JWAP, Lameris JS, Sanders FBM, and van Schelven IH: Good Clinical Practice. ed. by Radiologie NVv, 2011.
50. Gegenfurtner A, Kok E, van Geel K, de Bruin A, Jarodzka H, Szulewski A, and van Merrienboer JJ: The Challenges of Studying Visual Expertise in Medical Image Diagnosis. Med Educ 2017; 51: 97-104.
51. van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JP, and Ten Cate TJ: How Visual Search Relates to Visual Diagnostic Performance: A Narrative Systematic Review of Eye-Tracking Research in Radiology. Adv Health Sci Educ Theory Pract 2016.
52. van der Gijp A, Vincken KL, Boscardin C, Webb EM, Ten Cate OT, and Naeger DM: The Effect of Teaching Search Strategies on Perceptual Performance. Acad Radiol 2017.

Chapter 9

Summary & Samenvatting

SUMMARY

Background

Radiological image interpretation is a complex skill and requires years of training to master. Diagnostic errors in radiology are frequent and can have significant consequences for patients. To improve education and performance in radiological image interpretation, it is key to understand visual diagnostic reasoning.

The role of medical images in clinical decision-making is increasingly important and medical imaging techniques are developing very quickly. One of the most crucial developments in radiology was the introduction of digital volumetric imaging, which allows radiologists to navigate through thin cross sections on a computer monitor instead of viewing two-dimensional hard-copy images on a light box. This development has changed the nature of the visual diagnostic task. How this development affects the visual diagnostic reasoning process and how we should align our training and assessment methods is largely unknown and is the subject of this thesis.

In this thesis, we aim to 1) investigate which knowledge and skills are important for the development of expertise in radiological 2D and volumetric image interpretation and 2) contribute to the development and evaluation of training and assessment methods to stimulate expertise development in radiological 2D and volumetric image interpretation.

Study Results

Expertise Development in Image Interpretation

The first part of the thesis uncovers several aspects of the nature of radiology expertise. In the first study, described in **chapter 2**, we developed a framework to describe the knowledge and skills that are required for 2D and volumetric image interpretation. The study was based on a mixed-method design with three phases that contributed to the developmental process. First, a preliminary version of the framework was created based on the existing literature on the knowledge and skills required for image interpretation. Second, this preliminary framework was optimized based on semi-structured interviews with selected experts. Third, the framework was adjusted based on verbal protocols of observers interpreting radiological 2D and volumetric images. We identified three main components of image interpretation: perception, analysis and synthesis. These components were further specified into sixteen knowledge or skill elements, such as 'knowledge of anatomy', 'using efficient search strategies', 'discriminating relevant from irrelevant findings' and 'deciding about advice or action'. The framework should be considered a simplified model of the highly complex nature of the image interpretation process and may evolve through time based on new insights or developments in clinical practice.

In **chapter 3**, differences in knowledge and skills used for 2D and volumetric image interpretation were investigated in a think-aloud study. The framework of chapter 2 was

used to characterize the knowledge and skills of the verbal protocols obtained during image interpretation. A within-subjects design was applied to compare knowledge and skills used by advanced medical students in 2D and volumetric image interpretation. We found significant differences between the two types of image interpretation: perception was dominant in volumetric image interpretation, while synthesis was dominant in 2D image interpretation. This may be related to the higher amount of image information in volumetric image interpretation evoking more perception processes than in 2D image interpretation, where visual input is limited and more cognitive processes can be dedicated to synthesis.

Chapter 4 is a review study, describing how visual search relates to diagnostic performance. Visual search is traditionally investigated with eye tracking. In this study, six electronic literature databases were searched using 'visual perception' OR 'eye tracking' AND 'radiology' and synonyms. Inclusion and exclusion criteria were applied. We performed a thematic analysis to analyze the study results, and integrated data with a textual narrative synthesis. The eye tracking literature of radiology expertise research shows that experts and high performers fixate on abnormalities faster than less experienced readers or low performers, and need less time for their overall visual search. During their visual searches, experts exhibited fewer eye fixations with longer saccades in between the fixations, as compared to inexperienced observers. These findings support the view that experts have the capability to collect relevant image information efficiently and take advantage of a large visual span. The findings also support global-focal search theories. These theories advocate that experts make more use of a fast holistic approach to images, whereas novices primarily apply a slow search-to-find approach. In addition, experts took advantage of some specific search patterns in specific tasks, for example a 'drilling' search pattern through chest CT scans (see summary of chapter 5 below for an explanation of 'drilling').

Radiology Training and Assessment Methods

In the second part of the thesis, we developed and evaluated two assessment methods and one training method. In **chapter 5**, we investigated if teaching expert search strategies to junior trainees could improve perceptual performance. We included two types of search strategies in the training: 'drilling' and 'scanning'. The study group of Drew et al. distinguished these two search patterns in expert radiologists searching for lung nodules in chest CT scans. Drillers tend to focus on one area of the image and scroll up and down, before moving on to the next area, whereas scanners scan the complete image before scrolling to the next cross-section. We performed a randomized cross over design. Nineteen junior radiology residents were randomized into two groups. All participants were asked to complete three lung nodule detection tests. All participants were first exposed to a baseline free search condition, followed by the drilling and scanning instruction condition or vice versa. A mixed design ANOVA was applied to compare the three search conditions. Search strategy instruction had a significant effect on both scroll behavior and perceptual performance. The scanning instruction had a negative effect on true positive rate performance, while the drilling instruction did not have a significant effect. A questionnaire revealed that the majority of participants already used a type of drilling for their free search, which can explain the lack of effect of the drilling instruction on the true positive rate. Nonetheless, the number of false positive findings decreased after the drilling instruction, indicating a positive effect of drilling instruction on performance.

