ASE-15-0187.R2

Research Report

Exploring eye movements of experienced and novice readers of medical texts concerning the central cardiovascular system in making a diagnosis

Henna Vilppu[1,2], Mirjamaija Mikkilä-Erdmann[1,2], Ilona Södervik[1,2], Erika Österholm-Matikainen[3]

[1]Department of Teacher Education, Faculty of Education, University of Turku, Turku, Finland

[2]Centre for Learning Research, Faculty of Education, University of Turku, Turku, Finland

[3]Medical Education Research and Development Centre (TUTKE), Faculty of Medicine, University of Turku, Turku, Finland

Running title: Exploring eye movements of medical text readers

Correspondence to: Ms. Henna Vilppu, Department of Teacher Education and Centre for Learning Research, University of Turku, FI-20014, Turku, Finland. E-mail: henna.vilppu@utu.fi

ABSTRACT

This study used the eye-tracking method to explore how the level of expertise influences reading, and solving, two written patient cases on cardiac failure and pulmonary embolus. Eye-tracking is a fairly commonly used method in medical education research, but it has been primarily applied to studies analyzing the processing of visualizations, such as medical images or patient video cases. Third-year medical students (n = 39) and residents (n = 13) read two patient case texts in an eye-tracking laboratory. The analysis focused on the diagnosis made, the total visit duration per text slide, and eye-movement indicators regarding task-relevant and task-redundant areas of the patient case text. The results showed that almost all participants (45/48) made the correct diagnosis of the first patient case, whereas all the residents, but only 17 students, correctly diagnosed the second case. The residents were efficient patient-case-solvers: they reached the correct diagnoses, and processed the cases faster and with a lower number of fixations than did the students. Further, the students and residents demonstrated different reading patterns with regard to which slides they proportionally paid most attention. The observed differences could be utilized in medical education to model expert reasoning and also to teach the manner in which a good medical text is constructed. Eye-tracking methodology appears to have a great deal of potential in evaluating performance and growing diagnostic expertise in reading medical texts. However, further research using medical texts as stimuli is required.

Keywords: Undergraduate medical education; internal medicine residency training, cardiovascular sciences; text processing; eye-tracking; patient cases; expertise

INTRODUCTION

This study used the eye-tracking method to explore how the level of expertise influences reading, and solving, two written patient case studies on the central cardiovascular system, one of the most important subjects of the anatomical sciences. Eye-tracking is a fairly commonly used method in medical education research, but it has primarily been used for the study of medical image perception (Krupinski, 2010; Bertram et al., 2013; Gegenfurtner et al., 2013; Jaarsma et al., 2014, 2015; Zumwalt et al., 2015) and, more recently, for the evaluation of dynamic stimuli, such as patient video cases (Jarodzka et al., 2012; Khan et al., 2012). Therefore, the contribution of the current study lies in using clinically relevant written material, requiring an understanding of basic anatomy and physiology, as a stimulus.

The everyday duties of a medical doctor involve the activation of basic anatomical, physiological, and clinical knowledge while processing medical texts and documents, such as epicrises, referrals, medical records, journal articles, etc. These medical texts are extraordinarily complex documents that have their own rules and structures (Charon, 2000). Further, rapid developments in biological science and the treatment of clinical diseases, as well as the growing impact of the Internet and other health information technologies, are leading to changes in medicine and healthcare. These force clinical practitioners to keep up with the latest scientific developments, which means researching, reading, evaluating, comprehending, and incorporating the new information in patient care practices (Kaufman et al., 2008). Thus, text processing and learning from texts are viewed as important skills in present-day medical practice and as highly relevant for anatomic education.

In a traditional medical curriculum, basic science knowledge, such as biochemistry, anatomy, and physiology, is taught during the first 2 years of medical school, and is followed by clinical studies relating to knowledge of diseases, investigative procedures, and therapeutic management, as well as appropriate clinical skills. This multidisciplinary nature of medical knowledge is one aspect that makes learning medicine a challenge for many, and sometimes it is hard for students to see the value of basic sciences in doctors' everyday routine (Custers and Cate, 2002; Wilhelmsson et al., 2010). There appears to be a consensus regarding the value of biomedical knowledge as a foundation for clinical knowledge, but the precise role of biomedical knowledge in clinical reasoning remains a subject of debate (Kaufman et al., 2008). Recent studies have indicated that the biomedical basic science knowledge taught at the beginning of medical studies plays a central role in students' clinical reasoning (Ahopelto et al., 2011; Nivala et al., 2013). For example, it has been suggested that biomedical understanding acts as a mediator in such reasoning, since students with misconceptions in biomedical knowledge also appear to perform poorly in clinical reasoning, while those with excellent clinical reasoning skills remarkably improve their biomedical knowledge during preclinical studies (Ahopelto et al., 2011). Further, Nivala et al. (2013) confirmed the long-term value of basic science studies in the preclinical phase by showing that performance in diagnostic pathology was predicted by students' prior knowledge of histology, even after a relatively long time delay between measurements.

On the development of medical expertise

During the last 40 years, researchers have attempted to formulate a theory of the development of medical expertise, and thus the manner in which expert doctors address medical problems (Schmidt and Rikers, 2007). In the light of empirical generalizations, experts appear to be faster than novices in performing the skills of their own domain; for example, they solve problems quickly with little error, have a superior short- and long-term memory, and see and represent problems at a deeper level than do novices (Chi et al., 1988; Schmidt and Boshuizen, 1993a). It has been suggested that domain-specific knowledge in the form of schemas is the primary factor that distinguishes experts from novices with regard to problem-solving skills (Sweller, 1988). Experts' cognitive schemas are increasingly sophisticated, enabling them to identify, store, and retrieve large meaningful chunks of domain-specific information (Kalyuga et al., 2012).

One explanation for expert-novice differences in the medical context has been offered by encapsulation theory (Boshuizen and Schmidt, 1992; Schmidt and Boshuizen, 1993a; Boshuizen and Schmidt, 2008), which suggests that medical expertise develops via three stages. At first, medical students build knowledge networks consisting of lower-order, biomedical concepts. Direct lines of reasoning are constructed between different concepts through knowledge accretion and validation. When students obtain practical experience in the clinic, these lines of reasoning are repeated and intermediate concepts are skipped to form higher-order concepts, thereby the knowledge becomes gradually encapsulated. In the last stage of expertise development, so-called illness scripts replace the encapsulated network. Illness scripts are even more sophisticated clusters of knowledge that contain only little knowledge of pathophysiological causes of symptoms and complaints, but a great deal of clinically relevant information regarding the enabling conditions of disease (Schmidt and Rikers, 2007; Boshuizen and Schmidt, 2008). Thus, according to encapsulation theory, the lines of reasoning become shorter during the development of medical expertise, assuming that the findings are routine and correspond with the schema. However, knowledge structures acquired in different phases of development settle into "layers", which can be accessed when the more recently acquired structures fail to produce an adequate solution to a clinical problem, when encountering uncommon, non-routine cases, and in cases of uncertainty (Schmidt et al., 1990).

