MASTER THESIS

THE USER EFFECTS OF USING TEXTUAL CUES TO INCREASE IMAGE VIEWING ATTENTION

D.S. Kornalijnslijper

FACULTY ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE
HUMAN MEDIA INTERACTION

EXAMINATION COMMITTEE
M. Theune
E.M.A.G. van Dijk
J. Karreman

UNIVERSITY OF TWENTE.

16-08-2012

Abstract

Textual cues are commonly used in illustrated text documents to link content written in words with relevant content depicted in illustrations. Cues could potentially benefit learning from illustrated text documents by encouraging the simultaneous presence of textual and pictorial information in readers' working memory and thus, help readers gain an integrated understanding of the presented concepts. However, little is known of the effect of textual cues on learning and reading behaviour. Here I describe two experiments that were conducted to examine the effect of textual cues embedded in illustrated text documents on recall and reading behaviour. Two types of textual cues were used to link the text to the image: simple and explicit cues. Simple cues are a short reference such as "see the map" or "see the illustration" while explicit cues provide more aid to the reader by describing the referenced concepts and objects, such as "Notice in the picture how the Tuareg territory is spread over the new nations". Cues were expected to increase the recall of information linked to the cues. In the recall performance experiment, 60 participants were presented with 12 presentation and were tested on recall. In the reading observations experiment, 5 participants were given the same presentations and their gaze was tracked with an eye tracker to gain a better understanding of their reading behaviour. Results from the recall performance experiment showed no significant differences in recall between simple cues-, explicit cues- and the control (no cues) condition. Further analysis suggested that English skills may have influenced how well the information had been recalled. However, overall results were not conclusive enough to confirm any relation between the presence of cues and recall. Results from the reading behaviour experiment showed that participants switched attention to the illustration more often without the aid of cues than with the aid of cues. Furthermore, participants employed different strategies for switching attention, suggesting that reading styles may vary widely between readers. Several factors might have contributed to the results in both experiments, such as: lack of complexity in the presentations, the participants' reading behaviour, the level of motivation and reading skills, and limitations of the experiment design. The implications of these results are discussed and a number of recommendations and suggestions are provided to guide further text design research.

Acknowledgement

Writing this thesis has been a personal achievement. It has been quite an intense and interesting experience, which would not have been possible without the support and advice of several important people who contributed to the completion of this thesis.

The help, advice and knowledge of my supervisor, Dr. Mariet Theune, was the corner stone of my progress. I thank her for sharing her ideas, correcting me on my mistakes and for her great patience during this project. Special thanks to the members of the committee, Betsy van Dijk and Joyce Karreman, for providing their input, advice on statistics and help around the Faculty of Behavioural Sciences. Special thanks to Rose for her sublime correction of my English.

I thank my girlfriend Adriana Stan for giving me support, advice, for reading the work, and above all, for giving me the motivation to see this project through and finish the thesis.

Last but not least, I thank my parents and sisters for their support and belief in my abilities during my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cues are a common and broadly used method of guiding readers' attention. Cues are found in books, papers, magazines, internet web pages and interactive study software. They are used in multiple representations of content such as written- or spoken text, illustrations and animations. Cues, also called signalling devices, aid the reader in selecting, organising and integrating the displayed information (Mautone & Mayer, 2001; Mayer, 2005a; Koning de, Tabbers, Rikers, & Paas, 2009). They come in various shapes and forms: written or spoken in words, sounds, arrows, connecting lines, colouring, flashing, boldfacing, underlining, italics, headings or outlines. Each cue type has its own way of standing out from the rest of the material, capturing the attention and then redirecting the reader to other content. In the redirected content the reader is either offered extra information or shown a different representation which further elaborates on the topic.

1.1 Textual Cues and Multimedia Presentation Systems

In this thesis I will discuss the effects of textual cues written in words. They are a common type of cue which are often found in illustrated text. Specifically, I will look at textual cues which are integrated in a short body of text and which refer and draw attention to illustrations. Text is quite flexible and different variations of the cues could be made as suggested by Peeck (1993). Cues could be simple e.g. "See the image" or explicit like written in the text of figure 1.2: "See (step 3-5) in the image how proteins are synthesized and move back in the nucleus or to the Gogli apparatus.". Seufert and Brunken (2004) categorise the simple cues as less-directive and explicit cues as directive. Less-directive cues let readers identify the essential information by themselves. E.g. a simple cue directs the reader only to the illustration without going into further details on the contents inside the illustration. Directive cues are explicit and direct the reader to the essential concept that needs to be studied. E.g. an explicit cue can elaborate on the textual content by directly referring to the illustrated content. In this study I will compare both simple and explicit cues and use them in illustrated-text presentations such as the example shown in figure 1.2.

I have chosen this type of cue in light of the studies done on a question answering (QA) system created within the Dutch research programme IMIX (Interactive Multimodal Information eXtraction) (Hooijdonk, Bosma, Krahmer, Maes, & Theune, 2011). The QA system automatically generates illustrated-text answers to medical questions asked by non-expert users. It uses text excerpts from medical encyclopaedia and includes a picture from a different source. Cues are not present in the excerpts and would need to be inserted. When such systems are aware of the content in both text and illustration then meaningful textual cues could be generated. Systems with natural language generators such as intelligent multimedia presentation systems (IMMP) (André, Müller, & Rist, 1996; André, 2000) could integrate cues during the generation process. High level document authoring systems such as ILLUSTRATE, the system for cooking recipes by Deemter and Power (2002) and M-PIRO

(Androutsopoulos, Oberlander, & Karkaletsis, 2007) could suggest to authors what to integrate as cues while they are creating a document or presentation. The automatic generation of a simple cue will be less complicated than generating an explicit cue.

Before I continue to discuss the use of textual cues I will first give a brief introduction on the processes that underlie the learning from words and illustrations.

1.2 Learning from words and illustrations

What words cannot describe is sometimes better shown with an illustration. Each representation has its own qualities for communicating information. Words are abstract and generic and images are concrete and specific (Fletcher & Tobias, 2005; Schnotz, 2005). Words can describe the dimensions of an object, its use and construction in general terms, but they cannot show the actual object. E.g. using the generic term chair informs the reader that this is an object that functions as a seat, but the reader would have no idea what the chair in question looks like. Using words to describe the chair's dimensions, construction and design and purpose could be a lengthy process and the reader would still use their own references and expectations to construct a mental picture of the item. This visualisation, however detailed the text, would still be an approximation of the chair itself. Adding a picture would help the reader see what the chair looks like. However, the illustration is specific and shows only one or a few examples. See for example the description of a chair in figure 1.1. Presenting both the text and the illustration helps the reader to gain a better and integrated understanding (Fletcher & Tobias, 2005).

Figure 1.1: The example of a chair. Source: Wikipedia

A chair is a raised surface used to sit on, commonly for use by one person. Chairs are most often supported by four legs and have a back; however, a chair can have three legs or could have a different shape. The design may be made of porous materials, or be drilled with holes for decoration; a low back or gaps can provide ventilation. The back may extend above the height of the occupant's head, which can optionally contain a headrest. Chairs can also be made from more creative materials, such as recycled materials like cutlery and wooden play bricks, pencils, plumbing tubes, rope and PVC pipe.

Most illustrated text will describe more complicated topics than a simple chair. Consider the presentation shown in figure 1.2. In order to reach a complete understanding the reader must process the contents of the presentation and integrate it with the prior knowledge stored in long term memory. Mayer's Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2005a) and Schnotz's Integrated Model of Text and Picture Comprehension (ITPC) (Schnotz, 2005) explain how the human mind processes and learns from multimedia. The models have different explanations on how information from different sensory modalities (sounds, images) enter the mind and are processed. However, overall both models are very similar and have key concepts in common. After information has entered the mind through the eyes and ears it enters the working memory. The working memory has a limited capacity and can hold information for a limited amount of time. These limitations force us to make decisions on what to pay attention to, how to build connections between information and how that information maps on our prior knowledge. Thus, we are actively engaged in selecting, organizing and integrating knowledge to create a coherent model that is a synthesis of the new information and what we already know. Seufert and Brunken (2004) explain the process of coherence formation in a number of steps. Firstly, learners need to understand each representation. Next, they have to analyse the essential concepts in the text and the illustration and then identify the relations between the concepts. Finally, all concepts have to be mapped together to form an integrated mental representation.

These processes require conscious control and cost resources in the limited working memory. The demand placed on working memory capacity and processing is called cognitive load (Schnotz & Kürschner, 2007). Schnotz and Kürschner explain the working of three types of cognitive load: intrinsic, extraneous and germane load. Intrinsic load is caused by the processing of essential material and relations necessary to achieve a learning objective such as understanding how a system works. Extraneous load is caused by processing of non-essential or extraneous material that is not relevant to the learning objective and is an unnecessary load due to bad design and organisation of the material. Extraneous load interferes with learning because it places load on the limited cognitive capacity and should be reduced as much as possible. Germane load is caused by an intensive learning effort that involves deeper processing to create a coherent mental organization of the material (Mayer, Hegarty, Mayer, & Campbell, 2005; Schnotz & Kürschner, 2007). The total cognitive load is the sum of these 3 types of load. When the intrinsic load is high and there is a lot of extraneous material then working memory capacity may be overloaded. Learning might then suffer because there is no capacity left for germane load.

When the text and illustration describe a complicated topic or describe various detailed aspects then selecting and organizing the material becomes more demanding. Seufert and Brunken (2004) note that readers with insufficient prior knowledge are often unable to handle the task. They will study the representation that is the most familiar to them and will only switch if they have problems understanding. Seufert and Brunken (2004) further note that prior knowledge helps us make sense of the presented material; it helps us in selecting essential material and offers structure to organize before integration in long term memory. Without this aid the number of elements in the material that require processing can overload the working memory. In contrast, readers with high prior knowledge need less mental effort and so have a lower cognitive load.

1.3 Benefits of Cues on cognitive processes

According to theories CTML and ITPC, textual and illustrative content can only contribute to a coherent mental model if both are simultaneously in working memory and before the information decays. Therefore, it is best to study related content from both representations in quick succession. Cues can aid bringing related content simultaneously in working memory by referencing and thus, 'connecting' related content spread in different representations. Cues could increase the chance that readers select related and essential content mixed in between other content. Hence, cues could reduce the cognitive load placed by selection and could help to mentally organize the material. This is similar to the effect of the contiguity principle (Mayer & Ross, 2002; Mayer, 2005b). The contiguity principle states that text and illustration should be presented together in space and time rather than after each other. The spatial contiguity principle says that text and illustration should be presented spatially close together rather than far apart. The temporal contiguity principle says that text and illustration should be presented with the least possible time between them and preferably simultaneously.

In addition, switching between the two representations without guidance requires visual searching which increases chances of readers viewing extraneous material (Koning de et al., 2009) and will cost extra time, all the while information decays from working memory. This is similar to what the spatial contiguity principle describes. When both representations are far apart then readers need to use cognitive resources for visual searching. Thus, the materials are less likely to appear in working memory at the same time (Mayer & Ross, 2002). Cues could reduce visual searching because they link the essential content.

Lastly, the study material might contain extraneous material which is not needed for the learning objective or the material might be organised in a complex way. As discussed above, this will increase extraneous load because the reader has to examine all material for relevancy, then identify and relate the essential material. Without placing cues, essential learning material may fade from working memory while the reader copes with the extraneous load. Cues do not prevent the reading of extraneous material, however, by placing the essential material in quick succession in working memory they can lower the interference from extraneous load. Mayer (2005b) calls this the signalling principle.

1.4 Mapping concepts and structures

In order to acquire knowledge from both text and illustration, the reader must create referential connections between the representations and must relate the build structure with the presented topic (Seufert & Brunken, 2004). They call this coherence formation and they distinguish between two types of strategies for mapping the contents in the text and illustration. The first strategy is to rely on surface features of the representation that indicate correspondence, e.g. the use of emphasis with boldfacing, italics or colouring. These features can serve as signals for mapping and may lead the reader to process the cued elements and structures more deeply. However, Seufert and Brunken note the possibility that these surface features are only superficially observed. The guidance of attention can lead to a narrow focus on the surface features. Readers may not actively think and therefore fail to engage deeper into the subject. The second strategy employs a deeper analysis of the elements in both representations by matching and mapping their relations through analysis, e.g. by understanding how a process or a system, described in a text and shown by an image, operates and how individual elements interact. According to Seufert and Brunken, this help is mostly given verbally, e.g. as instructions. Peeck (1993) discusses this type of verbal aid in light of increasing students' attention to illustrations. For example, a teacher could instruct the students to not forget to look at the pictures. However, Peeck explains that simply telling a student to study will not likely lead to deeper learning. His suggestion is to use more specific instructions to increase the effects on learning from illustrations. Instructions could also be elaborated by telling a student what to observe in general or what to observe in particular. Assignments, such as telling a student to do something with the illustration (comparing, tracing, labelling), are even more effective, according to Peeck.

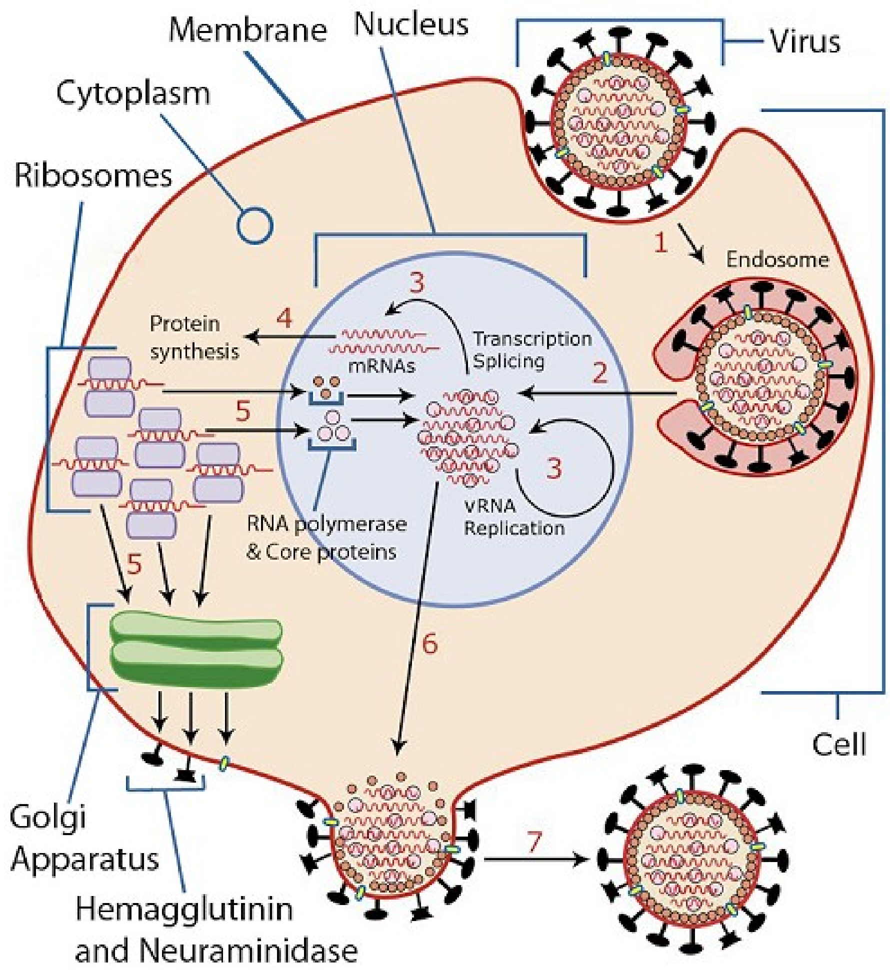
In this study I have used two types of textual cues. Both types of cues do not draw attention by means of surface features. They are embedded in the text without distinct visual features to separate them from other words. Simple cues are similar to what Peeck (1993) describes as simply telling a student to study the illustrations. The simple cues only indicate that certain information which is present in the illustration. The explicit cues are similar to the "more specific instructions"; they have the same goal as the simple cues however, they explicitly mention the related items and concepts. Hence, explicit cues aid the reader more than simple cues at selecting essential content in both representations. Explicit cues and to a lesser degree simple cues can be categorized as encouraging and facilitating the second structure mapping strategy of deeper analysis.

Figure 1.2: A presentation about the influenza virus.

The influenza virus can only replicate in living cells and contains RNA genome, RNA polymerase, and viral proteins. Infection and replication of Influenza is a multi-step process: firstly the virus has to bind to and enter the cell, produce new copies of viral proteins and RNA, assemble these components into new viral particles and finally exit the host cell. Influenza viruses bind through hemagglutinin onto sialic acid sugars on the surfaces of cells and the cell imports the virus into a structure called an endosome.

Once inside the cell, the acidic conditions in the endosome cause two events to happen: the hemagglutinin protein fuses the viral envelope with the endosome's membrane, then ion channels allow protons to move through the viral envelope and acidify the core of the virus, which causes the virus to dissemble and release the viral RNA (vRNA) and core proteins. The vRNA molecules, core proteins and RNA polymerase are then released into the cytoplasm; see the image (step 1 and 2) how the virus enters the cell and dissembles. These core proteins and vRNA form a complex that is transported into the cell nucleus, where the RNA polymerase begins transcribing vRNA. The vRNA is either exported into the cytoplasm where viral proteins will be synthesized by ribosomes, or the vRNA remains in the nucleus.

Newly synthesized viral proteins are either secreted through the Golgi apparatus onto the cell surface (e.g. hemagglutinin and neuraminidase molecules) or are transported back into the nucleus to bind vRNA and form new viral genome particles. See (step 3-5) in the image how proteins are synthesized and move back in the nucleus or to the Gogli apparatus. VRNAs, RNA polymerase, and other viral proteins are assembled into a viral particle. Hemagglutinin and neuraminidase molecules cluster into a bulge in the cell membrane. The vRNA and viral core proteins leave the nucleus and enter this membrane protrusion. Please look in the image (step 6) where the particles move from the nucleus to the membrane protrusion. The mature virus buds off in a sphere of host cell membrane, acquiring the hemagglutinin and neuraminidase molecules within the membrane.



1.5 Related cue research

A large body of research has been done on textual cues within text, called signal devices. However, these signals refer to elements within one representation and not to others, such as illustrations. Examples of text signalling devices that guide attention are: titles and headings, which label dominant themes, and boldfacing, italics, and colouring that draw attention on words. Other text signals provide organisation, such as: headings, enumeration, overviews and previews, pointer phrases and logical connection phrases. These signals have been reported to improve recall of relevant information and improve transfer for problem solving tests (Mautone & Mayer, 2001). However, these findings are not directly applicable to relational cues between representations. Switching within text can require a jump in concept; however, the information is still shown in the same abstract format and with similar wording. On the other hand, the switch from textual information to pictorial information requires integrating abstract knowledge with concrete knowledge and is not processed the same way in working memory (Mayer & Ross, 2002; Mayer, 2005b). Furthermore, I assume text signalling devices are forming a unity with the text whereas textual cues that refer to illustrations interrupt the reader and draw attention to a different representation.

Surprisingly, research on textual cues between multiple representations has been limited. Reinking, Hayes, and McEneaney (1988) reported at the time of writing that "the effects of cueing attention to graphic aids (illustrations) by including written cues embedded in text is yet unknown". Reinking et al. started an experiment to see if cues were of benefit for poor readers that might not know when to switch to illustrations. They researched: 1. the effect of text with pictures preceded by a general instruction, 2. text with pictures with explicit cues embedded and 3. the combination of a general instruction and embedded explicit cues. The general instruction stood above the main body of text and told readers to "Look at the picture to help yourself better understand and remember the information in this passage". The explicit cues were placed at the end of a paragraph and were similar to the explicit cues mentioned above: "Note in Figure 1A the layers of cinders, ash, and lava that make up this volcano." In addition, the explicit cues were set in bold and boxed with a black border whereas, the general instructions were not. Participants were 7th and 8th grade students at high school and divided in two groups: poor and good readers. Text comprehension was measured immediately after reading a topic, by means of multiple choice questions of which half could be explicitly found in the text and half implicitly. Furthermore, recall of the information presented in the graphics was tested in a post test exercise. Results from the text comprehension test indicated that in general cues and instructions improved comprehension significantly for poor readers, but not for good readers. However, explicit cues were more effective than the general instruction. Explicit cues improved graphic recall for both good and poor readers but benefited poor readers more. The comprehension gap between poor and good readers was bigger in the no cues condition. Given the increase of comprehension, Reinking et al. concluded that explicit cues increase attention to illustrations. However, it is not clear from this experiment whether the increased attention would have been the same if the explicit cues had not been bold faced and boxed. Boldfacing and boxing are both visually acute methods of drawing attention similar to colouring and flashing. Thus, it is not clear if the cues would have been as effective without the visually enhancing features.

Mayer, Steinhoff, Bower, and Mars (1995) conducted three experiments in which they tested whether distance between text and illustration was of influence (the spatial contiguity principle) and the influence of adding annotations to the illustration. Annotations were placed underneath the illustration, and repeated the same cause and effect chain information from the main body of text. Hence, they did not reference the illustration but rather they repeated information from the text. In addition to annotations, labels were added to the illustration. The participants were college students and they were divided into high and low prior knowledge groups. The results showed that the participants in the annotated group performed better in problem solving. The distance between the text and illustration was found to have no effect. In their conclusion, Mayer, Steinhoff, et al. claimed that annotated illustrations can function as a signal for readers to select relevant words and images. Annotated illustrations serve as structural summarisers and organise the material, and they function as elaborative cues that help readers to connect elements from the two representations. The annotations are relevant to textual cue research but are different from the textual cues used in my experiment; the annotation are not embedded in the main text body and do not reference the illustrations.

McTigue (2009) conducted an experiment to assess the combined use of textual cues embedded in text and of labels in illustrations. Labels were placed in a white box which made them stand out from the illustration. The textual cues were explicit in their instruction and directed the reader to look at the illustration: "Look again at the diagram, at number 6, to see where steam is turned to water". The number used in the cue text corresponded with a numbered location in the illustration. The participants were 6th grade students with an average age of 12 years. Text comprehension was tested with two types of questions: fact recall questions and inferential questions which required application beyond the text. Results showed that cues did not provide substantial benefits to students and showed only minor results. McTigue (2009) discusses a number of possible causes for the results: i) The 6th grade students might have lacked the skills to comprehend the information in the text and illustration, skills such as those needed to interpret abstract concepts and interpret diagrams. ii) Results might have been limited by the type of questions asked in the comprehension test and the participants' motivation. Another possibility, not mentioned in this paper, is that offering both labels and cues at the same time might have been more aid than necessary, because the visual properties of the labels might have drawn enough attention and made additional cueing from the text redundant.

Other related work includes the use of surface features to aid coherence formation (Seufert & Brunken, 2004) such as flashing and colouring related elements, and using arrows. Although these types of cues attempt to achieve similar goals as textual cues, relating elements in different representations, they draw attention and link content in a different way. Hence, they will only be discussed briefly. Boucheix and Guignard (2005) tested the effect of colouring in the illustration, arrow pointing in the illustration, the use of speed indicating icons, and textual cues simultaneously in one condition. Participants were better at explaining the contents of the presented material but no effect was observed for recall and comprehension. Jeung, Chandler, and Sweller (1997) investigated the effect of flashing related elements and found that flashing was only beneficial if the visual searching was high. Kalyuga, Chandler, and Sweller (1999) found beneficial effects from colouring related elements. Folker, Ritter, and Sichelschmidt (2005) tested colouring and found no improvement in retention. However, the participants in the cued condition were faster in processing and switched more often between text and illustration. Reduced eye fixation times, measured with an eye tracking device, suggested better processing. Jamet, Gavota, and Quaireau (2008) tested colouring and saw a floor effect probably due to low prior knowledge. An improved recall was observed due to cues and the subjective perception of processing difficulty was improved. In a self-paced experiment, Tabbers, Martens, and Merriënboer (2010) tested the colouring of diagram elements in web based materials and they found only a weak improvement of retention. These results were lower than expected when compared to prior research. They concluded that self pacing might have lowered the extraneous load for all conditions and hence cues had less effect. Koning de et al. (2009) reviewed prior cue research with the aim of using cues in instructional animations. They concluded that prior research on cues that link multiple representations showed an improvement of retention performance. However, results on transfer performance were inconsistent and for cues to be effective, instructional material seemed to need a considerable complexity.