The first assessment method is the stepwise-reasoning approach, described in **chapter 6**. This method was used to identify different types of errors in the image interpretation process of medical students. In this question format, the visual diagnostic reasoning process was assessed in steps: perception with a hotspot question, analysis with a multiple-choice or long menu list containing characteristics of the abnormality and synthesis with a multiple-choice or long menu list of diseases. Radiology test results of 109 medical students were evaluated and errors were categorized in perception, analysis and synthesis errors by two observers. More than three quarters of all errors could be related to one of the three image interpretation components. Furthermore, additional performance information of the learner was revealed: partial knowledge or latent errors were identified in 25% of the patient cases. A latent error was present when a correct diagnosis was given, even though perception or analysis were incorrect. Partial knowledge was identified when an abnormality was correctly detected and/or analyzed, while an incorrect diagnosis was given.

In **chapter 7** we introduced a digital simulation-based assessment method for testing image interpretation skills of radiology residents. A digital application with advanced image manipulation tools and question types was used to simulate the image interpretation task. The assessment method was designed to increase the authenticity of image interpretation skill assessment. The method was applied to the national Dutch Radiology Progress Test (DRPT) for residents. Three test administrations of the DRPT were used to evaluate the implementation and development process. Authenticity of the digital tests was evaluated and compared with previous paper-based tests. It was a challenge to meet the standards of radiology residents who are used to working with advanced equipment. Image manipulation options, ambient lighting, time limits and question types had to be optimized in several developmental steps to reach a higher perceived authenticity than in paper-based testing.

Conclusions

The results of these studies have contributed to the understanding of the development of visual diagnostic expertise. We have identified knowledge and skills that are required for 2D and volumetric image interpretation, and showed that the knowledge and skills required for 2D and volumetric image interpretation differ. Visual search patterns differ across various levels of expertise and we found that search strategy training can improve volumetric image interpretation performance of trainees.

The results also contribute to the development of radiology training and assessment methods by providing several tools. The framework of knowledge and skills can be used as a guideline for developing radiology education, for example as a template for learning objectives or test blueprints. For teaching and assessing volumetric image interpretation skills, it is important to realize that the construct differs from 2D image interpretation. Two assessment methods were developed: a method that can be used to gain insight into error types and performance of learners and a method that can increase the authenticity of image interpretation assessment. These methods can be used in radiology education to stimulate the development of image interpretation expertise.

NEDERLANDSE SAMENVATTING

Achtergrond

Radiologie is een medisch specialisme dat zich bezighoudt met het in beeld brengen van het menselijk lichaam met verschillende beeldvormende technieken. Radiologen worden getraind om ziekten uit te sluiten of te bevestigen op basis van deze beelden (beeldinterpretatie). Het interpreteren van radiologische beelden is een complexe vaardigheid die vele trainingsjaren vereist. Fouten in beeldinterpretatie komen regelmatig voor en kunnen verstrekkende gevolgen hebben voor patiënten. Voor het verbeteren van onderwijs in de radiologie is het belangrijk om het verkrijgen van expertise in beeldinterpretatie te begrijpen.

De rol van medische beelden in de klinische besluitvorming neemt toe en de beeldvormende technieken ontwikkelen zich snel. Eén van de meest ingrijpende ontwikkelingen in de radiologie was de introductie van de digitale volumetrische beelden. Dit zijn datasets van vele, dunne dwarsdoorsneden van het lichaam, waar radiologen doorheen kunnen scrollen op een monitor. Dit verschilt van oudere technieken waarbij radiologen dwarsdoorsneden naast elkaar op een lichtbak bekeken (2D beelden). Deze technische vooruitgang heeft de visuele diagnostische taak veranderd. Hoe deze ontwikkelingen het visueel diagnostisch proces beïnvloeden en hoe we de onderwijs- en toetsmethoden hierop aan moeten sluiten is grotendeels onbekend en is onderwerp van dit proefschrift.

Dit proefschrift heeft als doel om 1) te onderzoeken welke kennis en vaardigheden belangrijk zijn voor het ontwikkelen van expertise in de radiologie en 2) bij te dragen aan de ontwikkeling en evaluatie van onderwijs- en toetsmethoden voor het bevorderen van expertise ontwikkeling in de radiologie. Beide doelen hebben betrekking op expertise in beeldinterpretatie van zowel 2D als volumetrische beelden.

Studieresultaten

Expertise ontwikkeling in beeldinterpretatie

In de eerste studie, beschreven in **hoofdstuk 2**, ontwikkelden we een raamwerk met kennis en vaardigheden die nodig zijn voor radiologische beeldinterpretatie. Het raamwerk is ontwikkeld in drie fasen, waarin het raamwerk steeds werd verbeterd. De eerste versie van het raamwerk werd gemaakt met behulp van gegevens uit de literatuur over kennis en vaardigheden die nodig zijn voor beeldinterpretatie. Deze versie werd vervolgens aan experts voorgelegd in interviews en werd verbeterd aan de hand van hun feedback. In de derde fase werd het raamwerk gespiegeld aan gegevens uit de praktijk, die werden verkregen door proefpersonen hardop te laten denken terwijl zij beelden beoordeelden. In het raamwerk worden drie componenten van het beeldinterpretatie proces onderscheiden: perceptie, analyse en synthese. *Perceptie* is 'het identificeren van radiologische bevindingen'*; analyse* is 'het beoordelen van de kenmerken van de radiologische bevindingen' en *synthese* is 'het combineren van de radiologische en klinische bevindingen voor de totstandkoming van een conclusie over de differentiaal diagnose en het beleid'. De componenten worden in het raamwerk verder gespecificeerd in zestien soorten kennis en vaardigheden, bijvoorbeeld "kennis van de anatomie", "gebruiken van efficiënte zoekstrategieën" en "onderscheid maken tussen relevante en irrelevante bevinden". Het raamwerk moet gezien worden als een vereenvoudigd model van het complexe beeldinterpretatieproces, dat onderhevig kan zijn aan veranderingen door nieuwe inzichten of ontwikkelingen.