It is generally assumed that there is a linear relationship between level of expertise and memory performance, and this has been shown in numerous domains (Vicente and Wang, 1998). However, the relationship does not appear to be as linear in the medical domain, but rather resembles the shape of an inverted U (Rikers et al., 2000; Kalyuga et al., 2012). This is often referred to as the so-called intermediate effect, according to which medical students of an intermediate level of expertise, that is, advanced students, remember more and explain the signs and symptoms of a clinical case more accurately than experts or novices (Schmidt and Boshuizen, 1993b; Rikers et al., 2000). In accordance with encapsulation theory, the intermediate effect can be explained on the basis of the qualitatively different forms of medical knowledge that students and experienced physicians make use of while diagnosing a patient. Experts appear to be capable of immediately recognizing the signs and symptoms of the case and linking them in order to reach a diagnostic conclusion, without explicitly referring to biomedical knowledge. However, students, having insufficient clinical experience, must engage in detailed biomedical reasoning in order to be capable of understanding the signs and symptoms, and to possibly arrive at a correct diagnosis (Kalyuga et al., 2012).

Eye-tracking, medical expertise, and reading research

Eye-tracking research has shown that allocation of attention is often influenced by expertise (Reingold and Sheridan, 2011). Thus, many of the eye-tracking studies conducted in the medical domain focus on differences between practitioners with different levels of expertise (Khan et al., 2012; Bertram et al., 2013; Gegenfurtner et al., 2013; Jaarsma et al., 2014, 2015). In an overview of visual expertise research (Reingold and Sheridan, 2011) showed that the superior performance of experts was associated with their superior encoding of domain-related configurations.

Moreover, experts appear to be capable of global processing, that is, simultaneous processing of large parts of an image, and they also fixate on relevant parts of the image faster, and spend more time looking at them than do novices. In addition to expertise, the level of the task difficulty also appears to influence eye movements, at least in visual search tasks (Rayner, 1998). For example, it has been concluded that fixation frequency in an area indicates the degree of importance, while fixation duration can be interpreted as an indicator of the complexity and difficulty of a visual display (Fitts et al., 1950). Further, in studies focusing on comprehension of visualizations, encapsulation theory (Boshuizen and Schmidt, 1992), among others, has been used to explain expertise differences. Since it suggests that the development of expertise leads to an organization of clinical concepts in the form of encapsulated chunks, these qualitative changes in memory structures should be revealed in verbal protocols (Gegenfurtner et al., 2013).

Although several studies on medical text reading have been conducted, relating to radiology (Sistrom and Honeyman-Buck, 2005; Krupinski et al., 2012), for example, those that integrate eye-tracking appear to be scarce. A systematic review of the literature using the keywords and phrases "eye-tracking" or "eye movement", and "text" or "text reading" and "medical students" or "medical", revealed few medical text reading studies that have utilized the eye-tracking method [PubMed (US National Library of Medicine, National Institutes of Health, Bethesda MD), Education Resource Information Center (ERIC) (EBSCO Industries, Inc., Birmingham, AL), on November 2015]. However, the recording of readers' eye movements is relevant in the study of spontaneous reading strategies, since eye movements are an integral part of normal reading (Rayner, 1998; Hyönä and Nurminen, 2006). Further, the method allows readers to freely inspect the text in a highly authentic environment without interruptions, and it yields a temporally and spatially precise record of where readers look in the text and for how long (Hyönä and Nurminen, 2006; Kliegl et al., 2006).

Reading is a complex process that integrates visual, attentional, language-related, and oculomotor processing (Kliegl et al., 2006). While reading, the eyes move continually as a sequence of rapid saccades and fixations, during which the eyes remain relatively still (Rayner, 1998; Kliegl et al., 2006; Holmqvist et al., 2011). Information intake largely occurs during fixations. It is a common belief that human eye movements indicate ongoing mental processes while people are interacting with visual environments, such as texts, illustrations, or animations. Several researchers have accepted the so-called eye-mind hypothesis (Just and Carpenter, 1980), which proposes a close relationship between the direction of gaze and the focus of attention. In the level of eye movements, this would mean that readers retain a fixation on a word for as long as the word is processed (Kliegl et al., 2006). However, there are also alternative views. For example, Kliegl et al. (2006) showed evidence for the distributed processing of words across fixation durations. Thus, it appears that, most of the time, the mind processes several words in parallel at different perceptual and cognitive levels.

The history of eye movement research on reading is long and rich, and reading of different texts has been studied from various perspectives, such as via reading strategies (Goldman and Saul, 1990; Hyönä et al., 2002; Hyönä and Nurminen, 2006), and text structure effects on eye-movement patterns (Hyönä and Niemi, 1990; Vauras et al., 1992; Klusewitz and Lorch, 2000; Wiley and Rayner, 2000). For example, it has been found that fixation duration is a successful predictor of reading comprehension, in that skilled readers' fixations are more brief than those of less skilled readers (Underwood et al., 1990). In addition, it appears that when the text becomes comparatively difficult, longer fixations are generated (Rayner, 1998; Rayner and Slattery, 2009). Further, highly important sentences seem to receive more visual attention than those which are unimportant (Hyönä and Niemi, 1990).

The current study

This study used eye-tracking methodology to examine how the level of expertise influenced differences between students and internal medicine residents in reading, and solving, two written patient cases concerning cardiovascular medicine. This particular topic was chosen because previous research has shown that students have difficulties in understanding the central cardiovascular system, even at university level (Michael, 1998; Michael et al., 1999, 2002; Ahopelto et al.,2011; Mikkilä-Erdmann et al., 2012). Further, the solving of clinical patient cases requires an understanding of the basic anatomy and physiology of the central cardiovascular system. Thus, it was considered that an examination of the two texts would illuminate medical students' and residents' learning of basic science and its representation in the clinical context. In addition, it was examined as to whether the participants focused more on the task-relevant parts of the text than on the irrelevant areas, in accordance with published studies on visual expertise research. Thus, the research questions were: (1) What kind of differences exist between medical students and residents in solving the patient cases? (2) How do medical students' and residents' reading processes differ in duration? (3) What kind of differences exist between medical students and residents in focusing on the task-relevant and task-redundant parts of the patient case texts? The given diagnosis, total visit duration per text slide, and eye-movement indicators (number of fixations and average duration of fixations) on task-relevant and task-redundant areas of the patient case texts were analyzed. On the basis of encapsulation theory and eye-tracking research, it was hypothesized that the residents would more often correctly solve the patient case than would the students, and would do so more quickly, and that they would also more effectively direct their attention to the task-relevant areas of the text than would the students, since the case was related to the residents' area of expertise.

