Original article

Visual expertise in paediatric neurology

Thomas Balslev a,b,*, Halszka Jarodzka c, Kenneth Holmqvist d, Willem de Grave e, Arno M.M. Muijtjens e, Berit Eika a, Jeroen van Merriënboer e, Albert J.J.A. Scherpbier f

a Aarhus University, Centre of Medical Education, Brendstrupgaardsvej 102, 8200 Aarhus N, Denmark
b Viborg Hospital, Department of Paediatrics, Heibergs Allé 4, PO Box 130, 8800 Viborg, Denmark
c Center of Learning Sciences and Technologies, Open University of The Netherlands, PO Box 2960, 6401 DL Heerlen, The Netherlands
d Lund University, Humanities Lab, Helgonabacken 12, Box 201, 221 00 Lund, Sweden
e Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
f Institute for Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Abstract

Background: Visual expertise relies on perceptive as well as cognitive processes. At present, knowledge of these processes when diagnosing clinical cases mainly stems from studies with still pictures. In contrast, patient video cases constitute a dynamic diagnostic challenge that may simulate seeing and diagnosing a patient in person.

Aims: This study investigates visual attention and the concomitant cognitive processes of clinicians diagnosing authentic paediatric video cases.

Methods: A total of 43 clinicians with varying levels of expertise took part in this cross-sectional study. They diagnosed four brief video recordings of children: two with seizures and two with disorders imitating seizures. We used eye tracking to investigate time looking at relevant areas in the video cases and a concurrent think-aloud procedure to explore the associated clinical reasoning processes.

Results: More experienced clinicians were more accurate in visual diagnosis and spent more of their time looking at relevant areas. At the same time, they explored data less, yet they built and evaluated more diagnostic hypotheses.

Conclusions: Clinicians of varying expertise analyse patient video cases differently. Clinical teachers should take these differences into account when optimising educational formats with patient video cases.

© 2011 European Paediatric Neurology Society. Published by Elsevier Ltd. All rights reserved.

Abbreviations: AOI, areas of interest; PVC, patient video case.

* Corresponding author. Aarhus University, Centre of Medical Education, Brendstrupgaardsvej 102, 8200 Aarhus N, Denmark. Tel.: +45 21 73 69 04; fax: +45 78 44 54 09.
E-mail address: balf@dadlnet.dk (T. Balslev).

1090-3798/$ – see front matter © 2011 European Paediatric Neurology Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejpn.2011.07.004
1. Introduction

Neurologic education is going through a period of change and uncertainty. Economic limitations and increasing service demands often seriously limit learners’ exposure to and hence opportunities to learn about clinical pictures and principles in adult as well as in child neurology. Therefore, there is a need for neurologic education research to clarify approaches to effective, efficient and lifelong learning. As diagnostic evaluation of patients involves visual examination and conceptual understanding of classification systems, it is pertinent for this research to elucidate what clinicians do and what they think when they make a diagnosis based on visual observations.

For experienced physicians, making a visual diagnosis may be a rapid and unconscious pattern recognition process involving effortless connecting of clinical information to knowledge stored in long term memory. Pattern recognition may also be supplemented by a more analytic approach requiring conscious analysis of clinical information.

In visual domains, the physician searches the stimulus visually (perception) and attempts to render an interpretation (cognition). Studies on visual expertise in medical still-image diagnosis show that experienced physicians compared to residents or novices are faster diagnostics, raise more hypotheses and are more successful in connecting perceptual and cognitive processing.

Next to meeting and diagnosing a patient in real life, diagnosing a patient video case (PVC) may be the most realistic situation with high authenticity. Video recordings of patients often aid in diagnosis, and can be used in telemedicine. PVCs may benefit learning and in neurology they are recognized as essential teaching tools. Although PVCs may accelerate learning, we know too little about the associated perceptive and cognitive processes of clinicians. This knowledge would be essential for the development of new video-based approaches to learning, particularly in visual domains.

In this study of clinicians’ perceptive and cognitive processes, we hypothesized that more experienced clinicians would be diagnostically more accurate and would use more of their time visually inspecting relevant areas. Likewise, concerning the cognitive processes, we hypothesized that more experienced clinicians would explore data less and would build and evaluate more diagnostic hypotheses compared to their more inexperienced colleagues.

2. Materials and methods

We enrolled 43 participants from the Northern Region of Denmark and the Medical School at Aarhus University. (Table 1) To reduce bias, invitations were given both verbally and in writing to all eligible participants in the area. The experience of a participant, measured as a continuous variable in years of paediatric training after graduation, was used as an indicator of expertise.