In **hoofdstuk 3** worden de kennis en vaardigheden die gebruikt worden bij het interpreteren van 2D en volumetrische beelden met elkaar vergeleken in een hardopdenk studie. Gevorderde medisch studenten beoordeelden zowel 2D als volumetrische radiologische beelden terwijl zij hun gedachten steeds hardop uitspraken. Het raamwerk van hoofdstuk 2 werd gebruikt voor het categoriseren van de kennis en vaardigheden. De kennis en vaardigheden voor 2D en volumetrische beeldinterpretatie werden met elkaar vergeleken. Er waren significante verschillen tussen de twee typen beeldinterpretatie: perceptie was dominant bij interpretatie van volumetrische beelden, terwijl synthese juist dominant was bij interpretatie van 2D beelden. Deze bevinding hangt mogelijk samen met het verschil in de hoeveelheid beeldinformatie die de studenten moesten verwerken. Omdat volumetrische beelden meer visuele informatie bevatten, speelt perceptie een belangrijkere rol dan bij de interpretatie van 2D beelden. 2D Beelden bevatten minder visuele informatie, waardoor er relatief gezien meer ruimte over is voor synthese.

Hoofdstuk 4 is een literatuurstudie naar het verband tussen visuele zoekpatronen en diagnostische prestaties. Visuele zoekpatronen worden traditioneel onderzocht met behulp van het volgen van oogbewegingen. We zochten in zes elektronische zoekmachines naar relevante onderzoeken en hebben de resultaten verzameld en op een rij gezet. De oogbeweging studies tonen aan dat de ogen van experts sneller naar de afwijking toe gaan dan de ogen van minder gevorderden, en dat de experts minder lang nodig hebben om het beeld te bekijken. De experts maken ook langere oogbewegingen en blijven minder vaak met hun ogen 'ergens hangen'. Deze bevindingen bevestigen de gedachte dat experts de capaciteit hebben om belangrijke informatie op een efficiënte manier te verzamelen en dat zij gebruik maken van een grote visuele reikwijdte. De bevindingen zijn ook in lijn met de *global-focal search* theorieën. Deze theorieën beweren dat experts het beeld vooral op een snelle en globale manier bekijken, terwijl beginners juist een langzame, focale zoekstrategie bezigen, waarbij zij stukje voor stukje het beeld afspeuren.

Radiology training and assessment methods

In **hoofdstuk 5** hebben we onderzocht of we de perceptie van radiologen in opleiding kunnen verbeteren door middel van training in zoek strategieën. We gebruikten hiervoor twee zoekstrategieën: '*drilling*' en '*scanning*'. De onderzoeksgroep van Drew en collega`s in de United States had eerder ontdekt dat deze twee zoekpatronen gebruikt werden door radiologen voor het zoeken naar longafwijkingen in CT scans. 'Drillers' focussen hun blik op één gebied in het dwarsdoorsnede beeld, terwijl zij door de scan heen scrollen, en vervolgens focussen ze op een ander gebied in het beeld, waar zij dan weer doorheen scrollen. 'Scanners' daarentegen bekijken het complete beeld voordat zij verder scrollen naar de volgende dwarsdoorsnede. In onze studie onderzochten we negentien eerste- en tweedejaars radiologen in opleiding. Ze kregen drie perceptie testen waarin ze naar longafwijkingen moesten zoeken. De eerste test was een baseline test waarin ze hun zoekstrategie zelf mochten bepalen. Voor de tweede en derde test kregen ze een zoekstrategie instructie, één keer een *drilling* instructie en de andere keer een *scanning* instructie. De zoekstrategie bleek een significant effect te hebben op het zoekgedrag en de perceptie van de proefpersonen. De *scanning* instructie had een negatief effect op het aantal juist positieven, terwijl de drilling instructie geen significant effect had. Een vragenlijst onthulde dat het grootste deel van de proefpersonen uit zichzelf al een soort drilling strategie hanteerde. Dit kan verklaren dat het effect van de drilling instructie op het aantal juist positieve beoordelingen uitbleef. Echter, de drilling instructie had wel een verlagend effect op het aantal vals positieve beoordelingen, wat gezien kan worden als een positief effect van de instructie op de perceptie.

In **hoofdstuk 6** wordt de '*stepwise-reasoning approach*' (stapsgewijze redenering) beschreven en onderzocht. Dit is een toetsmethode die werd gebruikt met als doel om meer inzicht te krijgen in waar in het beeldinterpretatie proces fouten optreden. De *stepwise-reasoning approach* is een vraag-format waarin het beeldinterpretatie proces wordt getoetst in stappen: perceptie, analyse en synthese. De testresultaten van 109 medisch studenten werden geanalyseerd en de gemaakte fouten werden gecategoriseerd in perceptie, analyse en synthese fouten. Meer dan driekwart van de gemaakte fouten kon worden gerelateerd aan één van de drie componenten. Daarnaast kon aanvullende informatie verkregen worden over de leerprestaties van de studenten: partiële kennis en latente fouten werden vastgesteld in een kwart van de casus. Een latente fout was aanwezig wanneer een correcte diagnose werd gegeven, terwijl de afwijking niet was herkend en/of niet goed was geanalyseerd. Partiële kennis werd geïdentificeerd wanneer een afwijking juist was herkend en/of geanalyseerd, maar een onjuiste diagnose werd gesteld.

In **hoofdstuk 7** introduceren we een toetsmethode voor het testen van beeldinterpretatie met behulp van een digitale simulatie. De ontwikkelde toetsmethode maakt gebruik van volumetrische en 2D beelden uit de praktijk, die door de kandidaten interactief bekeken konden worden zoals dit ook in de klinische praktijk gebeurt (scrollen, vensterinstelling aanpassen, reconstructies maken). Hiervoor werd een digitale applicatie gebruikt met geavanceerde beeldbewerkingsmogelijkheden. De toetsmethode werd ontwikkeld om de authenticiteit van het toetsen van beeldinterpretatie in de specialistenopleiding radiologie te verbeteren. De methode werd toegepast op de voortgangstoets (VGT) radiologie. Het implementatie en ontwikkelingsproces van drie toetsafnames werd geëvalueerd. De authenticiteit van de digitale toets werd vergeleken met eerdere papieren toetsen. Vanwege de technische beperkingen was het een uitdaging om tegemoet te komen aan de hoge standaarden die worden ondervonden in de klinische praktijk. Vele factoren moesten geoptimaliseerd worden om de authenticiteit te verbeteren, zoals de beeldbewerkingsmogelijkheden, het omgevingslicht, de tijdslimieten en de vraagtypen.