MATERIALS AND METHODS

Participants

Participation in the study was voluntary, and 39 third-year medical students and 13 internal medicine residents were recruited. The third-year medical students were chosen because they had just started their clinical studies, after 2.5 years of biomedical studies. Thus, considering their lack of clinical experience, they were thought to accurately represent a group of novices. The internal medicine residents were selected because they represented a more experienced group, since they had completed their 6-year basic medical degree studies and had already began their specialization training. Thus, their clinical experience was far more extensive than that of the students. The participating students represented 32% of the third-year medical student population, with the total size of the cohort being 122. The residents represented the total number of residents currently working in the internal medicine ward. The age of the students varied from 22 to 35 years (mean 24.3 ± 2.5). The residents' experience in specialization training was an average of 4 years, with a range of 1 to 7 years. All the participants had gone through an extremely selective entrance exam (in Finland, only approximately 15% of applicants are accepted by medical schools) and they were native Finnish speakers. Of the participants, 35 were women (27 students and 8 residents), and 17 were men (12 students and 5 residents). Approval for the study was obtained from the ethics review board of the University of Turku, Turku, Finland.

Setting and apparatus

The eye-tracking laboratory was established on medical school premises for the medical students and in the internal medicine clinic at the university hospital for the residents. The residents took part in the study during their working hours at the clinic. Thus, the research was carried out in a relatively authentic setting. The participants read two patient cases on a high-resolution 24" computer display operating at 60 Hz at a resolution of 1,920 x 1,200 pixels, in which the Tobii T60XL Eye Tracker (Tobii Technology, Inc., Falls Church, VA) was integrated. The display had infrared cameras that recorded the participants' eye movements during the text reading. The accuracy of the eye tracker was 0.6°. To ensure that the reading situation was as comfortable and as natural as possible, no supporting chinrest was used, since the eye tracker allows even large head movements. (Note that the accuracy of the eye tracker and its robustness toward head movements are defined by the manufacturer (Tobii Technology, Inc., Falls Church, VA) and were not tested by the authors themselves.)

At the beginning of the study, each participant was seated approximately 60 cm from the screen, and the eye tracker was calibrated. The apparatus was placed on an electrically adjustable table, and the participants sat on a manually adjustable chair, for easier calibration of the apparatus and a comfortable reading position. Instructions regarding the session were then given. The participants were told that they would have to read two patient case texts, each of which was divided across three PowerPoint slides. After the participants left each textual slide, they would see a question slide prompting them to diagnose the case. The students wrote their diagnosis on a sheet of paper, and the residents gave an oral answer that was recorded (the purpose of the difference was only to save the residents' time, as they were taking part in the study during their busy working hours). The participants were told that they could not go back after they had left a text slide. Thus, the answer to the question concerning the diagnosis was given without seeing the text slide. There was no time limitation on reading the patient case texts or answering the questions, and the participants were told to move at their own pace. Between reading the two patient cases, the apparatus was recalibrated in order to ensure the quality of the data. The participants were given an opportunity to state whether they had any concerns about the session before the experiment started. When the session began, the instructor moved to the back of the room.

Materials

Two patient case texts were written by the authors and reviewed by two cardiology specialists. The first patient case was about cardiac failure and the second case concerned pulmonary embolus. The texts were in Finnish, but English translations of the first patient case text can be found in Appendix 1. The Finnish text of patient case 1 comprised 199 words, and that of patient case 2 consisted of 225 words. The font used in the slides was Arial 17 pt. The idea of the patient case texts was to simulate the phases of a patient encounter in a healthcare center, which is why the cases were divided into three PowerPoint slides. The first slide contained the medical history of the patient, that is, the preliminary knowledge regarding the beginning and progression of the medical condition, as provided by the patient. The second slide gave information about the status, that is, the examination of the patient in the doctor's office. Finally, the third slide contained text regarding the patient's laboratory results and any X-rays that were taken.

The patient case texts were written without using particular terminology, in order to avoid disruption of the students' reading processes as a result of unfamiliar wording. Further, the memorizing of details concerning the case results was not required, and the text also included some interpretation of the results, such as "*her blood pressure is 110/70 mmHg, thus normal*". Each text slide included three types of sentence: key sentences that were essential for solving the case, supplementary sentences that contained information in addition to key sentences and which helped the reader rule out incorrect diagnoses, and irrelevant sentences that were not important or contained misleading information concerning the patient case. The sentences were divided into these categories by the authors, one of whom is a biologist, and the division was also checked by a medical specialist. Details of the two texts are shown in Table 1.

[Insert Table 1 approximately here, please.]

Data analysis

First, the diagnoses given by the participants were digitized: written answers given by the students were scored as either correct or incorrect, and the oral answers provided by the residents were first transcribed and then scored in the same manner as the students' answers. The participants' reading processes were analyzed using Tobii Studio, version 2.2.8 (Tobii AB, Danderyd, Sweden). An example of the eye movements that were observed during reading is shown in Figure 1. In addition, the numerical data from Tobii Studio were transferred to IBM Statistical Package for the Social Sciences (SPSS) software, version 22 (IBM Corp., Armonk, NY), which was used for further analyses. The entire slide, key sentences, and irrelevant sentences on each slide were defined as areas of interest (AOIs), that is, the regions in the stimulus from which the authors were interested in gathering data (Holmqvist et al., 2011). For example, in the first slide of the first patient case the sentences "*She tells you that she is on prolonged sick leave because she suffers from back pain caused by rheumatism. Further, she* *talks about increasing difficulty with her memory."* would constitute an AOI of irrelevant sentences (Appendix 1). It has been found that highly important sentences receive more visual attention than unimportant sentences (Hyönä and Niemi, 1990), so the analyses focused on key sentences and irrelevant sentences, and the supplementary sentences were excluded. The following indicators were chosen: total visit duration per slide, number of fixations, and average fixation duration. Total visit duration indicates the duration of all visits within an active AOI (in this case, the entire text page). This was considered the best metric of the processing time, including the reading time and possible thinking time, before moving to the question slide. The number of fixations (fixation count) refers to the number of times the participant fixated on an AOI or an AOI group, for example, key sentences on slide 1. Finally, the average fixation duration represents the mean length of a fixation (in milliseconds) of an individual on each AOI. Four students had poor eye-tracking data in the first patient case, and five students had poor eye-tracking data in the second patient case. Thus, the number of students that were included in the eye-movement analyses was 34 or 35. Descriptive statistics and Shapiro-Wilk normality tests showed that the data were not normally distributed, so the authors had to conduct non-parametric statistical tests (Mann–Whitney U tests, Friedman tests, and Wilcoxon signed-rank tests). The data were analyzed using IBM SPSS software, version 22 (IBM Corp., Armonk, NY).