A large body of prior research has been conducted on cues in general. However, the amount of research on textual cues has been low and results have been mixed. Commonly, textual cues are part of the main text body and are not graphically enhanced to draw more attention. They rely on the reader to find the cue while reading the text. Reinking et al. (1988) used graphically enhanced cues. The work of Mayer, Steinhoff, et al. (1995) was focused on annotations but annotations are not quite the same as cues. The work of McTigue (2009) used textual cues as they are commonly used. However, the results were not conclusive and might have been influenced by the young participants' reading skill. In this paper I will assess textual cues which are embedded in a body of text. The cues will not be graphically enhanced and no other textual or visual aids will be added.

The goal is to research the effects of referencing through words without the use of visually prominent features. I will assess two types of textual cues: simple and explicit cues. As discussed above, the simple cues reference the illustration as a whole and do not reference specific concepts. The explicit cues reference concepts and items found in the illustration, which are relevant to the read text.

1.6 Objective and Research questions

In this study I will assess the effect of textual cues which are embedded in a main text body, on the recall and subjective experience of presented information. In addition, I will record the eye movement of a small number of participants to assess if participants follow the directions of the cues. The presented information consists of brief illustrated-text presentations and is presented on an internet website. The cues will come in two variations; simple and explicit. Simple cues will only refer to the illustration as a whole without mentioning specific concepts or items. The reader must then search for the essential material within the illustration. Explicit cues will guide the reader by referring to essential concepts or items in the illustration. See section 1.1 for examples. The questions I seek to answer are:

- Do simple textual cues to illustrations improve the recall of the information presented in illustrated text presentations?
- Do simple textual cues affect the subjective perception of the presentation difficulty, language difficulty, question difficulty and the participants' level of English?
- Do explicit textual cues to illustrations improve the recall of the information presented in illustrated text presentations?
- Do explicit textual cues affect the subjective perception of the presentation difficulty, language difficulty, question difficulty and the participants' level of English?
- Is there a noticeable difference in the recall of information between simple and explicit cues?
- Do simple textual cues cause participants to study a picture in illustrated text presentations?
- Do explicit textual cues cause participants to study a picture in illustrated text presentations?

1.7 Thesis outline

In chapter 2 I will discuss the recall experiment, by describing the methods used, the processing of the data, the results from the experiment and discussing the results. In chapter 3 I will discuss the reading observation experiment. Finally, in chapter 4 the thesis ends with a conclusion and recommendations for future research.

Chapter 2

Recall performance experiment

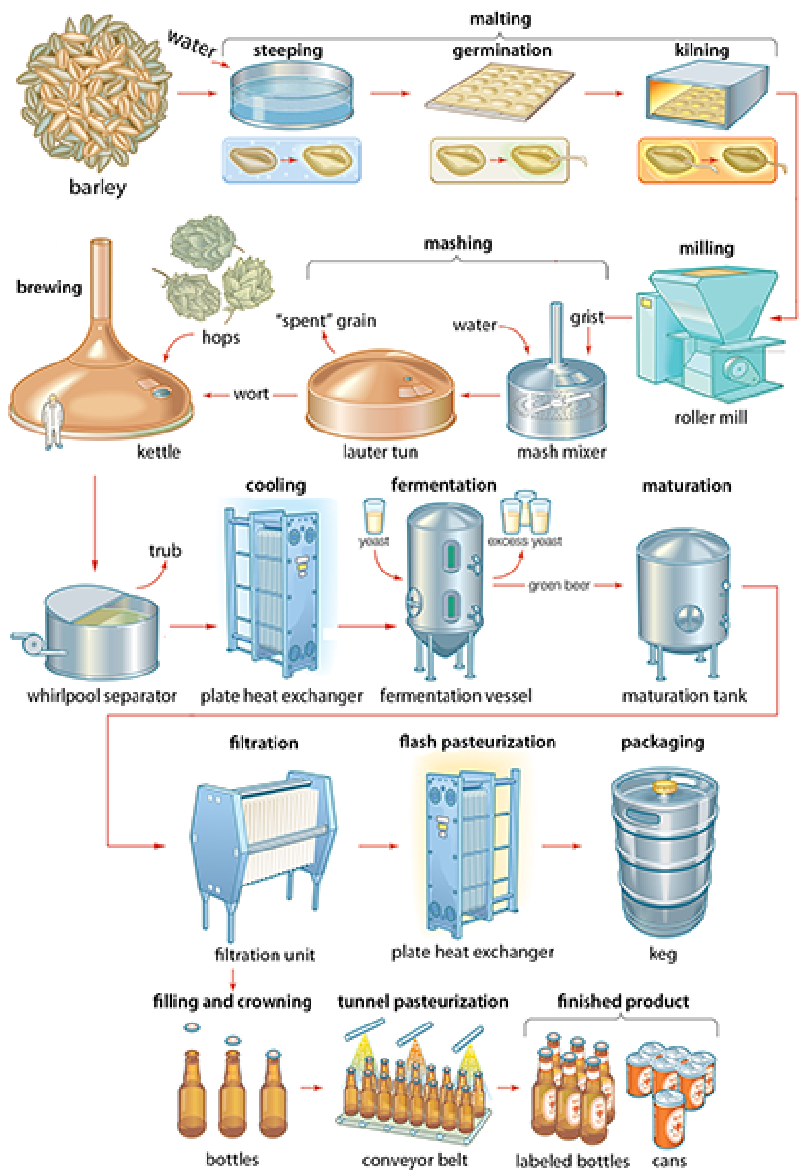
In this chapter I describe an experiment which assessed how well participants recalled information from 12 illustrated text presentations they had to study. Recall was tested using multiple choice questions which followed after seeing a presentation. In addition, participants' subjective rating was measured with post-test questions. The variable between participants was the use of textual cues in the text (no cues, simple cues and explicit cues). The cues were embedded in the main text body of the presentations. The no cues condition served as the control condition. The contents of the presentations described diverse topics ranging from astronomy, biology, gastronomy, geology, sociology to technology. A presentation consisted of a text with an average length of 350 words and was accompanied by an illustration. The used illustrations were chosen to help participants comprehend the text by depicting the written concepts and relations, e.g. showing in steps how the influenza virus infects and reproduces in a human cell, the physical location of Tuareg conflicts in neighbouring African countries or the steps of the beer brewing process. Participants were allowed to read presentations and answer questions at their own pace. The complete experiment was written in English to allow participation of Dutch and international students.

Figure 2.1: A presentation about the beer brewing process with simple cues inserted.

The process of making beer is known as brewing. The aim is to convert grain starches to sugar, extract the sugar with water, and then ferment it with yeast to produce the alcoholic beverage. The brewing starts with malting. Malting is a 3 step process where the barley is made ready by releasing the starches, see the image. Malting begins with steeping (immersing barley in water) for 45 hours. As the grain imbibes water its volume increases and a white root sheath (a chit) breaks through the husk. The chitted barley is then removed from the steep. In the 2nd step, germination, the grain is spread out in a room for 5 days. Germination causes the secretion of enzymes that convert the starch into sugars and the proteins into amino acids. The grain is then called green malt. In the third step, kilning, green malt is dried which stops enzyme activity but leaves 50% in an active state; see the image.

For efficient extraction of water the malt must be milled. The aim is to retain the husk relatively intact while breaking up the brittle starch into particles. The milled malt (grist) is mashed thus converting the starches into sugars that can be fermented. The malt is mixed together with water in a vessel to create a cereal mash. The leftover sugar rich water is then strained through the bottom of the cereal mash in a process known as lautering. At this point the liquid is called wort. After separation, the wort is transferred to a kettle for boiling, to stop enzyme activity, and to obtain the bitter taste of the added hops. (See the image.) The kettle boil sterilizes the wort, evaporates undesirable aromas, and precipitates insoluble proteins (known as trub).

At the end of the boil the wort settles in a vessel called a "whirlpool" by separating out the solid particles. The wort is then cooled in a heat exchanger. Then the cooled wort is moved into a fermentation tank where yeast is added. During the fermenting process the sugars turn into alcohol, carbon dioxide, and other components. During maturation, residual sugars ferment generating carbon dioxide which is vented and purges undesirable compounds. Filtration gives beer a natural polish and colour and stabilizes the flavour. Most beers that are packed in bottles, metal cans or metal kegs are pasteurized by heating.



2.1 Method

2.1.1 Participants

Participants were gathered from a pool of students from the faculty of behavioural sciences at the University of Twente. Because the presentations described complex information and were written in English, participants were requested to join only if they had a good level of English, one that could be expected from university students. In addition, participants should not be colour blind to avoid complications while viewing the illustrations.

2.1.2 Materials

The experiment consisted of 12 illustrated-text presentations and 1 example that was shown. All materials can be found in appendix A.

Layout

To minimize the number of variables I kept the layout simple and the same for all presentations. Each presentation contained of a single illustration and a piece of text written in English language. The text body was placed on the left half of the presentation and the image on the right. The text was aligned justified, thus flushed to its left and right border. The text needed to fit on a computer screen without scrolling to avoid unnecessary distractions while reading. The limiting factors were the vertical resolution of the used computer screen and the used font size. I chose a vertical size based on my personal experience with students' laptop screen sizes. Modern laptops are common among students and most have at least a vertical resolution of 800 pixels. On average, the text was 350 words long.

Type of illustrations

The chosen illustrations clarified or added information to the written text and were not merely decorational. Prior research has shown that informative and constructive illustrations that add useful information or help to clarify concepts are more effective for learning purposes (Mayer, Steinhoff, et al., 1995; Carney & Levin, 2002; Hooijdonk, Vos, et al., 2007). If the illustration has no added value other than being decorational, then a cue would only distract the reader from reading essential content.

Modifications

To avoid confusion while reading, I checked the illustrations for objects or labels which were not described in the text. The participants should not be left to wonder about unexplained contents. When an illustration contained unexplained contents then either the text or the illustration was altered to create one integrated presentation. Furthermore, I made certain that the text and illustrations would support the placement of at least three textual cues to reference the illustration from the text. The reason for this choice is explained in paragraph 2.1.2.

Topics

To avoid transfer of knowledge between presentations, and to lessen the possible effect from prior knowledge, diverse topics were selected from various knowledge domains. These were the astronomy, biology, gastronomy, geology, sociology and technology domains. A similarity in the domain was allowed. For example the topics

'Oceanic depths and divisions', 'The thermohaline circulation', and 'The mid-ocean ridge system' all describe something which happens in the oceans. However, the topics are all diverse and do not overlap. The selected topics always describe a process, a situation, or the organization of a network or an object. This allows for the text and illustration to have enough complexity to add relevant content to the presentation, which cannot be captured at a single glance.

Origin of the materials

All presentations were assembled from various internet resources which are referenced in Appendix section A.1. Not all sources supplied both a text and an illustration which met the requirements described above. Texts and illustrations sometimes originated from different sources and were altered to match each other.

The selected topics were the following:

1. The structure and development of hurricanes (example presentation),
2. The Tuareg and Tuareg conflicts,
3. African migration routes to western Europe,
4. The mid-ocean ridge system,
5. Replication of the influenza virus,
6. The various beef cuts from a cow,
7. The rock cycle,
8. Oceanic depths and divisions,
9. The thermohaline circulation,
10. The beer brewing process,
11. The malaria parasite and its replication,
12. An anatomy of the Sun,
13. The working of the fluorescence microscope.

Adding cues

Three textual cues were added to the main text body of each presentation in the simple and explicit cues conditions. With only one cue per presentation, participants might have recognized a relation between the cue and the queried content in the question. To avoid that participants would focus on the cues to answer the questions 3 cues were added and for consistency this was done for every presentation. The simple textual cues were phrases like: "See the picture", "See the image" or "See the map". The explicit textual cues were designed to relate essential content in the text and illustration. They described more details than the simple cues: e.g., *"Notice in the middle right of the image the merozoites moving from the exo-erythrocytic phase to the red blood cells in the erythrocyte phase."* The cues should not break the flow of text. Their location in the text must guide the attention of the reader to the illustration at the right moment. This was immediately after a piece of text that described a part of the topic that was also depicted in the illustration.

In addition, the cued information was relevant to the question that followed after the presentation. Only 1 of the 3 cues assisted in answering the question. The explicit cues were designed to not give away the answer to the question. For the example presentation in figure 2.2 the following question could be asked: *"Where are the most Tuareg confederations located?"* A textual cue that would give away the answer would be as follows: *The picture shows that the Tuareg are organized into loose confederations which are mostly located in Niger.* A more appropriate cue only directs the reader to study the essential contents: *"The picture shows that the Tuareg are organized into loose confederations which are highlighted in green."*

Figure 2.2: A presentation about the Tuareg and the Tuareg conflicts without inserted cues.

The Tuareg are a Berber nomadic pastoralist people who once controlled the caravan trade routes across the Sahara. They are nominally Muslim. They live in the semi-arid Sahel and arid Sahara in an area that overlaps southern Algeria, southwestern Libya, Mali, and Niger, and in fewer numbers, in Burkina Faso and Nigeria. Their economy is based on livestock, agriculture and trade. They are referred to as "the people of the veil" or "he blue people of the Sahara" in reference to the indigo turbans worn by men, which stain their skin and define their identity. The Tuareg confederations are the Kel-Ahaggar, Ajjer, Ayr, Adagh, Ataram, Denneg, Owey, and Gres. Most of them are located in Niger. The picture shows that the Tuareg are organized into loose confederations which are highlighted in green.

The twentieth century saw profound changes to their way of life: the end of colonial French rule and the creation of new countries; devastating, repeated droughts that decimated herds of livestock; and political marginalization and rebellions. Their social organization and economy have substantially transformed, and most Tuareg have given up their nomadic lifestyle. When African countries achieved widespread independence in the 1960s, the traditional Tuareg territory was divided among Niger, Mali, Algeria, Libya, and Burkina Faso. Notice in the picture how the Tuareg territory is spread over the new nations. Competition for resources has since led to conflicts between the Tuareg and neighbouring African groups, especially after the French colonization and independence.

In Mali, a Tuareg uprising surfaced in the northern mountains in the 1960s, following Mali's independence. Several Tuareg joined. The army of the recently formed state of Mali suppressed the revolt. Resentment among the Tuareg fueled the 2nd uprising from 1990 to 1996 in the aftermath of a clash outside a prison in Tchin-Tabaraden, Niger. Tuaregs claimed autonomy for their traditional homeland and caused several clashes in both countries that are marked as stars on the map. Sporadic fighting in the Mountains around Arlit continued. The tourist center of Agadez and the uranium mining town of Arlit were evacuated of foreigners and armed by the Niger Army. In Mali, the rising began when Tuareg separatists attacked government buildings around Gao. Deadly clashes followed with deaths numbering well into the thousands. Negotiations initiated by France and Algeria led to peace agreements. In 2007, a new surge in violence occurred.

2.1.3 Pre-experiment trials

An earlier version of the materials proved to be too easy for the participants. During a pre-experiment trial two male students completed the test and answered all questions correctly regardless of the use of cues. The ceiling effect was probably caused by a number of problems:

1. All the presentations covered similar topics from the medical domain,
2. The students said to have some prior knowledge of the medical domain,
3. The text length was about 140 words and thus, there was little information to memorize.

To correct for the ceiling effect, I redesigned the materials as described above in section 2.1.2. Diverse knowledge domains were used to minimize the effects of prior knowledge. The texts were made longer, and more difficult, topics were selected to increase the amount and complexity of the memorized information. In a second pre-experiment trial with renewed materials four participants (2 male and 2 female) were included. Two participants complained about the length of the experiment (The average time spent was about 1 hour.). The results were mixed; with some participants scoring very well while others scored badly.

2.1.4 Equipment

Participant pool system

The participants signed up for the experiment via the 'proefpersonenpool' (participants pool) website of the faculty for behaviour sciences at the University of Twente. The participants received 1 university credit after completion. Failure to attend the experiment without prior explanation was punished with a reduction of 5 university credits.

Web-based experiment system

The experiment website used a system for presenting web-based questionnaires and experiments called WWStim (Veenker, 2003). The system does have its limitations. The standard set-up does not allow the experimenter to first show the stimuli and then the question. An alternative implementation from standard use made this possible. However, this disabled the option to randomize the presentations. I further adapted the implementation with a time registration system which recorded the time of opening and closing of a web page. With the time stamp information I got an indication of the reading time per presentation and answering time per question.

2.1.5 Procedure

A link leading from the participant pool system guided the participants to the experiment website. A welcome page instructed participants that they would need to study 12 illustrated text presentations. Furthermore, it explained all the steps necessary to complete the experiment and informed the participants that the experiment would last about one hour. There was no time constraint set on the experiment. No details were given on the purpose of the experiment to avoid guessing on the experiment variables. A focus on the experiment variables might have influenced the participants to look for cues and could have caused a different reading behaviour.

After the introduction page followed a personal details form. Participants were asked to give their:

- age
- gender (male or female)

- mother tongue (Arabic, Chinese, Dutch, English, French, German, Italian, Japanese, Spanish or Other)
- reading level of English (mother tongue, proficient/excellent, advanced, intermediate, elementary or beginner)
- email address (optional)

Filling in the email address was optional and was only intended for students that wished to receive more information on the experiment.

Next an instruction page followed. Participants were instructed to read and study an example presentation and then 12 experiment presentations. In addition, they were informed that each presentation would be followed by a question on the content of the presentation.

The example presentation contained no cues regardless of the experiment condition. Each presentation page requested participants to study the presentation and to click on continue when they would be ready. Each presentation page was followed by a question page with 1 question and 4 multiple choice answers. It was not possible to view the presentation or to return to the presentation page while answering a question. Participants could continue to the next presentation by pressing the next button. However, they could only continue after choosing an answer. After the example, an intermediate instruction page was shown that informed the participants that the example was finished and that 12 presentation were to follow. In addition, the page instructed participants to read and study at their own pace.

After reading all presentations and answering all questions, participants were asked to answer the post-test questions. The participants were shown an instruction page that informed them that all presentations had been shown and that next five evaluation questions would follow. In addition, the page informed them that these were the last questions of the experiment. Participants had to answer a question before being allowed to continue to the next. After answering all questions, the participants were thanked for their participation. It was possible to leave a comment or ask questions via a comments box.

2.1.6 Post-test questions

The post-test questions were designed to measure how participants subjectively experienced the difficulty of the presentations, the difficulty of the language used in the presentations and the easiness of the presentation questions. Furthermore, the participants were asked again to assess their level of English which might have changed after reading the presentations. Lastly, to measure discomfort participants were asked to indicate how often they considered stopping the experiment before reaching the end of it.

These are the post-test questions:

1. *Considering the presentations that you have read, how many presentations did you consider difficult to study?*
   (a) *none,*
   (b) *1 to 3,*
   (c) *4 to 6,*
   (d) *7 to 9,*
   (e) *10 to 12.*

2. *Considering the presentations that you have read, how difficult did you find the language used in the presentations?*

   Not difficult □ □ □ □ □ □ □ *Very difficult*

3. *Considering the questions following each presentation, how many questions do you think were easy to answer?*

   (a) *none,*

   (b) *1 to 3,*

   (c) *4 to 6,*

   (d) *7 to 9,*

   (e) *10 to 12.*

4. *Considering the presentations that you have read, how do you assess your level of English?*

   Excellent □ □ □ □ □ □ □ *Basic*

5. *Considering the experiment as a whole, how many times did you consider stopping before reaching the end of the experiment?*

   (a) *never considered,*

   (b) *1 time,*

   (c) *2 to 5 times,*

   (d) *more than 5 times.*

2.1.7 Statistical tests

The results from the experiment were tested using generally accepted and used statistical tests. Choosing the right test depends on the type of tested data (interval, categorical or ordinal data), the number of tested groups, and results from pre-tests, such as Levene's test of Homogeneity of Variance and the Shapiro-Wilk test.

Levene's test of Homogeneity of Variance is intended to test for statically significant differences of variance. The Shapiro-Wilk test is intended to test the distribution of data. A statically significant result from the Shapiro-Wilk test means data is not normally distributed.

When testing the difference between two independent groups on an interval or ratio scale, I used the independent two-sample Student's t-test. When testing between more than two independent groups, I used the one-way ANOVA test. To reliably perform the ANOVA test, a number of assumptions have to be satisfied:

- The groups are independent and random,
- The data should be normally distributed,
- The groups have the same variances.

To meet the first assumption, the tested groups were always independent and the participants were selected randomly. To meet the second assumption, I tested for normality using the Shapiro-Wilk test. To meet the third assumption, I tested for equality of variance with Levene's test of Homogeneity of Variance. If either test was significant, then I used a non-parametric test.

Tests between two groups on an ordinal scale were done using a non-parametric test. When testing between two groups, I used the Mann-Whitney U test. When there were more than two groups, I used the Kruskal-Wallis test.

To test the difference in ranges (the minimum and maximum values) between groups, I used Levene's test of Homogeneity of Variance.

Tests between groups on a nominal scale were done using the Pearson's Chi-Squared test. If any of the observed counts were smaller than 10 or if the expected count was less than 5, then I used Fisher's Exact test. The Fisher's Exact test I will abbreviate from here on as FET.

To test the correlation between variables, I used Pearson's r test for normally distributed data and Spearman's rho test for not normally distributed data.

All tests were conducted using a two-sided alpha level of 0.05.

2.2 Processing of the data

The gathered data needed to be processed prior to analysing the results. The experiment set-up allowed participants to participate via the Internet and in any location of their liking. This limited the barrier for participation and made it easier to gather participants. However, the situation in which the participants studied the presentations was not under control and this influenced the experiment. Incomplete data sets and unforeseen behaviour were excluded from the results. In this chapter, I will discuss in which cases these data sets were excluded from the results.

The experiment set-up did not restrict the number of times a participant could start the experiment. Thus, for each participation a new dataset was recorded. Recording started when the personal data was entered and continued up to the last post-test question. If a participant broke off the experiment mid-way then the data set was registered as incomplete. The personal data, the experiment data, and the post-test data was registered within each dataset. The experiment data consisted of 12 answers given on the presentation questions and a time stamp taken at the transition to the next page in the experiment. The timestamps show approximately how long each page was open. This gives an indication of how much time the participants spent reading a presentation or how long a question was pondered upon. The view time of a page was calculated by subtracting the ending time stamp of the previous page from the ending time stamp of the current page. The recorded time is distorted by the web page load time and can be distorted by plausible distractions during the experiment, e.g. for making a cup of tea or answering a phone call. Since the experiment was not held in a laboratory under controlled conditions, distractions cannot be ruled out. Finally, a data set contains the post-test data which consists of 5 answers to the post-test questions.

Below I specify a number of requirements on each data set. When a data set did not fall within the requirements, it was discarded from the analysis.

1. **Incomplete data**
   A participation was required to have a complete data set with all experiment data. Those that did not, I considered to be incomplete and were discarded from the analysis.

2. **Minimum total reading time**
   Experiment time varied per participant and is dependent on the individual reading and answering speed. Especially reading speed can vary depending on the used reading technique. Liu (2005) reports that people reading from a screen use different techniques than when reading from printed media. In their survey

113 participants between the age of 30 and 45 years reported to spend more time on browsing, scanning, keyword spotting, one-time reading and non-linear reading than 10 years prior to the experiment. They read more selectively and did less in-depth reading. Morkes and Nielsen (1997) showed similar findings. In 3 experiments they tested different reading techniques for the web. They conclude web readers scan, search for keywords and like concise pages. Web readers like to get quickly to the point. Similar reading techniques are skimming and scanning. Skimmers 'scan' for relevant information and skip irrelevant parts with the aim to absorb information faster than while reading at normal speed. Reported normal speeds vary between 199 to 250 words per minute (wpm) (Masson, 1982; Muter & Maurutto, 1991; Dyson & Haselgrove, 2000; Hewitt, Brett, & Peters, 2007; Duggan & Payne, 2009). Experiments on skimming while reading from a computer screen (Muter & Maurutto, 1991; Dyson & Haselgrove, 2000; Duggan & Payne, 2009, 2011) observe various speeds ranging from 460 to 600 wpm. Based on results from Muter and Maurutto (1991) and a pilot study of 22 students, Hewitt et al. (2007) concluded that comprehension suffers at reading speeds above 8 words per second (480 wpm). They note this to be a conservative boundary and that the speed and comprehension may vary per student. In their experiment they measured the reading speed of students per note size. The notes varied from 0 - 24 to 600 - 624 words with incremental steps of 25 words. They found that when the text got longer more students read faster. When reading the shortest notes 0.2% of the student read faster than 8 wps and 51.9% for the longest notes. Not only document length but also the amount of characters per text line influences the reading speed (Dyson & Haselgrove, 2001). It is probable that skimming and scanning were used by participants in this experiment which was held on the web and read from a computer screen. One could argue that readers participating in a web experiment will behave differently when they read regular web content. However, the computer screen remains of influence as concluded by Liu (2005). Skimming and scanning participants should not be excluded from this experiment. However, as suggested by Hewitt et al. (2007), participants reading too fast might have missed essential information. Some speedy participants might not have read at all. Including those results might undesirably skew the results.