The study took place during September 2009 at the Centre for Medical Education (MEDU), Aarhus University, Aarhus, Denmark. Participants were compensated with a ten-Euro movie ticket voucher.

We used authentic PVCs to stimulate and examine visual search patterns along with the associated cognitive processes. The material consisted of four PVCs in an audio video interleave format (-avi), (Table 2) and each participant watched all of the PVC’s. The experiment consisted of individual sessions approximately 30 min long. A presentation of the patient’s age, gender and a few words indicating the nature of their problem was given via on-screen text prior to each PVC. The PVCs were looped automatically. Each PVC lasted 26–49 s and depicted a single infant. The PVCs included conditions sufficiently difficult to challenge the students and the residents, yet common and typical enough that an expert would be expected to reach a correct diagnosis.

Attention to the dynamic features of the movements and their change over time was crucial for diagnosing cases. The original sound, which was not necessary to reach a correct diagnosis, was removed from the PVCs to prevent interference with the verbal report.

To study the visual search strategies, we implemented eye tracking. In non-medical domains, important differences in perception between experts and novices studying dynamic stimuli have been documented. For instance, experts have been shown to attend more to relevant aspects of the stimulus than novices. Likewise, novices performed more slowly and not as accurately and were distracted by salient, but irrelevant features. However, in the more complex medical domain and specifically in the clinically relevant domain of authentic patient video cases, we lack knowledge regarding perception

<table>
<thead>
<tr>
<th>Table 1 – Demographic data of the 43 participants.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Medical students within 18 months of graduation</td>
</tr>
<tr>
<td>Residents</td>
</tr>
<tr>
<td>Experts</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number of participants (eligible participants)</td>
</tr>
<tr>
<td>Average age (range), years</td>
</tr>
<tr>
<td>Number of females (%)</td>
</tr>
<tr>
<td>Duration of postgraduate paediatric experience, years</td>
</tr>
</tbody>
</table>
and the related cognitive processes. Eye movements were recorded with an SMI High Speed eye-tracking system with a temporal resolution of 240 Hz and the iView X 2.2 software. The stimulus material was presented via Experiment Center 2.2. The eye-tracking data were analysed with BeGaze 2.3 software. The verbal data were recorded using iView Sound recording 1.0 software via a flat microphone (IMG) attached to the eye-tracking computer.

To analyse the eye-tracking data, so-called areas of interest (AOIs) were created. The AOIs accurately specified areas on the patient video (e.g., an arm, the face or the legs), and were defined by two domain experts. (Table 2) These AOIs were dynamic, i.e., they moved across the screen, changed in shape and size, and could appear and disappear depending on the focus of the video camera. The AOIs were defined as polygons superimposed on the PVCs surrounding each relevant area of the video (e.g., a leg of the patient) in the BeGaze software. Each time a relevant body part came out of focus and reappeared, a new AOI was created. These AOIs were adapted in shape and size for each video frame for the entire, looped video. Afterwards, the coordinates for the polygons were exported as .txt files and looped by a self-programmed MatLab code. The number of loops depended on the longest viewing times of all participants. Percentages of time looking at AOIs relative to the total duration of the looped PVCs were calculated.

Clinical diagnostic reasoning is a problem-solving process of interpreting information and refining hypotheses to identify processes causing a patient’s illness. We chose a concurrent think-aloud design to tap into these cognitive processes. After a brief instruction and familiarization with giving verbal reports, participants could think aloud without systematic changes to the sequential structure of their thought processes. The think-aloud protocols were transcribed verbatim.

Using a consensus procedure, two domain experts decided upon the criteria for the diagnostic accuracy and scored it (Table 2) from the transcripts. The verbal protocols were investigated using a coding system originally developed for use in paediatrics. (Table 3) Two raters analysed all transcripts with four PVCs from 43 participants. The clauses were categorized according to the following categories: data exploration, theory building and theory evaluation. Inter-rater reliability analysis revealed generalisability coefficients equal to 0.94, 0.91 and 0.85, indicating that the categorization showed sufficient reliability for all three categories.

Pretraining for thinking aloud was standardized, and the participants were instructed as follows: “Please take a look at the way the infant behaves/moves. Think aloud all the time. Mention a likely diagnosis when or if you think of it. The video is repeated until you say stop.” Eye-tracking and recording of concurrent thinking aloud were then performed with PVCs 1, 2, 3, and 4.