[Insert Figure 1 approximately here, please.]

RESULTS

Almost all of the participants correctly solved the first patient case; only four (10%) of the third-year medical students failed to make a correct diagnosis. However, only 17 (44%) students solved the second patient case, thereby showing that it was more difficult. All the residents reached a correct diagnosis for both patient cases. For patient case 1, the majority of the participants made the correct diagnosis after reading the second slide, regarding the patient's status, whereas for patient case 2, all of the residents and most of the students that solved the case had already made the correct diagnosis after reading the first slide (Table 2). However, there were no statistically significant difference between the students and residents with regard to the part of the slide show at which they made the correct diagnosis.

[Insert Table 2 approximately here, please.]

Second, the processing times, which were measured using the total visit durations on each of the textual slides, were examined. As hypothesized, the residents processed each slide statistically significantly faster than did the students in both patient case 1 and patient case 2 (patient case 1, slide 1: $Z = -3.10$, $P < 0.01$, $r = -0.43$; slide 2: $Z = -4.26$, $P < 0.001$, $r = -0.59$; slide 3: $Z = -5.06$, $P < 0.001$, $r = -0.70$; patient case 2, slide 1: $Z = -3.96$, $P < 0.001$, $r = -0.55$; slide 2: $Z = -4.81$, $P < 0.001$, $r = -.67$; slide 3: $Z = -4.94$, $P < 0.001$, $r = -0.68$ (see Figure 2). The effect sizes varied between -0.43 and -0.70, thus between medium and large (Cohen, 1969). The students' reading times increased toward the end of the text of both cases (patient case 1: 56.20 ±26.63, 106.71 ±84.11, 110.17 ±71.96; patient case 2: 85.85 ±46.37, 100.17 ±55.56, 102.07 ±84.05), but this type of processing pattern was not found among the residents (patient case 1: 36.39 ±12.41, 46.07 ±16.73, 39.77 ±14.19; patient case 2: 43.60 ±16.43, 36.84 ±13.66, 26.48 ±8.00). In fact, the residents had the exact opposite pattern with regard to patient case 2, as their reading times decreased toward the end of the case presentation.

[Insert Figure 2 approximately here, please.]

When the total visit durations were compared within the groups, Friedman tests showed statistically significant differences between the students' reading times regarding patient case 1 [x^2 (2) = 44.67, $P < 0.001$; median levels for slides 1, 2, and 3 were 53.67, 85.47, and 88.34, respectively], and both students' [x^2 (2) = 7.59, $P < 0.05$; median levels for slides 1, 2, and 3 were 75.63, 86.36, and 86.91, respectively] and residents' [x^2 (2) = 7.54, $P < 0.05$; median levels for slides 1, 2, and 3 were 45.50, 39.77, and 28.32, respectively] reading times regarding patient case 2. *Post hoc* analyses were performed using separate Wilcoxon signed-rank tests on the different combinations of slides. In patient case 1, significant differences between slides 2 and 1 (Z = -5.26, $P < 0.001$, r = -0.84) and between slides 3 and 1 (Z = -5.37, $P < 0.001$, r = -0.86) were found among the students. Thus, the students spent more time viewing the second and third slide than they did the first slide. In patient case 2, significant differences in reading slides 2 and 1 (Z = -2.83, $P < 0.01$, r = -.45) were observed among the students. and between slides 3 and 1 (Z = -2.69, $P < 0.01$, r = -0.75) and slides 3 and 2 (Z = -2,69; $P < 0.01$, r = -0.75) among the residents. Therefore, the students viewed the second slide for longer than the first slide, whereas the residents viewed the first and the second slide statistically significantly longer than the third slide.

Next, the number of fixations on the task-relevant (key sentences) and task-redundant (irrelevant sentences) areas of the patient cases were analyzed. The Mann–Whitney U-test showed statistically significant differences in the number of fixations between students and residents in the key and irrelevant sentences in both cases and on every slide (patient case 1: slide 1, key sentences, $Z = -3.64$, $P < 0.001$, $r = -0.53$, irrelevant sentences $Z = -2.41$, $P < 0.05$; $r = -0.35$; slide 2, key sentences $Z = -4.50$, $P < 0.001$; $r = -.65$, irrelevant sentences $Z = -3.47$, $P < 0.01$; $r = -.50$; slide 3, key sentences $Z = -4.65$, $P < 0.001$; $r = -.67$, irrelevant sentences $Z = -4.92$, $P < 0.001$; $r = -.71$; patient case 2: slide 1, key sentences $Z = -4.37$, $P < 0.001$; $r = -0.64$, irrelevant sentences $Z = -2.37$, $P < 0.05$; $r = -0.35$; slide 2, key sentences $Z = -4.88$, $P < 0.001$; $r = -.71$, irrelevant sentences $Z = -4.50$, $P < 0.001$; $r = -0.66$; slide 3, key sentences $Z = -4.62$, $P < 0.001$; $r = -.67$, irrelevant sentences $Z = -4.69$, $P < 0.001$; $r = -0.68$ (see Table 3). As shown in Table 3, the residents had the smallest number of fixations on both key sentences and irrelevant sentences on each slide. The reading of the task-relevant and task-redundant areas of the text in relation to the number of fixations was also compared within both subgroups. As there were different numbers of key sentences and irrelevant sentences in the slides, the number of fixations was first divided by the number of characters within each sentence type on each slide (see Table 1). In patient case 1, the Wilcoxon signed-rank test showed that students fixated statistically significantly more on the key sentences in the last slide ($Z = -4.28$; $P < 0.001$, $r = -0.72$), and that the residents fixated more on the irrelevant sentences in the first slide ($Z = –2.27$, $P < 0.05$, $r = -0.63$) and on the key sentences in the last slide ($Z = -2.41$, $P < 0.05$; $r = -0.67$), but no statistically significant differences were found within the remaining slides. In patient case 2, the students fixated more on key sentences in the first ($Z = -3.19$, $P < 0.01$; $r = -0.55$) and last slides ($Z = -4.90$, $P < 0.001$; $r = -0.84$), whereas they fixated more on irrelevant sentences in the second slide

(Z = -3.36, $P < 0.01$; r = -0.58). In addition, the residents fixated more on irrelevant sentences in the second slide (Z = -3.18, $P < 0.01$; r = 0.88), and on key sentences in the third slide (Z = -2.34, $P < 0.05$; r = -0.65).

[Insert Table 3 approximately here, please.]