Considering the various reported speeds in prior research, I set the maximum reading speed at 600 wpm. By counting the total word length of all 12 presentations and based on this maximum reading speed, a minimum total reading time was calculated for each experiment condition. See table 2.1 for the minimum reading times per condition. Those participants that read faster than the minimum reading speed might have skipped over cues and did not engage in meaningful learning. Thus, the data sets of those participations were discarded from the analysis. Note that the participant was not necessarily excluded. See the next requirement for more information on participants with multiple participations.

Table 2.1: The word length and minimum reading time per condition.

	Word count	Min. reading time
No cues	3962	6 min. 36 sec.
Simple cues	4070	6 min. 47 sec.
Explicit cues	4494	7 min. 29 sec.

3. **Excessive reading or answering time**
The recorded reading and answering times should not be excessive. I consider a recorded time excessive if it is 2.5x standard deviation above the average reading time or, respectively, answering time within the participated condition. These registered times distinctly deviate from the rest of the data and skew the result. The method of excluding scores which are deviating more than 2.5x from the mean is commonly accepted in statistics. The occurrence of these very long reading or answering times can be due to various reasons. It might be the case that a participant truly read very long or needed a long time to think over an answer. However, a more plausible explanation is a distraction during the experiment. Unfortunately, this cannot be verified. To avoid skewing of the results, data sets were excluded from the results if they

contained times that exceed excessively as described above.

4. **Multiple participations**
For various reasons a participant might have participated in the experiment multiple times. For example, a participants' Internet connection might have disconnected or the experimenter requested a new attempt. Preferably, the first participation was used for the analyses because, during later attempts, the participant might have remembered information from the earlier attempts. Also, the participant might have changed reading style, for example by skipping parts that were already understood or memorized. Furthermore, due to the experiment set-up, every participation was assigned sequentially to any of the 3 conditions. Thus, a participant could have participated in multiple conditions. The effects of one condition could then be transferred to another condition. Hence, given these possible unwanted effects, extra requirements are needed to decide which participation to include in the analysis.
   (a) Only one data set per participant is included in the analysis and all others are discarded,
   (b) The data set of the first participation is used in analyses if it meets the requirements mentioned above in points 1, 2 and 3,
   (c) If the first participation was discarded then the next participation is considered until a data set is found which meets the requirements,
   (d) In addition, to minimize the effects due to the transfer of knowledge between participations, discarded prior participations must satisfy one of the following conditions:
      i. In earlier participations the participant did not actually start the experiment and had not studied any of the presentations or,
      ii. The participant read faster than the total minimum reading time or,
      iii. About two weeks had passed since the last discarded participation.

Summary
Following the above requirements only one data set per participant was included in the analysis.

- Each data set was complete,
- The total reading time was longer or equal to the minimum total reading time,
- Each data set did not have excessive reading or answering times,
- In case of multiple participations, the effects due to the transfer of knowledge between participations was minimized.

Included data sets.

A total of 60 data sets were included in the analysis.

Discarded data sets.

In total 56 data sets have been discarded. See table 2.2 for a complete overview per requirement.

34 data sets were not complete. From those, participants in 16 participations never started the experiment and only signed up. In 18 participations not all 12 questions were answered and thus, the participation must have been interrupted and broke off the experiment prematurely. Of those that got interrupted, 4 participants immediately restarted with a new attempt and skipped the parts they had already completed. Unfortunately, due

to sequential condition assignment their new attempt took place in a different condition than their failed attempt. As a result, the two partial data sets are from two different conditions. Therefore, the two data sets cannot be combined and both attempts were discarded. Participants in 17 participations read the presentations faster than the minimum reading time. 3 participations were discarded because reading or answering time deviated above 2.5x standard deviation from the mean. 1 participation was redundant because the same participant had already participated successfully. 1 participation was discarded because an earlier attempt was held less than two weeks earlier.

Table 2.2: The number of discarded data sets per type of requirement.

	Frequency (N=56)
Experiment not started	16 (29%)
Incomplete experiment data	18 (32%)
Less than minimal reading time	17 (30%)
Excessive reading or answering score(s)	3 (5%)
Redundant valid data set	1 (2%)
Time difference between attempts	1 (2%)

2.2.1 Reported level of English

Group sizes for the levels Proficient/Excellent, Elementary or Beginner were small which would cause them to be ignored in the analysis. See table 2.3 for the levels as reported. Since most participants were concentrated in the Advanced and Intermediate levels I decided to merge the 6 English levels into 2 new levels so that the smaller groups are included in the analysis. The levels Mother tongue, Proficient/Excellent and Advanced were combined in a new group named the ‘Higher’ level. Levels Intermediate, Elementary and Beginner were combined in a group named the ‘Lower’ level. See table 2.6 for the combined levels.

Table 2.3: Correct answers per condition and reported level of English

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mother tongue		(n=0)	(n=0)	(n=0)	(n=0)
	Mean	-	-	-	-
	Standard deviation	-	-	-	-
	Range	-	-	-	-
Proficient/Excellent		(n=5)	(n=1)	(n=1)	(n=7)
	Mean	8.20	4	6	7.29
	Standard deviation	1.30	-	-	1.98
	Range	7-10	4-4	6-6	4-10
Advanced		(n=9)	(n=12)	(n=15)	(n=36)
	Mean	7.88	7.79	6.88	7.42
	Standard deviation	2.98	2.89	3.03	2.73
	Range	5-10	3-11	2-12	2-12
Intermediate		(n=4)	(n=6)	(n=5)	(n=15)
	Mean	5.75	6.67	7.00	6.53
	Standard deviation	1.71	1.21	1.23	1.36
	Range	4-8	6-8	6-9	4-9
Elementary		(n=1)	(n=0)	(n=0)	(n=1)
	Mean	8	-	-	8
	Standard deviation	-	-	-	-
	Range	8-8	-	-	8-8
Beginner		(n=0)	(n=1)	(n=0)	(n=1)
	Mean	-	7	-	7
	Standard deviation	-	-	-	-
	Range	-	7-7	-	7-7

2.3 Results

The results that have been gathered are divided in 3 parts: First, I will present the participant data and show the test results for homogeneity of the conditions. Homogeneity of the conditions is favourable to exclude possible unwanted differences in the results due to group differences. Second, I will present the results from the recall test (the number of correct answers on the presentation questions) and present the test results on the influence of the cues. After which, I assessed if differences within the participants influenced the recall results. For example by comparing differences between men and women. Third, I will present the post-test results and test for differences between the conditions. After which I will assess the relation between the post-test results and the recall results. Out of convenience, in the remainder of this document I will abbreviate "the number of correct answers on the presentation questions" as "the number of correct answers".

2.3.1 Participants

Before analysing the results I tested whether the participants were divided homogeneously over the three conditions. 60 participants met the requirements discussed in section 2.2 of which 19 participants were in the no cue condition, 20 participants were in the simple cue condition and 21 participants were in the explicit cue condition. I compared the groups with the personal information reported by the participants before starting the experiment. The tables are presented in the appendix.

- Gender (see table B.1),
- Average age (see table B.2),
- Mother tongue (see table B.3),
- Reported level of English (see table B.4).

No significant differences between the 3 conditions were found in terms of participants gender (2/3 were female and 1/3 male), average age (21 years) and mother tongue (50% Dutch and 50% German). There was no significant difference in the reported level of English between the conditions. In the no cues condition, about 1/4 of the participants reported a Proficient/Excellent level of English, 1/2 reported an Advanced level and 1/4 reported an Intermediate level. In the simple cues condition almost 2/3 reported an Advanced level and almost 1/3 reported an Intermediate level. In the explicit cues condition over 2/3 reported an Advanced level and 1/4 reported an Intermediate level. The remainders were spread over other levels.

These results show that the participant groups had similar properties. This was a desired outcome to avoid possible unwanted differences in the results due to group differences.

2.3.2 Influence of Cues

The first test assesses the difference in the number of correct answers between the experiment conditions. The number of correct answers compared to the no cue condition decreases when simple cues are used and even more with the explicit cues. However, the differences were not statistically significant ($F(2,57) = 0.29$, $p = .75$). A second test using Levene's test for equality of variances, to assesses the difference in the ranges (the minimum and maximum values) between the conditions, showed no statically significant differences ($F(2,57) = 1.17$, $p = .32$). See table 2.4. The total average shows that participants had 60%, 7 out of 12, answers correct. Each question had 4 possible answers giving a guessing participant a chance of 1 on 4. Participants might have increased their chances for guessing by excluding unlikely answers based on prior knowledge.

Table 2.4: Correct answers per condition (all data sets)

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mean	7.47	7.10	6.90	7.15
Standard deviation	1.71	2.55	2.68	2.34
Range	4-10	3-12	2-12	2-12

Summary

Participants in the three conditions scored about the same. Participants in the no cues condition scored slightly better than those in the simple cues condition and those in the explicit condition again worse. The range of scores in the simple and explicit cues condition seem more spread than the range of the no cues condition. Analysing the mean number of correct answers and their range, no statically significant differences were observed between the three conditions. Overall, participants had about the same number of correct answers in all three conditions.

Because no differences between the conditions were found it might be interesting to explore the possibility of differences within the conditions due to other factors. In the following tests I will look at differences in the number of correct answers due to:

- Gender,
- Mother tongue,
- The reported level of English,
- Reading and answering time,

2.3.3 Influence of Gender

Below I present the results from testing on the influence of gender. In three tests I assess if gender makes a difference in the number of correct answers.

In the first test, I tested the difference in the number of correct answers between men and women in general. Women scored higher than men, see table 2.5. However, the difference is not statistically significant (t(58) = -0.56, p = .58).

In the second test, I tested the difference of the number of correct answers between men and women in each of the 3 conditions. Scores on the number of correct answers in the no cue and simple cue condition were almost the same for women and men, see table 2.5. The difference between men and women, for both conditions, was not significant: no cues (t(17) = -0.24, p = .82) and simple cues (t(18) = -0.13, p = .90). Scores on the number of correct answers in the explicit cue condition were higher for women than for men. However, the difference is not significant (t(4.53) = -0.45, p = .67). Levene's test indicated unequal variances between men and women in the explicit cue condition (p = .002), so degrees of freedom were adjusted from 19 to 4.53.

In the third test, using the same groups, I assessed for each gender the difference in the number of correct answers between the experiment conditions. Levene's test indicated unequal variances between the experiment conditions in the group of male participants (p = .002). Thus, the Kruskal-Wallis test was used to assess the male participants. The number of correct answers for both men and women are lower in the cued conditions than in the no cue condition. However, the differences between the conditions for the male participants were not statistically significant (H(2) = 0.56, p = .76). Furthermore, the differences in the female group were not statistically significant (F(2,39) = 0.16, p = .85).

Table 2.5: Correct answers per gender

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Male		(n=6)	(n=7)	(n=5)	(n=18)
	Mean	7.33	7.00	6.20	6.89
	Standard deviation	1.03	3.16	4.44	2.95
	Range	6-9	3-11	2-2	2-11
Female		(n=13)	(n=13)	(n=16)	(n=42)
	Mean	7.54	7.15	7.13	7.62
	Standard deviation	1.98	2.30	2.03	2.06
	Range	4-10	3-12	3-12	3-12

Summary

When analysing the number of correct answers using three tests, no statically significant differences were observed between males and females. There is no indication that gender influences the number of correct answers.

2.3.4 Influence of English Level

The next tests assess whether the participant's level of English makes a difference in the number of correct answers. Participants were asked before the experiment to specify their level of English. The available answers were; Proficient/Excellent, Advanced, Intermediate, Elementary and Beginner. See table 2.3. These levels were combined to two new levels 'Higher' and 'Lower' as discussed in section 2.2.1. Table 2.6 shows an overview of the mean number of correct answers and standard deviation per combined English level and experiment condition.

In the first test, I assessed the differences between Higher and Lower level in general. The difference between the two levels is not statistically significant (t(54,46) = 1.37, p = .18). Levene's test showed a statistically significant difference between the variances (p = 0.01). Thus, the degrees of freedom were adjusted to 54,46.

In the second test, I tested the differences in the number of correct answers between the Higher and Lower levels for each experiment condition. Levene's test showed a statistically significant difference between the variances of the Higher and Lower level in the simple cues condition (p = .02). Thus, the test on the simple cue condition was done using the Kruskal-Wallis test. Participants in the no cues condition of the Higher level group have statistically significant more correct answers than those of the Lower level group (t(17) = 2.12, p = .049). The differences between the Higher and Lower level in the simple cue and explicit cue conditions were not statistically significant: simple cue (t(16.55) = 0.62, p = .54) and explicit cue (t(19) = -0.09, p = .93).

In the third test, I tested the differences in the number of correct answers between the experiment conditions for both levels Higher and Lower. See table 2.6. Levene's test showed a statistically significant difference between the variances in the Higher level (p = .04). Thus, the test on the 'Higher' level was done using the Kruskal-Wallis test. Both tests on the Higher and Lower level did not show a statistically significant difference: Higher (H(2) = 1.21, p = .55) and Lower (F(2,14) = 0.44, p = .65).

Table 2.6: Correct answers per condition and reported level of English combined

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Higher		(n=14	(n=13)	(n=16)	(n=43)
	Mean	7.93	7.31	6.88	7.35
	Standard deviation	1.49	3.09	3.03	2.63
	Range	5-10	3-12	2-12	212
Lower		(n=5)	(n=7)	(n=5)	(n=17)
	Mean	6.20	6.71	7.00	6.65
	Standard deviation	1.79	1.11	1.23	1.32
	Range	4-8	5-8	6-9	4-9

Summary

Participants in the no cue condition in the Higher English level group had statistically significant more correct answers than their counterparts in the Lower English level group. No other statistically significant differences were observed in the other two conditions.

2.3.5 Influence of Mother Tongue

The next tests assess whether the participant's mother tongue makes a difference in the number of correct answers. Participants were asked before the experiment to specify their mother tongue. See table 2.7 for an overview of the mean number of correct answers and standard deviation per reported English level and experiment condition.

I first I tested the difference in the number of correct answers between the mother tongues in general. Category 'other' has only one participant. The differences in the number of correct answers between the mother tongues is not statistically significant (F(2,57) = 0.77, p = .47).

In the second test, I tested the difference of the number of correct answers between the mother tongues in each of the 3 conditions. Dutch participants score better than German participants in the no cues condition. This trend continues less strongly in the simple cues condition and disappears in the explicit condition. German participants score slightly better than Dutch participants in the explicit cues condition. Differences between Dutch and German participants in the no cues condition were statistically significant (t(17) = 2.53, p = .02). No statistically significant differences in the number of correct answers were found between the mother tongues in the simple cues and explicit cues conditions: simple cue (t(18) = 0.68, p = .51) and explicit cue (F(2,18) = 0.02, p = .98). A second test excluding the category 'other' in the explicit cues condition was not significant either (t(18) = -0.19, p = .85).

In the third test, I tested the differences in the number of correct answers between the experiment conditions for each mother tongue. No test was conducted for category 'other' because it only appeared in one condition and thus, no difference could be tested. Tests on the Dutch group and German group showed no statistically significant differences: Dutch (F(2,29) = 0.81, p = .46) and German (F(2,24) = 0.17, p = .84).

Table 2.7: Correct answers per mother tongue

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Dutch		(n=10)	(n=14)	(n=8)	(n=32)
	Mean	8.30	7.36	6.75	7.50
	Standard deviation	1.34	2.85	3.37	2.62
	Range	6-10	3-12	2-12	2-12
German		(n=9)	(n=6)	(n=12)	(n=27)
	Mean	6.56	6.50	7.00	6.74
	Standard deviation	1.67	2.76	2.41	1.99
	Range	4-9	4-8	3-11	3-11
Other		(n=0)	(n=0)	(n=1)	(n=1)
	Mean	-	-	7.00	7.00
	Standard deviation	-	-	-	-
	Range	-	-	7-7	7-7

Summary

After analysing the number of correct answers using three tests, Dutch participants in the no cues conditions are shown to have statically significant more correct answers than their German counterparts in the same condition. No other statically significant differences were observed.

2.3.6 Relation between Mother Tongue and reported English Level

Both the analyses of the reported English level and mother tongue showed statically significant differences within the no cues condition. Thus, it could be that there is a relation between the mother tongues Dutch and German and the English levels Higher and Lower. It could be that the differences in the number of correct answers observed in section 2.3.5 are due to the level of English.

The differences between mother tongue and the reported English level were not statistically significant in all three conditions: no cues (p = .63), simple cues (p = 1.00)and explicit cues (p = 1.00). The analyses was done using a FET test. The number of Dutch and German participants in the no cues condition are fairly equal within each level of English. If the level of English was causing the difference in the number of correct answers, I would expect to see more Dutch than German participants in the Higher level and vice versa in the Lower level. However, this is not the case. The Dutch and German participants are fairly equally distributed over the English levels. The difference in the number of correct answers in the no cues conditions between Dutch and German participants is thus not explicitly due to the reported level of English.

Table 2.8: Mother tongue versus reported English level

No cues		Higher (n=14)	Lower (n=5)	Total (N=19)
	Dutch	8 (57%)	2 (40%)	10 (53%)
	German	6 (43%)	3 (60%)	9 (47%)
Simple cues		(n=13)	(n=7)	(N=20)
	Dutch	9 (69%)	5 (71%)	14 (70%)
	German	4 (31%)	2 (29%)	6 (30%)
Explicit cues		(n=16)	(n=5)	(N=21)
	Dutch	6 (38%)	2 (40%)	8 (38%)
	German	9 (56%)	3 (60%)	12 (57%)
	Other	1 (6%)	0	1 (5%)

Summary

Dutch and German participants in the no cues condition are fairly equally distributed over the reported levels of English. Furthermore, tests show no statistically significant differences. Thus, the English level is not perse the cause of the difference in the number of correct answers between Dutch and German participants in the no cues condition.

2.3.7 Influence of Reading and Answering time

In this section I test whether reading and answering time can predict the number of correct answers. While participants were performing the experiment the total viewing time per presentation and answering page was recorded. See tables 2.9 and 2.10. A Shapiro-Wilk test gave statistically significant results for both the reading ($p < .001$) and answering time ($p < .001$). The distribution of the time values is asymmetric and is positively skewed (the majority of the data is on the left side of the mean). Because the data is skewed I will present both the mean and the median.

There is no direct relation visible between reading time and cues. On average participants with no cues read as long as those with explicit cues. However, the median shows that most participants in the explicit cues were faster. Participants from the simple cues condition were slower than both no cues and explicit cues participants. Differences in the reading time between the conditions were not statistically significant ($H(2) = 4.24$, $p = .12$). The patterns in the answering time are similar with the reading time. Again, the average of the no cues and explicit cues conditions are similar but the median shows most explicit cues participants were faster. Simple cues participants have about the same answering time as the no cues participants. Differences in the answering time between the conditions were not statistically significant (H(2) = 1.41, p = .49).

Table 2.9: Reading time

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mean	00:27:06	00:35:51	00:27:49	00:30:16
Median	00:23:59	00:27:58	00:18:49	00:23:11
Range	00:06:51-01:07:27	00:08:55-01:36:21	00:07:59-01:22:44	00:06:51-01:36:21

Table 2.10: Answering time

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mean	00:05:59	00:06:13	00:05:29	00:05:53
Median	00:05:00	00:04:40	00:03:55	00:04:09
Range	00:02:07-00:17:37	00:01:36-00:16:54	00:01:49-00:20:35	00:01:36-00:20:36

In the next tests I assessed if the number of correct answers was related to the reading time and answering time in general. The number of correct answers per condition and in general can be viewed in table 2.4. The number of correct answers is positively correlated with the reading time (Spearman's rho = 0.27; p = .04), although the relation was not very strong. The number of correct answers is positively correlated to the answering time (Spearman's rho = 0.34; p = .01), however the relation was again not very strong.

Next, I tested if the number of correct answers was related to the reading time and answering time within each experiment condition. There was no significant correlation in any of the three conditions between the number of correct answers and the reading time: no cues (Spearman's rho = 0.27; p = .26), simple cues (Spearman's rho = 0.16; p = .50) and explicit cues (Spearman's rho = 0.43; p = .05). The results for the explicit cues condition is only weakly significant and it shows a positive but weak correlation. This suggests there might be a weak relation between the number of correct answers and the reading time. The median reading time of the explicit cues participants is about 22% faster than the no cues participants and 33% faster than those in the simple cues condition. A possible explanation is that explicit cues participants reached reading speeds close to scanning and started skipping information. This certainly can have impacted the amount of recalled information. In the other conditions skipping might have been less to a degree that it had no significant impact on the amount of recalled information. The same test with the reading time showed no significant correlation in any of the three conditions: no cues (Spearman's rho = 0.18; p = .47), simple cues (Spearman's rho = 0.39; p = .09) and explicit cues (Spearman's rho = 0.32; p = .15).

Summary

Differences between the conditions in reading and answering time were not statistically significant. Comparing the number of correct answers with the reading and answering time in general, taken all conditions together, showed weak correlations. Thus, there seems to be a weak relation between the amount of time that participants spend on reading and answering and the amount of correct answers they give. However, it is not guaranteed that if a participant studies longer or thinks longer that they will get better results. Within the experiment conditions these correlations were not statistically significant. The weak significance and correlation in the explicit cues condition suggests a possible relation between the reading speed and the number of correct answers. It could be that explicit cues allowed participants to read faster. However, there was no significant difference in reading time between the conditions. Thus, my results do not support this hypothesis.

2.3.8 Post-test results

I will now present the post-test results. 5 questions were asked on the following topics:

- Presentation difficulty,
- Presentation language difficulty,
- Ease of the questions,
- Rating of English level,
- Inclination to give up.

For each question I tested for differences in the results between and within the conditions. Following, I tested the relation between the post-test results and the recall test results between and within the conditions. By comparing how the participants answered the post-test questions and the number of correct answers they gave in the recall experiment.

Presentation difficulty

Participants were asked to specify how many of the presentations they considered difficult. The available answers were; 'none', '1 to 3', '4 to 6', '7 to 9' and '10 to 12' presentations. Table 2.11 shows the results of the participants' ratings on presentation difficulty. I tested the differences in presentation difficulty ratings over all experiment conditions. The result from the Fisher's Exact test did not reveal a statistically significant difference (p = .74).

Table 2.11: Frequency of presentation difficulty per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
none	0	2 (10%)	0	2 (3%)
1 to 3	1 (5%)	1 (5%)	3 (14%)	5 (8%)
4 to 6	6 (32%)	7 (35%)	8 (38%)	21 (35%)
7 to 9	8 (42%)	6 (30%)	8 (38%)	22 (37%)
10 to 12	4 (21%)	4 (20%)	2 (10%)	10 (17%)

Next, I tested if there was relation between the number of correct answers and the presentation difficulty ratings. See table 2.12 for an overview of the number of correct answers per presentation difficulty rating and experiment condition. First, I tested the differences in the number of correct answers between the presentation difficulty ratings in general. No statistically significant differences between the presentation difficulty ratings were found (H(4) = 3.84, p = .43). Levene's test showed that variances were significantly different (p = .046) and thus, the test was done using the Kruskal-Wallis test. Next, I tested the differences in the number of correct answers between presentation difficulty ratings within each experiment condition. No statistically significant differences were found: no cues (F(3,15) = 0.53, p = .67), simple cues (F(4,15) = 1.21, p = .35) and explicit cues (F(3,17) = 2.00, p = .15). Next, I tested the differences in the number of correct answers per presentation difficulty rating between the experiment conditions. No statistically significant differences were found: '1 to 3' (F(2,2) = 0.57, p = .64), '4 to 6' (F(2,18) = 1.68, p = .21), '7 to 9' (F(2,19) = 0.86, p = .44) and '10 to 12' (F(2,7) = 0.58, p = .94). The 'none' rating was not tested due to the zero group sizes in the no cues and explicit cues conditions.

