Multimodal learning strategies of pupils with various abilities

Johan Ek
Department of Cognitive Science
Lund University
Supervisor: Jana Holsanova

Books and learning materials are often multimodal to their character; with content depicted in text, images, tables, diagrams and so on. This multimodality has shown to motivate students, improve retention and knowledge creation. The question is how this fragmented information is approached, being read and combined into meaning – and is this behaviour similar among different ability groups?

Both eye-tracking and verbal retrospective protocols were used in this pilot study, and in addition to a quantitative eye-tracking analysis a new analytical method was developed combining eye-tracking and verbal data. By segmenting the pupils' verbal retrospective data, in which they describe their reading process, smaller functional time units was transposed onto the eye-tracking data, giving a more detailed window to the cognitive processes of the pupils.

Similar results with previous studies were found such as a general low attention on images and a higher attention to relevant information. However in contrary with previous research higher ability pupils did not attend to relevant content to a higher degree. And in our more search-oriented task, more integrative saccades between text and images, was an unsuccessful strategy in answering the questions.

Although problems of the newly developed dual analytical method; such as large differences in the verbal data, lag between gaze and speech, and the few participants in this pilot study – we see possibilities in the method in cognitive, educational and usability research.

Keywords: multimodality, eye-tracking, design, learning, cognition, ability, relevant info, text and images, verbal protocols, strategies, task-oriented reading.

Introduction

This thesis is a part of the research project *Multimodal Learning* conducted at *Lund University Humanities Lab*. The area of research concerns how students interact with multimodal school books, their comprehension of the material and their utilisation of text and images.

The work conducted is an exploratory pilot study using both quantitative eye-tracking measures and qualitative verbal retrospective protocols to examine the cognitive processes of pupils performing task oriented reading. How do the pupils utilise text and images of the learning material when solving the tasks? Are there differences between varying ability groups concerning attention on text or images, reading behaviours and strategies?

The focus of the study lies in developing an analytical method where eye-tracking data are combined with verbal retrospective protocols to gain a deeper understanding of the pupils cognitive processes. The study consists of two parts: a quantitative eye-tracking analysis and an explorative and hypothesis generating analysis where eye-tracking and verbal data are combined using the new method.

Background

Multimodal communication

A large part of the learning process in elementary and secondary school is to look up and answer questions in books. From explicit questions such as historical dates to more complex concepts and ideas where the pupil has to construct meaning from many different sources. With the arrival of computers and Internet, the variety of information and sources is endless, stretching between different modalities e.g. images, texts, movies, sounds, interactive models, articles and so forth. Still however, the traditional textbooks are the main source of information and central pedagogical component during individual exercises.

While textbooks in many cases may appear very traditional, their content consists of an increased amount of multimodal information: text, illustrations, photos, diagrams, fact boxes, captions and different typographic elements and layouts. This development is in line with modern media in general such as newspapers, magazines and adverts (Underwood, 2005).

In many cases the fragmented structures and layouts

of modern media is designed in a multi sequential way, i.e. there is no obvious linear order to read in as the layout offers many possible starting points and reading paths (Holsanova et al, 2006). One might browse using images and headlines only, or when a deeper comprehension is needed read certain paragraphs more closely depending on the current goal. A typical layout today offers many entry points for the reader to start reading from (Figure 1).

The main goal of information design in learning materials is of course to optimise readability and usability. As with all design the designer can inform the user to an effective usage by giving clues or affordances (Norman, 2002). We need signs of what it is for, what is happening, and what alternative actions are (Greeno, 1994).

Clues in e.g. graphic design include colour coding, typographic norms such as headlines, italic keywords, indents to communicate rhetorical clusters of the text and layouts to structure the underlying rhetoric of the information. Good graphic design supports the reader's ability to process and comprehend the content of the information (Holsanova & Nord, 2010).

Do readers in general and of varying abilities differ in what, in which order and how they read and utilise the affordances given by the design? Despite the huge production of all sorts of multimedia instructions, instructional designers seem to base their design choices more on intuitive ideas, estetics and traditions than on sound research results. The design of educational material, and information in general, may need to be re-evaluated and to a higher degree taking the reader's actual behaviour into consideration.


CIÊNCIA & TECNOLOGIA ARQUITETURA

Ecologia nas alturas

O que fará da nova sede do Bank of America, em Nova York, o maior prédio verde do mundo

One Bryant Park
208 metros
55 andares

Como é a arquitetura ecológica do One Bryant Park

Reaproveitamento de água A água das pias, da chuva e do ar-condicionado será capturada para ser usada nas descargas dos banheiros e para irrigar o teto verde. A água capturada da chuva (95% do que cai sobre o prédio) deixa de ir para o sistema de esgoto da cidade. Quando sobrecarregado, o esgoto transborda e joga lixo no Rio Hudson

A TORRE cultiva a água da chuva. Em seguida, ela escoa para dentro de um tanque que a armazena

Teto verde Para diminuir o aquecimento em torno do prédio, o topo do edifício terá uma área coberta com plantas

Energia própria Uma usina no edifício vai gerar 70% da eletricidade a partir de gás natural, menos poluente que o carvão, usado nas termelétricas americanas

Mictórios Não usarão água. A urina escorre por um reservatório selado com glicerina para evitar mau cheiro. Dali vai para o esgoto

Tanques de gelo Máquinas instaladas no subsolo do prédio produzirão gelo à noite, quando a energia é mais barata, e abastecerão o ar-condicionado durante o dia

Materiais recicláveis O concreto usado no One Bryant Park é uma mistura de cimento com 45% de um material feito com sobras de minérios

55% Cimento

45% Sobras de minérios

Transporte público O One Bryant Park terá poucas garagens para carros e dará acesso a 17 linhas de metrô. Quem quiser ir pedalando para o trabalho encontrará lugares para estacionar a bicicleta e chuveiros para tomar banho

1 ILUMINAÇÃO NATURAL O prédio tem janelas panorâmicas. O objetivo é diminuir a iluminação artificial e proporcionar um ambiente mais agradável para os funcionários. Paredes duplas de vidro ajudarão a evitar que o prédio esquente, diminuindo o trabalho do ar-condicionado

2 AR-CONDICIONADO Os dutos serão no chão, e não no teto. O ar frio sai por orifícios no chão e sobe à medida que esquenta

3 SISTEMA DE LUZES Serão usadas luzes LED, mais eficientes que as lâmpadas incandescentes. Sensores instalados nas salas acenderão as luzes apenas quando houver pessoas e a iluminação natural diminuir

4 AR LIMPO O One Bryant Park promete devolver para Nova York um ar mais limpo do que o que entrou no edifício. Um sistema de filtros retirará 95% das partículas do ar. Sensores monitorarão o nível de gás carbônico nas salas e automaticamente aumentarão a quantidade de ar fresco

Figure 1. An example of multimodal content in a newspaper; with images, captions, annotations, headlines and information boxes. (http://www.flickr.com/photos/vergotti/)

Multimodality and learning

Multimedia learning is a research area concerned with the learning effects of combining words with pictures. Commonly the term multimedia is associated with digital information presented in for instance sounds, images or videos. Richard Mayer, educational psychologist and originator of the *Multimedia learning theory*, defines multimedia as to present both words (spoken or printed) and pictures (illustrations, videos, animations) (Mayer, 2005).

The main idea of the theory is that optimal learning occurs when both verbal and visual materials are presented simultaneously. Mayer sees two possible explanations to this phenomenon. First is the idea that the mind consists of two separate input and processing systems: a verbal and a visual. This is consistent with Baddely and Hitch's model of working memory as two parallel and simultaneously working systems (Gazzaniga et al, 2009). By presenting information in both channels the cognitive capacity is maximized.

The second aspect is the properties of the verbal and visual representations. Some types of information are more suited as visual representations compared to verbal and vice versa. Visual representations provide for example a visual-spatial layout of the information in correspondence to the objects in the represented world, i.e. a geographical map, which in many cases are complicated to describe verbally. Another principle is that best effects are seen when text and images are connected by the reader to a uniform mental model.

In school settings the use of illustrations has shown to attract and motivate students, improve retention and facilitate active construction of new knowledge as well as integration with existing (Cook et al, 2008). Four general functions of illustrations have been modelled by Levie & Lentz (Hannus & Hyönä, 1999): *attention guiding*, *affective*, *cognitive*, and *compensatory*.

The first, *attention guiding*, concerns the assumption that readers pay more attention to reading material accompanied with illustrations. Second, the motivation to read will rise as the illustrations are alleged to have an *affective function*. Evidence strongly suggests these are significant factors in learning. However, any motivational factors are not applicable on the text content in general; illustrations

need to be directly connected to the text content to make a difference in learning of the actual content. One might although ask if irrelevant but exciting images may not be a factor when getting motivated to actually open a less intriguing book. This is a factor probably not seen in an experimental setup.

The third, *cognitive aspect*, is the increase of comprehension and memory as illustrations make it easier to process the textual content; and secondly the different types of media create multiform mental representations of a higher detail. According to the *dual-coding theory* (Paivio, 1986), illustrations encourage readers to create pictorial representations to support the propositional ones assuming to be created from the text content.

Lastly, the fourth function: poor readers are expected to gain more effect from illustrations as they have a *complementary* way of input which helps the semantic processing of the text.

Well-designed multimodal presentations can therefore, according to multimedia theory, increase the level of engagement and active learning of the students. By engaging in an active cognitive processing; students divide their attention to different modalities e.g. pictures and text, organising them into a coherent inner mental representation and integrate the representations with prior knowledge. Giving the right clues on how the information is best approached, the deep reading and comprehension can be enhanced (Holsanova et al, 2008). For example Koran & Koran (1980) showed improved results for low reasoning able students in the 7th and 8th grade when schematic diagrams illustrated the structure of the textual information. They reasoned that low ability students do not spontaneously create mental structures of the text information without the supporting diagram (Hannus & Hyönä, 1999).

One might however ask: Does not the increasing usage of images, graphics and modern layouts make it more complicated to read? A successful creation of a coherent mental model requires the reader to connect fragmented multimodal information which might be hard, especially if no explicit references are made between relevant and associated areas. As Mayer himself states this is the case:

"Extraneous material competes for cognitive resources in working memory and can divert attention from the important material, can disrupt the process of organising the material, and can prime the learner to organise the material around an inappropriate theme." (Mayer, 2001, p. 113)

The challenge is to choose the correct representation for a certain type of information and intended usage. Just to add pictures in general has a low likelihood of positive effects as stated by the coherence principle of multimedia theory; extraneous material should only harm the learning of relevant information. Presenting the information with suboptimal affordances such as disconnected layouts or unsystematic typography could lead to a *split-attention effect* impairing the creation of a unified mental model (Mayer, 2005).