Conclusies

De resultaten van deze studies dragen bij aan het begrip over de ontwikkeling van expertise in beeldinterpretatie. We hebben kennis en vaardigheden geïdentificeerd die belangrijk zijn voor het interpreteren van 2D en volumetrische beelden. We hebben aangetoond dat gebruik van deze kennis en vaardigheden verschilt tussen het interpreteren van 2D en volumetrische beelden. Zoekpatronen verschillen tussen expertise levels en het aanleren van een zoekstrategie kan de beeldinterpretatie verbeteren.

De resultaten dragen ook bij aan de ontwikkeling van onderwijs- en toetsmethoden in de radiologie door het aanreiken van verschillende instrumenten. Het raamwerk van kennis en vaardigheden kan gebruikt worden als een richtlijn voor het ontwikkelen van radiologie onderwijs, bijvoorbeeld als template voor leerdoelen of voor een toetsmatrijs. Twee toetsmethoden werden ontwikkeld die in de praktijk toegepast kunnen worden: een methode die gebruikt kan worden om beter inzicht te krijgen in typen fouten en leerprestaties, en een methode die de authenticiteit van toetsing van radiologie toetsen kan verbeteren. Deze methoden kunnen worden gebruikt om het ontwikkelen van expertise in beeldinterpretatie te bevorderen.

Appendix

DANKWOORD

In 2011 werd ik tijdens een weekenddienst in het Gelre Ziekenhuis in Apeldoorn gebeld door een collega: hij wist echt iets heel moois voor mij - een promotieplek over onderwijs in de radiologie. Een week later had ik een gesprek bij professor Van Schaik en daarna was het verzoek of ik binnen een week wilde laten weten of ik aan de promotie wilde beginnen. Het was geen moeilijke keuze. Als dochter van twee docenten heb ik een voorliefde voor onderwijs en het vak radiologie had mij in de eerste fase van mijn opleiding gegrepen. Een perfecte combinatie dus.

Ik heb geen moment spijt gehad van deze keuze. Het promotietraject heeft mij heel veel mogelijkheden gegeven, die ik dankbaar heb aangegrepen. Ik heb kansen gekregen waar ik aan het begin van het traject geen rekening mee had gehouden en ik heb ontzettend veel geleerd van de verschillende uitdagingen die je tegenkomt tijdens het promoveren.

Dit proefschrift is tot stand gekomen dankzij hulp van velen. In de eerste plaats dankzij alle studenten, coassistenten, arts-assistenten en experts die mee hebben gedaan aan de onderzoeken in dit proefschrift. Dankzij hen zijn we meer te weten gekomen over hoe radiologische kennis en vaardigheden zich ontwikkelen en hoe we het onderwijs in de radiologie kunnen verbeteren.

Ik had dit traject nooit kunnen volbrengen zonder alle geweldige begeleiders en collega's om mij heen. Jullie enthousiasme, expertise en vertrouwen hebben mij enorm gemotiveerd de afgelopen jaren.

Ik mag mijzelf gelukkig prijzen met heel veel lieve familie en vrienden. Met jullie samen zijn, daar krijg ik enorm veel energie van! Ik ben jullie dankbaar voor alle liefde die ik van jullie krijg!

Een aantal mensen zou ik graag in het bijzonder willen bedanken:

Promotor, Prof. dr. Van Schaik, Jan, bedankt voor alle mogelijkheden en kansen die je me hebt gegeven. Je bent iemand die mogelijkheden ziet en geen obstakels. Je hebt een enorme drive om te blijven innoveren waar je dat nodig acht. Jij maakt je steeds opnieuw hard voor verbeteringen in het radiologieonderwijs, en krijgt het steeds voor elkaar. Ik denk dat niet iedereen ziet wat jij allemaal doet en wat je allemaal hebt bereikt, doordat je niet graag in de belangstelling staat. Ik heb het wel gezien en ik bewonder jou hierom. Jij hebt mij steeds alle ruimte gegeven om mij te ontwikkelen en daar ben ik je heel dankbaar voor.

Promotor, Prof. dr. Ten Cate, Olle, het was een voorrecht om bij jou te mogen promoveren. Je bent een zeer integere, kundige, scherpe en intrinsiek gemotiveerde wetenschapper. Jouw ideeën worden wereldwijd toegepast in het medisch onderwijs. Jouw werk is goed doordacht en raak. De feedback die ik van jou krijg is steeds scherp en creatief, met oog voor het lezerspubliek. Daarnaast is de hoeveelheid werk die jij verzet bewonderenswaardig. Toen ik voorstelde om een tijdje bij UCSF onderzoek te gaan doen, was je meteen enthousiast. Bedankt dat je mij die kans hebt gegeven. Ik vind het bijzonder dat ik zoveel van jouw expertise heb mogen leren.

Co-promotor, Marieke, ongelofelijk wat een energie heb jij! En dan steeds blijven zeggen dat ik zoveel energie heb… Een artikel lag zelden langer dan een paar dagen bij jou op het bureau. Je kan ontzettend goed schrijven en bent een enorm gedreven onderzoeker. Ondanks jouw drukke agenda heb jij altijd wel tijd om even te bellen, mailen of af te spreken. Jouw onderwijskundige expertise is echt een hele belangrijke pijler in ons team, en jouw positieve instelling en enthousiasme werken aanstekelijk. Ik ben ontzettend blij dat je erbij bent. Bedankt voor al jouw expertise, toewijding en energie.

Co-promotor, Irene, onze geschiedenis gaat terug naar mijn studententijd, toen ik als student onderzoek deed en bij jou op de kamer kwam. Ik lustte de dropjes die jij niet lustte dus dat was een mooie basis als kamergenoten. In de tussentijd is er veel veranderd en heb ik je mogen leren kennen als een enorm gedreven persoon, net zoals je zus. Jij gáát ergens voor, dat vind ik mooi. Als co-promotor was je dan ook heel resultaatgericht en had jij altijd met een schuin oog de planning van de artikelen in de gaten. Je stelde kritische vragen en waakte ervoor dat de artikelen ook voor radiologen goed leesbaar waren. Jouw doelgerichtheid en jouw kritische blik hebben mij gedreven en bijgestuurd. Bedankt voor deze sturing tijdens mijn promotietraject.