Table 2.12: Correct answers per condition and presentation difficulty

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
none		(n=0)	(n=2)	(n=0)	(n=2)
	Mean	0	7	0	7
	Standard deviation	-	5.66	-	5.66
	Range	-	3-11	-	3-11
1 to 3		(n=1)	(n=1)	(n=3)	(n=5)
	Mean	8	12	10	10
	Standard deviation	-	-	2.65	2.36
	Range	8-8	12-12	7-12	7-12
4 to 6		(n=6)	(n=7)	(n=8)	(n=21)
	Mean	8.17	6.29	6.88	7.05
	Standard deviation	2.32	1.50	1.81	1.94
	Range	4-10	4-8	4-9	4-10
7 to 9		(n=8)	(n=6)	(n=8)	(n=22)
	Mean	7.13	7.50	5.88	6.77
	Standard deviation	1.46	2.59	3.14	2.47
	Range	5-9	3-11	2-11	2-11
10 to 12		(n=4)	(n=4)	(n=2)	(n=10)
	Mean	7	6.75	6.50	6.80
	Standard deviation	1.41	2.22	0.71	1.55
	Range	6-9	4-9	6-7	4-9

Summary When analysing the presentation difficulty rating no statistically significant differences were found between the experiment conditions. Furthermore, when assessing the relation between the presentation difficulty rating and the number of correct answers no statistically significant differences were found.

Language difficulty

Participants were asked to rate on a 7 point rating scale how difficult they found the language used in the presentations. A rating of 1 was classified as 'not difficult' and a rating of 7 as 'very difficult'. The frequency of ratings for language difficulty are shown in table 2.13.

I tested the differences in language difficulty ratings over all experiment conditions. The result from the Fisher's Exact test did not reveal a statistically significant difference (p = .81)

Table 2.13: Language difficulty per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
1 = not difficult	0	0	0	0
2	2 (5%)	2 (10%)	1 (5%)	5 (8%)
3	2 (5%)	4 (20%)	3 (14.3%)	9 (15%)
4	3 (32%)	4 (20%)	3 (14.3%)	10 (17%)
5	6 (42%)	6 (30%)	11 (52%)	23 (38%)
6	6 (21%)	3 (15%)	3 (14.3%)	12 (20%)
7 = very difficulty	0	1 (5%)	0	1 (2%)
Mean	4.63	4.35	4.57	4.52
Standard deviation	1.34	1.39	1.08	1.26

Next, I tested if there was relation between the number of correct answers and the language difficulty ratings. See table 2.14 for an overview of the number of correct answers per language difficulty rating and experiment condition. I tested the differences in the number of correct answers between the language difficulty ratings in general. No statistically significant differences were found (F(5,54) = 0.57, p = .73).

I tested the differences in the number of correct answers between language difficulty ratings within each experiment condition. No statistically significant differences were found: no cues (H(4) = 4.95, p = .29), simple cues (F(5,14) = 1.48, p = .26) and explicit cues (F(4,16) = 1.60, p = .22). Levene's test showed that variances between language difficulty ratings in the no cues condition were significantly different (p = .04) and thus, the test was done using the Kruskal-Wallis test.

Next, I tested the differences in the number of correct answers per language difficulty rating between the experiment conditions. No statistically significant differences were found: 2 (F(2,2) = 3.58, p = .22), 3 (F(2,6) = 0.00, p = 1.00), 4, (F(2,7) = 4.32, p = .06), 5 (F(2,20) = 2.30, p = .13) and 6 (F(2,9) = 0.11, p = .89). No tests were conducted on ratings 1 and 7 due to the low number of participants in these groups.

Table 2.14: Correct answers per condition and language difficulty

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
1 = not difficult		(n=0)	(n=0)	(n=0)	(n=0)
	Mean	0	0	0	0
	Standard deviation	-	-	-	-
	Range	-	-	-	-
2		(n=2)	(n=2)	(n=1)	(n=5)
	Mean	5.50	9.50	12	8.40
	Standard deviation	2.12	2.12	-	3.21
	Range	4-7	8-11	12-12	4-12
3		(n=2)	(n=4)	(n=3)	(n=9)
	Mean	7	7	7	7
	Standard deviation	0	2	1.73	1.50
	Range	7-7	4-8	6-9	4-9
4		(n=3)	(n=4)	(n=3)	(n=10)
	Mean	7	4.50	8.33	6.40
	Standard deviation	2.65	1.29	1.16	2.32
	Range	5-10	3-6	7-9	3-10
5		(n=6)	(n=6)	(n=11)	(n=23)
	Mean	8.50	7.50	6	7.04
	Standard deviation	1.05	3.02	2.49	2.51
	Range	7-10	3-12	2-11	2-11
6		(n=6)	(n=3)	(n=3)	(n=12)
	Mean	7.50	8	7	7.50
	Standard deviation	1.64	2.65	4	2.36
	Range	5-9	6-11	3-11	3-11
7 = very difficult		(n=0)	(n=1)	(n=0)	(n=1)
	Mean	0	8	0	8
	Standard deviation	-	-	-	-
	Range	-	8-8	-	8-8

Summary When analysing the language difficulty rating no statistically significant differences were found between the experiment conditions. Furthermore, when assessing the relation between the language difficulty rating and the number of correct answers no statistically significant differences were found.

Ease of the questions

Participants were asked to answer how many of the experiment questions they found easy to answer. The available answers were 'none', '1 to 3', '4 to 6', '7 to 9' and '10 to 12'. The frequency of ratings for the ease of questions are shown in table 2.15.

I tested the differences in ease of questions rating over all experiment conditions. The result from the Fisher's Exact test did not reveal a statistically significant difference (p = .12).

Table 2.15: Ease of questions difficulty per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
none	3 (16%)	3 (15%)	0	6 (10%)
1 to 3	8 (42%)	11 (55%)	17 (81%)	36 (60%)
4 to 6	7 (37%)	4 (20%)	2 (10%)	13 (22%)
7 to 9	1 (5%)	1 (5%)	1 (5%)	3 (5%)
10 to 12	0	1 (5%)	1 (5%)	2 (3%)

Next, I tested if there was a relation between the number of correct answers and the ease of the questions ratings. See table 2.16 for an overview of the number of correct answers per ease of the questions rating and experiment condition.

Table 2.16: Correct answers per condition and rating for the ease of questions

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
none		(n=3)	(n=3)	(n=0)	(n=6)
	Mean	6.67	7.33	0	7
	Standard deviation	2.08	1.16	-	1.55
	Range	5-9	6-8	-	5-9
1 to 3		(n=8)	(n=11)	(n=17)	(n=36)
	Mean	7.88	5.73	6.47	6.56
	Standard deviation	1.81	2.10	2.45	2.30
	Range	5-10	3-9	2-11	2-11
4 to 6		(n=7)	(n=4)	(n=2)	(n=13)
	Mean	7.86	9.50	6.00	8.08
	Standard deviation	0.90	2.38	0	1.80
	Range	7-9	7-12	6-6	6-12
7 to 9		(n=1)	(n=1)	(n=1)	(n=3)
	Mean	4	8	11	7.67
	Standard deviation	-	-	-	3.51
	Range	4-4	8-8	11-11	4-11
10 to 12		(n=0)	(n=1)	(n=1)	(n=2)
	Mean	-	11	12	11.50
	Standard deviation	-	-	-	0.71
	Range	-	11-11	12-12	11-12

I tested the differences in the number of correct answers between the ease of the questions ratings in general. Differences between the ratings were statistically significant ($F(4,55) = 3.30$, $p = .02$). Tukey post-hoc comparisons indicated that the participants who rated '1 to 3' had significantly less correct answers than the participants that rated '10 to 12', ($p = .02$). However, with only 2 participants rating '10 to 12' the group is too small to draw any conclusions. Comparison between the other ratings did not reveal statistically significant differences.

I tested the differences in the number of correct answers between ease of the questions ratings within each experiment condition. Differences between the ratings in the simple cues condition were statistically significant (F(2,15) = 3.52, p = .03). Tukey post-hoc comparisons indicated that the participants who rated '1 to 3' had significantly fewer correct answers than the participants that rated '4 to 6'. Comparison with the participants that rated 'none' did not reveal statistically significant differences. Results for ratings '7 to 9' and '10 to 12' were not compared due to the small group sizes. Differences between the ratings in the others conditions were not statistically significant: no cues (F(3,15) = 2.24, p = .13) and explicit cues (F(3,17) = 2.80, p = .07)

Next, I tested the differences in the number of correct answers per ease of the questions rating between the experiment conditions. No statistically significant differences were found: 'none' (F(1,4) = 0.24, p = .65), '1 to 3' (F(2,33) = 2.18, p = .13) and '4 to 6' (H(2) = 5.75, p = .06). Levene's test indicated statistically significant differences between variances in rating '4 to 6' (p < .01) thus, the test on rating '4 to 6' was done with the Kruskal-Wallis test. No tests were conducted on ratings '7 to 9' and '10 to 12' due to the small group sizes.

Summary When analysing the ease of questions ratings between the experiment conditions no statistically significant differences were found. When assessing the relation between the ease of questions rating and the number of correct answers in general, participants in the group of rating '1 to 3' were found to have a statistically significant lower number of correct answers than participants in the group of rating '10 to 12'. The number of participants in the group of rating '10 to 12' is very small. Within the simple cues condition, participants in the group of rating '1 to 3' were found to have a statistically significant lower number of correct answers than participants in rating '4 to 6'. No other statistically significant differences were found. Furthermore, the group sizes in the simple cues condition are quite small. Most participants cluster in one rating which makes comparison with the number of correct answers difficult.

Rating of the level of English

Participants were asked, considering the presentations they had read, to re-assess their level of English, this time on a 7 point rating scale. A rating of 1 was classified as 'Excellent' and a rating of 7 as 'Basic'. The frequency of participants' ratings on their level of English can be seen in table 2.17.

I tested the differences in the level of English rating between the experiment conditions. The result from the Fisher's Exact test did not reveal a statistically significant difference (p = .55).

Table 2.17: Rating of the level of English per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
1 = Excellent	0	1 (5%)	1 (5%)	2 (3%)
2	2 (11%)	2 (10%)	2 (10%)	6 (10%)
3	7 (37%)	6 (30%)	5 (24%)	18 (30%)
4	1 (5%)	5 (25%)	7 (33%)	13 (22%)
5	5 (26%)	1 (5%)	3 (14%)	9 (15%)
6	3 (16%)	2 (10%)	2 (10%)	7 (12%)
7 = Basic	1 (5%)	3 (15%)	1 (5%)	5 (8%)
Mean	4.16	4.05	3.90	4.03
Standard deviation	1.50	1.76	1.45	1.55

Next, I tested if there was a relation between the number of correct answers and the rated post-test English level. See table 2.18 for an overview of the number of correct answers per English level rating and experiment condition. I tested the differences in the number of correct answers between the English level ratings in general. No statistically significant differences were found (F(6,53) = 1.45, p = .21). I tested the differences in the number of correct answers between English level ratings within each experiment condition. No statistically significant differences were found: no cues (H(5) = 5.18, p = .39), simple cues (F(6,13) = 0.52, p = .79) and explicit cues (F(6,14) = 1.32, p = .31). Levene's test showed that variances between language difficulty ratings in the no cues condition were significantly different (p = .04) and thus, the test was done using the Kruskal-Wallis test. Next, I tested the differences in the number of correct answers per English level rating between the experiment conditions. No statistically significant differences were found: 2 (H(2) = 0.29, p = .86), 3 (F(2,15) = 0.23, p = .79), 4 (F(2,15) = 0.80, p = .48), 5 (F(2,6) = 0.53, p = .61), 6 (F(2,4) = 3.27, p = .14) and 7 (F(2,2) = 0.89, p = .53). Levene's test indicated that differences in variances were statistically significant for rating 2 (p < .001) thus, the test on rating 2 was done with the Kruskal-Wallis test. No test was conducted on rating 1 due to the low number of participants in this group.

Table 2.18: Correct answers per condition and English level rating

		No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
1 = Excellent		(n=0)	(n=1)	(n=1)	(n=2)
	Mean	0	11	12	11.50
	Standard deviation	-	-	-	0.71
	Range	-	11-11	12-12	11-12
2		(n=2)	(n=2)	(n=2)	(n=6)
	Mean	7	5.50	8.50	7
	Standard deviation	0	3.54	3.54	2.61
	Range	7-7	3-8	6-11	3-11
3		(n=7)	(n=6)	(n=5)	(n=18)
	Mean	7.71	7.50	6.80	7.39
	Standard deviation	1.80	2.67	2.59	2.23
	Range	5-10	4-12	4-11	4-12
4		(n=1)	(n=5)	(n=7)	(n=13)
	Mean	10	6.80	6	6.62
	Standard deviation	-	3.27	2.77	2.93
	Range	10-10	3-11	2-9	2-11
5		(n=5)	(n=1)	(n=3)	(n=9)
	Mean	7	6	8	7.22
	Standard deviation	2.12	-	1	1.72
	Range	4-9	6-6	7-9	4-9
6		(n=3)	(n=2)	(n=2)	(n=7)
	Mean	8	6.50	4.50	6.57
	Standard deviation	0	2.12	2.12	1.99
	Range	8-8	5-8	3-6	3-8
7 = Basic		(n=1)	(n=3)	(n=1)	(n=5)
	Mean	5	7.33	7	6.80
	Standard deviation	-	1.53	-	1.48
	Range	5-5	6-9	7-7	5-9

Compared with pre-test reported level of English In general terms, the post-test ratings are quite spread and centered around 4 points. Thus, after reading the presentations participants did not think their level of English to be great and not bad either. It looks like they felt ok with their level of English and that it was adequate for the test. The English level reported by participants before the experiment seems to be slightly more optimistic. 77% reported to have an advanced level of English. 25% of the participants reported to have an intermediate level which in my opinion can be compared to rating of 4 points in the post-test level of English. Thus, it seems participants adjusted their level of English based on their experience with the experiment.

Summary When analysing the post-test English level rating no statistically significant differences were found between the experiment conditions. Furthermore, when assessing the relation between the post-test English level rating and the number of correct answers no statistically significant differences were found. Compared to the pre-test reported English level participants seem to be less optimistic about their level of English.

Inclination to give up

Participants were asked to rate how many times they considered stopping the experiment before reaching its end. The available answers were 'never considered', '1 time', '2 to 5 times' and 'more than 5 times'. The frequency of participants' rating on this 'inclination to give up' can be seen in table 2.19.

I tested the differences in the inclination to give up between the experiment conditions. The result from the Fisher's Exact test did not reveal a statistically significant difference (p = .85).

Table 2.19: Inclination to give up per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
never considered	6 (32%)	8 (40%)	8 (38%)	22 (37%)
1 time	3 (16%)	5 (25%)	6 (27%)	14 (23%)
2 to 5 times	9 (47%)	6 (30%)	7 (33%)	22 (37%)
more than 5 times	1 (5%)	1 (5%)	0	2 (3%)

Next, I tested if there was relation between the number of correct answers and the ratings given on the inclination to give up in general. No statistically significant differences between the various ratings for the inclination to give up were found (F(3,56) = 0.43, p = .73). See table 2.20 for an overview of the number of correct answers per rating for the inclination to give up.

Table 2.20: Correct answers per inclination to give up

		Total (N=60)
never considered		(n=22)
	Mean	7.41
	Standard deviation	2.68
	Range	2-12
1 time		(n=14)
	Mean	7
	Standard deviation	2.11
	Range	3-11
2 to 5 times		(n=22)
	Mean	6.86
	Standard deviation	2.53
	Range	3-12
more than 5 times		(n=2)
	Mean	8.50
	Standard deviation	0.71
	Range	8-9

Summary When analysing the rating for the inclination to give up no statistically significant differences were found between the experiment conditions. Furthermore, when assessing the relation between the rating for the inclination to give up and the number of correct answers, in general, no statistically significant differences were found.

2.4 Discussion

In this experiment I tested the effects of cues by measuring the amount of correct answers given by the participants on questions regarding the presentations they had studied. If cues would aid participants in comprehending the presentations then I would expect to see an in increase in the number of correct answers. The opposite might also be possible when, for example, cues unnecessarily split the attention of the reader between the text and illustration (Schnotz, 2005); drawing the reader's attention might not be necessary when readers have sufficient prior knowledge and the illustration is redundant to the study task.

When comparing the number of correct answers between the conditions no significant differences were found; regardless of whether participants were shown simple cues, explicit cues or no cues at all. Scores were very close and similar. On average the participants had just over half of the questions correct: 7 out of 12. Ratings given in the post-test results show similar results; no statistically significant differences were observed between the conditions. Furthermore, participants found a little over half of the presentations difficult and the language was rated moderately difficult. These average ratings from the post-test result confirm the average score of 7 correct answers.

Differences were found in relation with the participant's level of English and mother tongue. Analyses of the combined English levels Higher and Lower showed a difference in the number of correct answers in the no cues condition. Participants in the Higher level had more correct answers than those in the Lower level. Similar results were shown in the differences between mother tongues. Dutch participants had more answers correct than German participants in the no cues condition. These results suggest Dutch participants and those with a Higher level of English were negatively affected by cues, or German participants and those with a Lower level of English benefited from cues.

The results found in the Level of English and mother tongue seem to contradict the comparison of the conditions in general. The former suggest some local effects due to cues while the latter suggest no effects at all. To explore these results I will discuss both, starting with the possible reasons why no differences were observed between the conditions.

2.4.1 Reasons for no found effects of cues

Below I present four possible reasons why no differences between the conditions were observed. The first explores the idea that the cues were not used at all. The second continues on the first and discusses the influence of participants' motivation. The third that cues were used but no effects were observed. The fourth explores the idea that no effects were observed due to the presentations. The fifth that perhaps there were unmeasured effects.

Cues were not used

The first possible explanation is that the participants in the simple and explicit cue conditions did not use the cues. This would make it impossible for cues to have any effects and could explain the majority of the results. Lorch (1989) concludes that "for a signalling device to facilitate performance - the reader must note the relevance of the signal and attend to the cued information". Why would the participants ignore the cues? It could be due to reading style or a reading preference. For example, participants might have preferred to read the text first and afterwards observe the illustration or, they thought the text too short to interrupt reading and studied the illustration later or not at all. Some might have started with the illustration before reading the text.

From the results can be seen that the reading speed was quite high, reaching close to scanning speed in some cases. The participants were university students. University students are trained in abstract thinking, are used to absorb large volumes of literature, are used to read under the stress of time, and often read with the intention to study. Readers skilled in abstract thinking will likely need illustrations less than those with less skills. In addition, their reading behaviour might be oriented to absorbing as much knowledge as possible and go through the material as fast as possible. This behaviour might include the skipping of cues and illustrations and could explain the general outcome of this study. As a consequence, the relation between the textual content around the cue and the contents of the illustration might then not be effectively established.

Motivation

The second possible explanation is that motivation might have influenced reading behaviour. One of the post questions asked participants to answer how often they had considered to stop the experiment before reaching the end. I added the question to measure a basic level of discomfort due to the experiment design. For example, the number of presentations and the length of the text might put some participants off. Quite a large group considered stopping the experiment; only 37% never considered quitting, 23% considered it once and 37% 2 to 5 times. Thus, I assume the participants found the experiment unpleasant or uninteresting. The participants all received credits for participating. It is possible they were motivated to receive the credits but not to fully engage in a lengthy experiment. This could have changed their behaviour compared to the behaviour one would expect if they would have studied on their own initiative. Moreno and Mayer (2007) point out those learners with a lack of motivation may not integrate and organize the new information even when the cognitive capacity is available. Furthermore, McTigue (2009) has shown concern for motivation in her paper. According to Schnotz and Kürschner (2007) motivation can impact the amount of germane load. Hence, it impacts learning effort and the conscious processing of information.

Cues were used but no observed effects

The third possible explanation is that participants did follow up on the cues and studied the illustration right after reading the cue. In that case using the cues did not affect the participants. It might be that the attention given to the illustrations was equal for participants that used the cues and those that did not. Participants could have studied just as long and intense. They had the time to study the presentations again and the effect of cues might have dissipated. This was also observed in Tabbers et al. (2010); participants had an hour time to study at their own pace, while earlier researching in lab situations and at a controlled pace did show results. Reading speed in this experiment was faster than expected but it still remains possible participants had enough time. This might have lowered extraneous load.

Alternatively, the participants may have thought the illustrative content not to be as important as the textual content. If the textual content was more important for memory recall then studying the illustrations was not strictly necessary. Participants might then still have used the cues and perhaps studied the illustrations more intensely but any effect from doing so was negligible or has not been measured by the experiment. This could indicate a low additional value of the illustrations.

Schnotz and Kürschner (2007) explain in their paper that a highly coherent illustrated text can offer too much help to readers with high expertise. Prior knowledge affects the perception of difficulty and hence the intrinsic load. A highly coherent presentation would be too easy for skilled learners and would not challenge their capacities. If the intrinsic and extraneous load were low then the cues might have been redundant. Low intrinsic load might have been caused by high prior knowledge, a too low complexity and a low amount of presented information. Low extraneous load might have been caused by the good organization of the presentations.

A variation in the experiment, and a suggestion for follow up studies, could be to create two additional conditions where in one the participants are only shown the text and in the other only the illustration. The amount of knowledge retrieved and recalled from the text and illustration could then be measured. It would give an indication of the relevance of the added illustration and text. A third condition, only consisting of questions and deprived of text and illustrations, could measure the impact of prior knowledge.

No effect due to setting

The fourth explanation is that cues had no effect due to the setting in which they were used. The text in the presentations was not long compared to an article or a book. The text length per presentation was on average 350 words; not more than half a page. Compared to studying a book, there is much less information to remember and it might have been possible to recall most of the textual content. Overall the cognitive load might have been too low. However, the average number of correct answers does not support this thought: 42% of the questions were wrongly answered. Having a little less than half of the questions wrong suggest that not everything was recalled. Mautone and Mayer (2001) note that cues might be more effective in complex texts that contain many topics. During experiment design complex topics were included; however, the text length might have been too short. When a reader is confronted with a large amount of information he might be enticed to follow up on a cue, hoping to gain a better or more complete understanding.

Another factor which might have influenced the reading style is the used medium for showing the presentations. The participants might have scanned the presentations and used a different reading technique. Research by Liu (2005) and Morkes and Nielsen (1997) suggests that readers scan text when presented on modern media. Perhaps, in more traditional media the reader takes more time and attention to read the material and use the cues more actively. A hurried style to scan an illustrated-text might cause cues to be overseen, ignored or illustrations to be only skimmed. The effects of cues would then be reduced significantly.

Not measured effects and design

A fifth explanation is that there were effects but they were not measured in the experiment. In this experiment I explicitly looked at knowledge that was both present in the text and the illustration. It remains unclear if participants who were presented the cues might have been able to recall more information from the illustrations than those who were not presented the cues. The suggestions I made above to alter the experiment with additional conditions could have aided in answering this question.

More results could perhaps have been gotten from questions that test the transfer of knowledge rather than recall (Mautone & Mayer, 2001). Mautone and Mayer note that the reader can recall a few phrases from a text but if transfer happened they could apply the newly gained knowledge to, for example, solve new problems. Alternatively, participants could explain or answer question in their own words rather than choosing from multiple choice answers (Boucheix & Guignard, 2005). These methods can show if a better integrated learning took place.