Increased complexity does require higher reading abilities – this is the conclusion drawn by Hannus & Hyönä (1999) to their results showing that comprehension scores were improved only for higher ability pupils (age 10) when illustrations were accompanying the text. No positive effect was found in the low-ability group. There may be a lower limit of when additional illustration is beneficial. An increased complexity of the information may increase the cognitive demands and negatively influence students with lower skills. This is in line with Sweller's *Cognitive load theory* where the limitations of working memory during learning need to be taken into consideration when designing information materials. Any extraneous cognitive load from attention diverting information harms the learning process (Mayer, 2005). Problems of e.g. split attention between different modalities can be decreased using layouts keeping images and text close to each other, minimizing the demands on working memory (Sweller et al, 1998).

Additionally one must keep in mind that visual processing in the same way as verbal, is a competence varying between individuals. Some people are better at processing visual information than others. Another influencing factor of this competence is the domain specific prior knowledge of the illustrated content where more prior knowledge minimizes the negative effects of complex material (Scheiter et al, 2010).

To sum up, well designed learning materials with illustrated content do increase learning by facilitating motivation, comprehension and the creation of mental models. However, the key is well designed. The illustrations need to be relevant to the textual content and the layout and design need to guide the user to connect the relevant parts. Otherwise the result will most likely be negative for less able students as the complexity and decoding increase the workload. Of importance for this study is how pupils of different abilities utilise illustrations, to what degree they concentrate on relevant information, integrate between modalities and in general how the information is approached and in what way it is read.

Previous eye-tracking research in multimodal reading

As there is a close relationship between gaze and attention, eye movements are well suited to unfold the underlying processes when reading (Rayner, 1998). How do the students look at the information, how much time is attended to different areas and how does the attention move? Compared to eye-tracking studies of text reading, empirical findings on reading behaviour when texts are combined

with images are scarce (Rayner et al, 2001). This might be surprising regarding the principles stated in e.g. multimedia theory concerning the benefits of images and the vast amount of images used in learning materials and in media in general.

In reading studies regarding plain text, conventional eye-tracking measures are for example fixation durations, look backs (regressions) and saccade lengths. This is a well researched area with plenty and good data. Such as the development of reading skills in children, with faster reading times, decreased fixation durations and fewer regressions. However this data is hard to compare directly with complex multimodal information. For example is the normal fixation duration when reading, between 200–300 ms, dependent on the given task. The fixation durations are influenced by reading styles such as skimming or deep reading, complexity of material and readability issues such as font styles or light conditions.

As mentioned, the main difference when reading a plain body text and complex multimodal information, is of course that you in the later case don't have to follow a straight sequential path. Instead, you can jump freely between different content more similar to searching. A number of studies have for example found that with an increased search difficulty (when distractors are similar to targets), fixation duration increases, more fixations are made and saccade sizes decreases (Rayner, 1998). Searching through various text paragraphs and illustrations looking for a particular answer or keyword would in the same line probably reveal longer fixation duration times compared to fluent reading.

So what mechanisms steer gaze behaviour when reading multimodal information? Current theories of human gaze control in active vision focus on two sources of input to the control system: stimulus-based (bottom-up) and goal-directed (top-down). Bottom-up mechanisms are influenced by the physical properties of the visual stimuli such as colour, contrast, spatiality and complexity; where we tend to notice and attend areas that contrast their surrounding areas. The gaze behaviour is in this manner a result of the visual input, and in our learning materials one might expect that e.g. colourful images, large contrasting headlines and shapes will attract attention. Top-down processes conversely are the result of cognitive processes such as previous experience with the material or the current goal of the task, e.g. where the pupils expect to find an answer. Generally the influence of bottom-up decreases with semantically meaningful scenes and active tasks (Henderson, 2003).

In previous research on multimodal information, two aspects have been in focus: the amount of time attended to different types of representations, e.g. pictures or text, and how the integration between the modalities is done. The attention time on different modalities answers the questions: Where do we look for information when reading an illustrated text? Do we use all types of information or just a specific type of images and so forth? The second aspect of integration can give answers on how we combine text and images and create meaning. Do we constantly switch between connecting images and text or is the input process more of a sequential type?

Hannus and Hyönä (1999) studied both of these aspects to discover differences in how pupils of varying ability utilise images in school books. They found no difference in the amount of attention to text versus illustrations between the groups – both high and low ability groups looked only minimal to illustrations (6% of the total study time). However they found that the skill to distinguish between relevant and irrelevant information were found to correlate with ability – higher ability students paid more attention to relevant areas. This of course is an important capability in all learning and especially in academic studies working with large volumes of information. Additionally, successful pupils to a higher degree integrated between relevant illustrated content and corresponding text; this would indicate positive effects in line with multimedia learning theories as the reader creates better mental models using both modalities. The authors conclude that this integrative process require a certain amount of knowledge and intellectual ability.

The low overall attention times on images is consistent with studies by Hegarty et al (1991) and Carroll et al (1992) where adult people looked at text and pictures consisting of technical drawing and comics respectively. The gaze behaviour in these two studies indicated that the pictures in general were attended to after the connecting text had been read. Secondly, the integrative saccades between text and images were low as the text or captions were interpreted before any inspection of the pictures was made – suggesting that comprehension is text directed (Rayner et al, 2001). This strategy seems natural given the usual design of printed information such as newspapers and magazines with captions describing the content and purpose of an image. It is probably more work to try to decipher the meaning of an image before any of the text content has been read. Commonly stated in newspaper design we generally first gaze to images (bottom-up driven), however any detailed inspection are performed after the processing of the corresponding text has been performed.

Additionally, often images and illustrations in newspapers and books are not relevant or informative. This issue of relevance was tried in an educational setting by Slykhuis et al (2005) where students looked at a PowerPoint slideshow including both complementary and decorative images with various relevance and references in the text. The results showed that complementary images referred in the text received significantly more attention from the students.

Previous experience with learning material can affect our overall reading behaviour, such as ignoring images.

Low dwell times on images and low integrative saccades was also found by Rayner et al (2001), when viewers looked at print advertisements of products they were planning to buy. The viewers also generally first looked at the large print, small print and lastly the picture. Most likely readers concentrate and stick to the areas that seem most relevant and informative according to their current goal.

In a similar study by Radach et al (2003), the participants were asked to view ads for a following questionnaire asking about interestingness and content. This task resulted in a higher degree of transitions between text and pictures. Perhaps is the evaluative task more complicated than the search oriented task used in Rayner et al (2001); and the participants to a higher degree needed to form a mental model of the ad including personal judgements of the content.

In a follow-up study by Rayner et al (2008), the goal of the viewer was used as an variable in the experiment. The participants were asked to view ads with the intention of rate the likeability or the effectiveness of the ads. These two different tasks resulted in a shift of attention from text to images, compared to the study from 2001. One might speculate that the task of rating the effectiveness or likeability of an ad, similiar to Radach et al (2003), is more a matter of evaluating the picture, copy or concept behind it, than reading the information of the product which is more relevant when purchasing it. However in contrast with Radach et al (2003) the viewers did not tend to jump between text and images.

As Yarbus showed in his well cited study, *Eye movements and Vision*, fixations over time and space reveal different strategies of information acquisition as the result of different tasks (Yarbus, 1967). The influence of a particular task is crucial for the gaze behaviour as the relevance of information is dependent on the aim of the user.

In addition, the properties of the information material (bottom-up factors) are of course influencing how it is being read. E.g. the absence of explicit references in the text to the images makes integrative saccades less likely. The influence of layout for attention on text and images (bottom-up guidance) has for example been studied by Holsanova et al (2008) concluding that formats using spatial closeness and sequential orders facilitate integrative saccades and increases reading time and comprehension compared to a separate layout where text and images are treated as two different units.

Other studies show how graphic cues such as colour coding promotes integrative saccades leading to a more effective learning due to a higher efficiency of locating corresponding information between illustrations and text (Ozcelik et al, 2009).

However as mentioned, increased complexity and transitions between text and images are not always positive. Reid and Beveridge (1990) showed that less successful learners to a higher degree spent time on and more often accessed illustrations. Successful students remembered more and less successful students remembered less when complementary images were used. Their conclusion drawn was that the frequent move of attention between text and illustrated areas harmed the learning process as the student did not manage to create a coherent meaning of the information (Hannus & Hyönä, 1999). A similar conclusion was made by Harber (1983) who observed negative effects of illustrations among disabled children between the 2nd and 4th grade; in contrast to positive effects in the normal achieving group. Although Harber did not use eye-tracking or some other measure of their attentional process he interpreted the result as an indication that the illustrations were more distractive than helpful for the disabled children (Hannus & Hyönä, 1999). A hypothesis is that the disabled children to a higher degree are driven by bottom-up factors and this is also in line with the hypothesis that reading multimodal information is a complex process requiring many active decisions of the reader to benefit from the information, such as evaluating the relevance of the information. Because pictures and text cannot be perceived simultaneously, the learner is forced to switch back and forth between the two and integrate them mentally. This integration process is cognitively demanding and at the expense of mental resources that could otherwise be allocated to other areas of the learning process. The effects of integrative saccades is also discussed in Holsanova et al (2008) were no correlation was found between integrative saccades and comprehensions, the positive effects of the integrated format was instead found in reading time, reading order and the number of fixations. It may not be the amount of transitions being made that is important but when and between what types of information. Such as the integrative saccades between associated and relevant content seen in Hannus and Hyönä (1999). A question of doing it in the right moment for the right reason?

An additional variable reflecting attention to an area is the ability to extract and comprehend the information. In reading, words or texts that are difficult to process, usually result in longer dwell times, longer fixation durations and a higher degree of fixations (Rayner, 1998). This correlation has also been found in advertisements with text and pictures where areas of a higher degree of complexity are looked at for a longer time (Radach et al, 2003). Graf et al have similarly found in software usability tests that long fixation duration times and longer saccade lengths correlate with higher cognitive strain reported by the users (Hansen, 1991). Eye movements recorded during solving maths and physics problems also show more and longer fixations with

increased complexity (Rayner, 1998).

To sum up the previous findings, reading multimodal learning materials is to a high extent text driven and attention on images are in general low as well as integrative saccades between modalities. The common view of images and illustrations as most helpful for less able pupils does not seem to match the reality. Contrary, illustrated content accompanying text in most referred studies is mostly beneficial for more advanced readers. They seem to have the cognitive resources to extract the positive effects from the often more complicated multimodal information – to evaluate the information, create meaning of different modalities and not getting distracted by extraneous information.