Cécile, mijn onderzoeksmaatje, tja waar te beginnen. Jij bent onlosmakelijk verbonden aan mijn proefschrift en mijn promotietijd. We hebben ontzettend veel samengewerkt aan verschillende subsidieprojecten, onderwijsontwikkelingen en artikelen. Het liefst bespraken we alles in detail met elkaar. Vaak kwamen we in deze discussies dan weer op allerlei nieuwe ideeën die we wilden gaan oppakken. Wat een inspiratie en energie kreeg ik daar altijd van! En nog steeds! Je hebt zo'n enorme drive en daarmee lukt het je om jouw verschillende passies met elkaar te combineren. Ik heb gezien hoe hard je hebt gewerkt (en nog steeds) en heb er enorm veel bewondering voor. We vinden allebei dat we elkaar heel goed aanvullen, en daar maken we dan ook heel veel en heel graag gebruik van. Ik hoop in de toekomst nog heel veel samen te kunnen werken en samen te kunnen delen.

Louise, mijn opleidingsmaatje, wat hebben wij veel samen meegemaakt. Vooral tijdens de eerste drie jaar van onze opleiding in Apeldoorn. We begonnen tegelijk aan onze opleiding en we waren meteen een twee-eenheid. Hoewel we niet op elkaar lijken, haalden velen ons vaak door elkaar. Wat hebben wij veel (en hard) gelachen samen, zowel binnen als buiten de verslagkamer… In de Komeet, de BIT en tijdens het handsfree bellen in de auto. Waar niet eigenlijk. Het was een intensieve tijd, met veel mooie momenten waar we nog vaak aan memoreren als we afspreken met de collega's van toen. Extra bijzonder vind ik het dat jij nu in de maatschap zit waar wij beiden door zijn opgeleid. Ik ben blij dat wij opleidingsmaatjes waren!

Koen en Chris, jullie kan ik eigenlijk alleen maar gezamenlijk bedanken. Wat een duo; en wat heb ik veel lol met jullie gehad de afgelopen jaren. Met jullie is het nooit saai! We begrijpen elkaar niet altijd meteen – de ICT taal en dokterstaal komen niet altijd overeen – en dat zorgt altijd voor veel hilariteit. Jullie zijn een geweldige combinatie samen en vullen elkaar feilloos aan. Ik vond het geweldig dat jullie in San Francisco langs zijn geweest. Zonder jullie aanwezigheid was het onderzoeksproject bij UCSF niet gelukt. Natuurlijk was het ook niet verkeerd om daar een tijdje samen op te trekken. We weten onze werkbezoeken altijd goed te combineren met de geneugten van het leven; goed eten en goede wijn. Als ik er een keer niet bij kan zijn, dan krijg ik meestal wel een foto doorgestuurd met de (Italiaanse) wijn waar de keuze op is gevallen. Ik ben blij dat we de komende tijd nog veel samen gaan werken!

De sprekers van het symposium, Geoff Norman, Halszka Jarodzka, David Naeger, Cécile Ravesloot, Laura Zwaan en Ellen Kok, bedankt voor jullie bijdrage aan het symposium. Jullie zijn allen zeer gewaardeerde sprekers en ik voel me vereerd dat jullie allen meteen enthousiast waren om te komen spreken.

Dave, a special thanks to you as my supervisor during my visiting scholarship at UCSF. It was a great pleasure to work with you. You are smart, creative and goal oriented. Our brainstorm sessions were very inspiring and always gave rise to several new ideas for papers and research projects. The way you inspire young doctors and how you give them their first experience with radiology is very special. I hope we will get the chance to continue our collaboration in the future.

De staf van de divisie beeld, jullie creëren een klimaat waarin de combinatie van patiëntenzorg, onderzoek en onderwijs wordt gestimuleerd en wordt ondersteund. Ik voel me bevoorrecht dat ik door jullie ben opgeleid en ben ontzettend blij dat ik jullie mag gaan versterken. Als ik aan de divisie beeld denk, en als ik erover praat met anderen, dan ben ik trots op de afdeling. Ik vind het bijzonder dat je je op onze afdeling kan ontwikkelen op een manier die bij je past. Als je ambitie hebt, dan krijg je de ruimte om hier iets mee te doen. Dat heb ik als ontzettend stimulerend ervaren, bedankt!

De neuro-sectie, jullie wil ik in het bijzonder bedanken voor jullie support de afgelopen jaren. De afronding van mijn proefschrift moest gecombineerd worden met de kliniek. Dat is gelukt mede dankzij jullie flexibiliteit, collegialiteit en de goede werksfeer. Iedereen staat steeds voor elkaar klaar om van alles voor elkaar over te nemen en met elkaar mee te kijken. De consensus meetings zijn zo'n beetje een dagelijkse aangelegenheid. Het niveau van de neuroradiologie ligt heel erg hoog en ik ben heel blij dat jullie mijn leermeesters zijn. Het is een feest om met jullie te mogen werken!

Dik, nog even een apart woordje aan jou gericht: jij hebt een bijzondere bijdrage gehad aan mijn promotietraject, vanuit de examencommissie en als differentiatie- en fellowshipopleider van de parallelle opleidingstrajecten. Ik waardeer je om jouw heldere kijk op zaken en jouw zuivere manier van redeneren. Ook waardeer ik je om jouw enorme betrokkenheid, interesse en droge humor. Ik heb ontzettend veel zin om samen met jou het onderwijs avontuur aan te gaan.