2.4.2 Influence of reported English level and mother tongue

Analyses of the combined English levels Higher and Lower showed a difference in the number of correct answers in the no cues condition. Participants in the Higher level (M=7.93,SD=1.49) had more correct answers than those in the Lower level (M=6.20,SD=1.79). The differences within the simple cues and explicit cues conditions were not statistically significant. Thus, the scores of the Higher and Lower groups got closer when confronted with cues. One or both groups were influenced and caused the scores to get closer. Unfortunately, the differences within these groups were not large enough to exactly determine which of the two groups was influenced; it is not possible to conclude if one groups was hindered or the other benefited.

In the no cues condition similar results can be observed for the mother tongue. Dutch participants (M=8.30,-SD=1.34) had statistically significant more answers correct than German participants (M=6.56,SD=1.67). Again, as seen in the English level tests, the differences in the simple cues and explicit cues conditions were not statistically significant. I am faced with the same dilemma. Either one or both group were influenced but it is not possible to determine which.

Mother tongue or level of English?

The difference between the Dutch and German participants could perhaps be explained by their reported level of English. A test on the relation between mother tongue and the reported level of English did not confirm this and showed no statistically significant differences. However, the reported English level is a subjective self-reflection by the participants who compare themselves with their peers. If there is a difference in how Dutch and Germans perceive their levels of English then they could be of different levels but subjectively still rate the same. The observed differences in the no cues condition could thus be due to a difference in the level of English. Are there reasons to believe Germans would be less skilled in English than Dutch people? Surveys completed by the European commission show differences between Germans and Dutch participants that could explain a difference in the English level. In a survey completed among European citizens, 56% of the German and 87% of the Dutch participants reported English as a language they spoke well enough to be able to have a conversation (Eurobarometer, 2006). In another survey participants were asked about their use of a language, other than respondents' own, to read or watch content on the Internet. 65% of the German and 72% of the Dutch participants reported English as one of the languages. The same question was asked about writing on the Internet. 32% of the German and 43% of the Dutch participants reported English as one of the languages (Eurobarometer, 2011). Thus, it seems the English language is more known and used among Dutch participants than German participants. This is also evident from movie entertainment shown in both countries; in the Netherlands most movies are shown in their original language while in Germany foreign voices are often replaced by German speaking voice actors.

I expect that readers with a good comprehension of the English language find it easier to study and understand a textual topic written in English. Those with higher English skills might scan through the text with confidence while those with a lower skill might need more concentration and find it more difficult to grasp the complete subject. McTigue (2009) in her review reports that young students with lower skills for the interpretation of illustrations have more difficulty extracting information. Hannus and Hyönä (1999) report less knowledge transfer effect for children with a low learning ability compared to high-ability children. The same could be true for text. This could explain the difference seen in the no cues condition. Schnotz (2005) explains that for poor readers illustrated content becomes more important than textual content. Schnotz predicts that poor readers will benefit more from illustrative content than good readers. In the cued conditions the lower skill participants might have been aided by cues and they can have used the images to gain a better understanding of the topic. Mayer, Steinhoff, et al. (1995) showed that readers with low experience, low prior knowledge, benefited from placing annotations while high experience readers saw no effects. Fletcher and Tobias (2005) in their review report the same for the adding of illustrations to text. Houts, Doak, Doak, and Loscalzo (2006) in their review of illustration in medical text conclude that adding illustrations is especially helpful for patients with low literacy skills. Those with a higher skill might have been distracted by the cues; the cues disrupted their usual reading methods causing more cognitive strain. Meanwhile their reading speed might not have been adjusted and the changing between text and illustration could have been less effective than first reading the complete text and then looking at the illustration. Earlier research shows that instructional support may interfere with learning (Seufert & Brunken, 2004; Fletcher & Tobias, 2005). Seufert and Brunken (2004), Schnotz (2005) report that, for those with high prior knowledge, added information which is unnecessary requires extra mental effort for visual searching between representations and extra cognitive capacity to assess the added value of new information. The findings in the above mentioned literature are not directly relevant to a difference in language skill but all mention a difference in skill and ability that causes differences between the groups.

Chapter 3

Reading observation experiment

In this chapter I describe an experiment where I used an eye-tracker to observe whether participants studied an illustration after reading a textual cue.

This experiment was set up like the recall performance experiment with the following changes:

- A small group of participants was used,
- The experiment was held in a laboratory under supervision of the experimenter,
- The material was not displayed on a website but instead screen shots of the website were shown,
- The participants answered the questions on paper instead of on the website,
- The participants' gaze was tracked using an eye-tracker.

Prior to the experiment a few open questions were considered: i) in which order will the participants study the presentation content? E.g. will they read the text first, or look at the illustration? Or they will switch attention between text and illustration? ii) will cues have any effect at all? Thus, I observed the participants' scan paths and the fixation points. A scan path shows the order in which participants studied the text and the illustration. The fixation points show when participants process information and where they focus their attention.

Eye-tracking

Eye-tracking experiment provides a path of participants' attention and can show the amount of processing being applied (Jacob & Karn, 2003). Its use and application has been summerized and explained by various researchers (Rayner, 1998; Jacob & Karn, 2003; Poole & Ball, 2005). There are various measurements which can be taken with an eye-tracker and those used in this experiment will be explained below. The eye-tracker records *gaze points* where the participant looked at for a certain time, and how long a particular gaze point was observed. A *fixation* is the time when the eye is relatively stationary on a gaze point. How long this lasts can vary with the task at hand. It is believed that during a fixation people are processing information. Poole and Ball (2005) explain that fixations can be interpreted differently depending on the context. A higher number of fixations on a single area can mean that the participant is interested. However, it can also mean they are confused or that the area is difficult to understand. In search tasks a higher number of fixations can mean uncertainty in recognizing a target item. The duration of a fixation is linked to a higher processing time. Poole and Ball (2005) advise to set a *fixation* threshold of at least 100ms. Jacob and Karn (2003) set the length between 100 and 200ms and Rayner (1998) between 200 and 300ms. *Saccades* are quick eye movements occurring between fixations. No information is obtained during a saccade because the eyes are moving too fast. The sequence of gaze points creates a trail which is called the *scan path*.

3.1 Method

3.1.1 Participants

The participants were recruited from the same students pool as mentioned in the recall performance experiment. The same criteria were set as in the recall performance experiment, see section 2.1.1. No constraints were set on the wearing of eye glasses or eye lenses. The documentation of the eye-tracker equipment stated this would not influence the recording of data.

3.1.2 Materials

The used materials were the same as in the recall performance experiment; however, the website from the recall performance experiment could not be used due to limitations of the eye-tracker recording software. Instead, screen shots of the website were presented. Paper forms were used to record the personal details and to answer the questions.

3.1.3 Equipment

The used eye-tracker was a Seeing Machines Facelab 4.5., which uses a non-invasive recording method. Cameras were used to track the position of the face and infra-red corneal reflection was used to track the eyes. The participants were placed about 0.4 meters away from the screen, which was placed on a desk.

To analyse the recorded gaze points, I used OGAMA version 4.0 beta 2.(Open Gaze And Mouse Analyzer). OGAMA is an open source program for recording and analysing eye-tracking data (Voßkühler, Nordmeier, Kuchinke, & Jacobs, 2008; Voßkühler, 2011).

3.1.4 Procedure

The participants were recruited similarly to the recall performance experiment. Once welcomed in the laboratory, the participants were explained the task they needed to perform. They were explained that the images they would see would be screen shots. A single click anywhere on the screen would show them the next page. They were given paper forms to answer all the questions. Once seated, the participants were told to take a comfortable position in which they could remain for one hour; however, small movements were allowed. Then the eye-tracker was calibrated on the participants face and eyes and the experiment was started. The remainder of the procedure continued as described in the recall performance experiment. Occasionally, participants needed to be reminded to sit straight in their seat so the eye-tracking system could follow their eyes.

3.2 Processing of the data

3.2.1 Precision

The eye-tracker equipment is quite sensitive and the results showed some issues. The first issue was the accuracy, which was not high enough to determine the exact sentence a participant was reading. The second issue was shifted gaze points, which could only be noticed only during the analysis of data, after the experiment was conducted. When gaze points are shifted, their location does not match the location of text paragraphs and of objects in the illustration. This can be seen by the analysing the position of the scan paths relative to the text paragraphs and objects in the illustration. The data shift probably happens due to calibration problems and/or the moving of the participant's head while reading the presentation. When the participant moves the head, the viewing angle changes and this would need to be corrected by the eye-tracker software. However, this did not seem to have happened consistently.

Figure 3.1 shows an example of gaze point shifting. The shape of the scan path has the shape of the paragraphs and matches the objects in the illustration. The distortion of the gaze point location became worse as the participant looked lower in the presentation. In most cases, one can estimate which pieces of text and illustrated objects were observed from the shape of the scan path. This means that the location at which the participant switched from the text to the illustration can at best be estimated within the range of a few text lines.

3.2.2 Discarded data

Data from participant 4 for presentation 8 (The thermohaline circulation) was too distorted for analyses and was discarded.

3.2.3 Counting attention switches

For each presentation, the number of times a participant switched from the text to the illustration (fixations) were counted; saccades over the illustration without eye fixation were not counted. The number of times that a switch was likely triggered by a cue, were counted separately. A switch was regarded to be triggered by a cue when:

- The participant's gaze has just passed the location of a cue,
- The participant looked at objects in the illustration that were directly related to the content of the cue.

Figure 3.2 shows an example of eye fixations when the participant switched from the text to the illustration. See table 3.2 for the results of the counting for different conditions.

Figure 3.1: This illustration shows an example of gaze point shifting. In this presentation, participant 1 was reading "the working of the fluorescence microscope". The blue lines represent the gaze path and the blue dots indicate the fixation points. In order to to visualize the gaze point shift, the paragraphs and the illustrated objects have been boxed in with a red line to show the position and shape of the read text and observed objects in the illustration; the gaze path has been boxed in with a yellow dotted line. At the top of the presentation, the shape and size of the yellow boxed gaze path corresponds to the red boxed text paragraphs and their their relative position is only slightly shifted. However, further down the slide, the shape of the gaze path is vertically stretched (yellow dotted line) and its position is dramatically shifted compared to the read paragraph (red box).

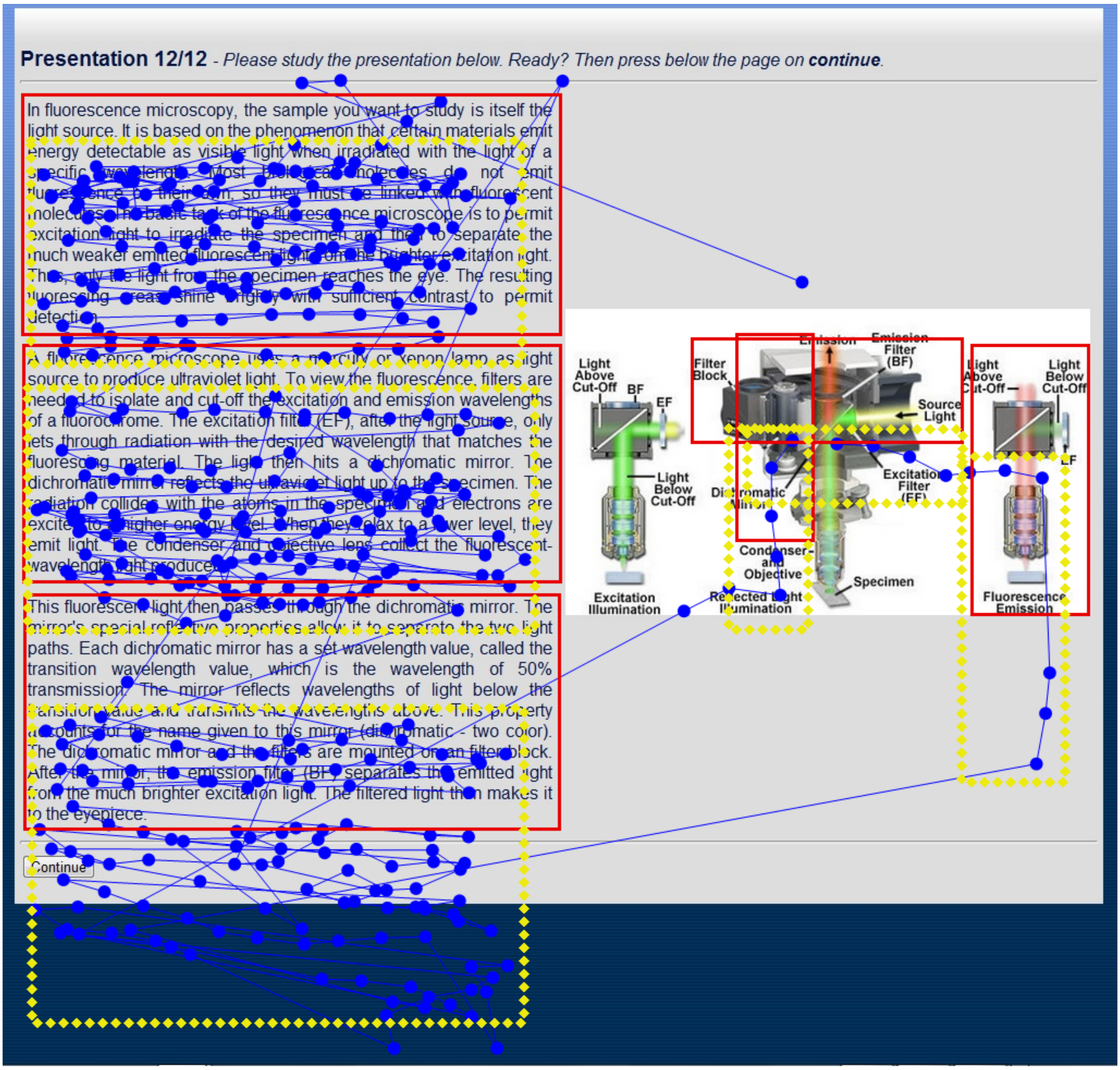
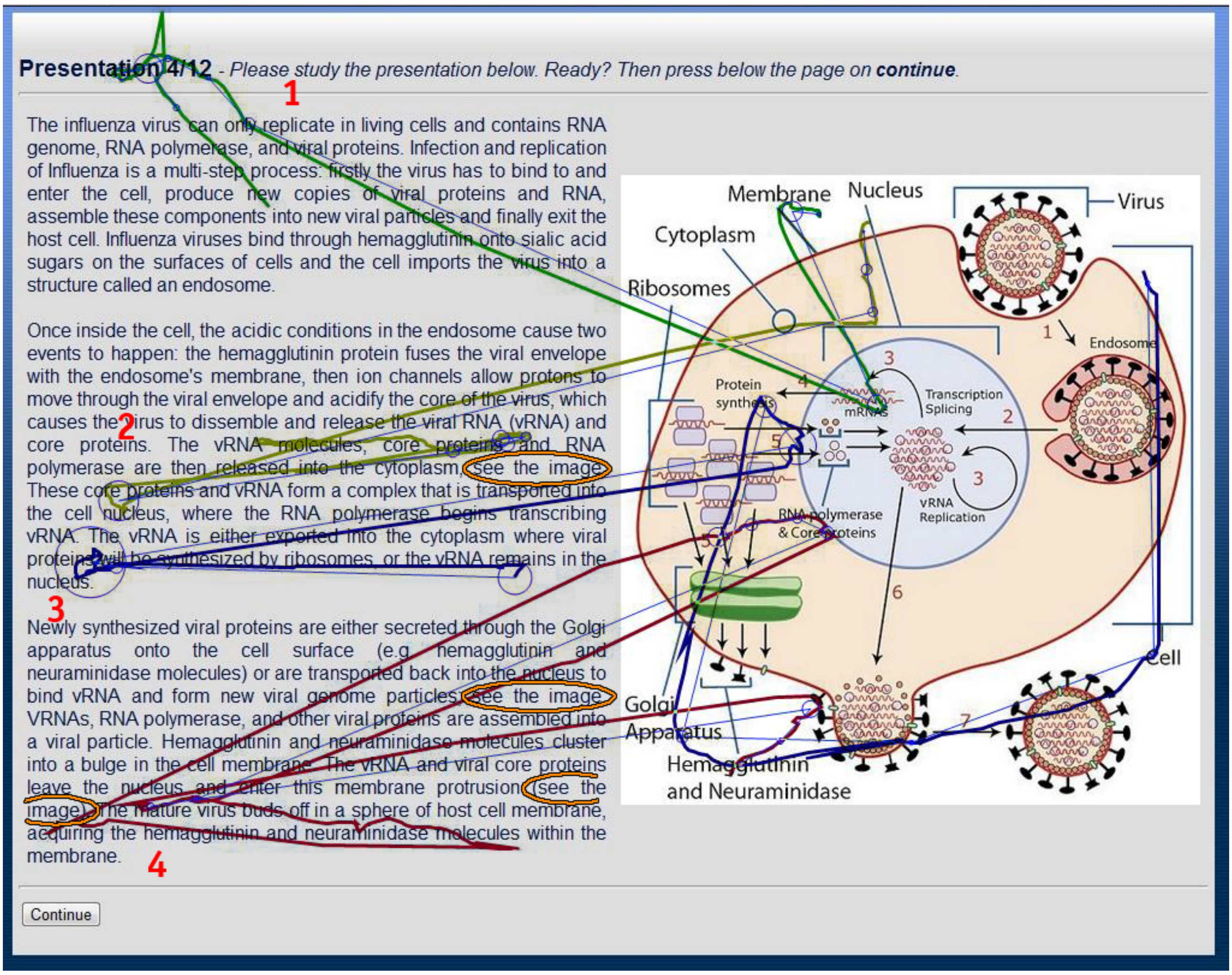


Figure 3.2: The illustration shows the switches (fixations) from text to illustration. In this presentation, participant 2 was reading the “replication of the influenza virus”. Switch 1 took place before reading the text. Switch 2 took place near the first cue ‘see the image’. The second cue ‘see the image’ is ignored. Switch 3 took place after reading paragraph 2. Switch 4 took place near the cue ‘(see the image)’.



3.3 Results

In this section I present the results from the reading observation experiment using the eye-tracker. First an overview of the participants is shown and then the results from the analysis are presented.

3.3.1 Participants

Table 3.1 shows an overview of the participants and their gender, age, the experiment condition, English level and Mother tongue.

Table 3.1: Participants overview

	Gender	Age	Condition	English level	Mother tongue
Participant 1	Male	21	No cues	Proficient/Excellent	Dutch
Participant 2	Male	21	Simple cues	Advanced	Dutch
Participant 3	Female	20	Simple cues	Advanced	German
Participant 4	Male	20	Explicit cues	Advanced	Dutch
Participant 5	Male	25	Explicit cues	Proficient/Excellent	Dutch

3.3.2 Attention switching to the illustration

In this section I will present for each participant, the number of times they have looked at the illustration. Each partcipant was shown 12 presentations.

Participant 1 had no cues in the presentations. The number of times the participant switched from the text to illustration was very low (average 1.2 per presentation). See table 3.2 for the average number of switches and standard deviation. For most presentations this participant observed the illustration after reading all text. Participant 1 had 8 out of 12 correct answers to the presentation questions.

Participant 2 had simple cues in the presentations and in total he switched attention to the illustration more often than participant 1 (average 4.2 per presentation). This participant followed on average between 1 or 2 cues per presentation. He always studied the illustration before reading the text. However, he also switched attention to the illustration nearly 3 times per presentation, without using cues. Participant 2 had 11 out of 12 correct answers to the presentation questions.

Participant 3 had simple cues in the presentations. In general, this participant had on average even more switches than participant 2 (average 6.4 per presentation). The number of switches related to cues is a little over 1 per presentation. This participant switched about 5 times to the illustration without the presences of cues and often looked at the illustration after reading a first few lines. Participant 3 had 7 out of 12 correct answers to the presentation questions.

Participant 4 had explicit cues in the presentations. In general, this participant switched attention to the illustration about once per presentation (average 1.2), which is similar to the reading behaviour of participant 1. However, his study behaviour was different. Participant 4 usually switched from text to the illustration while reading instead of before or after reading the text. In about half of the 12 presentations the participant looked at the illustration after reading a cue and half of the presentations without reading a cue. Participant 4 had 5 out of 12 correct answers to the presentation questions.

Participant 5 had explicit cues in the presentations and switched often to looked at the illustrations (average 10.3). In most cases he studied the illustration before reading the text or read the first line and then look at the illustration. This participant had the highest amount of switches after reading a cue which was on average almost

3 times per presentation and almost 8 switches which were not cued. This participant was highly engaged in relating contents from the text and illustration. Participant 5 had 12 out of 12 correct answers to the presentation questions.

Table 3.2: Attention switches to illustration and the number of correct answers per per participant

		Not cued switch		Cued switch		Total switches		Correct answers
		Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	
No cues								
	Participant 1	1.2	0.4	-	-	1.2	0.4	8
Simple cues								
	Participant 2	2.7	1.4	1.5	0.8	4.2	1.9	11
	Participant 3	5.1	2.5	1.3	1.1	6.4	2.8	7
Explicit cues								
	Participant 4	0.6	0.5	0.6	0.8	1.2	0.6	5
	Participant 5	7.7	3.6	2.7	0.8	10.3	3.8	12

No trend is visible in regard to the number of correct answers and the number of attention switches. In any case, the number of participants in this experiment was too low to draw solid conclusions on the relation between the two.

Not all participants followed the same cues or switched to look at the illustration at the same time. Figures 3.3 to 3.7 show the participants reading a part of the Tuareg presentation. Participant 1 had no cues in the presentations and does not switch to look at the illustration. Participants 2 and 3 had simple cues in the presentations. Participant 2 ignored the cue '*(see the map)*' and kept on reading while participant 3 switched to look at the illustration. Participants 4 and 5 had explicit cues in the presentations. Participant 4 ignored the cue '*Notice in the picture how the Tuareg territory is spread over the new nations*' while participant 5 switched to look at the illustration.