Mentioned eye-tracking measures such as dwell time (attention) and integrative saccades (transitions between text and images) reveals when and how long an area is focused upon, and additionally saccade lengths and fixation duration times can reveal the current cognitive load. The problem is that these measures make it difficult to deduce why e.g. the reader looks at images or why the reader integrates between modalities (Pernice & Nielsen, 2009). Is integrative transitions a good measure for an integrative meaningful processing or is it mere a result of diverted attention? Eye-tracking methodology is limited in gaining insights into deeper cognitive processes of e.g. strategic problem solving. None of the above studies used complementary data such as verbal protocols to investigate the cognitive processes more deeply. This may be a way to find if, how and when different actions are performed by the pupils and how to interpret the above measures during these sequences.

Verbal protocols

A commonly method in e.g. usability studies to get user input is concurrent verbal reports a.k.a think-aloud protocols, where the user verbalises the process at the same time as it is performed (Nielsen, 1993). The user formulates verbally statements such as what she is doing, how she perceives things, when problems arise and why certain actions are done. The main idea of the method is to *"show what the users are doing and why they are doing it while they are doing it, in order to avoid later rationalisations"* (Conyer, 1995).

Still commonly used, the method has some obvious drawbacks. The most problematic is how much the actual verbalisation affects the cognitive processes the method is trying to gain insight into. Applied in complicated problem-solving tasks, the method has shown to reduce probability of insight (Knoblich et al, 2005). This indicates that the verbalisation alters the simultaneous thinking process. It can also influence the performance of the current task impacting variables such as time and results making it a less good choice in performance measurements (Pernice & Nielsen, 2009). The mental process of speech production is also closely tied to eye movements (Griffin, 2004), affecting the eye-tracking data with e.g. more fixations per element and longer fixation times (Ehmke & Wilson, 2007). Another, more noticeable issue, is that thinking out loud seems very unnatural to most people making it hard to keep up a steady stream of utterances as the task is performed. This is in many cases most problematic for expert users as they perform the actions quickly and may not even consciously know what they are doing as the processes are automated (Nielsen, 1993). This is of great concern in such a highly automated process as reading and searching in a book.

An alternative to concurrent reporting is retrospective reports, i.e. the user or participant describes and explains the process after the task has been completed, removing any influence on the actual task. A disadvantage on the other hand is the risk of forgetfulness or a selective memory such as only remembering successful strategies or to rationalize ones not always completely rational actions.

A method advocated by Hansen (1991) and Van Gog et al (2005) is by combining the retrospective method with external memory cues as support during the verbalisation. In e.g. usability studies video recordings of the interaction process are often used to cue the memories of external activities and the internal mental processes that can be derived from them.

In the same manner as video recordings, the user can watch recordings of their own eye movements of the previous interactions superimposed on the original stimuli. As a large part of our gaze behaviour is more or less unconscious the gaze recordings are a good method to actually see how you did complete a task. This method has shown to be very informative. Critics of this method however may argue to what extent the user remembers his own process and how the gaze recording patterns can be interpreted and used to create new explanations of what is happening on the screen. The comments will express how the user interprets the actions after the task is done and will reflect knowledge of the task and solution they did not possess in the original situation (Hansen, 1991).

Overall, the method of retrospective protocols is best suited for our fast paced and in many cases automatic process where we don't want to influence the gaze behaviour recordings.

Combining verbal protocols and eye-tracking data

The same way as our eyes move as a "spotlight of attention" between information, our cognitive attention is divided into and moved between verbal utterances. Although we in our day-to-day conversations experience language as well-structured, fluent and a result of a coherent flow of thoughts, this is not the case. The linguist Wallace Chafe has shown that we, during verbal description, focus on one idea at the time due to our cognitive limitations. This segmented structure of the verbal output makes it possible

to gain insight into the underlying processes (Holsanova, 2008). Additionally, the time boundaries of these segments can provide us with time stamps of when an idea segment starts and ends, which can be coordinated with the data collection method of eye-tracking. The eye-tracking data consisting of fixations and saccades can thereby be divided into smaller units, using the time stamps from the verbal data, reflecting the different "functions" performed by the user. We thereby get a functional segmentation in time.

However, these start and end times given from the verbal protocols may not be in temporal sync with the ET-data. E.g. Holsanova (2008) shows how there seldom is a 1:1 match between verbal and visual attention. Even in a real-time, e.g. in verbal descriptions of images; the speech production is usually lagging behind the visual focus. If an object (e.g. an apple) is fixated and described, the verbal account normally occurs 2–3 seconds after the fixation. In our case luckily, we are only interested in the wider functional units and not specific words, which gives us a wider analytical window. Secondly the participants describe a recording replayed in a quarter of the original speed making it easier to keep up with the movements of the eyes.

The aim of this combined analytical phase of verbal and eye-tracking data is to find more detailed insights into the underlying cognitive processes of the pupils when reading multimodal information. Can the segmentation in function and time unfold issues such as: How do the pupils use the different available modalities and how is their cognitive phases structured? Different strategic processes are used in various cases such as reading to locate a particular fact, skimming to get the essence, or to evaluate the content. Using this method we can combine the quantitative eye-tracking measures with the pupils account of what they are doing in a given moment. Measures such as dwell time or transitions may be more useful if we know certain actions or cognitive functions are being performed. This method has not been used before and this part of the thesis are exploratory to its nature aiming to generate hypothesis and evaluate the method for future research.

Intentions and research questions

We intend to discover differences in how the different ability pupils chose to navigate and read the information. As this thesis is divided into one experimental part using quantitative data and one explorative part using both quantitative and qualitative data, the nature of the intentions and questions differ. The four hypothesis listed is quantitative and to some extent replicating the methods and designs of previous studies, however with a new educational context of science books with a high ecological validity.

We intend to:

1. Find a positive correlation between increased ability and score result.

2. Overall attention (dwell time) on images is assumed to be low compared to attention on text as in previous referred studies.

3. Find that proportionally more attention (dwell time) on relevant areas correlate with higher ability.

4. Find a difference in the amount of integrative saccades (transitions between text and images) between the three groups and their possible positive or negative effects.

The second explorative part using the new method of combing verbal retrospective protocols and eye-tracking data is not hypothesis driven but builds upon the findings and questions arising from previous and the current quantitative study. Can more detailed insights reveal as the eye-tracking data is segmented into smaller time units of varying functional type? What data is found during certain actions of the pupils and can this clarify the intentions of eye-tracking measures such as e.g. integrative transitions?

A major part is the development and evaluation of the method itself to see what can be gained from its usage, possible limitations and future improvements.

Material and method

Participants

A total of twelve native Swedish speaking 8th-grade pupils (6 females, 6 males, age = 14) were recruited from a secondary level school in Lund. The pupils were selected by their ability to cooperate and interest in participating, and thereafter divided into 3 groups, high, medium and low ability, based on their overall school performance according to grades and teacher assessments. The groups were evenly divided between boys and girls. One issue mentioned by the supervising teacher during the experiment was the problem of finding low ability students motivated to participate in the experiment. Both the particular school and the city of Lund has an overall high performance and grade point average compared to Sweden and surrounding regions. This may show up as a decreased difference between the ability groups. All participants had normal or corrected-to-normal vision.

Equipment

Eye movements were recorded at Lund University Humanities Lab and monitored via an SMI Remote RED 250 eye tracker estimating the point of gaze based on the pupil and corneal reflection. The eye tracker has a spatial resolution of 1° and eye position was sampled at 120 Hz. The stimuli were presented on a 22-inch LCD monitor with a resolution of 1680×1050 pixels running at 60 Hz and at the viewing distance of approximately 700 mm.

The data was recorded with the iView X 2.4 software and participants were calibrated using a 9-point RED calibration routine with validation. The experiment leader controlled and observed the different steps of the procedure, such as calibration and data recording, using a dual screen setup. The accuracy of calibration was below 1° for all of the participants.

The pupils interacted with the computer, i.e. selecting answers, using a mouse. Sounds from the verbal retrospective protocol were recorded using an unobtrusive omnidirectional table microphone.

Stimuli

Five book pages were selected and digitised from an existing science book used in the secondary levels of Swedish schools to increase the ecological validity of the study. Each page described different topics in the areas of biology and chemistry: the function of blood, photosynthesis, blood cells, bloodstreams and biofuels. The pages consisted of multimodal information such as headings, texts, illustrations, captions and annotations. The illustrations and photos were linked to the text to a various degree (e.g. some were more of aesthetically