Finally, the average fixation durations on the task-relevant (key sentences) and task-redundant (irrelevant sentences) areas of the text were studied (see Table 4). Overall, the residents appeared to have shorter average fixation durations on both the task-relevant and task-redundant sentences. However, the Mann Whitney U-test showed no statistically significant differences between the groups, with the exception of the irrelevant sentences in the third slide in patient case 1 (Z = -2.26, $P < 0.05$, r = -0.33) and the key sentences in the third slide in patient case 2 (Z = -2.37, $P < 0.05$, r = -0.35). Finally, the average duration of fixations was examined within the groups for both of the patient cases. In the first case, the Wilcoxon signed-rank test showed that both the students (Z = -2.28, $P < 0.05$, r = -.38) and the residents (Z = -2.62, $P < 0.01$, r = -0.73) fixated longer on irrelevant sentences than on the key sentences in the first slide. However, with regard to the third slide both the students (Z = -4.06, $P < 0.001$, r = -0.69) and the residents (Z = -3.11, $P > 0.01$, r = -0.86) fixated statistically significantly longer on the key sentences than on the irrelevant sentences. In the second case, the students fixated longer on the irrelevant sentences in the second slide (Z = -4.54, $P < 0.001$, r = -.78), but on the key sentences in the third slide (Z = -4.56, $P < 0.001$, r = -.78). Within the residents' group, the only difference in patient case 2 was found with regard to the second slide, whereby they fixated longer on the irrelevant sentences than on the key sentences (Z = -2.83, $P < 0.01$. r = -0.78).

[Insert Table 4 approximately here, please.]

In summary, the residents appeared to be highly efficient patient case-solvers. Their expertise was shown in accuracy of their diagnoses (Table 2), shorter total visit durations (Figure 2) and lower number of fixations (Table 3) , compared to the students. The overall viewing patterns of students and residents appeared to be different, since the students' reading times increased toward the end of the text in both patient cases, while the residents demonstrated the opposite pattern in the second patient case (Figure 2).

DISCUSSION

The present study investigated the manner in which the level of expertise influences the reading, and solving, of two written patient cases on cardiovascular medicine. Third-year medical students and internal medicine residents' reading processes were analyzed using the eye-tracking method. The analyses focused on the diagnosis made, total visit duration per slide, and eye-movement indicators (number of fixations and average duration of fixations) on task-relevant and task-redundant areas of the patient case texts.

Almost all of the participants correctly diagnosed the first patient case (10% of the students did not make a correct diagnosis). However, less than half of the students correctly diagnosed the second patient case, showing that it was clearly more difficult. Again, all of the residents were capable of making the correct diagnosis. In both of the cases and on every slide, the residents processed the text remarkably more quickly than did the students. Therefore, the residents appeared to need less time than the students to produce accurate diagnoses. The results showing the experts' problem-solving efficiency are supported by evidence from both cognitive and visual expertise studies. For example, Rikers et al. (2000, 2004) and Verkoeijen et al. (2004) showed that experts diagnosed clinical cases faster and more accurately than did students. Further, in studies utilizing eye-tracking in the examination of medical visualizations, shorter overall viewing times or time-on-task have been reported for experts compared to novices (Krupinski, 1996; Krupinski et al., 2006; Jaarsma et al., 2015). One explanation for the superiority of the residents might be derived from encapsulation theory (Boshuizen and Schmidt, 1992; Schmidt and Boshuizen, 1993a; Boshuizen and Schmidt, 2008), according to which frequent exposure to clinical cases leads to the development of higher-level encapsulating concepts. Through encapsulation, the lines of reasoning gradually become shorter, thus the solving of routine cases becomes faster. Therefore, the residents could be expected to possess higher-level encapsulated knowledge networks that enable faster problem-solving compared to the students.

Further, in the first patient case, the majority of the participants made the correct diagnosis after the second slide, but in the second case, most of the students that solved the case and all of the residents gave a correct answer after reading only the first slide. It is suggested that the residents' knowledge of enabling conditions might have helped them to already rule out other diseases, and to focus on those that were most likely, right at the beginning of the case (Schmidt and Rikers, 2007), thus they processed the second and third slides even faster than they did the first. Some features of the text in the first slide might have triggered encapsulated concepts, which guided residents' problem-solving (Boshuizen et al., 1995; Rikers at al., 2004; Schmidt and Rikers, 2007). It is interesting that in the first case, most of the participants made the correct diagnosis after the second slide, but the students continued to spend more absolute time on the third slide. In the second case, the residents decreased their reading time immediately after the first slide, at which point they had made the correct diagnosis. Thus, for the residents, reaching a decision on diagnosis appeared to decrease their reading time, but the students did not manifest a similar pattern. It might be that, although the students had made the correct diagnosis, they were not as certain of their decision as were the residents. Therefore, they might have continued going through alternative diagnoses, whereas the residents stuck to their initial diagnosis and searched for confirmatory information in the subsequent slides.

At the eye movements' level, it was found that the residents had fewer fixations on both the task-relevant and task-redundant areas of the text compared to the students. Further, Manning et al. (2006) and Bertram et al. (2013) found that fewer fixations on a relevant area indicated a more advanced level of development, although these studies used visual stimuli. When reading of task-relevant and task-redundant areas was compared among residents and students, no consistent pattern was found. Finally, average fixation durations on task-relevant and task-redundant areas of the texts were examined. Although the descriptive statistics showed that the residents had shorter average fixation durations on both sentence types overall, no consistent statistically significant differences were found between the two groups.

To summarize, the results indicate that the residents were fast and accurate processors of patient case texts. The residents' superior processing compared to third-year students might be derived from wider studies in medicine and accumulated clinical experience, rather than being an aspect that has been taught.

Limitations of the study

In order to improve the research design, several issues must be considered in the future. The patient case text format used was somewhat unfamiliar to both of the subgroups, since the text simulated a patient encounter in a healthcare center. Although the texts contained elements and a structure similar to authentic case reports, they also contained sentences that helped the reader to position him/herself in the doctor's role. This might have affected the participants' behavior. Therefore, the differences between the results concerning the two patient cases might be partly explained by the participants becoming accustomed to the setup. However, the use of an equally unfamiliar text format placed the participants on the same level; using authentic case reports might have given the residents an advantage because of their clinical experience. Further, the experimental study design meant that the participants were perhaps not reading the cases as they would normally do, since it was not possible to go back to the previous slide in both case studies. Thus, this might have affected their reading patterns. Although the case topics were chosen on the basis that they would be familiar to the students from their previous studies, the study was not controlled for whether some of the students had, for example, practical experience of the topics from attending rounds in their clinical studies. The addition of another type of text, such as a medical journal article, would have provided valuable information regarding the participants as readers in general. Moreover, all student participants were a homogenous group of high achievers who had passed a selective entrance exam and who voluntarily took part in the study during their free time. Thus, the sample of students was highly selective. Inclusion of another control group, such as first-year medical students, would have strengthened the study design. Finally, to more comprehensively test the encapsulation theory, a recall protocol, or *post hoc* explanations of the diagnoses, should be added to the design in order to obtain further information about the participants' lines of reasoning (Rikers et al., 2000; Verkoeijen et al., 2004; Schmidt and Rikers, 2007; Gegenfurtner et al., 2013; Jaarsma et al., 2014, 2015).