Figure 3.3: A partial scan path showing Participant 1 (no cues) reading the Tuareg presentation

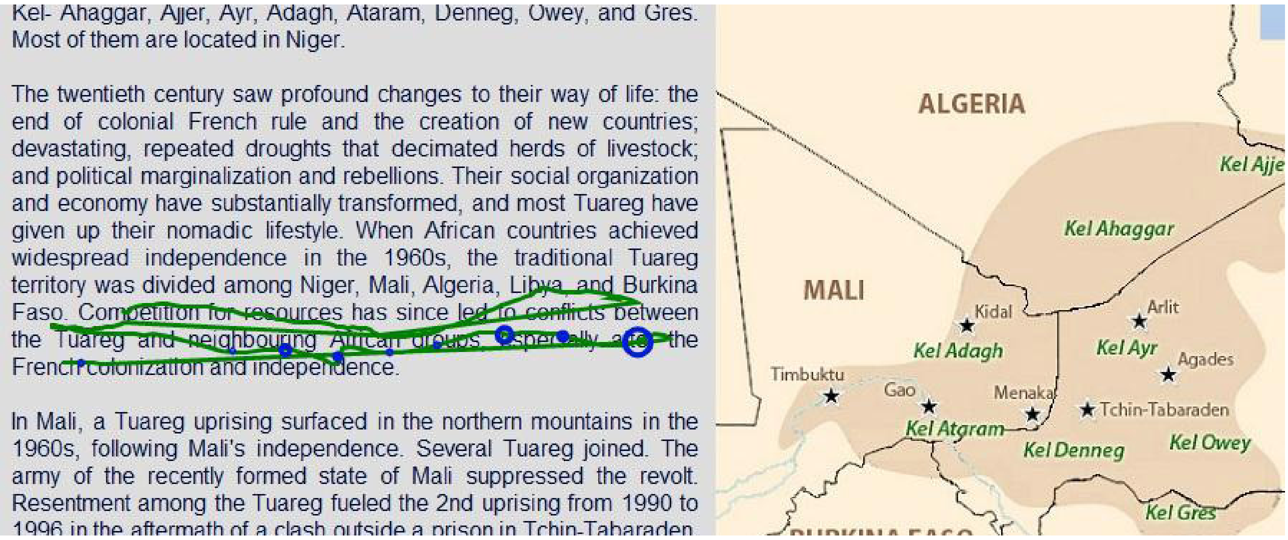


Figure 3.4: A partial scan path showing how participant 2 (simple cues) ignores the cue '(see the map)' in the Tuareg presentation

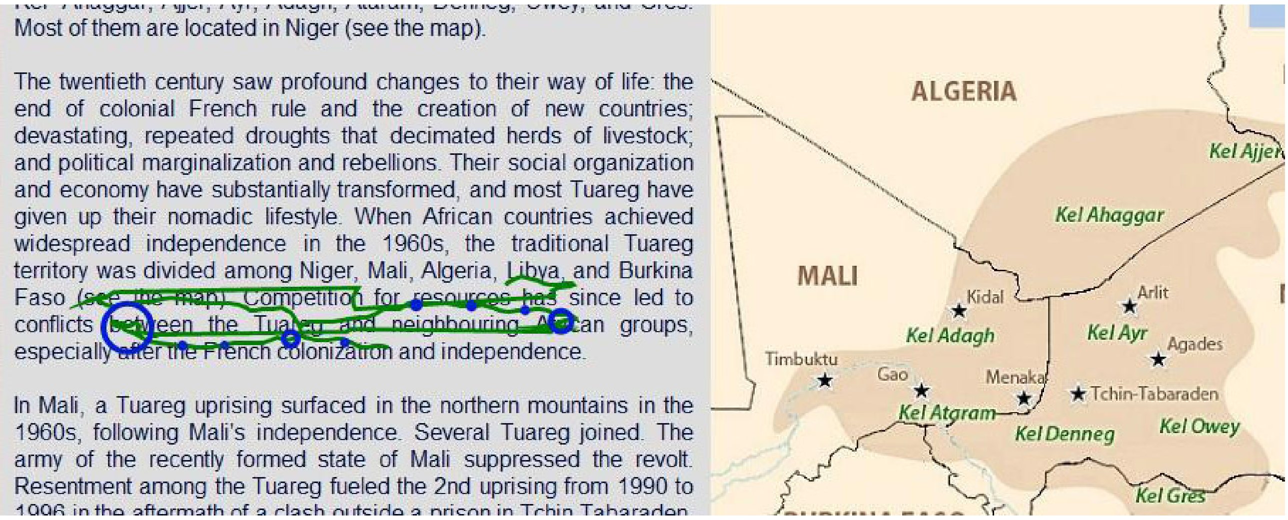


Figure 3.5: A partial scan path showing how participant 3 (simple cues) follows the cue '(see the map)' in the Tuareg presentation

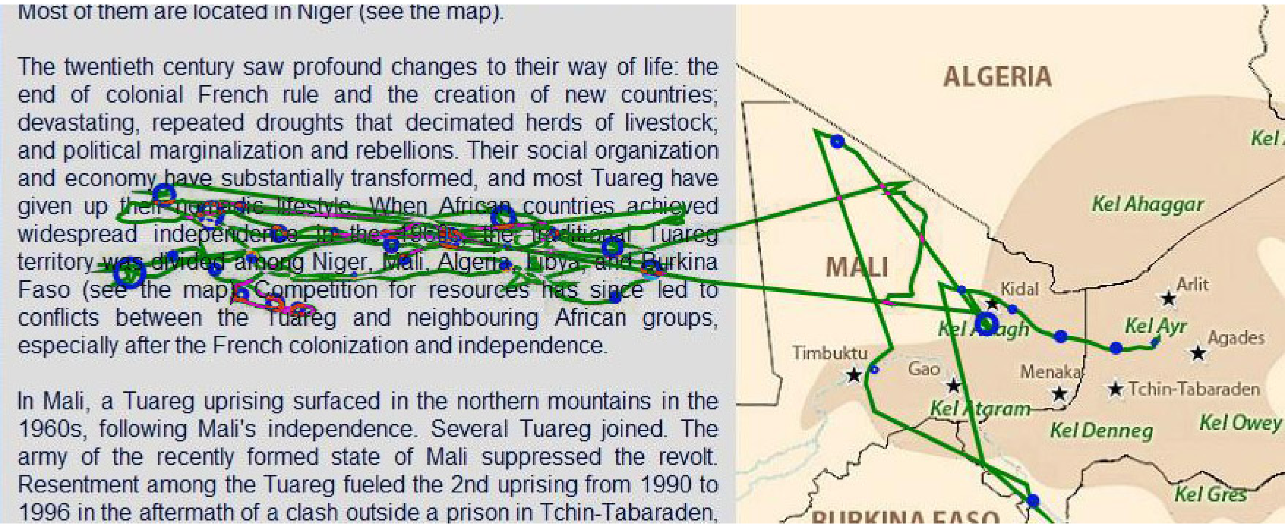


Figure 3.6: A partial scan path showing how participant 4 (explicit cues) ignores the cue 'Notice in the picture how the Tuareg territory is spread over the new nations' in the Tuareg presentation

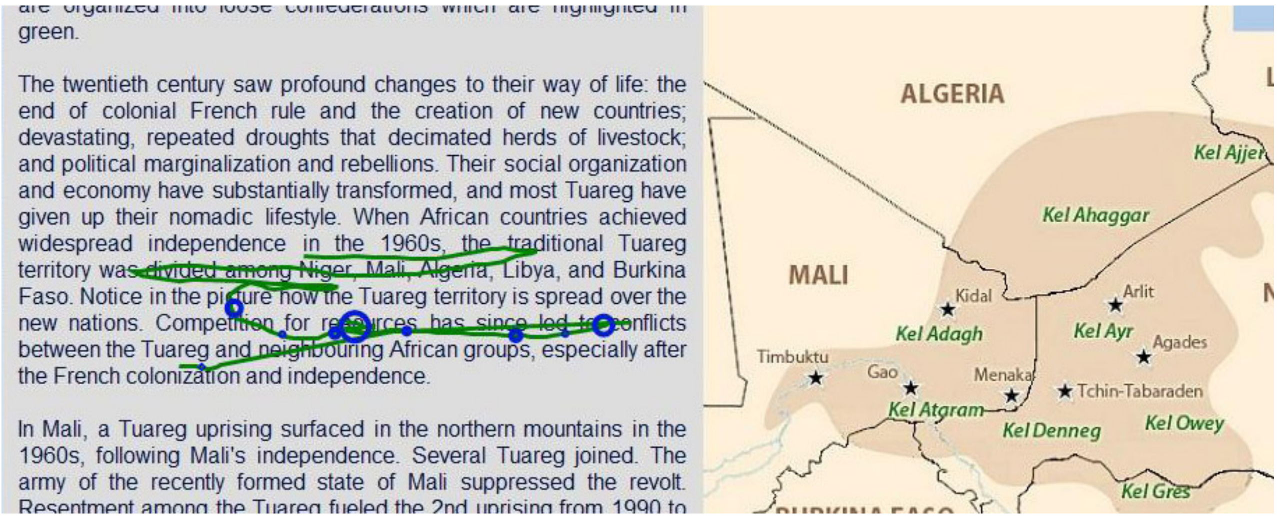
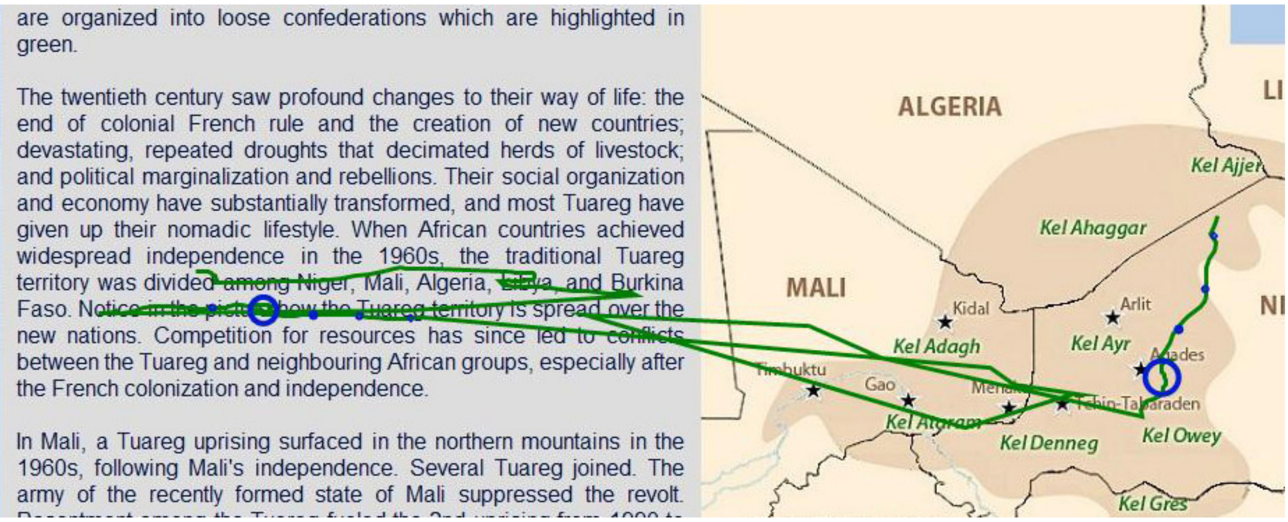


Figure 3.7: A partial scan path showing how participant 5 (explicit cues) follows the cue 'Notice in the picture how the Tuareg territory is spread over the new nations' in the Tuareg presentation



3.4 Discussion

The number of participants in this experiment is too low to draw solid conclusions. Also, the reliability of the data limits the interpretation of these results. Among these participants, it seems that when cues are present the participants tend to use some of them. Although inconsistently, textual cues do seem to draw the reader's attention to watch the illustration.

The recall performance experiment does not show clear benefits from cues. In light of the reading observation experiment, this could be explained by the higher number of attention switches without the use of cues compared to the number of attention switches with the use of cues (in 3 out of 4 participants in the cue experiment conditions). Participants switched attention often without the stimulation of a cue and thus, the effect of the cued switches might have become negligible. This might also explain why no differences were observed between the no cues condition and simple and explicit cues conditions in the recall performance experiment: the participants might have switched attention without the aid of cues leading to a reduced effect of cues on the recall performance. What remains unclear from the reading observation experiment and still remains possible, is that cues might stimulate participants to study the illustration better.

Different reading styles could be observed among the 5 participants. Some looked at illustration only a few times while others often switched between the text and illustration. Furthermore, there were differences in when participants looked at the illustration for the first time; one participant had a habit to usually study the illustration after reading the text, one while reading the text, and three others usually did it before reading the text or after reading the first few lines. These differences in the moment of attention switch could potentially influence the learning process. For example: looking at the illustration before reading the text might make it easier to relate the contents of both representations. However, during the first glance, the reader might not entirely understand what he is looking at. Thus, with the knowledge of the textual content, studying the illustration after reading might be effective as well. Alternatively, switching while reading might be the best method assuring that the content of both representations is in working memory at the same time. However, the switches might cause a loss of focus and be less effective than reading the text in one go. Considering the variability seen within this small experiment, future research on this reading behaviour could aid text design and the process of learning from illustrated text.

Chapter 4

Conclusion and Recommendations

In this work I have described the effect of simple and explicit textual cues on the recall of information presented in the form of illustrated text presentations and the effect of simple and explicit textual cues on reading behaviour. Sixty students participated in the recall performance experiment: 19 in the no cue condition, 20 in the simple cue condition and 21 in the explicit cue condition. Five students participated in the reading observation experiment. In both experiments students were given 12 illustrated text presentations to study. In the recall performance experiment, the number of correct answers given to questions regarding the presentations was used as a measure of their ability to recall information. In the reading observation experiment, students' eye-movement was recorded to observe whether cues would cause participants to study the illustration.

When comparing the number of correct answers between conditions in the recall performance experiment no significant differences were found. Further analysis of the participants' level of English indicated that English skills may influence how well the information has been recalled. Although the current results are not conclusive enough to establish with certainty that cues had an effect on recall, prior research has demonstrated the influence of participants' skill on comprehension. In light of these studies, I will discuss the possible explanations for the current results and make suggestions for future research.

i) Lack of complexity. The experiment presentations might have lacked complexity to challenge the participants, due to several reasons: a) Prior knowledge could have helped the readers to easily select and organize the content without the use of cues. b) The presentations might have been very well organized (easy to read and clear illustrations), thus lowering the complexity of the material. c) The presentations might have been too short and therefore easy to recall by the readers. These 3 factors have been shown to reduce intrinsic and extraneous load and the motivation to use cues might be higher when complexity demands it. Previous research has shown that participants with lower prior knowledge benefited from the presence of cues, whereas participants with higher prior knowledge did not or were even impaired.

Recommendation 1. For future research of textual cues, I recommend matching the complexity of the presentation/ topic to the participants' prior knowledge, by increasing the intrinsic load of the material. Extraneous load should normally be decreased. However, it could be artificially increased to test cues in conditions of high cognitive load. Extraneous load can be increased either by disorganizing the structure of the presentations and/or by adding extraneous material. Furthermore, separately measuring the amount of knowledge learned from text and illustration can show the added value of each representation.

ii) Reading behaviour and motivation. a) Prior research suggests that reading behaviour from modern media, such as internet web sites, is less concentrated than from more traditional media, such as books; web readers scan, search for keywords, and like concise pages. The participants might have skipped the cues while scanning the presentations. b) Participants' motivation in the recall performance experiment might have influenced reading behaviour in a similar way, e.g. readers with a low motivation might skip cues or skim over an illustration. Participants from the reading observation experiment completed their experiment in a laboratory under the supervision of the experimenter. This might have motivated them to study the presentations better. Considering that these participants only read about half of the cues in the presentations, participants in the recall performance experiment might have used the cues even less. c) In the reading observation experiment, 3 out of the 4 participants that studied the presentations with cues, switched attention to the illustration more often without the aid of cues than with the aid of cues. Assuming this behaviour also took place in the recall performance experiment, the impact of the cues in the presentations might have been negligible. The participants' own tendency to switch attention might be more important than cue-aided switching.

Recommendation 2. For future textual cue research, I recommend taking into account the participants reading behaviour when setting up an experiment. Laboratory controlled experiments might increase peer pressure to read presentations with attention, whereas in uncontrolled experiment settings motivation/attention cannot be controlled. In non-laboratory experiments, instructions or assignments could be provided to stimulate learning.

iii) Reading skills. a) Participants with higher reading skills and prior knowledge might have found cues redundant: their experience might have been high enough to select and organize the concepts in the presentations without using the cues. In addition, prior research suggests that coherence formation aids can impair learners with high prior knowledge. This negative effect might be generalized to highly skilled readers. b) English language skills might have influenced the results. Participants with lower English language skills might not understand new words, or abstract concepts. Consequently, they might have used the illustrations, which show a concrete example of the discussed concept, to guess the meaning of an unknown word or a sentence. Therefore, adding the cues might have helped those with lower English language skills to recognize related concepts in the text and illustration and reduce their extraneous load. These effects might be generalized to other cases of impaired skills.

Recommendation 3. Future textual cue research should examine the effect of reading skills on the use of cues. So far it remains unclear whether highly skilled readers use cues and why readers with lower or impaired skills might benefit from cues. The effect of cues on high/low skilled readers should be investigated as well.

iv) Insufficient or inadequate experiment design. While recall might be unaffected by cues, deep brain processing of the material can still happen. Deep processing should lead to a better understanding of the topic but not necessarily to improved recall of the major information. This has not been tested in this experiment. Other measures include: knowledge transfer tests, verbal explanations, subjective ratings of learning difficulty and delayed post tests. These measures might offer a broader overview of the absorbed and used information.

Recommendation 4. For future research of textual cues, I recommend including a broad set of knowledge tests to cover the various ways in which knowledge is transferred and stored.

Eye-tracking experiments used to analyse reading behaviour could aid future research in understanding how cues impact the reading of illustrated text. Preliminary results presented here suggest that cues do cause readers to study the illustrations. However, readers also switched attention to the illustration without the presence of a cue. These results should be interpreted with caution due to the low number of participants.

In summary, I recommend that both the impact of cues on reading behaviour and the impact of skills (such as prior knowledge and reading skill) on the recall of information, should be further investigated in text design research. Learners read and experience illustrated text in different ways, depending on their prior overall knowledge. Hence, illustrated text and cues should be designed and adjusted to the knowledge and skill levels of the participants. Future research should examine in further detail why people with certain levels of skills require more or less aids. Last, future research should measure the amount of learned information by using a variety of memory tests in combination with eye tracking to observe the influence of reading behaviour on knowledge transfer.

Bibliography

André, E. (2000). The generation of multimedia presentations. In R. Dale, H. Somers & H. Moisl (Eds.), *Handbook of natural language processing* (pp. 305–327). USA: Marcel Dekker, Inc.

André, E., Müller, J., & Rist, T. (1996). WIP/PPP: Knowledge-Based methods for automated multimedia authoring. In *In proceedings of EUROMEDIA96* (pp. 95–102). PUB-SCS. doi:10.1.1.54.7718

Androutsopoulos, I., Oberlander, J., & Karkaletsis, V. (2007). Source authoring for multilingual generation of personalised object descriptions. *Natural Language Engineering*, *13*(3), 191–233. doi:10.1017/S1351324906004268

Boucheix, J., & Guignard, H. (2005). What animated illustrations conditions can improve technical document comprehension in young students? format, signaling and control of the presentation. *European Journal of Psychology of Education*, *20*(4), 369–388. doi:10.1007/BF03173563

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students learning from text. *Educational Psychology*, *14*(1), 5–26. doi:10.1023/A:1013176309260

Deemter, K. v., & Power, R. (2002). High-Level authoring of illustrated documents. *NATURAL LANGUAGE ENGINEERING*, *9*, 2003. doi:10.1017/S1351324902003054

Duggan, G. B., & Payne, S. J. (2009). Text skimming: the process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, *15*(3), 228–242. doi:10.1037/a0016995

Duggan, G. B., & Payne, S. J. (2011). Skim reading by satisficing: evidence from eye tracking. In *Proceedings of the 2011 annual conference on human factors in computing systems* (p. 1141 to 1150). CHI '11. New York, NY, USA: ACM. doi:10.1145/1978942.1979114

Dyson, M. C., & Haselgrove, M. (2000, June). The effects of reading speed and reading patterns on the understanding of text read from screen. *Journal of Research in Reading*, *23*(2), 210. doi:10.1111/1467-9817.00115

Dyson, M. C., & Haselgrove, M. (2001, April). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, *54*, 585–612. doi:10.1006/ijhc.2001.0458

Eurobarometer. (2006). *Europeans and their languages* (tech. rep. No. 243/ Wave 64.3). European Commission. EU. Retrieved from

Eurobarometer. (2011). *User language preferences online* (tech. rep. No. 313). European Commission. EU. Retrieved from

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (first, pp. 117–133). New York: Cambridge University Press.

Folker, S., Ritter, H., & Sichelschmidt, L. (2005). Processing and integrating multimodal material: the influence of Color-Coding. In *Proceedings of the 27th annual conference of the cognitive science society* (690–695). Mahwah, NJ: Lawrece Erlbaum.

Hannus, M., & Hyönä, J. (1999, April). Utilization of illustrations during learning of science textbook passages among low- and High-Ability children. *Contemporary Educational Psychology*, *24*(2), 95–123. doi:10.1006/ceps.1998.0987

Hewitt, J., Brett, C., & Peters, V. (2007, November). Scan rate: a new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, *21*(4), 215–231. doi:10.1080/08923640701595373

Hooijdonk, C. v., Bosma, W., Krahmer, E., Maes, A., & Theune, M. (2011). Experiments in multimodal information presentation. In A. Bosch & G. Bouma (Eds.), *Interactive multi-modal Question-Answering* (pp. 89–115). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from

Hooijdonk, C. v., Vos, J. d., Krahmer, E., Maes, A., Theune, M., & Bosma, W. (2007, October). On the role of visuals in multimodal answers to medical questions. In *Proceedings of the 2007 conference of the IEEE professional communication society* (pp. 1–15). USA: IEEE Computer Society Press. doi:10.1109/IPCC.2007.4464073

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006, May). The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, *61*(2), 173–190. doi:10.1016/j.pec.2005.05.004

Jacob, R. J. K., & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: ready to deliver the promises. *Work*, *2*(3), 573–605.

Jamet, E., Gavota, M., & Quaireau, C. (2008, April). Attention guiding in multimedia learning. *Learning and Instruction*, *18*(2), 135–145. doi:10.1016/j.learninstruc.2007.01.011

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, *17*(3), 329–345. doi:10.1080/0144341970170307

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*(4), 351–371. doi:10.1002/(SICI)1099-0720(199908)13:4%3C351::AID-ACP589%3E3.0.CO;2-6

Koning de, B., Tabbers, H., Rikers, R., & Paas, F. (2009, June). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educational Psychology Review*, *21*(2), 113–140. doi:10.1007/s10648-009-9098-7

Liu, Z. (2005). Reading behavior in the digital environment: changes in reading behavior over the past ten years. *Journal of Documentation*, *61*, 700–712. doi:10.1108/00220410510632040

Lorch, R. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, *1*(3), 209–234. doi:10.1007/BF01320135

Masson, M. E. (1982). Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 400–417. doi:10.1037/0278-7393.8.5.400

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*(2), 377–389. doi:10.1037/0022-0663.93.2.377

Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (first, pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2005b). Principles of reducing extraneous procressing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (first, pp. 31–48). New York: Cambridge University Press.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied*, *11*(4), 256–265. doi:10.1037/1076-898X.11.4.256

Mayer, R. E., & Ross, B. H. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. Volume 41, pp. 85–139). Academic Press. Retrieved from

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*(1), 31–41. doi:10.1007/BF02300480

McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology*, *34*(2), 143 –153. doi:DOI:10.1016/j.cedpsych.2008.12.003

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. doi:10.1007/s10648-007-9047-2

Morkes, J., & Nielsen, J. (1997). Concise, SCANNABLE, and objective: how to write for the web. Retrieved from URL: http://www.useit.com/papers/webwriting/writing.html

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour & Information Technology*, *10*(4). Retrieved from

Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction*, *3*(3), 227–238. doi:10.1016/0959-4752(93)90006-L

Poole, A., & Ball, L. J. (2005). Eye tracking in Human-Computer interaction and usability research: current status and future. In *Prospects, chapter in c. ghaoui (Ed.): encyclopedia of Human-Computer interaction. pennsylvania: idea group, inc.* doi:10.1.1.95.5691

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Reinking, D., Hayes, D. A., & McEneaney, J. E. (1988, September). Good and poor readers' use of explicitly cued graphic aids. *Journal of Literacy Research*, *20*(3), 229–247. doi:10.1080/10862968809547641

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (first, pp. 49–69). New York: Cambridge University Press.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. doi:10.1007/s10648-007-9053-4

Seufert, T., & Brunken, R. (2004). Supporting coherence formation in multimedia learning. In *Instructional design for effective and enjoyable computer-supported learning* (pp. 138–147). Tübingen: Knowledge Media Research Center. doi:10.1016/S0959-4752(02)00022-1

Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. G. (2010, December). Multimedia instructions and cognitive load theory: effects of modality and cueing. *British Journal of Educational Psychology*, *74*(1), 71–81. doi:10.1348/000709904322848824

Veenker, T. (2003, May). WWStim: a CGI script for presenting web-based questionnaires and experiments. Utrecht: Utrecht University, Department of Linguistics. Retrieved from URL: http://www.let.uu.nl/~Theo.Veenker/personal/projects/wwstim/doc/en/

Voßkühler, A. (2011, April). OGAMA (Open gaze and mouse analyzer). Berlin: Freie Universität Berlin. Retrieved from URL: http://www.ogama.net/

Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008, November). OGAMA (Open gaze and mouse analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior Research Methods*, *40*(4), 1150–1162. doi:10.3758/BRM.40.4.1150

Appendix A

Materials

A.1 Text and image sources

- Example: The structure and development of hurricanes
  - http://science.howstuffworks.com/nature/natural-disasters/hurricane2.htm
  - http://en.wikipedia.org/wiki/Tropical_cyclone
  - http://environment.about.com/od/globalwarming/a/hurricanecauses.htm
  - http://fcit.usf.edu/florida/teacher/science/mod2/tropical.cyclone.html
- 1: The Tuareg and Tuareg conflicts
  - http://en.wikipedia.org/wiki/Tuareg_people
  - http://africa.si.edu/exhibits/tuareg/who.html
  - http://www1.american.edu/ted/ice/tuareg.htm
  - http://en.wikipedia.org/wiki/Tuareg_Rebellion_(1990-1995)
  - http://en.wikipedia.org/wiki/Tuareg_Rebellion_(2007-2009)
  - http://anthrocivitas.net/forum/showthread.php?p=126519
- 2: African migration routes to western Europe
  - http://news.bbc.co.uk/2/hi/europe/6228236.stm
  - http://www.migrationinformation.org/feature/display.cfm?id=484
- 3: The mid-ocean ridge system
  - http://en.wikipedia.org/wiki/Mid-ocean_ridge
- 4: Replication of the influenza virus
  - http://www.news-medical.net/health/Types-of-Influenza.aspx
  - http://en.wikipedia.org/wiki/Influenza
- 5: The various beef cuts from a cow

  - http://www.ehow.com/list_6814736_parts-cow-used-meat_.html
  - http://en.wikipedia.org/wiki/Beef
  - http://www.primermagazine.com/wp-content/uploads/2008/08/Beef/Beef_cow.jpg
- 6: The rock cycle
  - http://en.wikipedia.org/wiki/Rock_cycle
- 7: Oceanic depths and divisions
  - http://en.wikipedia.org/wiki/Ocean
- 8: The thermohaline circulation
  - http://en.wikipedia.org/wiki/Thermohaline_circulation
- 9: The beer brewing process
  - http://www.britannica.com/EBchecked/topic/58378/beer/236418/The-brewing-process
  - http://en.wikipedia.org/wiki/Brewing
- 10: The malaria parasite and its replication
  - http://en.wikipedia.org/wiki/Malaria
  - http://www.pnas.org/content/104/29/11865/F1.expansion.html
- 11: An anatomy of the Sun
  - http://fusedweb.pppl.gov/cpep/Chart_Pages/5.Plasmas/SunLayers.html
  - http://www.lpi.usra.edu/education/skytellers/sun.shtml
  - http://en.wikipedia.org/wiki/Sun
- 12: The working of the fluorescence microscope
  - http://nobelprize.org/educational/physics/microscopes/fluorescence/index.html
  - http://micro.magnet.fsu.edu/primer/techniques/fluorescence/anatomy/fluoromicroanatomy.html

A.2 Presentations

Figure A.1: The structure and development of hurricanes (example presentation) - Presentation without cues

Hurricanes develop in warm tropical regions where the water is at least 27Âř C. They also require moist air and converging equatorial winds. A hurricane's low-pressure center of relative calm is called the eye. The area surrounding the eye is called the eye wall, where the most violent winds occur. The bands that circulate outward from the eye are the rain bands.