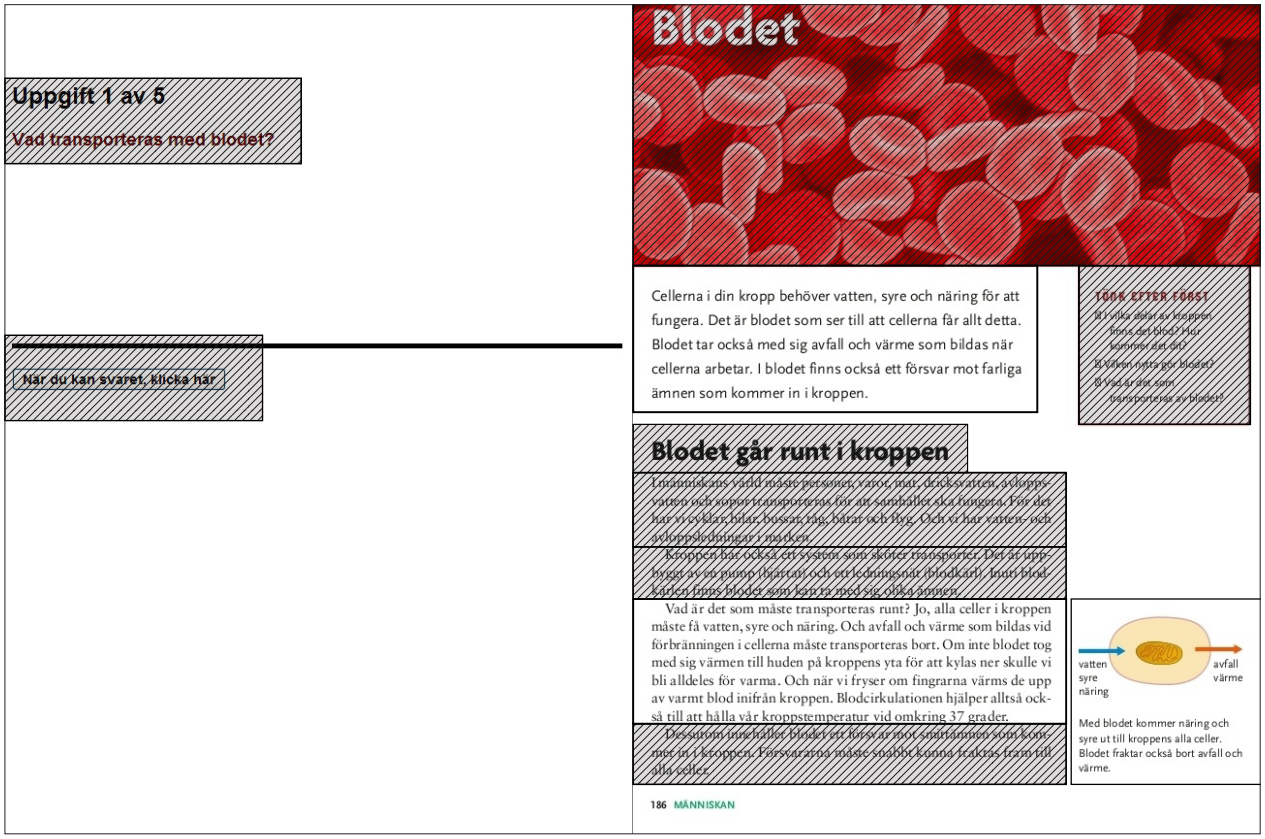

Figure 2. Example of the learning material stimuli used in the study: the question is in the upper left corner, below is the button to get to the answering page, and to the right side of the spread is the information taken from an existing educational book. The rectangles defines the AOIs used in the analysis where the unshaded ones represents relevant segments. The pupils had not seen the material before.

than informative character) but there were no explicit references to the illustrations in the text or any use of colour coding or pointers to guide the user – except of course spatial closeness between e.g. illustrations and captions.

The reading material presented on the screen for the participants consisted of pages laid out as a book spread with the question/task and answering button to the left and the connecting book page material to the right (Figure 2).

The rectangles defines the AOIs used in the analysis where the unshaded ones represents relevant segments.

Clicking the answering button reveales the next spread consisting of four multiple choice answers to the right with one correct answer, two partial correct answers and one incorrect answer (Figure 3). The word lengths of the different answers were kept in a close range and in randomized order for every participant minimizing possible influence of order sequence. The material was implemented in HTML and running on Internet Explorer 8.

Procedure

At the beginning of the experiment, each participant was instructed to seat themselves comfortable in front of the computer and thence calibrated until an acceptable accuracy been reached. The experimenter could also check the accuracy of the calibration at any time during the experiment and recalibrate if necessary. The experimental setup consisted of three phases (Figure 4):

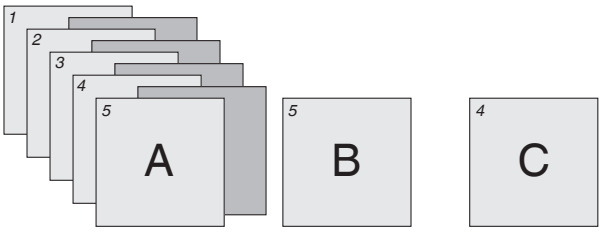

Figure 4.

A. The student is presented with 5 screens with one question and corresponding material in each. After each screen an answering page with 4 multiple choice answers prompts the pupil for an answer.

B. The student watches a previously recorded video of a user performing a verbal retrospective of task 5.

C. The student performs a verbal retrospective while watching her recorded gaze path during the solution of task 4.

A. Completing the reading and answering of the five tasks. In this phase the pupil was presented with on-screen instructions and all pupils could independently move through the different pages at their own pace; stepping through a total of five spreads of questions and book pages, and answering four of these five questions.

B. A video instruction of how to perform a verbal retrospective. The last spread was used as an example of how

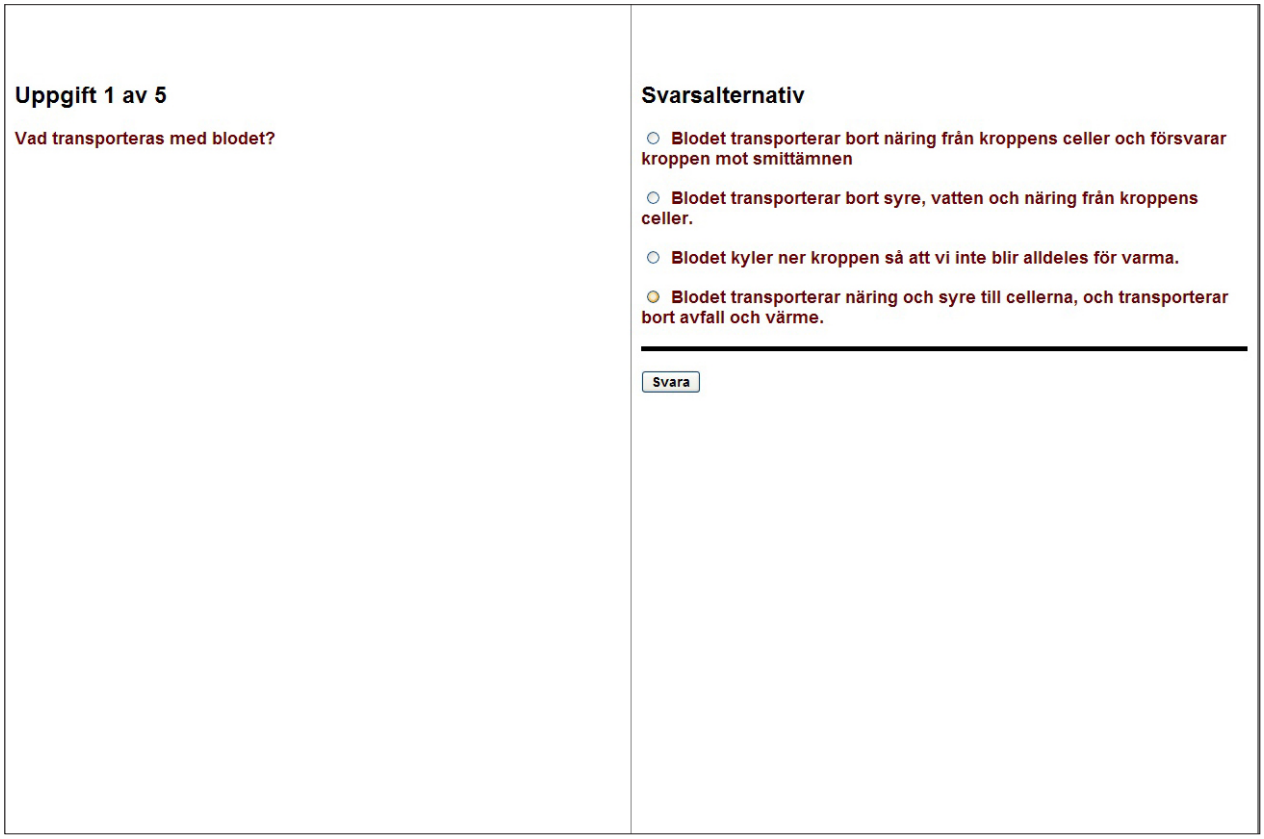

Figure 3. Example of the answering page. 1 correct answer, 2 partial correct answers and 1 incorrect answer.

to perform the final verbal retrospective protocol. As all of the pupils were naïve of eye tracking and verbal protocols; a previously recorded movie was presented of how a participant describes the process with the help from the participant's own superimposed gaze path on the fifth spread, running at a speed of 25%. The idea behind this step was for the participants to familiar themselves of the appearance and presence of gaze paths and how to describe and talk about their inner thinking processes without any feedback from the experimenter during the retrospective. All pupils quickly comprehended the meaning of the digital fixation marks.

C. Performing a verbal retrospective of task number 4. All pupils (with the exception of one pupil feeling uncomfortable talking to the computer) performed the retrospection of the fourth task. They were formulating freely, without guiding or feedback from the experiment leader. The time taken to complete these three phases ranged between 15–40 minutes (M = 26; SD = 7.6). The longest times were due to technical problems, Internet Explorer froze during five of the recordings (P02, P03, P07, P10, P11) requiring a restart of Internet Explorer. These sessions were continued at the last functioning step after a recalibration of the eye tracking equipment. No data was lost as the problems all occurred during transitions between different screens.

Eye-tracking data analysis

Eye-tracking data were analysed using SMI's BeGaze software by dividing the different pages into different regions (areas of interest (AOI)) of text and illustrative components. All though the definition and process of selecting relevant and irrelevant information is not clearly described in the earlier referred studies, we have no such problem in our current study as the particular question for each task give boundaries of what information is vital to answer the question. AOIs defining the relevant segments were selected by a process where five adults, not involved in the study, were asked to highlight the relevant segments of texts and illustration according to the associated question. Segments highlighted by at least 3 of 5 persons were defined as relevant (Figure 1). The sizes and the borders of the AOIs were selected with respect to the accuracy of the eye tracking equipment and the textual and rhetorical clusters of the layout, e.g. an AOI consisted of one headline or a text passage typographically signalled with an indent.

Verbal retrospective data analysis

The retrospective verbal data was recorded, transcribed and segmented into idea units. Each line in the transcript represents a new verbal focus, expressing the content of active mind. The verbal focuses are usually divided by a pause or hesitation when new active information replaces the old.

These verbal focuses can in turn be clustered into super foci, i.e. a group of verbal utterances with the same thematic content expressing the overall idea of the verbal expressions. This was done reading the transcript while listening to the audio recording. Three people involved in the project listened to the recording and read the transcript to divide the utterances from the pupils into idea segments, as described in Holsanova (2008). An example of this segmentation is seen in Table 1. The horizontal lines represent the borders of our segmentation of the verbal foci into super foci, or idea units.