No risks were associated with the use of eye-tracking equipment or otherwise by the participation in this study. Verbal and written information was provided before informed consent was sought and documented. Data were anonymised prior to analysis and used only for research purposes. Ethical approval was sought through the Regional Danish Ethical Committee, and the study was exempted from formal ethical approval. Eye-tracking data and investigator remarks were screened for invalid data, and only 10 eye-tracking datasets (5.8%) were excluded from analysis.

3. Results

A total of 12% of the medical students, 45% of the residents and 94% of the experts were able to make a correct diagnosis in the cases. Duration of postgraduate experience was highly correlated with diagnostic accuracy. The diagnostic accuracy (% correct) increased by 37.2 when experience increased by a factor of 10. (Table 4) In the logarithmic representation of years of experience (Figs. 1 and 2), the regression coefficient indicates the increase (or decrease) of a variable when years of
experience increases with a factor of 10. Note that an offset of 0.25 year was chosen in order to avoid problems with the logarithm (the logarithm of 0 is undefined). So for all participants the number of years of experience is increased by 0.25, resulting for all students in a value of 0.25.

The total time used for the task was not associated with duration of postgraduate experience. Remarkably, more experienced clinicians spent more of their time attending to relevant AOIs. (Fig. 1).

The cognitive processes of the participants in terms of the relative frequencies of data exploration, theory building and theory evaluation are illustrated graphically. (Fig. 2) For an increase of factor 10 in years of experience, the percentage of exploration clauses decreased by 16.9, whereas the percentages of building and evaluating increased by 7.8 and 9.7, respectively (Table 4). The total number of clauses did not significantly change with years of experience. More experienced participants explored data less, built more theory and evaluated more theories. The regression equation explains a considerable amount of the total variation of diagnostic accuracy, data exploration, hypothesis building and evaluation, as is indicated by the R² values, which range between 50 and 72 per cent.

Total time on task did not change significantly with years of experience, but percentage time looking at an AOI significantly increased by 8.5 when experience increased by a factor 10, although the explained variance was modest (R² = 12%).

**4. Discussion**

Not surprisingly, this study shows that more experienced clinicians have high diagnostic accuracy; i.e., they have visual expertise. They are simply effective diagnosticians, and the finding adds to the validity of our measure of diagnostic accuracy.

The main findings of the present study are the concomitant perceptive and cognitive data, revealing how more experienced clinicians compare to less experienced clinicians. Experienced clinicians use relatively more of their time looking at AOIs in the PVCs. With decreasing duration of clinical experience, clinicians use less and less of their time looking at relevant areas, and more and more of their time searching other areas. Thus, it seems that experienced clinicians process visual data more efficiently; i.e., their visual search is more effective.

**Table 4 – Linear regression of cognitive variables, diagnostic accuracy, time on task and percentage of time spent looking at areas of interest. In this logarithmic representation of years of experience, the regression coefficient (b) indicates the increase (or decrease) of a variable when years of experience increases by a factor 10.**

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Regression coefficient (b)</th>
<th>Standard regression coefficient (beta)</th>
<th>Significance (one-sided p)</th>
<th>Rsquare (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clauses</td>
<td>5.8</td>
<td>0.17</td>
<td>p = 0.15 (NS)</td>
<td>3</td>
</tr>
<tr>
<td>% Exploration</td>
<td>−16.9</td>
<td>−0.83</td>
<td>p &lt; 0.0005</td>
<td>69</td>
</tr>
<tr>
<td>% Building</td>
<td>7.8</td>
<td>0.73</td>
<td>p &lt; 0.0005</td>
<td>53</td>
</tr>
<tr>
<td>% evaluating</td>
<td>9.7</td>
<td>0.70</td>
<td>p &lt; 0.0005</td>
<td>50</td>
</tr>
<tr>
<td>Diagnostic accuracy (% correct)</td>
<td>37.2</td>
<td>0.85</td>
<td>p &lt; 0.0005</td>
<td>72</td>
</tr>
<tr>
<td>Time on task [s]</td>
<td>−1.4</td>
<td>−0.03</td>
<td>p = 0.47 (NS)</td>
<td>0</td>
</tr>
<tr>
<td>% Time looking at AOI</td>
<td>8.5</td>
<td>0.35</td>
<td>p = 0.013</td>
<td>12</td>
</tr>
</tbody>
</table>

**Fig. 1 – Time on task, % of time looking at AOI and diagnostic accuracy as a function of years of (log) postgraduate experience. Overall values for four PVCs are indicated per participant (dots) and the regression line (solid line) and 95% confidence interval (broken lines) are presented.**
The combination of perceptual and cognitive data adds a new perspective to theory of how expertise develops. The cognitive data show that with longer experience, clinicians do not need to invest so much effort exploring data. Instead, they can focus their clinical reasoning processes on active building and evaluation of diagnostic hypotheses. Through effortless matching of the new case with illness scripts stored in the extensive, coherent knowledge base in long term memory,\textsuperscript{27} they are able to make a probable diagnosis with high diagnostic accuracy.