De maatschap radiologie in Apeldoorn, jullie hebben een belangrijke bijdrage gehad aan mijn ontwikkeling als radioloog. Ik vind jullie een hele bijzondere opleidingsgroep, met veel oog voor kwaliteit en persoonlijke aandacht. Jullie geven veel prioriteit aan opleiden en proberen bij iedereen alles eruit te halen wat erin zit. In Apeldoorn wordt je zogezegd ook 'breed opgeleid' wat betekent dat er ook veel extra-curriculaire activiteiten zijn. Ik heb mijn opleidingstijd in Apeldoorn ervaren als een hele intensieve periode waarin we een hechte band met elkaar hebben opgebouwd. Ik denk dan ook nog vaak met heel veel plezier terug aan mijn tijd in het Gelre!

Leden en oud-leden van de examencommissie, bedankt voor de goede samenwerking bij het digitaliseren van de voortgangstoets (VGT). Jullie enthousiasme en flexibiliteit waren de sleutel voor het slagen van het project. Ik ben erg dankbaar voor al het extra werk dat jullie hebben verzet om het project te laten slagen. De examencommissie van de NVvR loopt voorop in toetsing in de specialistenopleiding. Ik ben dan ook heel blij dat ik inmiddels ook deel mag uitmaken van deze mooie commissie en hoop dat we samen nog veel verbeterslagen zullen maken.

SURF projectleden, ik wil iedereen die mee heeft gewerkt in de SURF subsidieprojecten heel erg bedanken voor de goede, vruchtbare samenwerking. De sfeer in de projectteams was heel positief, en iedereen was bereid om een stapje extra te zetten als dat nodig was. Deze instelling heeft geleid tot hele mooie projectresultaten, waar jullie trots op kunnen zijn. Bedankt voor de mooie tijd en hopelijk ontmoeten we elkaar weer in toekomstige projecten!

SURF en haar medewerkers, bedankt voor jullie bijdragen aan de mooie subsidieprojecten die mede door jullie steun mogelijk zijn gemaakt. Bedankt dat de deur altijd open stond voor vragen, en voor het delen van alle kennis en kunde.

Bobby, Mirja en Larissa, bedankt voor jullie bijdragen als (destijds) studenten onderwijskunde. Jullie zijn inmiddels allemaal zeer verdienstelijk afgestudeerd en allemaal gaan promoveren! Zonder uitzondering hebben jullie prachtige onderzoeksprojecten afgerond. Bobby, jij bent intussen zelf gepromoveerd en bent mijn statistische vraagbaak geworden. Je bent zo goed in statistiek dat je zelf nieuwe statistische methoden bedenkt. En het bijzondere is, dat je het ook nog heel goed kan uitleggen! Dat is een bijzondere combinatie en daar maak ik dankbaar gebruik van. Bedankt voor al jouw hulp en hopelijk kunnen we nog veel samenwerken in de toekomst.

Mede-onderzoekers, bedankt voor jullie collegialiteit, behulpzaamheid en gezelligheid. De gezamenlijke lunches, koffie-momentjes en andere festiviteiten hebben mijn promotietijd extra leven in geblazen.

Kamergenootjes van de vissenkom, bedankt voor de mooie jaren! De eerste jaren heb ik periodes intensief in de vissenkom doorgebracht en hebben we veel met elkaar gedeeld, zowel mooie als verdrietige momenten. Ik heb veel mooie herinneringen aan de vissenkom: submissie-chips, de vele rondjes thee, en de vissenkomuitjes. De laatste jaren kwam ik soms even binnenvliegen of een dagje met mijn koptelefoon op typen, omdat de uurtjes die ik aan mijn proefschrift kon besteden kostbaar waren. Het was fijn dat daarmee rekening werd gehouden en dat er hard gewerkt kon worden in de vissenkom! Bedankt voor alle gezelligheid en support.

Josephine, voor jou een speciaal bedankje: voorafgaand aan je radiologie-opleiding heb je een tijdje meegewerkt aan onderwijs- en onderzoeksprojecten van Cécile en mij. Je bent een heel harde werker, die altijd aan anderen denkt. Je hebt niet alleen heel veel data verzameld, je dacht ook ontzettend goed mee in de uitvoering van de onderzoeken. Jouw hulp was onmisbaar! Daarnaast ben je een ontzettend sociaal persoon met oog voor de medemens. Bedankt voor jouw hulp en betrokkenheid!

Alle aios waarmee ik heb gewerkt, bedankt voor de goede samenwerking, flexibiliteit, collegialiteit en ook heel veel gezelligheid in de afgelopen jaren. Veel van jullie zijn niet alleen collega's, maar ook vrienden geworden door de tijd die we samen hebben gewerkt. Iedereen is inmiddels uitgewaaid, maar gelukkig blijven we elkaar nog zien op congressen en andere gelegenheden. De eetclub: het is heerlijk om met jullie samen verhalen op te halen van onze opleidingstijd in Apeldoorn. De avonden met jullie zijn goud waard! Laten we dat erin houden!

De onderwijsassistenten, jullie wil ik enorm bedanken voor jullie hulp en medewerking in de verschillende SURF projecten. Jullie hebben er mede voor gezorgd dat deze projecten

geslaagd zijn en dat er meerdere verbeteringen in de onderwijspraktijk konden worden doorgevoerd.

Deelnemers van CTL, ik heb verschillende generaties meegemaakt en was vanwege de combinatie met de kliniek wisselend aanwezig. Ik heb de discussies tijdens CTL ervaren als een verrijking. Het was mooi om iets mee te krijgen van alle andere onderzoeken in het medisch onderwijs in Utrecht en goed om het onderzoek in een groter geheel te kunnen zien. Allen bedankt voor het discussiëren, meedenken en inspireren! Eugène Custers, bedankt voor jouw jarenlange inzet voor CTL en OIP, en voor het meelezen van enkele van mijn stukken.

The PhD students for Medical Education at UCSF, thank you for having me in your doctoral seminars. It was a pleasure to participate in the program during my time at UCSF: I was impressed by the level of the discussions and the level of your scientific work and I felt honored to be part of it. Patricia O`Sullivan, Pat, a special thanks to you for your hospitality and for giving me the opportunity to participate in all sorts of Medical Education meetings at USCF.