CONCLUSION

Despite the limitations, the current study indicates that the eye-tracking method has great potential in assessing growing expert performance with written medical texts as stimuli. The most important finding related to medical education is that the residents required less time and fewer fixations to produce more accurate diagnoses compared to the third-year medical students. In addition, the students and residents appeared to demonstrate different reading patterns. For the residents, making a decision regarding the diagnosis seemed to decrease their reading time of the following slides, but the students appeared to increase their reading time toward the end of the case. The observed differences in reading patterns could be utilized in medical teaching by demonstrating to students the importance of the medical history and patient status in clinical reasoning, rather than laboratory results, which mainly confirm or refute the working diagnosis. Information from medical text reading studies could also be included in teaching how a good medical text, such as a referral, medical record, or medical case report, is constructed, and this could even help students to produce accurate diagnoses and enhance patient safety. Finnish legislation determines what must be written in medical records and the manner in which it must be written (MSAH, 2009). However, despite these instructions, the practice of writing patient records appears to be somewhat incoherent.

The present study showed that even third-year medical students can demonstrate nascent expertise and the construction of clinical understanding, since almost all of them were capable of correctly diagnosing the first patient case, and almost half of them also correctly diagnosed the second, more complex, patient case, although they had only just finished their preclinical study phase and begun their clinical studies. Thus, the traditional curriculum divided into the 2–2.5-year preclinical phase, consisting of basic science studies, such as anatomy and physiology, and the following clinical phase appeared to offer support for the students' clinical reasoning. However, from the encapsulation theory perspective (Boshuizen and Schmidt, 1992; Schmidt and Boshuizen, 1993; Boshuizen and Schmidt, 2008), students would benefit from frequent exposure to clinical cases in order to develop higher-order encapsulating concepts that would lead to faster and more accurate reasoning. Moreover, returning to the basic sciences in the clinical phase of medical school has proven beneficial in assimilating biomedical science into clinical practice (Spencer et al., 2008). It is suggested that senior medical students are better able to appreciate the relevance of basic science concepts after having had exposure to clinical medicine (Spencer et al., 2008). However, further studies that include different medical texts, as well as participants with more diverse levels of expertise, are required.

ACKNOWLEDGEMENTS

The authors wish to thank the Academy of Finland for its financial support for the LeMed project (Learning of Medical Expertise), 128892. Further, the first author wishes to thank the Finnish Cultural Foundation for a personal research grant that helped reporting this study. Additionally, the authors thank the Faculty of Medicine at the University of Turku for their co-operation during the data-gathering process. Lastly, the authors are grateful for the medical students and residents who took part in this study.

NOTES ON CONTRIBUTORS

HENNA VILPPU, M.A. (Educ.), is a doctoral candidate in the Department of Teacher Education and Centre for Learning Research, Faculty of Education at the University of Turku, Finland.

MIRJAMAIJA MIKKILÄ-ERDMANN, Ph.D. (Educ.), is a professor for educational sciences in the Department of Teacher Education and Centre for Learning Research, Faculty of Education at the University of Turku, Finland. She is teaching educational sciences for primary school and subject teachers, as well as university teachers and her research interest is in learning of complex phenomena and teaching in higher education, such as medical education.

ILONA AHOPELTO, M.A. (Biol.), is a doctoral candidate in the Department of Teacher Education and Centre for Learning Research, Faculty of Education at the University of Turku, Finland.

ERIKA ÖSTERHOLM-MATIKAINEN, M.A. (Educ.), is an educational planner in the Medical Education Research and Development Centre (TUTKE), Faculty of Medicine at the University of Turku, Finland. Her research interest is in shared metacognition and self-regulation in learning processes.

LITERATURE CITED

Ahopelto I, Mikkilä-Erdmann M, Olkinuora E, Kääpä P. 2011. A follow-up study of medical students' biomedical understanding and clinical reasoning concerning the cardiovascular system. Adv Health Sci Educ Theory Pract 16:655–668.

Bertram R, Helle L, Kaakinen JK, Svedström E. 2013. The effect of expertise on eye movement behaviour in medical image perception. PLoS One 8:e66169.

Boshuizen HP, Schmidt HG. 1992. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognit Sci 16:153–182.

Boshuizen HP, Schmidt HG. 2008. The development of clinical reasoning expertise. In: Higgs J, Jones MA, Loftus S, Christensen N (Editors). Clinical Reasoning in the Health Professions. 3rd Ed. Philadelphia, PA: Butterworth-Heineman. p 113–121.

Boshuizen HP, Schmidt HG, Custers EJ, van de Wiel MW. 1995. Knowledge development and restructuring in the domain of medicine: The role of theory and practice. Learn Instr 5:269–289.

Charon R. 2000. Reading, writing, and doctoring: Literature and medicine. Am J Med Sci 319:285–291.

Chi MTH, Glaser R, Farr MJ (Editors). 1988. The Nature of Expertise. 1st Ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 464 p.

Cohen J. 1969. Statistical Power Analysis for the Behavioral Sciences. 1st Ed. New York, NY: Academic Press. 415 p.

Custers EJ, Cate OT. 2002. Medical students' attitudes towards and perception of the basic sciences: A comparison between students in the old and the new curriculum at the University Medical Center Utrecht, The Netherlands. Med Educ 236:1142–1150.

Fitts PM, Jones RE, Milton JL. 1950. Eye movement of aircraft pilots during instrument-landing approaches. Aeronaut Engng Rev 9:24–29.

Gegenfurtner A, Siewiorek A, Lehtinen E, Säljö R. 2013. Assessing the quality of expertise differences in the comprehension of medical visualizations. Vocat Learn 6: 37–54.

Goldman SR, Saul EU. 1990. Flexibility in text processing: A strategy competition model. Learn Indiv Differ 2:181–219.

Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, van de Weijer J. 2011. Eye Tracking. A Comprehensive Guide to Methods and Measures. 1st Ed. New York, NY: Oxford University Press. 537 p.

Hyönä J, Lorch RF Jr, Kaakinen JK. 2002. Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. J Educ Psychol 94:44–55.

Hyönä J, Niemi P. 1990. Eye movements in repeated reading of a text. Acta Psychol 73:259–280.

Hyönä J, Nurminen AM. 2006. Do adult readers know how they read? Evidence from eye-movement patterns and verbal reports. Brit J Psychol 9:31–50.

Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, Boshuizen HP. 2014. Expertise under microscope: Processing histopathological slides. Med Educ 48:292–300.

Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, Boshuizen HP. 2015. Expertise in clinical pathology: Combining the visual and cognitive perspective. Adv Health Sci Educ Theory Pract 20:1089–1106.