In the research setting used here, the time on task for all participants is similar (Fig. 2). We speculate whether the PVCs used were so difficult that the experts did not finish the task quickly but instead purposefully kept looking for additional data to evaluate or confirm a diagnostic hypothesis. For instance, in PVC 2 or 3 (Table 2), an expert may carefully and purposefully pay attention to an AOI, waiting for signs of seizure recurrence in that particular AOI. Experts are known to be able to continuously monitor for, and if necessary, build and evaluate alternative hypotheses.\textsuperscript{5} By simultaneously tracking visual attention and cognitive processes, this study shows how clinicians with increasing experience are able to focus their attention more while building and evaluating hypotheses.

Conversely, the cognitive processes of less experienced clinicians are characterized by extensive data exploration. They try to integrate the data; they stumble over many points that need clarification or they indicate there is insufficient data.\textsuperscript{24} They cannot easily build hypotheses, let alone evaluate them. We know from think-aloud studies that medical students have extensive knowledge, but this knowledge is qualitatively different and not so easily accessible.\textsuperscript{5}

Some noteworthy conclusions can be drawn from the combined perceptual and cognitive data, adding new perspectives on clinical reasoning with complex moving pictures. Firstly, experts’ visual search is effective and accurate, strongly suggesting a non-analytical approach. Secondly; our study shows experts are highly skilled in evaluating diagnostic hypotheses. This skill certainly is of importance in domains like paediatric neurology, in which alternative diagnoses mimicking each other occur frequently.\textsuperscript{4} Thirdly; less experienced clinicians are less effective in visual search. They do not always possess illness scripts that match, and thus their diagnostic accuracy is lower. They are therefore more likely to make diagnostic errors. Fourth; less experienced clinicians do not possess the ability to evaluate hypotheses, i.e., monitor for alternative diagnoses.

The validity of the eye-tracking data is high because of the high quality and accuracy of the eye-tracking equipment used and the skilfulness of the operator. The stringency of the think-aloud protocol adds to the trustworthiness of the concurrent think-aloud reporting.

Although this study included clinicians from 5 hospitals, and therefore must have some external validity, all of the medical students were associated with a single medical school. In spite of this limitation, we do not expect generalisability of the medical student results to be low. We succeeded in recruiting 12 out of 16 eligible specialists in paediatric neurology, yet a smaller proportion of the eligible residents and medical students were included. This difference might introduce potential sampling bias, as those participating could be relatively motivated or skilled diagnosticians. If this source of potential bias was present, the regressions in Table 4 would be underestimated. We believe the size of the bias would be small, as the medical students were within 18 months of graduation and therefore were expected to possess similar diagnostic skills.

The findings may help clinical teachers, using patient video cases in a broad variety of specialties, to better understand why learners pick up – or alternatively miss – important details in a representation, or could not commence the relevant reasoning processes.\textsuperscript{28,29} A study on medical students

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{Cognitive processes identified in the transcripts (total number of clauses, percentage of clauses on data exploration, theory building and theory evaluation) as a function of years of (log) postgraduate experience. Overall values in four PVCs are indicated per participant (dots) and the regression line (solid line) and 95\% confidence interval (broken lines) are presented.}
\end{figure}
identified conditions that video cases used for teaching should meet: the content of video cases should be in line with learners’ previous knowledge, cases should be short and should be viewed in a structured, purposeful manner. Examples of simple strategies using PVCs in clinical settings might include asking residents or medical students cognitive explorative questions, such as: ‘what is important to note in this patient video case?’ or signalling modelling behaviour as follows: ‘listen to what I would think while examining this case’. In conclusion, when more experienced paediatricians analyse a PVC, they spend more time looking at relevant areas and explore data less, yet they build and evaluate more diagnostic theories. These differences have implications for the development of effective teaching environments with PVCs in clinical settings.

Acknowledgements

This work was partly funded by The Quality Development Foundation for Education of Residents, Aarhus, Denmark. Consultant Thomas Thelle, Regional Hospital, Viborg, Denmark is thanked for assistance in designation of AOIs in the PVCs.

REFERENCES

30. De Leng BA, Dolmans DHJM, van de Wiel MWJ, Muijtjens AMM, van der Vleuten CPM. How video cases should be used as authentic stimuli in problem-based medical education. Medical Education 2007;41:181—8.