Het NVvR bureau, jullie hebben een hele belangrijke rol gespeeld bij het SURF project aangaande de digitalisering van de VGT . Mede dankzij jullie is de digitale VGT nu de nieuwe onderwijspraktijk. Ik vond het een hele fijne samenwerking. Ivonne, ik vind het heel knap hoe jij je de materie in korte tijd hebt eigen gemaakt, terwijl we middenin de transitie zaten. Je bent een goede, stabiele basis waar je op kan bouwen en jouw werk is precies en doordacht. Ik ben blij dat ik weer met je mag samenwerken vanuit de examencommissie.

Het onderwijs-secretariaat, Linda en Amanda, ik kan me eigenlijk niet herinneren hoe het was voordat jullie er waren. Voor mij zijn jullie het onderwijs-secretariaat en zijn jullie dat altijd geweest. Jullie hebben in vele verschillende onderwijsprojecten en lopende onderwijszaken een heel belangrijke rol. Dit geldt voor het lokale, regionale en landelijke onderwijs. De kennis en ervaring die jullie op dat front hebben is heel waardevol en ik ben blij dat ik daar gebruik van heb kunnen maken. Ondanks dat jullie het heel druk hadden hebben jullie mij geholpen bij de organisatie van het symposium, wat heel erg fijn was. Bedankt voor al jullie hulp!

Het stafsecretariaat, ook jullie zijn voor mij een vaste baken op het werk. Jullie staan altijd klaar voor anderen. Als ik bij jullie binnenloop, geven jullie me het gevoel dat ik de eerste ben die dag (wat natuurlijk niet zo is want men loopt aan de lopende band binnen). Dat is een uitstraling die uniek is en die ik enorm waardeer. Daar mogen jullie trots op zijn!

Het bedrijfsbureau, ik heb de afgelopen jaren veel met jullie te maken gehad vanwege de vele subsidieprojecten. Aangezien ik zelf nogal weinig verstand heb van financiën ben ik erg blij dat jullie mij daarbij ondersteund hebben! Nieuwe termen als matching, overhead en derde geldstroom kregen door jullie betekenis. Marieke Krispijn, voor jou speciale dank voor het ondersteunen van de SURF projecten. Dat was een behoorlijke klus en zonder jou had ik het niet gekund.

Medewerkers van de fotografie/multimedia, het is altijd een feestje om bij jullie binnen te lopen en jullie staan altijd voor een ander klaar. Jullie hebben mij op vele momenten geholpen bij allerlei uiteenlopende zaken in mijn promotietraject. Ook nu bij de organisatie van het symposium waren jullie weer onmisbaar, voor het maken van foto's, de website en uitnodigingen. Bedankt dat jullie er zijn!

De 57000, wat moet een promovendus zonder jullie? Als de computer je in de steek laat, staan jullie altijd klaar om de boel zo snel mogelijk op te lossen. De magie van het opnieuw opstarten is ongelofelijk, maar als dat niet helpt, dan weet ik het echt niet meer en ben ik blij dat ik jullie heb! Bedankt!

Vrienden en vriendinnetjes, vrienden van vroeger, de studententijd, de jaarclub, mijn roeiploeg en van de schaatsclub, jullie zijn een heel belangrijk deel van mijn leven. Eén van mijn grootste hobby's is afspreken met jullie, en dan het liefst rond borreltijd. De afgelopen maanden is het er wat minder van gekomen door de afronding van mijn proefschrift en ben ik wat minder proactief geweest in onze vriendschappen. Toch bleven jullie mij allemaal trouw en waren jullie er voor me. Na mijn promotie zal ik er ook weer meer voor jullie zijn, ik kijk ernaar uit!

Lieve familie, aanhang en kids, wat is het fijn als je familie ook je vrienden zijn, en je ze precies zo uit had gekozen als dat had gekund. We hebben allemaal een druk bestaan maar toch plannen we regelmatig samen borrels, etentjes of vakanties en dat vind ik heel waardevol. Jullie zijn heel belangrijk voor mij. Jullie zijn altijd heel betrokken geweest bij het promotietraject, leefden met mij mee en waren geïnteresseerd in alle verhalen over artikelen, presentaties en subsidieaanvragen. Rutger, ik vind het heel bijzonder dat jij mijn kaft wilde ontwerpen. Jouw oog voor fotografie en design komt heel mooi naar voren in het ontwerp. Bedankt voor deze mooie bijdrage aan mijn proefschrift.

Pap en mam, jullie hebben mij altijd gestimuleerd om alles uit mezelf te halen. Maar zonder het mij op te leggen, altijd vanuit het positieve en door mij aan te moedigen datgene te kiezen wat ik zelf graag wilde. Het onderwijs zit in de genen en ik wilde eigenlijk juf worden. Dat is het uiteindelijk toch niet geworden. Maar het onderwijs blijft kriebelen, en ik heb een manier gevonden om dat in mijn werk een prominente plek te geven. Jullie hebben mij altijd geholpen bij het maken van keuzes, en mij gesteund in de weg die ik inging. Jullie zijn oprecht geïnteresseerd in wat ik doe. Ik ben dankbaar voor jullie onvoorwaardelijke liefde. Ik had me geen betere ouders kunnen wensen.

Herman, zonder jou was het mij zeker niet gelukt. Jij kent mij echt, je waardeert mij zoals ik ben, en je weet mij door moeilijke periodes heen te helpen. Ik geniet van ons leven en de grapjes en gewoonten die we samen hebben. Je bent de beste en de liefste. Je weet niet half hoeveel ik van je hou.

LIST OF PUBLICATIONS

Journal Publications

Van der Gijp, A., Van der Schaaf, M.F., Van der Schaaf, I.C., Huige, J.C.B.M., Ravesloot, C.J., Van Schaik, J.P.J., Ten Cate, Th. J. Interpretation of radiological images: towards a framework of knowledge and skills. *Advances in Health Sciences Education.* 2014; 19(4): 565-80.