Jarodzka H, Balslev T, Holmqvist K, Nyström M, Scheiter K, Gerjerts P, Eika B. 2012. Conveying clinical reasoning based on visual observation via eye-movement modelling examples. Instr Sci 40:813–827.

Just MA, Carpenter PA. 1980. A theory of reading: From eye fixations to comprehension. Psychol Rev 87:329–354.

Kalyuga S, Rikers R, Paas F. 2012. Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. Educ Psychol Rev 24:313–337.

Kaufman DR, Keselman A, Patel VL. 2008. Changing conceptions in medicine and health. In: Vosniadou S (Editor). International Handbook of Research on Conceptual Change. 1st Ed. New York, NY: Routledge. p 295–327.

Khan RS, Tien G, Atkins MS, Zheng B, Panton ON, Meneghetti AT. 2012. Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? Surg Endosc 26:3536–3540.

Kliegl R, Nuthmann A, Engbert R. 2006. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. J Exp Psychol Gen 135:12–35.

Klusewitz MA, Lorch RF Jr. 2000. Effects of headings and familiarity with a text on strategies for searching text. Mem Cognit 28:667–676.

Krupinski EA. 1996. Influence of experience on scanning strategies in mammography. Proc SPIE 2712: 95–101.

Krupinski EA. 2010. Current perspectives in medical image perception. Atten Percept Psychophys 72:1205–1217.

Krupinski EA, Hall ET, Jaw S, Reiner B, Siegel E. 2012. Influence of radiology report format on reading time and comprehension. J Digit Imaging. 25:63–69.

Krupinski EA, Tillack AA, Richter L, Henderson JT, Battacharayya AK, Scott KM, Graham AR, Descour MR, Davis JR, Weinstein RS. 2006. Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with expertise. Hum Pathol 37:1543–1556.

Manning D, Ethell S, Donovan T, Crawford T. 2006. How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 12:134–142.

Michael JA. 1998. Students' misconceptions about perceived physiological responses. Adv Physiol Educ 19: S90–S98.

Michael JA, Richardson D, Rovick A, Modell H, Bruce D, Horwitz B, Hudson M, Silverthorn D, Whitescarver S, Williams S. 1999. Undergraduate students' misconceptions about respiratory physiology. Adv Physiol Educ 22:S127–S135.

Michael JA, Wenderoth MP, Modell HI, Cliff W, Horwitz B, McHale P, Richardson D, Silverthorn D, Williams S, Whitescarver S. 2002. Undergraduates' understanding of cardiovascular phenomena. Adv Physiol Educ 26:72–84.

Mikkilä-Erdmann M, Södervik I, Vilppu H, Kääpä P, Olkinuora E. 2012. First-year medical students' conceptual understanding of and resistance to conceptual change concerning the central cardiovascular system. Instr Sci 40:745–754.

MSAH. 2009. Ministry of Social Affairs and Health. Decree on Patient Documentation (Sosiaali - ja terveysministeriön asetus potilasasiakirjoista). Ministry of Social Affairs and Health, Helsinki, Finland. 30.3.2009/298.

Nivala M, Lehtinen E, Helle L, Kronqvist P, Paranko J, Säljö R. 2013. Histological knowledge as a predictor of medical students' performance in diagnostic pathology. Anat Sci Educ 6:361–367.

Rayner K. 1998. Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422.

Rayner K, Slattery TJ. 2009. Eye movements and moment-to-moment comprehension processes in reading. In: Wagner RK, Schatschneideder C, Phythian-Sence C. (Editors). Beyond Decoding: The Behavioral and Biological Foundations of Reading Comprehension. 1st Ed. New York, NY: The Guilford Press. p 27–45.

Reingold EM, Sheridan H. 2011. Eye movements and visual expertise in chess and medicine. In: Liversedge SP, Gilchrist ID, Everling S. (Editors). Oxford Handbook on Eye Movements. 1st Ed. Oxford, UK: Oxford University Press. p 528–550.

Rikers RM, Loyens SM, Schmidt HG. 2004. The role of encapsulated knowledge in clinical case representations of medical students and family doctors. Med Educ 38:1035–1043.

Rikers RM, Schmidt HG, Boshuizen HP. 2000. Knowledge encapsulation and the intermediate effect. Contemp Educ Psychol 25:150–166.

Schmidt HG, Boshuizen HP. 1993a. On acquiring expertise in medicine. Educ Psychol Rev 5:205–221.

Schmidt HG, Boshuizen HP. 1993b. On the origin of intermediate effects in clinical case recall. Mem Cognit 21:338–351.

Schmidt HG, Norman GR, Boshuizen HP. 1990. A cognitive perspective on medical expertise: Theory and implications. Acad Med 65:611–621.

Schmidt HG, Rikers RM. 2007. How expertise develops in medicine: Knowledge encapsulation and illness script formation. Med Educ 41:1133–1139.

Sistrom CL, Honeyman-Buck J. 2005. Free text versus structured format: Information transfer efficiency of radiology reports. Am J Roentgenol 185:804–812.

Spencer AL, Brosenitsch T, Levine AS, Kanter SL. 2008. Back to the basic sciences: An innovative approach to teaching senior medical students how best to integrate basic science and clinical medicine. Acad Med 83:662–669.

Sweller J. 1988. Cognitive load during problem-solving: Effects on learning. Cognit Sci 12:257–285.

Underwood G, Hubbard A, Wilkinson H. 1990 Eye fixations predict reading comprehension: The relationship between reading skill, reading speed, and visual inspection. Lang Speech 33:69–81.

Vauras M, Hyönä J, Niemi P. 1992. Comprehending coherent and incoherent texts: Evidence from eye movement patterns and recall performance. J Res Read 15:39–54.

Verkoeijen PP, Rikers RM, Schmidt HG, van de Wiel MW, Kooman JP. 2004. Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. Med Educ 38:617–627.

Vicente KJ, Wang JH. 1998. An ecological theory of expertise effects in memory recall. Psychol Rev 105:33–57.

Wiley J, Rayner K. 2000. Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. Mem Cognit 28:1011–1021.

Wilhelmsson N, Dahlgren LO, Hult H, Scheja M, Lonka K, Josephson A. 2010. The anatomy of learning anatomy. Adv Health Sci Educ Theory Pract 15:153–165.

Zumwalt AC, Iyer A, Ghebremichael A, Frustace BS, Flannery S. 2015. Gaze patterns of gross anatomy students change with classroom learning. Anat Sci Educ 8:230–241.

APPENDIX 1. The patient case 1 (Cardiac failure) translated into English.

Key sentences are in bold; Supplementary sentences are underlined; and *Irrelevant sentences* are in italics. First and last lines of each slide are not classified since their purpose is to help the reader to position himself or herself to the task and give a framework for the task.

SLIDE 1

You are working as a doctor at a health care center in city X.