Thereafter the segments were categorized into various functional categories of what the content of the utterance described. Importantly, the categorization only was based upon explicit utterances of the pupils, trying to avoid any interpretation of what an utterance may represent in a certain context. This was an iterative process where new categories were created when needed and previous categorizations were updated to the new list. Finally the 28 cat-

subject	cst	cet	com
4	*4558*	*5790*	*mm*
4	*5790*	*8315*	*först så kollade ja på uppgiften*
4	*8315*	*11415*	*såg va de va för en uppgift ja skulle göra*
4	*11415*	*14548*	*och så*
4	*14548*	*16305*	*kollade ja igenom teckningarna*
4	*16305*	*18371*	*men de va inte så mycke som*
4	*18371*	*20231*	*såg relevant ut så som dom sa*
4	*20231*	*22746*	*å så såg ja att de stod stora kretsloppet där*
4	*22746*	*24291*	*å lilla kretsloppet*
4	*24291*	*27290*	*så då ville ja läsa lite mer om de*
4	*27290*	*29083*	*och så*
4	*29083*	*32564*	*i å med att ja hade sett frågan så*
4	*32564*	*33450*	*äh*
4	*33450*	*35150*	*så kollade ja på skillnaden där*
4	*35150*	*36881*	*att dom ha olika blodkärl*
4	*36881*	*39790*	*utgår från olika delar av hjärtat*
4	*39790*	*41158*	*och*
4	*41158*	*42603*	*ja ja läste de ett par gången*
4	*42603*	*43801*	*för å verkligen*
4	*43801*	*47395*	*förstå va de va ja hade läst*
4	*47395*	*49565*	*ähm*
4	*49565*	*50883*	*och*
4	*50883*	*52723*	*så kollade ja på teckningen igen*
4	*52723*	*54065*	*för å se*
4	*54065*	*54980*	*vad*
4	*54980*	*58476*	*vilka delar de va ja hade läst om*
4	...	...	...

Table 1. Wavy lines represent the segmented idea units.

egories created in this process were reduced and combined into 7 super categories whereas one was a null category (consisting of irrelevant types of utterances not related to the retrospective) not used in the analysis (Table 2).

To test the reliability of the functional categories during the process of defining the utterances, an interrater reliability analysis using the Kappa statistic was performed to determine the consistency among the raters. The mean interrater reliability for the 3 raters was found to be Kappa = 0.70 ($p < .001$). This result is interpreted as a substantial agreement according to Landis & Kock (1977).

As each utterance has a start and end time, a temporal segmentation can be transposed onto the eye-tracking data. The temporal structures of the functional categories are visualized as a colour coded time line for ocular inspection (Figure 5). The lengths of the time lines are normalized to a standard length to accommodative for an easier comparison of possible patterns.

Func. Category	E.g. Utterance
Reading task:	*"först läste jag frågan eller"*
	"såg va de va för uppgift ja skulle göra"
	"å så läste jag uppgiften igen för å kolla exakt vad de va ja skulle svara på"
Reading text:	*"och fortsätter skum skumläsa lite"*
	"ja ja läste de ett par gången för å verkligen förstå va de va ja hade läst"
	"ja läste nog inte hela texten"
	"ja läste mest den översta delen av texten"
	"och så ögnade jag igenom den undre texten också"
Looking at picture:	*"tittar på nedre bilden"*
	"för där såg man liksom en bild"
	"så titta ja på den de hjärtat me pilarna"
	"så kollade ja på teckningen igen för å se vad vilka dela de va ja hade läst om"
Looking back:	*"sen kontrollerade ja de en gång till liksom om de va nåt mer ja kunde hitta"*
	"sen kolla ja rubriken igen"
	"och så läste ja om den igen för å va extra säker"
Remembering:	*"å försökte komma ihåg de"*
	"å försökte memorera va de va ännu mer"
	"försökte minnas allting"
	"läsa å sen komma ihåg"
Evaluting info:	*"å de va rättså viktigt"*
	"såg ja inte så mycket så ja hoppade"
	"men de va inte så mycke som såg relevant ut så som dom sa"
	"för jag trodde att den informationen där uppe var viktigast"

Table 2. Examples of utterances in the 6 functional categories.

Results

1. Score results

The pupils were credited with one point for each correct answer, resulting in a maximum score of four points. The mean scores (1. M = 3.25 (SD = 1.5); 2. M = 2.25 (SD = 1.7); 3. M = 3.00 (SD = 0.8)) of the three different groups, using a one-way ANOVA with the ability level as a between subject variable, showed no significant difference between the groups (Figure 6).

2. Dwell time on text and images

The AOIs in this analyse are grouped into the categories text and images. As seen in Figure 7 no significant differences were found in dwell times on text and illustration between the three groups. Moreover no differences in total reading time were found; better students did not answer the questions faster. Among all groups the mean dwell time on images were 27.5 (SD = 20.8) seconds compared to 260.1 (SD = 87.9) on text segments.

3. Dwell time on and first fixation of relevant areas

Given the difference in screen size taken up by relevant and irrelevant AOIs a pixel control analysis was made by adjusting the dwell times to the amount of space actually taken up by the various AOI segments of the spreads getting a dwell time per pixel measure (Rayner et al, 2001). The size in pixels of relevant text segments were 10% smaller than irrelevant text segments; increasing the dwell time on relevant segments slightly when adjusted for pixel size (from a mean dwell percentage of 60 to 67%). Also the relevant illustrated segments were smaller but to a higher degree, 30%, increasing the mean proportional dwell time on relevant areas from 43 to 62%. The pixel compensation did though not affect any levels of significance compared to the original times.

A two-tailed paired samples t-test confirmed that the

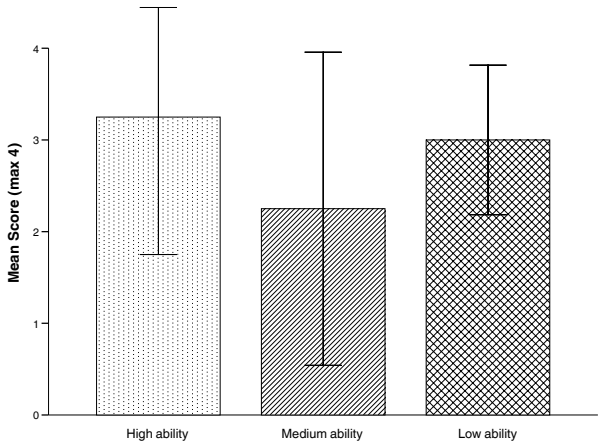

Figure 6. Mean score for the three groups (maximum score = 4).

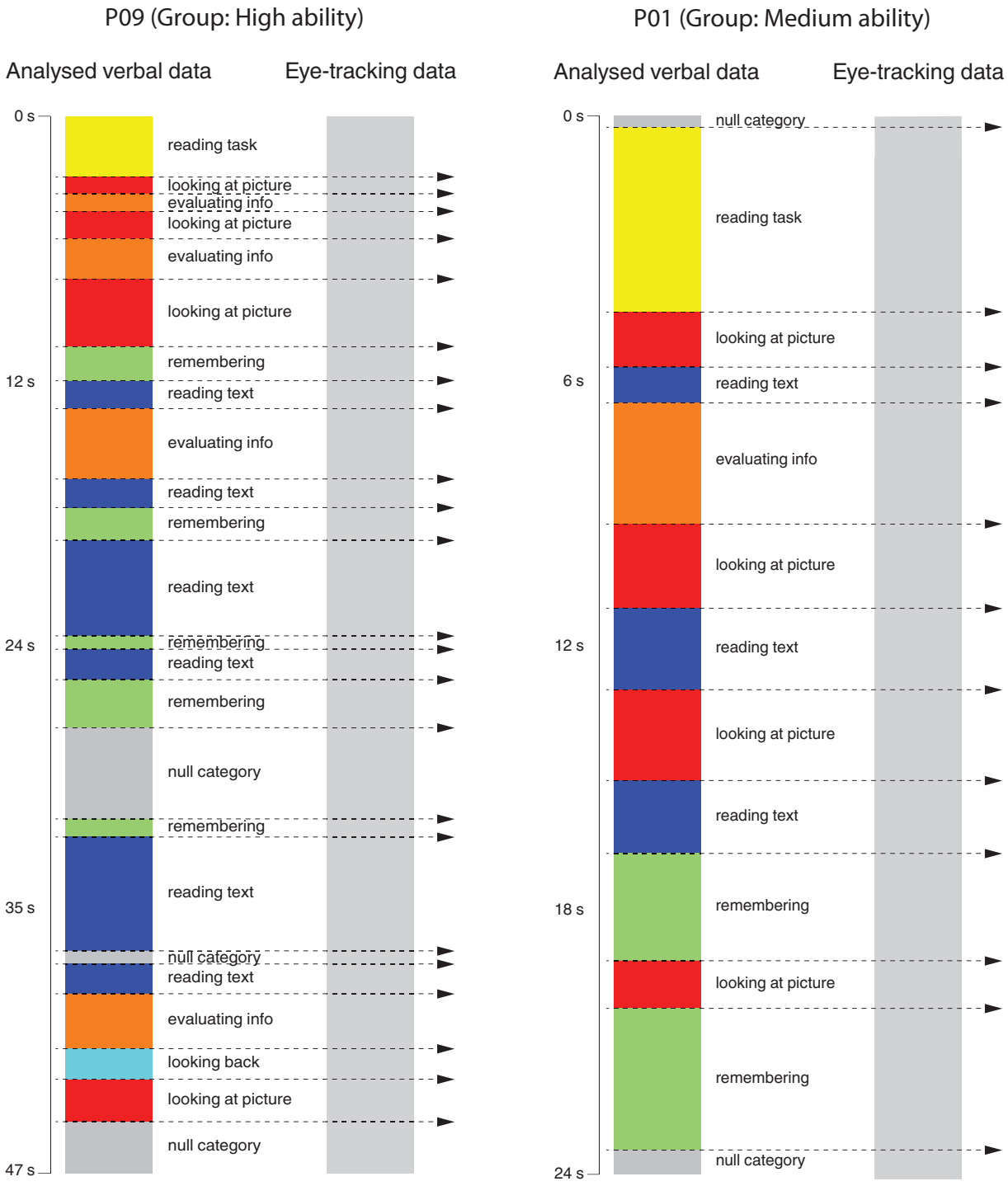

Figure 5. As each verbal utterance in the retrospective protocol has a start and end time, a temporal segmentation can be transposed onto the eye-tracking data. The temporal structures of the functional categories are visualized as a colour coded time line for easier ocular inspection. Example of the functional segmentation of the verbal data (to the left) and the temporal borders (arrows) adopted onto the eye-tracking data (to the right) in 2 pupils.