Van der Gijp, A., Ravesloot, C.J., Van der Schaaf, M. F., Van der Schaaf, I.C., Huige, J.C.B.M., Vincken, K.L., Ten Cate, Th. J., Van Schaik, J.P.J. Volumetric and two-dimensional image interpretation show different cognitive processes in learners. *Academic Radiology.* 2015; 22(5): 632-9.

van der Gijp, A., Ravesloot, C.J., Jarodzka, H., van der Schaaf, M.F., van der Schaaf, I.C., van Schaik, J.P.J., ten Cate, Th.J. How visual search relates to visual diagnostic performance: a narrative systematic review eye-tracking research in radiology. *Advances in Health Sciences Education.* 2017; 22(3): 765-787.

Van der Gijp, A., Vincken, K.L., Boscardin, C., Webb, E.M., ten Cate, Th.J., Naeger, D.M. The effect of teaching search strategies on perceptual performance. *Academic Radiology.* 2017; 24(6):762-767.

van der Gijp, A., Cécile J. Ravesloot, C.J., ten Cate, Th.J., van Schaik, J.P.J., Webb, E.M., Naeger, D.M. Tests, quizzes and self-assessments: how to construct a high quality exam. *AJR.* 2016; 1: 1-5.

van der Gijp, A., Webb, E.M., Naeger, D.M. 2017. How radiologists think: understanding fast and slow thought processing and how it can improve our teaching. *Academic Radiology.* 2017; 24.

Ravesloot, C.J., **Van der Gijp, A.,** Van der Schaaf, M., Huige, J.C.B.M., Vincken, K.L., Mol, C.P., Bleys, R.L.A.W., Ten Cate, Th.J., Van Schaik, J.P.J.. Support for external validity of radiological anatomy tests using volumetric images. *Academic Radiology.* 2015; 22(5):640-5.

Ravesloot C.J., **Van der Gijp, A.,** Vincken, K.L., Van Schaik, J.P.J. 2014. [Improved radiology testing with VQuest]. *Medisch Contact* **20**: 1010-11. Dutch.

Ravesloot, C.J., Van der Schaaf, M.F., Kruitwagen, C.L.J.J., **Van der Gijp, A.,** Rutgers, D.R., Haaring, C., Ten Cate, Th.J., Van Schaik, J.P.J.. Predictors of knowledge and image interpretation skill development in radiology residents. *Radiology. 2017 Apr 11. Epub ahead of print.*

Ravesloot, C.J., Van der Schaaf, M., Van Schaik, J.P.J., Ten Cate, Th.J., **Van der Gijp, A.,** Mol, C., Vincken, K.L.. Volumetric Images Improve Testing of Radiological Image Interpretation Skills. *European Journal of Radiology.* 2015; 84(5):856-61.

Den Harder, A. M., Frijlingh, M., Ravesloot, C.J., Oosterbaan, A.E., **Van der Gijp, A.**. The importance of human computer interaction in radiology e-learning. *Journal of Digital Imaging.* 2016; 29(2):195-205.

Stuijfzand, B.G., Van der Schaaf, M.F., Kirschner, F.C., Ravesloot, C.J., **Van der Gijp, A.,** Vincken, K.L.. Medical Students' Cognitive Load in Volumetric Image Interpretation: Insights from Human-Computer Interaction and Eye Movements. *Computers in Human Behavior*. 2014; 62: 394-403.

Ravesloot, C.J., **Van der Gijp, A.,** Van der Schaaf, M.F., Huige, J.C.B.M., Ten Cate, Th. J., Vincken, K.L., Mol, C.P., Van Schaik J.P.J. Identifying error types in visual diagnostic skill assessment. *Diagnosis. 2017; 4(2)*

Van der Gijp, A., Ravesloot, C.J., Tipker, C.A., de Crom, K., Rutgers, D.R., van der Schaaf, M.F., van der Schaaf, I.C., Mol, C.P., Vincken, K.L., ten Cate, Th.J., Maas, M., van Schaik, J.P.J. Increasing authenticity of simulation-based assessment in diagnostic radiology. *[Revisions submitted Simulation in Health Care]*

Book Publications

Van der Gijp, A., Ravesloot, C.J., Van der Schaaf, M.F., Tipker, C.A., de Crom, K., Van den Berg, J.W., Vincken, K.L., Mol, C.P., Van der Schoot, P.J., Van der Schaaf, I.C., Rutgers, D.R., Maas, M., Van Schaik, J.P.J. 2014. [Recommendations for implementation of digital assessment with images]. Utrecht, The Netherlands: Digitalis. Dutch.

Ravesloot, C.J., **Van der Gijp, A.,** Van der Schaaf, M.F., Vincken, K.L., Mol, C.P., Bleys, R.L.A.W., Huige, J.C.B.M., Ten Cate, Th.J., Van Schaik, J.P.J. 2014. [Digital assessment of radiological image interpretation]. Utrecht, The Netherlands: Digitalis. Dutch.

CURRICULUM VITAE

Anouk van der Gijp was born on the 18th of October 1983 in Apeldoorn, the Netherlands. She completed high school in 2001 (cum laude, Stedelijk Gymnasium Apeldoorn). That same year, she moved to Utrecht to study medicine at the University of Utrecht. She received her medical degree in 2008 and started her radiology residency. During the first three years of her residency, Anouk was trained in Gelre Ziekenhuizen in Apeldoorn, under the supervision of dr. J.W.C. Gratama. During the remaining two years of her residency, she was trained in the University Medical Center Utrecht (UMCU) under the supervision of Prof. dr. J.P.J. van Schaik and dr. R.A.J. Nievelstein. In 2011, Anouk started a PhD track in Medical Education at the UMCU under the supervision of Prof. J.P.J. van Schaik, Prof. Th.J. ten Cate, dr. M.F. van der Schaaf en dr. I.C. van der Schaaf. During her PhD training, she was project leader and member in several funded projects that aimed to improve radiology education. In 2015, Anouk worked in the USA as a visiting scholar at the Radiology Department of the University of California, San Francisco (UCSF). She completed several research projects and was a guest teacher at UCSF medical school under the supervision of D.M. Naeger, MD, Associate Professor of Radiology. She returned to the Netherlands in 2016 to finish radiology training and start a fellowship in neuroradiology at the UMCU under the supervision of dr. D.R. Rutgers. After completing her fellowship in the summer of 2017, Anouk joined the radiology staff of the UMCU as a faculty member.