The process of hurricane generation begins when water evaporates at the sea surface, and the resulting warm, moist air rises into the atmosphere. Because warm air is less dense than cold air, the warm ocean air is able to rise upwards through the cooler layers of the atmosphere above it. As the air rises, the moisture condenses to form clouds and thunderstorms. The energy released from the condensation, a process called the latent heat of condensation, warms the cool atmosphere above causing more air to rise. This process continually provides heat and moisture. Additional warm, moist air rises from the ocean surface to replace the air above and creates a relative area of high pressure. The rising air causes the pressure at the surface to continue to fall. Pressure forces dictate winds to diverge out of this high pressure zone which helps to further deepen the low surface pressure. Resulting surface winds blow in towards the low-pressure center and circulate forming the familiar cyclonic wind pattern of a hurricane. Wind speeds increase as cooler air rushes underneath the rising warm air. Faster winds are able to evaporate more water vapor from the ocean surface to fuel the storm. This positive feedback loop continues for as long as conditions are favorable for hurricane development.

A hurricane's vertical structure influences the amount of precipitation produced by a storm: increased latent heating warms the environment inside the clouds relative to the air outside of the clouds, creating an unstable situation as warm air tends to rise above the cold air. This allows clouds to grow taller, discharge more precipitation, and forms the rain bands.

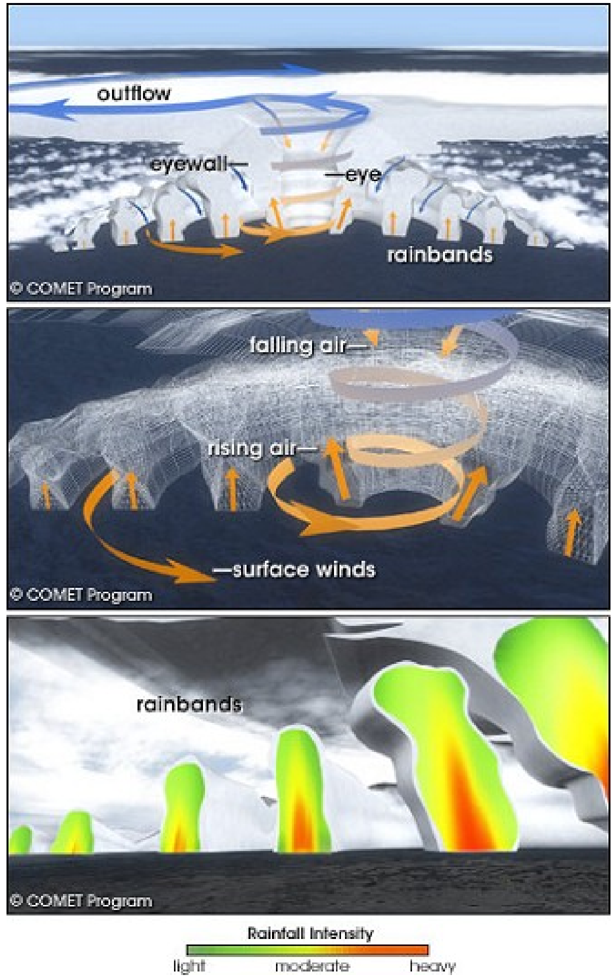


Figure A.2: 1: The Tuareg and the Tuareg conflicts - Presentation without cues

The Tuareg are a Berber nomadic pastoralist people who once controlled the caravan trade routes across the Sahara. They are nominally Muslim. They live in the semi-arid Sahel and arid Sahara in an area that overlaps southern Algeria, southwestern Libya, Mali, and Niger, and in fewer numbers, in Burkina Faso and Nigeria. Their economy is based on livestock, agriculture and trade. They are referred to as "the people of the veil" or "the blue people of the Sahara" in reference to the indigo turbans worn by men, which stain their skin and define their identity. The Tuareg confederations are the Kel-Ahaggar, Ajjer, Ayr, Adagh, Ataram, Denneg, Owey, and Gres. Most of them are located in Niger.

The twentieth century saw profound changes to their way of life: the end of colonial French rule and the creation of new countries; devastating, repeated droughts that decimated herds of livestock; and political marginalization and rebellions. Their social organization and economy have substantially transformed, and most Tuareg have given up their nomadic lifestyle. When African countries achieved widespread independence in the 1960s, the traditional Tuareg territory was divided among Niger, Mali, Algeria, Libya, and Burkina Faso. Competition for resources has since led to conflicts between the Tuareg and neighbouring African groups, especially after the French colonization and independence.

In Mali, a Tuareg uprising surfaced in the northern mountains in the 1960s, following Mali's independence. Several Tuareg joined. The army of the recently formed state of Mali suppressed the revolt. Resentment among the Tuareg fueled the 2nd uprising from 1990 to 1996 in the aftermath of a clash outside a prison in Tchin-Tabaraden, Niger. Tuaregs claimed autonomy for their traditional homeland and caused several clashes in both countries. Sporadic fighting in the Mountains around Arlit continued. The tourist center of Agadez and the uranium mining town of Arlit were evacuated of foreigners and armed by the Niger Army. In Mali, the rising began when Tuareg separatists attacked government buildings around Gao. Deadly clashes followed with deaths numbering well into the thousands. Negotiations initiated by France and Algeria led to peace agreements. In 2007, a new surge in violence occurred.

Figure A.3: 2: The Tuareg and the Tuareg conflicts - On the left the text with simple cues and on the right the text with explicit cues

Simple cues	Explicit cues
The Tuareg are a Berber nomadic pastoralist people who once controlled the caravan trade routes across the Sahara. They are nominally Muslim. They live in the semi-arid Sahel and arid Sahara in an area that overlaps southern Algeria, southwestern Libya, Mali, and Niger, and in fewer numbers, in Burkina Faso and Nigeria. Their economy is based on livestock, agriculture and trade. They are referred to as "the people of the veil" or "the blue people of the Sahara" in reference to the indigo turbans worn by men, which stain their skin and define their identity. The Tuareg confederations are the Kel-Ahaggar, Ajjer, Ayr, Adagh, Ataram, Denneg, Owey, and Gres. Most of them are located in Niger (see the map).	*The Tuareg are a Berber nomadic pastoralist people who once controlled the caravan trade routes across the Sahara. They are nominally Muslim. They live in the semi-arid Sahel and arid Sahara in an area that overlaps southern Algeria, southwestern Libya, Mali, and Niger, and in fewer numbers, in Burkina Faso and Nigeria. Their economy is based on livestock, agriculture and trade. They are referred to as "the people of the veil" or "the blue people of the Sahara" in reference to the indigo turbans worn by men, which stain their skin and define their identity. The Tuareg confederations are the Kel-Ahaggar, Ajjer, Ayr, Adagh, Ataram, Denneg, Owey, and Gres. Most of them are located in Niger. The picture shows that the Tuareg are organized into loose confederations which are highlighted in green.*
The twentieth century saw profound changes to their way of life: the end of colonial French rule and the creation of new countries; devastating, repeated droughts that decimated herds of livestock; and political marginalization and rebellions. Their social organization and economy have substantially transformed, and most Tuareg have given up their nomadic lifestyle. When African countries achieved widespread independence in the 1960s, the traditional Tuareg territory was divided among Niger, Mali, Algeria, Libya, and Burkina Faso (see the map). Competition for resources has since led to conflicts between the Tuareg and neighbouring African groups, especially after the French colonization and independence.	*The twentieth century saw profound changes to their way of life: the end of colonial French rule and the creation of new countries; devastating, repeated droughts that decimated herds of livestock; and political marginalization and rebellions. Their social organization and economy have substantially transformed, and most Tuareg have given up their nomadic lifestyle. When African countries achieved widespread independence in the 1960s, the traditional Tuareg territory was divided among Niger, Mali, Algeria, Libya, and Burkina Faso. Notice in the picture how the Tuareg territory is spread over the new nations. Competition for resources has since led to conflicts between the Tuareg and neighbouring African groups, especially after the French colonization and independence.*
In Mali, a Tuareg uprising surfaced in the northern mountains in the 1960s, following Mali's independence. Several Tuareg joined. The army of the recently formed state of Mali suppressed the revolt. Resentment among the Tuareg fueled the 2nd uprising from 1990 to 1996 in the aftermath of a clash outside a prison in Tchin-Tabaraden, Niger. Tuaregs claimed autonomy for their traditional homeland and caused several clashes in both countries (see the map). Sporadic fighting in the Mountains around Arlit continued. The tourist center of Agadez and the uranium mining town of Arlit were evacuated of foreigners and armed by the Niger Army. In Mali, the rising began when Tuareg separatists attacked government buildings around Gao. Deadly clashes followed with deaths numbering well into the thousands. Negotiations initiated by France and Algeria led to peace agreements. In 2007, a new surge in violence occurred.	*In Mali, a Tuareg uprising surfaced in the northern mountains in the 1960s, following Mali's independence. Several Tuareg joined. The army of the recently formed state of Mali suppressed the revolt. Resentment among the Tuareg fueled the 2nd uprising from 1990 to 1996 in the aftermath of a clash outside a prison in Tchin-Tabaraden, Niger. Tuaregs claimed autonomy for their traditional homeland and caused several clashes in both countries that are marked as stars on the map. Sporadic fighting in the Mountains around Arlit continued. The tourist center of Agadez and the uranium mining town of Arlit were evacuated of foreigners and armed by the Niger Army. In Mali, the rising began when Tuareg separatists attacked government buildings around Gao. Deadly clashes followed with deaths numbering well into the thousands. Negotiations initiated by France and Algeria led to peace agreements. In 2007, a new surge in violence occurred.*

Figure A.4: 2: African migration routes to western Europe - Presentation without cues

Thousands of Africans try to make the journey to Europe each year as illegal migrants - risking people smugglers, deserts, sea crossings and the possibility of being sent home, all for the dream of a better life. The main departure points are: (1) the west coast of Africa (Northern Mauritania, Western Sahara and southern Morocco) from where most head for the Canary Islands, (2) Northern Morocco to cross into Ceuta and Melilla or cross the Strait of Gibraltar to Spain, and (3) Tunisia and Libya for boats heading for Italy's island of Lampedusa.

First the migrants must cross great distances. The trans-Saharan journey is made in several stages, and takes anywhere between 1 month and several years. Migrants often settle temporarily in towns located on migration hubs to work. The majority of migrants enter Northern Africa overland from Agadez in Niger despite the existing alternative routes. From Agadez, migration routes bifurcate to the Sebha oasis in Libya and to Tamanrasset in southern Algeria. From Sebha, migrants move to Tripoli and other coastal cities or to Tunisia; from the coast, migrants travel by boat to either Malta or the Italian islands of Lampedusa, Pantalleria, and Sicily. From Tamanrasset in Algeria, migrants move to the northern cities or enter Morocco.

In reaction to intensified border patrolling in the Strait of Gibraltar, migrants in Morocco have moved southward to the Western Sahara to get to the Canary Islands. An increasing number of West African migrants circumvent the central Saharan routes via the western edge of the continent by sailing from Mauritanian, Cape Verdean, Senegalese, and other West African coasts to the Canary Islands; they also travel over land from Mauritania to Morocco via the Western Sahara.

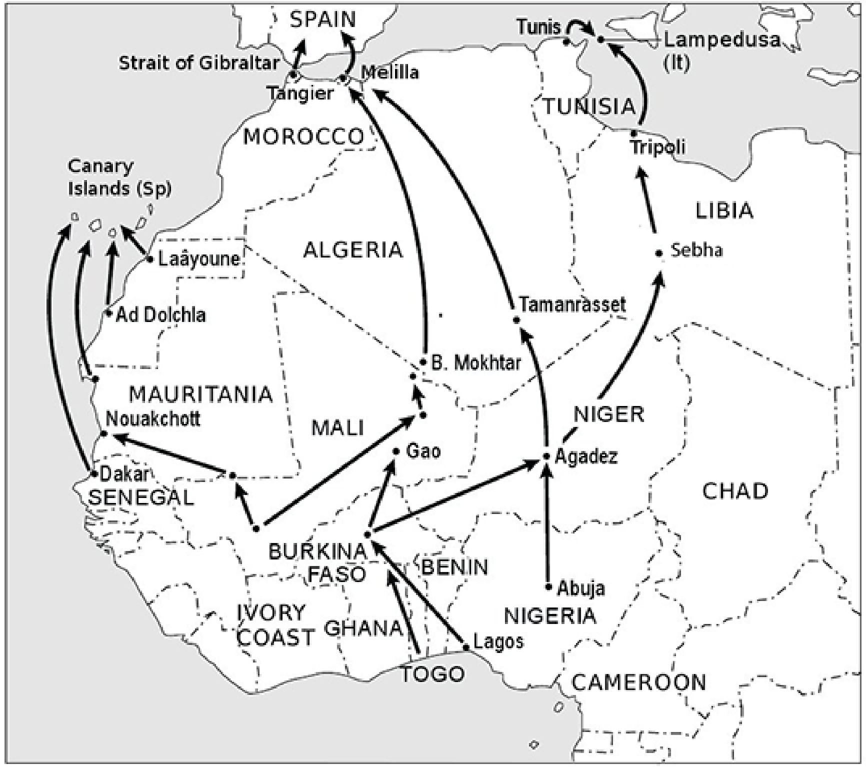


Figure A.5: 2: African migration routes to western Europe - On the left the text with simple cues and on the right the text with explicit cues

Thousands of Africans try to make the journey to Europe each year as illegal migrants - risking people smugglers, deserts, sea crossings and the possibility of being sent home, all for the dream of a better life. The main departure points are: (1) the west coast of Africa (Northern Mauritania, Western Sahara and southern Morocco) from where most head for the Canary Islands, (2) Northern Morocco to cross into Ceuta and Melilla or cross the Strait of Gibraltar to Spain, and (3) Tunisia and Libya for boats heading for Italy's island of Lampedusa. (See the map.)

First the migrants must cross great distances. The trans-Saharan journey is made in several stages, and takes anywhere between 1 month and several years. Migrants often settle temporarily in towns located on migration hubs to work. The majority of migrants enter Northern Africa overland from Agadez in Niger despite the existing alternative routes. From Agadez, migration routes bifurcate to the Sebha oasis in Libya and to Tamanrasset in southern Algeria, see the map. From Sebha, migrants move to Tripoli and other coastal cities or to Tunisia; from the coast, migrants travel by boat to either Malta or the Italian islands of Lampedusa, Pantalleria, and Sicily. From Tamanrasset in Algeria, migrants move to the northern cities or enter Morocco.

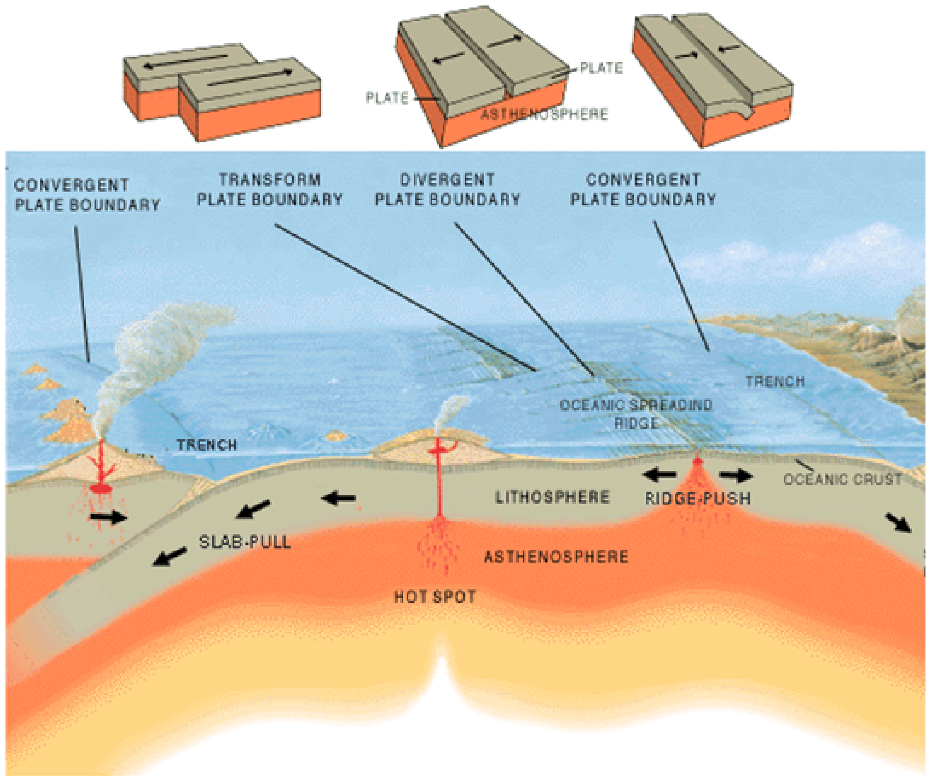
In reaction to intensified border patrolling in the Strait of Gibraltar, migrants in Morocco have moved southward to the Western Sahara to get to the Canary Islands. An increasing number of West African migrants circumvent the central Saharan routes via the western edge of the continent (see the map) by sailing from Mauritanian, Cape Verdean, Senegalese, and other West African coasts to the Canary Islands; they also travel over land from Mauritania to Morocco via the Western Sahara.

Thousands of Africans try to make the journey to Europe each year as illegal migrants - risking people smugglers, deserts, sea crossings and the possibility of being sent home, all for the dream of a better life. The main departure points are: (1) the west coast of Africa (Northern Mauritania, Western Sahara and southern Morocco) from where most head for the Canary Islands, (2) Northern Morocco to cross into Ceuta and Melilla or cross the Strait of Gibraltar to Spain, and (3) Tunisia and Libya for boats heading for Italy's island of Lampedusa; see the destination points in the image on the western and northern coastline and the routes leading to them.

First the migrants must cross great distances. The trans-Saharan journey is made in several stages, and takes anywhere between 1 month and several years. Migrants often settle temporarily in towns located on migration hubs to work. The majority of migrants enter Northern Africa overland from Agadez in Niger despite the existing alternative routes. From Agadez, migration routes bifurcate to the Sebha oasis in Libya and to Tamanrasset in southern Algeria. In the image you can see the two routes leading North. From Sebha, migrants move to Tripoli and other coastal cities or to Tunisia; from the coast, migrants travel by boat to either Malta or the Italian islands of Lampedusa, Pantalleria, and Sicily. From Tamanrasset in Algeria, migrants move to the northern cities or enter Morocco.

In reaction to intensified border patrolling in the Strait of Gibraltar, migrants in Morocco have moved southward to the Western Sahara to get to the Canary Islands. An increasing number of West African migrants circumvent the central Saharan routes via the western edge of the continent, as you can see from the sea routes on the map, by sailing from Mauritanian, Cape Verdean, Senegalese, and other West African coasts to the Canary Islands; they also travel over land from Mauritania to Morocco via the Western Sahara.

Figure A.6: 3: The mid-ocean ridge system - Presentation without cues



The mid-ocean ridge system is the largest single volcanic feature on the Earth, encircling it like the seams of a baseball. It consists of thousands of individual volcanoes or volcanic ridge segments which periodically erupt. At the ridges, two divergent plates move away from each other as hot magma emerges from the asthenosphere in the upper mantle. The magma oozes forth as lava onto the ocean floor and into the crust at and near rifts along the ridge axes. The lava cools and attaches itself to the trailing edge of each plate, forming new ocean floor crust in a process known as sea-floor spreading. The rocks making up the crust below the sea floor are youngest at the axis of the ridge and age with increasing distance from that axis. New magma of basalt composition emerges at and near the axis because of decompression melting in the underlying asthenosphere.

Moving away from the mid-ocean ridge, ocean depth progressively increases; the greatest depths are in ocean trenches. As the oceanic crust moves away from the ridge axis, the rock in the underlying mantle cools and becomes more rigid. The crust and the relatively rigid rock below it make up the oceanic lithosphere. There are two formation processes of the oceanic ridge, ridge-push and slab-pull, thought to be responsible for the spreading seen at mid-ocean ridges. Ridge-push occurs when the growing bulk of the ridge pushes the rest of the tectonic plate away from the ridge, often towards a subduction zone. At the subduction zone, "slab-pull" comes into effect. This is simply the weight of the tectonic plate being subducted (pulled) below the overlying plate dragging the rest of the plate along behind it.

The oceanic crust is made up of rocks much younger than the Earth itself: most oceanic crust in the ocean basins is less than 200 million years old. As the crust is spreading the ocean floor is renewed and the older crust is recycled back into the mantle where plates converge. The rate at which the mid-ocean ridge creates new material is known as the spreading rate, and is generally measured in mm/yr.

Figure A.7: 3: The mid-ocean ridge system - On the left the text with simple cues and on the right the text with explicit cues

The mid-ocean ridge system is the largest single volcanic feature on the Earth, encircling it like the seams of a baseball. It consists of thousands of individual volcanoes or volcanic ridge segments which periodically erupt. At the ridges, two divergent plates move away from each other as hot magma emerges from the asthenosphere in the upper mantle; see the picture. The magma oozes forth as lava onto the ocean floor and into the crust at and near rifts along the ridge axes. The lava cools and attaches itself to the trailing edge of each plate, forming new ocean floor crust in a process known as sea-floor spreading. The rocks making up the crust below the sea floor are youngest at the axis of the ridge and age with increasing distance from that axis. New magma of basalt composition emerges at and near the axis because of decompression melting in the underlying asthenosphere.