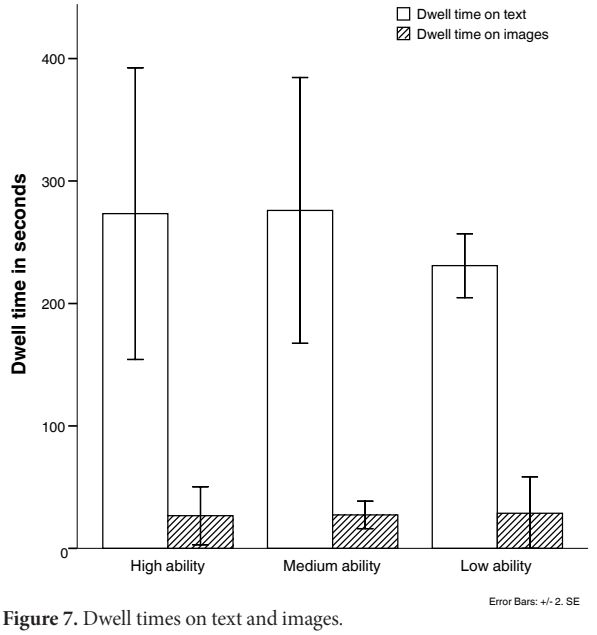

Figure 7. Dwell times on text and images.

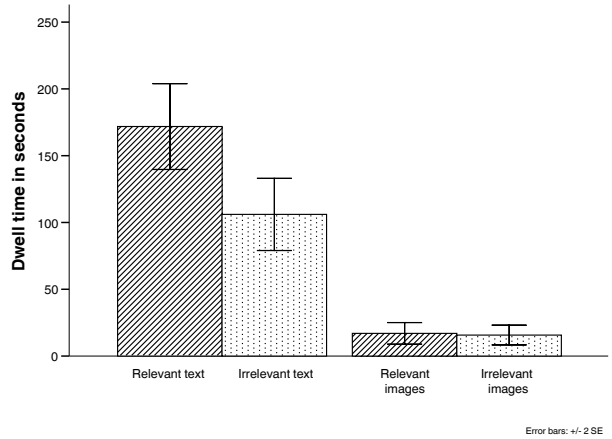

Figure 8. Dwell times on relevant and irrelevant AOIs.

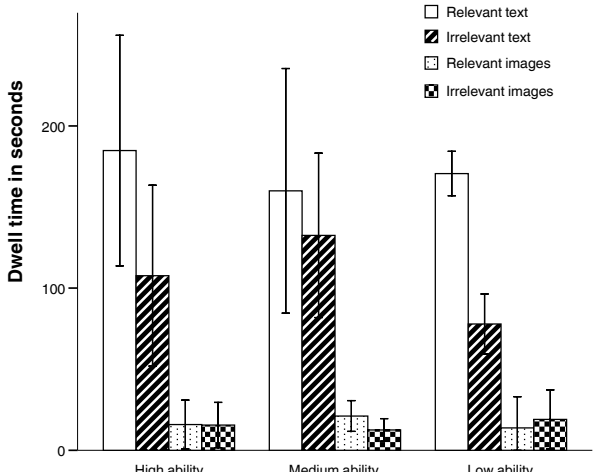

Figure 9. Dwell times relevant and irrelevant AOIs divided per group.

pupils had a significant ($t(10) = 5.27$; $p < 01$) longer dwell time on relevant text segments (M = 106; SD = 46.78) than irrelevant text segments (M = 171.8; SD = 55.6) (Figure 7). However divided into the three groups only the third group had a statistically significant difference on relevant segments due to large variations of group 1 and group 2 (Figure 9).

On images, no difference was found between relevant and irrelevant content: the dwell times were in general low and the variations were large as certain pupils neglected a large portion of the images.

As well, no significant differences were found between the groups in the time to first fixate on a relevant areas. On text areas: 1. M = 5.4 (SD = 9.4); 2. M = 4.2 (SD = 7.9); 3. M = 9.1 (SD = 10.3). On image areas: 1. M = 31.9 (SD = 23.4); 2. M = 13.6 (SD = 18.3); 3. M = 27.7 (SD = 30.5).

4. Amount of transitions between text and images

The mean amount of transitions between text and images for group 1, 2 and 3 were respectively (Figure 10): M = 7.50 (SD = 1.92); M = 14.50 (SD = 3.70); M = 9.25 (SD = 4.03). An one-way ANOVA showed F to be significant beyond the .05 level: $F(2, 9) = 4.74$; $p < .05$. Adjusted R Squared = .405. A Tukey post hoc test revealed a significant difference between group 1 and 2 ($p < .05$). Figure 11 shows the text/image transitions compared to the total amount of AOI-transitions.

5. Combined verbal and eye-tracking data

Using the data from the verbal analysis the eye-tracking data are divided into the 6 functional super categories (the null category was excluded), giving data on the various moments of the task solving.

Figure 12 shows transitions between AOIs in each functional category divided by group.

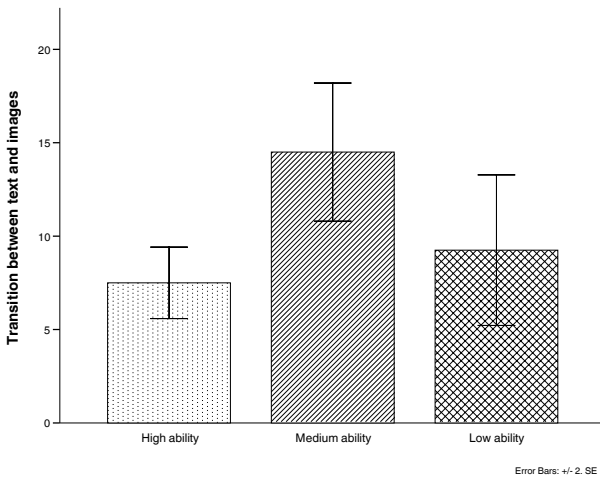

Figure 10. Amount of transitions between text and image AOIs.

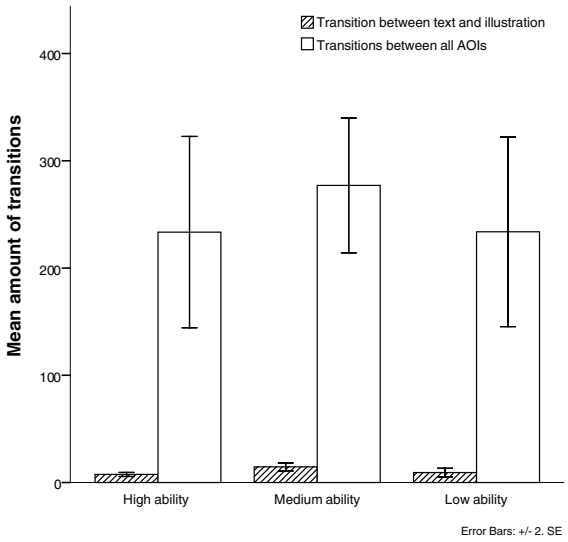

Figure 11. Amount of text/image-transitions compared to total amount of transitions.

Figure 13 shows the mean fixation dispersions, i.e. the mean variance of the y and x coordinates of the fixations per category.

Figure 14 shows the fixation durations of the three different groups on the content: text and images. The mean fixation duration times for the groups differ between 338–466 ms for text and 310–385 ms for images. No significant differences were found between the groups. Figure 15 shows the same fixation duration data separated into the 6 functional categories.

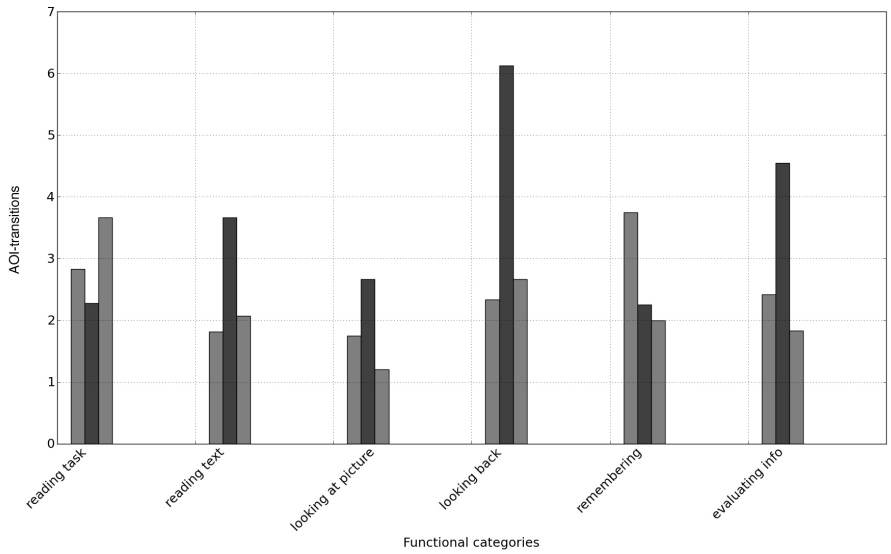

Figure 12.
Amount of transitions per functional category (High, medium and low ability).

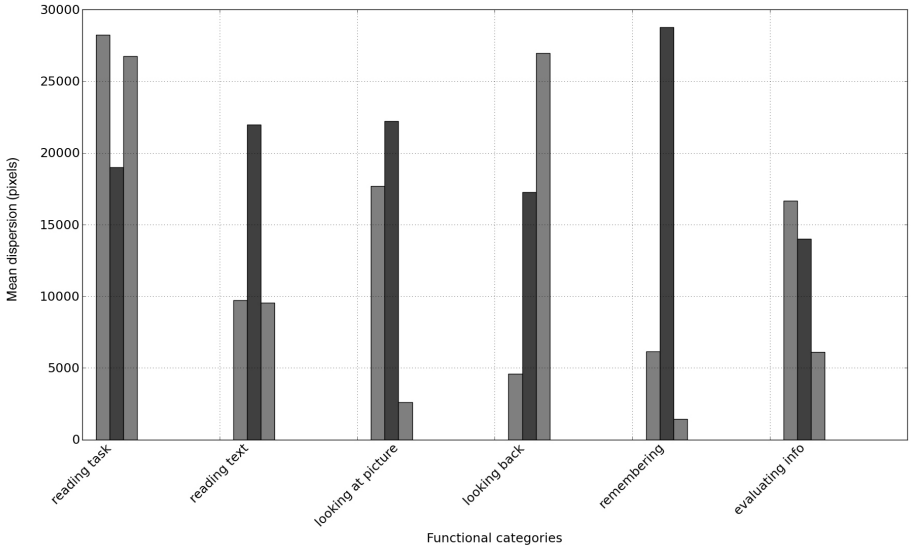

Figure 13.
Fixation dispersion per functional category. (High, medium and low ability)

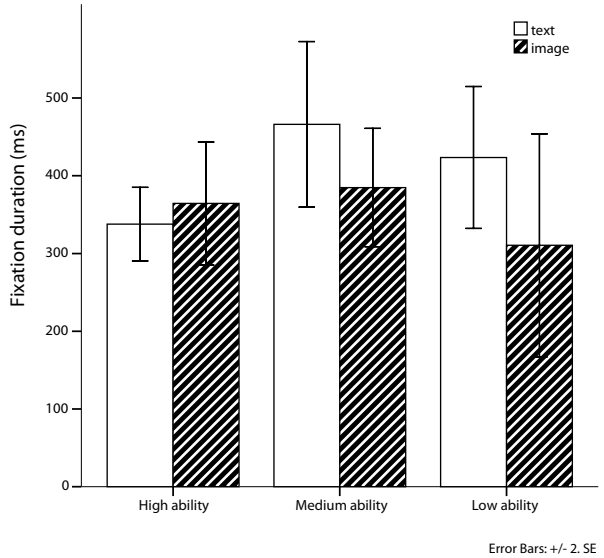

Figure 14. Mean fixation duration of the 3 groups on text and images.