A 65-year-old woman comes to your doctor's practice. **She describes that the reason she came to the doctor is increasing shortness of breath that arises in even minor strain. She has felt tired recently, and mentions that her digestion has not been working normally.** *She tells you that she is on prolonged sick leave because she suffers from back pain caused by rheumatism. Further, she talks about increasing difficulty with her memory.* Otherwise, she has been healthy. Uses alcohol rarely. **Smokes about 10 cigarettes daily.**

That is all that comes up in the anamnesis.

SLIDE 2

After the anamnesis, you start doing a medical examination.

Patient's general condition seems satisfactory**. However, you notice that she looks pale and exhausted.** *She is slightly overweight.* **Her pulse is unstable, approximately 100 beats per minute.** Her blood pressure is 110/70 mmHg, thus normal. *Fundi of the eyes are normal.* **Her legs are a little swollen.** The stomach is not tender, and the calves are soft. *The thoracic area is quite tender.* **A systolic heart murmur in the tip of the heart, radiating toward the arm pit. In addition, rales from both lungs.**

Nothing else comes up in the patient's status.

SLIDE 3

You start to skim through the patient's laboratory and other test results.

The laboratory results show that sedimentation rate is normal. Hemoglobin is slightly elevated. The hematocrit is also slightly elevated. Electrolytes are inside the reference rate. The kidney function is normal. Creatine phosphokinase is at the lower boundary of normal. *CRP is 10 mg/ml, thus on the upper boundary of normal.* **A compensated respiratory alkalosis is found in the blood gas analysis. EKG is normal.** *Cholesterol is 6.5 mmol/l, thus slightly elevated.* **Thorax X-ray shows congestion in the lungs and an enlarged heart**. *Volume of the lungs is normal.*

Other test results are not available at the moment.

TABLES

Table 1.

Proportions of task-relevant (key sentences) and task-redundant (irrelevant sentences) areas in the two patient case texts.

		Case 1 Cardiac failure			Case 2 Pulmonary embolus		
		Slide 1	Slide 2	Slide 3	Slide 1	Slide 2	Slide 3
Key	Sentences (n / % of all sentences)	3 / 30	5 / 38	3 / 21	4 / 36	3 / 21	2,5 / 25
	Words (n / % of all words)	27 / 44	33 / 46	14 / 21	46 / 52	23 / 30	13 / 22
	Characters (n / % of all characters)	204 / 39	231 / 43	141 / 25	350 / 50	148 / 26	136 / 26
Irrelevant	Sentences (n / % of all sentences)	2 / 20	3 / 23	3 / 21	2 / 18	2 / 14	2 / 20
	Words (n / % of all words)	15 / 25	11 / 15	15 / 23	20 / 22	12 / 16	14 / 24
	Characters (n / % of all characters)	151 / 29	115 / 21	101 / 18	157 / 22	115 / 20	110 / 21

Table 2.

The part of the slide show at which the participants reached the correct diagnosis

Participants (n)	Correct diagnosis reached after reading			Correct / incorrect diagnoses n (%)
	1st slide n (%)	2nd slide n (%)	3rd slide n (%)	
Case 1 (Cardiac failure)				
Students (39)	7 (18)	20 (51)	8 (21)	35 (90) / 4 (10)
Residents (13)	1 (8)	7 (54)	5 (39)	13 (100) / 0 (0)
Case 2 (Pulmonary embolus)				
Students (39)	7 (18)	6 (15)	4 (10)	17 (44) / 22 (56)
Residents (13)	13 (100)	0 (0)	0 (0)	13 (100) / 0 (0)

Table 3.

Students' and residents' number of fixations within different types of sentences in patient cases 1 and 2

Case	Slide	Sentences	Students with a correct diagnosis (Case 1: n = 32; Case 2: n = 15)	Students with an incorrect diagnosis (Case 1: n = 3; Case 2: n = 19)	Residents (n = 13)
			M±*SD*	*M*±*SD*	*M*±*SD*
1	1	Key sentences***	87±40	83±16	56±17
		Irrelevant sentences*	64±30	74±29	46±15
	2	Key sentences***	194±155	157±69	73±28
		Irrelevant sentences**	84±70	87±43	42±18
	3	Key sentences***	144±105	185±174	48±19
		Irrelevant sentences***	73±41	63±28	26±9
2	1	Key sentences***	183±91	169±106	86±25
		Irrelevant sentences*	68±33	67±45	43±21
	2	Key sentences***	96±44	115±80	30±12
		Irrelevant sentences***	93±40	106±55	43±14
	3	Key sentences***	113±98	163±95	40±18
		Irrelevant sentences***	59±44	87±53	24±8

Note. Statistical comparisons are made between the whole group of students and residents, *$p < 0.05$, **$P < 0.01$, ***$P < 0.001$. Supplementary sentences were not included in the analyses.

Table 4.

Students' and residents' average duration of fixations (ms) within different types of sentences

Case	Slide	Sentences	Students with a correct diagnosis (Case 1: n = 32; Case 2: n = 15)	Students with an incorrect diagnosis (Case 1: n = 3; Case 2: n = 19)	Residents (n = 13)
			M±*SD*	*M*±*SD*	*M*±*SD*
1	1	Key sentences	260.41±34.40	282.77±35.23	242.24±36.31
		Irrelevant sentences	271.83±35.09	278.37±29.86	266.72±51.39
	2	Key sentences	272.56±39.45	277.67±5.17	255.70±38.33
		Irrelevant sentences	274.15±42.31	284.80±38.52	271.05±33.78
	3	Key sentences	294.35±55.65	325.13±59.34	266.98±35.66
		Irrelevant sentences*	263.10±40.64	260.27±16.75	238.48±33.10
2	1	Key sentences	253.89±28.53	263.22±35.90	248.32±30.93
		Irrelevant sentences	261.55±34.74	257.43±29.14	251.20±38.37
	2	Key sentences	235.05±24.38	248.95±31.24	231.05±23.45
		Irrelevant sentences	271.11±35.80	271.02±30.01	260.49±26.52
	3	Key sentences*	301.80±70.48	287.91±48.13	251.35±23.10
		Irrelevant sentences	241.97±27.62	255.36±32.10	241.22±30.39

Note. Statistical comparisons are made between the whole group of students and residents, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Supplementary sentences were not included in the analyses.

FIGURE LEGENDS

Figure 1. A capture from Tobii T60XL Eye Tracker with text and overlying eye-gaze pattern. The spots represent fixations: the bigger the spot, the longer the fixation. Lines between spots represent saccades, i.e. the fast oculomotor movements between fixations. *Note.* The eye-gaze pattern is on the original Finnish text. The English translation is provided next to the original text.

Figure 2. Means and standard deviations of total visit durations (seconds) per slide among students and residents in patient cases 1 and 2. *Note.* $*P < 0.05$, $**P < 0.01$, $***P < 0.001$