Discussion

In general the results were non-significant in describing any differences between the groups; the individual variations between the pupils are large. Our expected linear change between the 3 groups was not evident, and to our surprise group number 2 stands out in most of the measures.

Score results (1)

As apparent of the mean scores, there is no correlation in the mean score with increasing ability and the within group deviations are large. Factors could be that the categorisation made of the pupils were not a good indication of performance in this particular task. A mentioned issue could be the problem of not getting the lowest ability pupils to participate in the study due to lack of motivation.

An additional factor could be the type of questions used in the task. E.g. the results in Hannus and Hyönä (1999) suggests that there is a more pronounced difference between high- and low-ability pupils in answering comprehension (explanative) than detailed questions (non-explanative). A factor that may be even more pronounced in this setup where the pupils have the task of finding the correct answers. Any possible differences between the groups are likely to be small and due to individual deviations more participants and amount of questions are probably needed for any significance.

Dwell time on text and images (2)

In line with the results from Hannus & Hyönä (1999), Radach et al (2003) and Rayner et al (2001) dwell times are low on images compared to text. No significant differences were found in how much time was attended to text/illustrations or their proportional relation. This was not expected as 8th grade pupils have good reading skills and any complementary input of images to compensate poor reading skills, as stated by the principle of Levie & Lentz, are not expected. Respectively, high dwell times can be the result of motivated students reading carefully and double checking all options or in contrast by a student who have trouble finding or understanding the relevant information. Hannus & Hyönä (1999) showed that the same dwell times were the result of fast readers reading the text multiple times and slow readers with fewer tendencies to re-read.

The general low dwell times on illustrated content in all groups may be the result of both the current task and the design of the textbook material (Rayner et al, 2008). Com-

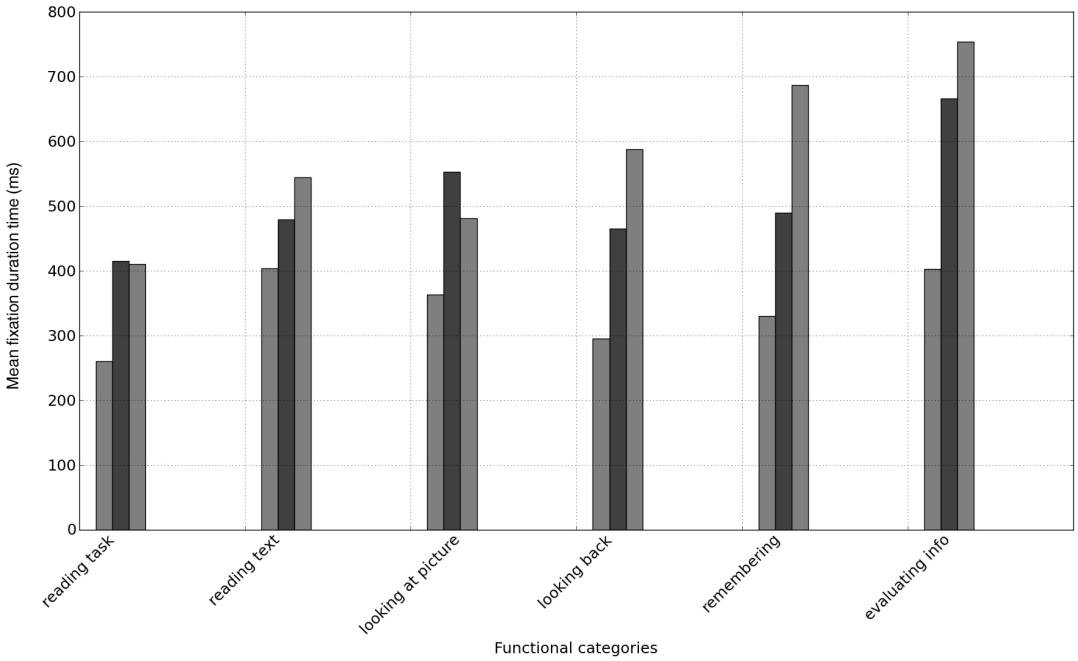

Figure 15. Fixation duration (ms) in functional segments per ability group (High, medium and low ability).

monly and criticized by Mayer (Fletcher & Tobias, 2005), a portion of the images served no instructional purpose at all, and given the goal of our task the majority of images were irrelevant. All questions are solvable without looking at any of the illustrations; this is not the case for the illustrations which only gives enough information in 3 of 5 questions. In addition the questions may have been biased towards textual information as opposed to formulations such as "where?", more likely to be found in images which excel at spatial representations.

Due to the design of the task both the question and answers are written in text and one might argue that using information of illustrated format is more cognitive demanding as its needs to be converted between the different modalities. It is probably easier to utilize the terms given in the question to search and match the text content. A possible strategy is, instead of forming a mental model and a full understanding of the question and content, to semantically match the question. This may put less demand on working memory than comparing inner mental models. If the questions or answers on the other hand would have been presented in a pictorial format we might have seen a shift in attention towards the illustrated content. If multimodal information processing is mainly guided by text content as suggested by Carroll (Rayner et al, 2001), then one would assume that if enough information is found in the text, there is no need to look at the images. Previous experience with typical school books could also steer away from images as their functional use often are scarce.

Furthermore, visualizations have shown to improve learning more in delayed than immediate tests which may decrease the advantage of images in our immediate multiple choice task. The positive effects of dual-coding may only be relevant when there are greater demands on memory and understanding (Scheiter et al, 2010). There may not be a need to create inner mental models in this particular task as a "true" understanding of the information is not needed to answer the question. There is no need for the students to save the information longer than the time it takes to skip to the next screen and click on the answer.

Dwell time on relevant areas (3)

As seen, the initial dwell time analysis on text and images showed no differences between the groups. On the other hand more importantly regarding the abilities to perform these kinds of typical school tasks may be to what extent the reader can distinguish between more and less relevant information.

However, our results show no significant difference between the groups regarding the proportional amount of dwell time on relevant segments. The only group with a significant higher dwell time on relevant segments were the lowest ability group. This result contrasts the results by Hannus and Hyönä (1999). One possible explanation for this result is the fewer participants used in this study, the smaller range of differences in ability and the older pupils. Another difference may lie in the task. The pupils task in our study is to find the answer to a known question, in contrast to Hannus and Hyöna (1999) where the students needed to read, memorize and hopefully understand the material to be able to answer the subsequent unknown multiple questions. As the students are not able to memorize all content it is crucial to select and concentrate on the information that is evaluated as relevant. This is a skill that probably rises in importance with increasing complicated and fragmented types of information, but is almost certainly less important in our more search oriented task. The pupils do not have to evaluate the relevance of the information as the question itself defines what particular information is relevant putting less demand on the intellectual abilities.

Secondly, as discussed above, higher ability students may be able to find the relevant segments quicker and more easily, discarding irrelevant paragraphs and illustrations – but may also be motivated to double-check everything. Pupils having problem understanding the relevant segments may on the other hand need to look at it more carefully as found in Radach et al (2003).

The results on the time to first fixate on relevant areas however show no indication of higher ability students finding relevant areas quicker, quite contrary to this the lowest scoring groups had the lowest times. Of course, a simple fixation, says nothing about whether the input area actually are comprehended and evaluated; rather it may be the result of a different overall scanning behaviour. Most likely the first fixation time on relevant areas is a result of the total amount of transitions.

Transitions between text and images (4)

The real-time gaze behaviour recorded in eye-tracking is crucial in studying how the different modalities are combined and read as a factor of time. Although no significant differences were found in dwell time of text or illustrations, there may vary in how the pupils move between the different modalities. Is the reading path for example sequential, e.g. reading all the texts first and thereafter starts attending the illustrations, or does the attention jump back and forth.

Surprisingly the results showed a significant difference between group 1, with least transitions between text and images, and group 2 with the highest amount of transitions. Group 3 placed in between. The amount of text/image-transitions are low compared to the total amount of transition between AOIs seen in Figure 11which is expected as the text is divided into AOIs for each paragraph. Also the text/image-transitions correlates with the total amount of transitions. Compared to the score result (remember: no

significant difference) we see a tendency for transitions to increase with decreasing score points. Apparently more transitions are not a good strategy solving this particular task. As the current task to some extent can be categorized as a search task, as the students do not need to read everything or to create a deeper understanding of the material, one might hypothesize that a higher degree of transitions indicates problems in finding the correct answer. This could both be a problem of evaluating the information or more a result of diverted attention as discussed by Harber (1983).

As discussed, a more complex task where different aspects or types of information are needed to form a correct answer or if the questions had been revealed post-reading; would probably result in more integrative transitions and greater cognitive emphasis on the creation of mental models emphasised in multimedia theory. Additionally, if there had been explicit textual or graphic references in the text guiding the reader to particular illustrated content or images had been more relevant, transitions would most likely increase as concluded by Slykhuis (2005) and Holsanova et al (2008).

Verbal results

Although the functional segments over time are hard to quantify, some very similar characteristics were found in how the students read the spread regarding attentional changes from text to illustrations. In most cases illustrations were attended to shortly in the beginning, as to get an overview of the complete spread. Or alternative, in the end just before the participant clicks the answer button. In the last case both the quick glance is common, but also a more detailed studying of the content. During reading of the text content only in rare cases any integrative saccades are made to the illustrated content as seen in the data.

A qualitative analysis did not find any differences between the groups in the types of functional processes and in what order they were used. The analysing process is quite difficult due to the fact that there are large individual differences in how much the pupils talk during the retrospective. This could be the result of both the influence of the experimental setup and personal ability to verbalise or motivate/rationalise ones actions in retrospect. No significant differences were found in the amount of utterances made by the groups or correlation to score result.

Functional categories and transitions/fixation dispersions

No pattern emerged when the AOI-transitions were segmented into the 6 functional categories. Both comparing groups and categories the data looks random (Figure 12). The only result that stands out is the large amount of transitions made by group 2 in the "looking back" category, where one might expect transitions to be high for all groups. Measuring AOI-transitions is however problematic in this case as the AOI segmentation of e.g. paragraphs can accumulate when reading/scanning the text passing over multiple AOI borders – a behaviour very different to making integrative saccades between different modalities. It is also hard to generalise the results as it is to a large extent depending on the selected AOI configuration.

An alternative measure is fixation dispersions, i.e. how large saccades are being made. This way only the eye behaviour is measured and not the current AOI setup (Figure 13). Although this measure is difficult to compare with AOI transitions it gives a complementary view of the reading behaviour. For example, the "reading task" category has high dispersion figures for all groups which is expected as the question is placed far from the learning material content. The behavioiur of looking back to the question results in a high dispersion value, unlike an AOI analysis where only 2 transitions would be detected. However in general the results looks very random.

A critical issue of this analytical method is the difference in the verbal data from the pupils, such as amount and type of utterances. These can influence the functional categorisation to a large degree as they vary in occurrence and time.

Functional categories and fixation durations

As fixation durations has been found to correlate with cognitive load, it is interesting to see if and what the functional segmentation of the eye-tracking data can reveal. Generally the fixation duration times (text: 338–466 ms and images: 310–385 ms) are high if you compare with the common reading span of 180–300 ms (Figure 14). This is expected as the task to a large degree is search oriented making the pupils jump from place to place instead of a fluent reading behaviour. All groups combined, no pattern is seen in the fixation duration times of the 6 different categories. An intuitive expectation would be to find higher times in categories such as "evaluation info" or "remembering" as they intuitive seem more cognitive demanding.

However, separating the ability groups, an interesting step pattern of the bars emerges (Figure 15). Fixation durations seem to increase with lower ability and interestingly the differences between the groups seem to increase in the expected more cognitively demanding categories. Perhaps differences only are to be found in the more demanding segments where less able students start to struggle, whereas the high ability group still find the task easy. This suggests for future studies to carefully select the level of task to amplify any differences.

Limitations of functional categorisation and eye-tracking measures

The development of the analytical method of combining eye-tracking data with verbal retrospective protocols have

unveiled some limitations. One cannot disregard from the inaccuracy of the temporal segmentation of the eye-tracking data. Inevitable, a lag in the verbalisation is unavoidable resulting in an AOI latency. One might argue for a temporal adjustment to counteract any latency; however any precise adjustment is probably difficult to achieve as any latencies to some extent will correlate with individual differences and to what type of function the participant is describing. Some processes on the screen are harder to identify then others, needing a longer visual input and a longer time to describe.

Additionally, it varied to what extent the pupils succeeded to verbalise their process, resulting in long pauses and a lack of data in certain categories. This is as mentioned a general problem in verbal protocols and may be to a higher extent taken into consideration when designing the experiment. Further, to improve the general accuracy of the time borders of the utterances, future coding of the verbal data should define all silent moments as "pauses" to a higher extent than the case in this study.

General discussion and future improvements

The generative theory of multimedia learning describes three processes for an improved learning: 1. Select and process verbal content. 2. Select and process visual content. 3. Combine the verbal and visual content into a coherent mental model. We have not seen indications for this positive combinatory process in the eye-tracking data. This could be attributed to our more search-oriented task, in contrast to a more comprehensive test. It seems that more integrative saccades between modalities more is a factor of an increased visual search.

A general point in the discussions of the result is the lack of any significant differences between the 3 groups. This is probably mainly due to the low number of pupils (n=12) and small differences of ability between the groups. Additionally, a more precise level of task difficulty, subject and reading material could enhance existing differences. One might argue against the use of 3 ability groups instead of just using 2 as previous studies. This is a valid argument especially with few participants, however using more than 2 groups may reveal new aspects. Reading and task solving behaviour may not exclusively change linearly with ability; perhaps in some cases the extremes may show more similarities than the mean pupil.

Although the usage of real and current educational material provides good validity, it makes it difficult to isolate factors influenced by the design. For example the low amounts of transitions between text and images are probably partly a result of a design not promoting it. It also complicates the process of comparing measures e.g. fixation duration times as the body, captions and headlines differ in type size and font. This can effect functional categories such as "looking at question" and "reading text" as the question is printed in larger size possibly influencing fixation duration and saccade lengths. In future studies, adjusting influencing design affordances, could make it easier to isolate certain mechanisms. It also provides the opportunity to design the stimuli according to the limitations of the eye-tracking equipment, making it possible to divide the content into more detailed AOIs of interest.

Another issue is to what degree the actual digitisation of the analog book influence the reading behaviour. Perhaps sitting in front of a computer, reading on a screen, using the mouse to interact – influence the pupils to read more like when browsing the web. Different results may be found if using real analog books instead.

The developed analytical method of combining eye-tracking and verbal data did provide us with the opportunity to divide the eye-tracking data into functional units. Although only the fixation duration time measures indicated some interesting results such as the cognitive strain in the functional units, we se positive in more evaluation and development of the method. Preferably with more participants and stimuli selected or designed on the basis of what measures to use.

A future development and use of the method is not only relevant for the field of educational design as in this thesis, but in education and design in general, usability, information architecture and visual communication.

Acknowledgements

The author wishes to thank the supervisor Jana Holsanova for all the help, co-supervisor Nils Holmberg also involved in the project and the participants of the eye-tracking supervision group for helpful feedback and motivation.

References

Conyer, M. (1995). User and usability testing – How it should be undertaken? *Australian Journal of Educational Technology*, 38-51.

Ehmke, C., & Wilson, S. (2007). Identifying Web Usability Problems from Eye-tracking Data. *People and Computers XXI – HCI... but not as we know it: Proceedings of HCI 2007*. British Computer Society.

Fletcher, J. D., & Tobias, S. (2005). The Multimedia Principle. i R. E. Mayer, *The Cambridge Handbook of Multimedia Learning* (pp. 117-134). USA: Cambridge University Press.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2009). *Cognitive neuroscience - the biology of the mind*. USA: W. W. Norton & Co.

Greeno, J. G. (1994). Gibson's Affordances. *Psychological Review, Vol. 101*, No. 2 , 336-342.

Griffin, Zenzi M. (2004). Why look? Reasons for eye movements related to language production. in Henderson & Ferreira, *The integration of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

Hannus, M., & Hyönä, J. (1999). Utilization of Illustrations during Learning and Science Textbook Passages among Low- and High-Ability Children. *Contemporary Educational Psychology 24* , 95-123.

Hansen, J. P. 1991. The use of eye mark recording to support verbal retrospection in software testing. *Acta Psychologica 76*, Elsevier, North-Holland pp. 31-49

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, Vol. 17,* No. 11 , 498-504.

Holsanova, J. (2008). Discourse, Vision, and Cognition. Amsterdam/Philadelphia: Human Cognitive Processes 23. John Benjamins Publishing Company.

Holsanova, J., & Nord, A. (2010). Multimodal design: Media structures, media principles and users' meaning making in newspapers and net papers. in H.-J. G. Bucher, & K. Lehnen, *Neue Medien – neue Formate (pp. 81-103)*. Frankfurt/New York: Campus Verlag.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2008). Reading Information Graphics: The Role of Spatial Contiguity and Dual Attentional Guidance. *Applied Cognitive Psychology. 22.*

Holsanova, J., Rahm, H., & Holmqvist, K. (2006). Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurments. *Visual communication*, 65-93.

Knoblich, G., Öllinger, M., & Spivey, M. J. (2005). Tracking the eye to obtain insight into insight problem solving. in G. Underwood, *Cognitive Processes in Eye Guidance* (pp. 355-375). Great Britain: Oxford University Press.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics 33*, 159–174

Mayer, R. E. (2005). Introduction to Multimedia Learning. in R. E. Mayer, *The Cambridge Handbook of Multimedia Learning* (pp. 1-19). USA: Cambridge University Press.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Nielsen, J. (1993). *Usability Engineering.* Academic Press Limited.

Norman, D. (2002). *The design of everyday things.* USA: Basic Books

Ozcelik, E., Karakus, T. K., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education 53*, 445-453.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford: Oxford University Press.

Pernice, K., & Nielsen, J. (2009). *Eyetracking Methodology – How to conduct and Evaluate Usability Studies Using Eyetracking.* Fremont, USA: Nielsen Norman Group.

Pozzer, L. L., & Roth, W. M. (2003). Prevalence, function, and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching 40*, 1089-1114.

Radach, R., Lemmer, S., Vorstiues, C., Heller, D., & Radach, K. (2003). Eye Movements in the processing of print advertisements. i J. Hyönä, & R. D. Radach, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 609-623). Oxford: Elsevier Science.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, Vol. 124, No. 3,* 372-422.

Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye Movement When Looking at Print Advertisements: The Goal of the Viewer Matters. *Applied Cognitive Psychology, 22(5)*, 697-707.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating Text and Pictorial Information: Eye Movements When Looking at Print Advertisements. *Journal of Experimental Psychology: Applied, 7 (3)*, 219-226.

Scheiter, K., Wiebe, E., & Holsanova, J. (2010). Theoretical and Instructional Aspects of Learning with Visualizations. in R. Z. Zheng, *Cognitive Effects of Multimedia Learning* (pp. 67-88). New York: Hershey.

Slykhuis, D. A., Wiebe, E. N., & Annetta, L. A. (2005). Eye-tracking Students' Attention to PowerPoint Photographs in a Science Education Setting. *Journal of Science Education and Technology, Vol. 14, Nos. 5/6*, 509-520.

Sweller, J., van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Underwood, G. (2005). Eye fixations on pictures of natural scenes: Getting the gist and identifying the components. i G. Underwood, *Cognitive Processes in Eye Guidance* (pp. 163-187). Great Britain : Oxford University Press.

van Gog, T., Paas, F., van Merriënboer, J. J., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied, Vol 11(4).*

Yarbus, I. A. (1967). *Eye movements and vision*. New York: Plenum Press.

LUND UNIVERSITY