Moving away from the mid-ocean ridge, ocean depth progressively increases; the greatest depths are in ocean trenches, see the picture. As the oceanic crust moves away from the ridge axis, the rock in the underlying mantle cools and becomes more rigid. The crust and the relatively rigid rock below it make up the oceanic lithosphere. There are two formation processes of the oceanic ridge, ridge-push and slab-pull, thought to be responsible for the spreading seen at mid-ocean ridges, see the picture. Ridge-push occurs when the growing bulk of the ridge pushes the rest of the tectonic plate away from the ridge, often towards a subduction zone. At the subduction zone, "slab-pull" comes into effect. This is simply the weight of the tectonic plate being subducted (pulled) below the overlying plate dragging the rest of the plate along behind it.

The oceanic crust is made up of rocks much younger than the Earth itself: most oceanic crust in the ocean basins is less than 200 million years old. As the crust is spreading the ocean floor is renewed and the older crust is recycled back into the mantle where plates converge. The rate at which the mid-ocean ridge creates new material is known as the spreading rate, and is generally measured in mm/yr.

The mid-ocean ridge system is the largest single volcanic feature on the Earth, encircling it like the seams of a baseball. It consists of thousands of individual volcanoes or volcanic ridge segments which periodically erupt. At the ridges, two divergent plates move away from each other as hot magma emerges from the asthenosphere in the upper mantle. Please look at the location of the astenosphere, indicated in red, and how it interacts with the ridge system. The magma oozes forth as lava onto the ocean floor and into the crust at and near rifts along the ridge axes. The lava cools and attaches itself to the trailing edge of each plate, forming new ocean floor crust in a process known as sea-floor spreading. The rocks making up the crust below the sea floor are youngest at the axis of the ridge and age with increasing distance from that axis. New magma of basalt composition emerges at and near the axis because of decompression melting in the underlying asthenosphere.

Moving away from the mid-ocean ridge, ocean depth progressively increases; the greatest depths are in ocean trenches which are on the left and right side of the picture where the plates touch. As the oceanic crust moves away from the ridge axis, the rock in the underlying mantle cools and becomes more rigid. The crust and the relatively rigid rock below it make up the oceanic lithosphere. There are two formation processes of the oceanic ridge, ridge-push and slab-pull, thought to be responsible for the spreading seen at mid-ocean ridges. Notice the directions of the plates and how ridge-push and slab-pull work. Ridge-push occurs when the growing bulk of the ridge pushes the rest of the tectonic plate away from the ridge, often towards a subduction zone. At the subduction zone, "slab-pull" comes into effect. This is simply the weight of the tectonic plate being subducted (pulled) below the overlying plate dragging the rest of the plate along behind it.

The oceanic crust is made up of rocks much younger than the Earth itself: most oceanic crust in the ocean basins is less than 200 million years old. As the crust is spreading the ocean floor is renewed and the older crust is recycled back into the mantle where plates converge. The rate at which the mid-ocean ridge creates new material is known as the spreading rate, and is generally measured in mm/yr.

Figure A.8: 4: Replication of the influenza virus - Presentation without cues

The influenza virus can only replicate in living cells and contains RNA genome, RNA polymerase, and viral proteins. Infection and replication of Influenza is a multi-step process: firstly the virus has to bind to and enter the cell, produce new copies of viral proteins and RNA, assemble these components into new viral particles and finally exit the host cell. Influenza viruses bind through hemagglutinin onto sialic acid sugars on the surfaces of cells and the cell imports the virus into a structure called an endosome.

Once inside the cell, the acidic conditions in the endosome cause two events to happen: the hemagglutinin protein fuses the viral envelope with the endosome's membrane, then ion channels allow protons to move through the viral envelope and acidify the core of the virus, which causes the virus to dissemble and release the viral RNA (vRNA) and core proteins. The vRNA molecules, core proteins and RNA polymerase are then released into the cytoplasm. These core proteins and vRNA form a complex that is transported into the cell nucleus, where the RNA polymerase begins transcribing vRNA. The vRNA is either exported into the cytoplasm where viral proteins will be synthesized by ribosomes, or the vRNA remains in the nucleus.

Newly synthesized viral proteins are either secreted through the Golgi apparatus onto the cell surface (e.g. hemagglutinin and neuraminidase molecules) or are transported back into the nucleus to bind vRNA and form new viral genome particles. VRNAs, RNA polymerase, and other viral proteins are assembled into a viral particle. Hemagglutinin and neuraminidase molecules cluster into a bulge in the cell membrane. The vRNA and viral core proteins leave the nucleus and enter this membrane protrusion. The mature virus buds off in a sphere of host cell membrane, acquiring the hemagglutinin and neuraminidase molecules within the membrane.

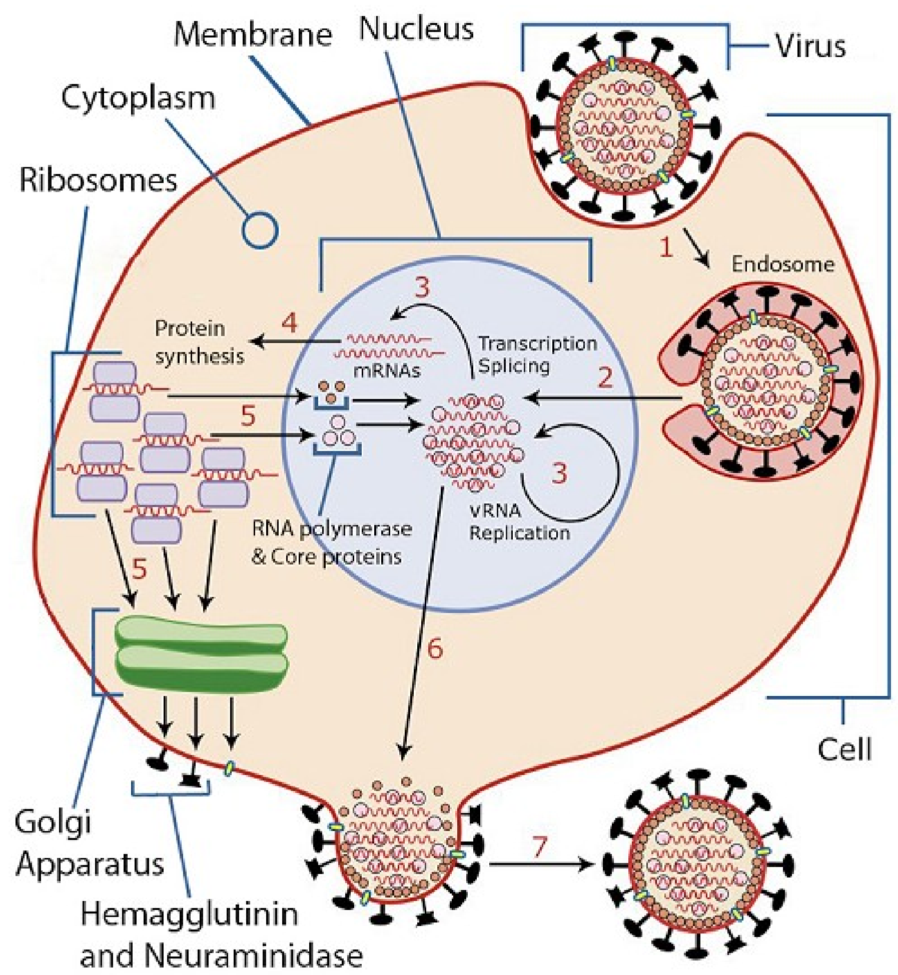


Figure A.9: 4: Replication of the influenza virus - On the left the text with simple cues and on the right the text with explicit cues

The influenza virus can only replicate in living cells and contains RNA genome, RNA polymerase, and viral proteins. Infection and replication of Influenza is a multi-step process: firstly the virus has to bind to and enter the cell, produce new copies of viral proteins and RNA, assemble these components into new viral particles and finally exit the host cell. Influenza viruses bind through hemagglutinin onto sialic acid sugars on the surfaces of cells and the cell imports the virus into a structure called an endosome.

Once inside the cell, the acidic conditions in the endosome cause two events to happen: the hemagglutinin protein fuses the viral envelope with the endosome's membrane, then ion channels allow protons to move through the viral envelope and acidify the core of the virus, which causes the virus to dissemble and release the viral RNA (vRNA) and core proteins. The vRNA molecules, core proteins and RNA polymerase are then released into the cytoplasm, see the image. These core proteins and vRNA form a complex that is transported into the cell nucleus, where the RNA polymerase begins transcribing vRNA. The vRNA is either exported into the cytoplasm where viral proteins will be synthesized by ribosomes, or the vRNA remains in the nucleus.

Newly synthesized viral proteins are either secreted through the Golgi apparatus onto the cell surface (e.g. hemagglutinin and neuraminidase molecules) or are transported back into the nucleus to bind vRNA and form new viral genome particles; see the image. VRNAs, RNA polymerase, and other viral proteins are assembled into a viral particle. Hemagglutinin and neuraminidase molecules cluster into a bulge in the cell membrane. The vRNA and viral core proteins leave the nucleus and enter this membrane protrusion (see the image). The mature virus buds off in a sphere of host cell membrane, acquiring the hemagglutinin and neuraminidase molecules within the membrane.

The influenza virus can only replicate in living cells and contains RNA genome, RNA polymerase, and viral proteins. Infection and replication of Influenza is a multi-step process: firstly the virus has to bind to and enter the cell, produce new copies of viral proteins and RNA, assemble these components into new viral particles and finally exit the host cell. Influenza viruses bind through hemagglutinin onto sialic acid sugars on the surfaces of cells and the cell imports the virus into a structure called an endosome.

Once inside the cell, the acidic conditions in the endosome cause two events to happen: the hemagglutinin protein fuses the viral envelope with the endosome's membrane, then ion channels allow protons to move through the viral envelope and acidify the core of the virus, which causes the virus to dissemble and release the viral RNA (vRNA) and core proteins. The vRNA molecules, core proteins and RNA polymerase are then released into the cytoplasm; see the image (step 1 and 2) how the virus enters the cell and dissembles. These core proteins and vRNA form a complex that is transported into the cell nucleus, where the RNA polymerase begins transcribing vRNA. The vRNA is either exported into the cytoplasm where viral proteins will be synthesized by ribosomes, or the vRNA remains in the nucleus.

Newly synthesized viral proteins are either secreted through the Golgi apparatus onto the cell surface (e.g. hemagglutinin and neuraminidase molecules) or are transported back into the nucleus to bind vRNA and form new viral genome particles. See (step 3-5) in the image how proteins are synthesized and move back in the nucleus or to the Gogli apparatus. VRNAs, RNA polymerase, and other viral proteins are assembled into a viral particle. Hemagglutinin and neuraminidase molecules cluster into a bulge in the cell membrane. The vRNA and viral core proteins leave the nucleus and enter this membrane protrusion. Please look in the image (step 6) where the particles move from the nucleus to the membrane protrusion. The mature virus buds off in a sphere of host cell membrane, acquiring the hemagglutinin and neuraminidase molecules within the membrane.

Figure A.10: 5: The various beef cuts from a cow - Presentation without cues

The meat available from a cow comes in a wide variety of cuts, each with its own distinct qualities. Tender steaks are highly prized and costly, while cuts of lesser quality often find their way into hamburgers, kebabs and jerky. The meat is first divided into primal cuts. These are basic sections from which steaks and other subdivisions are cut. Since the animal's legs and neck muscles do the most work, they are the toughest; the meat becomes more tender as distance from hoof and horn increases.

The chuck is the lower neck and upper shoulder of the cow. A large amount of connective tissue makes this area tougher than other upper half cuts, and the meat ideally needs to be stewed, slow-cooked, braised or ground. The benefit of this is that connective tissue contains collagen, which melts as the meat is cooked and spreads its flavor throughout the dish. The rib area is used for premium steaks, most notably the highly regarded rib eye steak. The whole cut is often referred to as "prime rib," highlighting the overall quality of the meat in this section. The loin is in front of the round and it runs from the backbone down to the upper belly. It is divided into three separate sections: the short loin, sirloin and tenderloin. These cuts provide some of the best steaks, including the T-bone steak, Porterhouse steak and filet mignon. The round, also known as the rump, is the equivalent of the animal's hind leg including the butt, ham and thigh. The round is lean and fairly tough, making it a prime piece for boiling and stewing.

The shank is the upper portion of the legs between the shoulder and the knee. Shank is generally used for stews and soups as it is one of the toughest cuts. The brisket is the breast area between the front legs. Brisket is used for braised beef brisket and in the production of corned beef. The plate and flank are both lower half cuts, running along the underside of the cow. As with other lower half cuts, both the plate and the flank are tough and best suited for moist cooking methods such as stewing and braising.

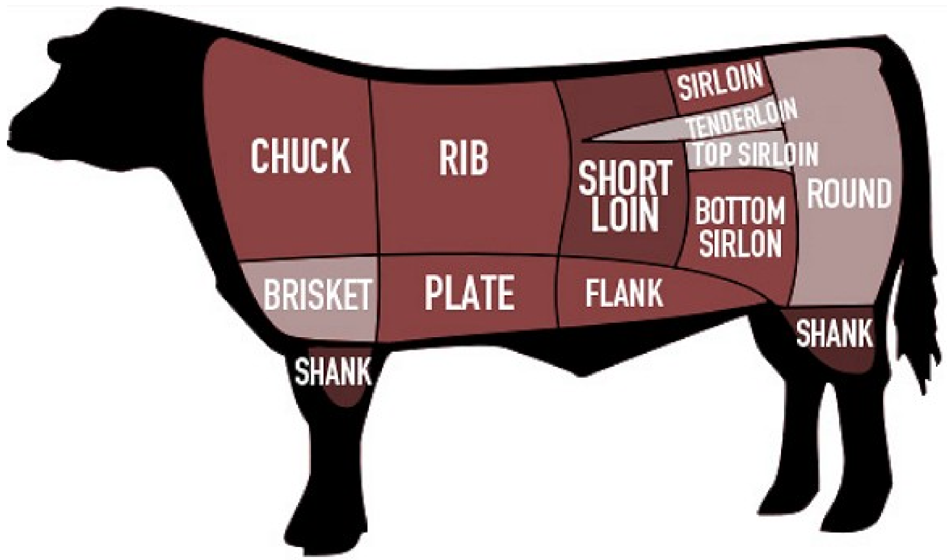


Figure A.11: 5: The various beef cuts from a cow - On the left the text with simple cues and on the right the text with explicit cues

The meat available from a cow comes in a wide variety of cuts, each with its own distinct qualities. Tender steaks are highly prized and costly, while cuts of lesser quality often find their way into hamburgers, kebabs and jerky. The meat is first divided into primal cuts. These are basic sections from which steaks and other subdivisions are cut. Since the animal's legs and neck muscles do the most work, they are the toughest; the meat becomes more tender as distance from hoof and horn increases.

The chuck is the lower neck and upper shoulder of the cow. A large amount of connective tissue makes this area tougher than other upper half cuts, and the meat ideally needs to be stewed, slow-cooked, braised or ground. The benefit of this is that connective tissue contains collagen, which melts as the meat is cooked and spreads its flavor throughout the dish. The rib area is used for premium steaks, most notably the highly regarded rib eye steak. The whole cut is often referred to as "prime rib," highlighting the overall quality of the meat in this section; see the image. The loin is in front of the round and it runs from the backbone down to the upper belly. It is divided into three separate sections: the short loin, sirloin and tenderloin. These cuts provide some of the best steaks, including the T-bone steak, Porterhouse steak and filet mignon. The round, also known as the rump, is the equivalent of the animal's hind leg including the butt, ham and thigh; see the image. The round is lean and fairly tough, making it a prime piece for boiling and stewing.

The shank is the upper portion of the legs between the shoulder and the knee. Shank is generally used for stews and soups as it is one of the toughest cuts. The brisket is the breast area between the front legs. Brisket is used for braised beef brisket and in the production of corned beef. The plate and flank are both lower half cuts, running along the underside of the cow (see the image). As with other lower half cuts, both the plate and the flank are tough and best suited for moist cooking methods such as stewing and braising.

The meat available from a cow comes in a wide variety of cuts, each with its own distinct qualities. Tender steaks are highly prized and costly, while cuts of lesser quality often find their way into hamburgers, kebabs and jerky. The meat is first divided into primal cuts. These are basic sections from which steaks and other subdivisions are cut. Since the animal's legs and neck muscles do the most work, they are the toughest; the meat becomes more tender as distance from hoof and horn increases.

The chuck is the lower neck and upper shoulder of the cow. A large amount of connective tissue makes this area tougher than other upper half cuts, and the meat ideally needs to be stewed, slow-cooked, braised or ground. The benefit of this is that connective tissue contains collagen, which melts as the meat is cooked and spreads its flavor throughout the dish. The rib area is used for premium steaks, most notably the highly regarded rib eye steak. The whole cut is often referred to as "prime rib," highlighting the overall quality of the meat in this section. Notice in the image the location of the chuck and the rib. The loin is in front of the round and it runs from the backbone down to the upper belly. It is divided into three separate sections: the short loin, sirloin and tenderloin. These cuts provide some of the best steaks, including the T-bone steak, Porterhouse steak and filet mignon. The round, also known as the rump, is the equivalent of the animal's hind leg including the butt, ham and thigh. Please look at the round and loin in reference to their location and how the loin is divided in sections. The round is lean and fairly tough, making it a prime piece for boiling and stewing.

The shank is the upper portion of the legs between the shoulder and the knee. Shank is generally used for stews and soups as it is one of the toughest cuts. The brisket is the breast area between the front legs. Brisket is used for braised beef brisket and in the production of corned beef. The plate and flank are both lower half cuts, running along the underside of the cow as can be seen in the picture below the Rib and Loin cuts. As with other lower half cuts, both the plate and the flank are tough and best suited for moist cooking methods such as stewing and braising.

Figure A.12: 6: The rock cycle - Presentation without cues

The rock cycle is a fundamental concept that describes the transitions among 3 main rock types: sedimentary, metamorphic, and igneous. Each type of rock is altered or destroyed when it is forced out of its equilibrium. Plate tectonics and the water cycle drive the rock cycle. Rocks that are pushed under the surface may melt into magma. If the magma no longer stays liquid, it will cool and solidify to form a rock called an igneous. Any of the three main types of rocks (igneous, sedimentary, and metamorphic rocks) can melt into magma and cool into igneous rocks.

Rocks exposed to the atmosphere become unstable and are subject to the processes of weathering and erosion. Rain, frost, carbonic acid and oxygen crumble the mass down into smaller fragments and resolve most of its ingredients into new products or carry it away in aqueous solution. The fragmented material or grains accumulate and are buried. A rock made up of grains fused together is a sedimentary rock. Sedimentary rocks can be formed from the lithification of smaller rock fragments, material generated by living organisms, or from chemically precipitated material from a mineral bearing solution due to evaporation.

Rocks exposed to high temperatures and/or pressures can be changed physically or chemically to form a different rock, called metamorphic. Regional metamorphism refers to the effects on large masses of rocks over a wide area, typically associated with mountain building events. Another type of metamorphism is caused when a rock is heated up, but not melted, by solidifying magma. This contact metamorphism results in a rock that is altered and re-crystallized. Any pre-existing type of rock can be modified by the processes of metamorphism.

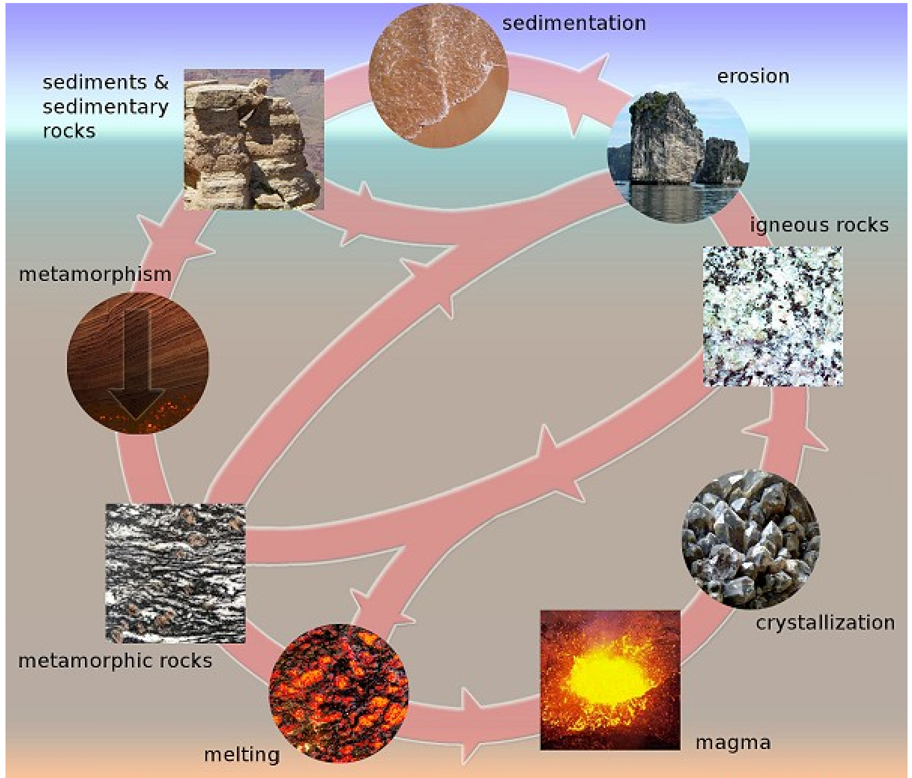


Figure A.13: 6: The rock cycle - On the left the text with simple cues and on the right the text with explicit cues

The rock cycle is a fundamental concept that describes the transitions among 3 main rock types: sedimentary, metamorphic, and igneous. Each type of rock is altered or destroyed when it is forced out of its equilibrium. Plate tectonics and the water cycle drive the rock cycle. Rocks that are pushed under the surface may melt into magma. If the magma no longer stays liquid, it will cool and solidify to form a rock called an igneous. Any of the three main types of rocks (igneous, sedimentary, and metamorphic rocks) can melt into magma and cool into igneous rocks (see the image).

Rocks exposed to the atmosphere become unstable and are subject to the processes of weathering and erosion. Rain, frost, carbonic acid and oxygen crumble the mass down into smaller fragments and resolve most of its ingredients into new products or carry it away in aqueous solution. The fragmented material or grains accumulate and are buried. A rock made up of grains fused together is a sedimentary rock ,see the image. Sedimentary rocks can be formed from the lithification of smaller rock fragments, material generated by living organisms, or from chemically precipitated material from a mineral bearing solution due to evaporation.

Rocks exposed to high temperatures and/or pressures can be changed physically or chemically to form a different rock, called metamorphic (see the image). Regional metamorphism refers to the effects on large masses of rocks over a wide area, typically associated with mountain building events. Another type of metamorphism is caused when a rock is heated up, but not melted, by solidifying magma. This contact metamorphism results in a rock that is altered and re-crystallized. Any pre-existing type of rock can be modified by the processes of metamorphism.

The rock cycle is a fundamental concept that describes the transitions among 3 main rock types: sedimentary, metamorphic, and igneous. Each type of rock is altered or destroyed when it is forced out of its equilibrium. Plate tectonics and the water cycle drive the rock cycle. Rocks that are pushed under the surface may melt into magma. If the magma no longer stays liquid, it will cool and solidify to form a rock called an igneous. Any of the three main types of rocks (igneous, sedimentary, and metamorphic rocks) can melt into magma and cool into igneous rocks. Please look in the image how igneous rock forms.

Rocks exposed to the atmosphere become unstable and are subject to the processes of weathering and erosion. Rain, frost, carbonic acid and oxygen crumble the mass down into smaller fragments and resolve most of its ingredients into new products or carry it away in aqueous solution. The fragmented material or grains accumulate and are buried. A rock made up of grains fused together is a sedimentary rock. Please follow in the image how sedimentary rock forms and changes. Sedimentary rocks can be formed from the lithification of smaller rock fragments, material generated by living organisms, or from chemically precipitated material from a mineral bearing solution due to evaporation.

Rocks exposed to high temperatures and/or pressures can be changed physically or chemically to form a different rock, called metamorphic. Notice in the image how rock can be metamorphosed. Regional metamorphism refers to the effects on large masses of rocks over a wide area, typically associated with mountain building events. Another type of metamorphism is caused when a rock is heated up, but not melted, by solidifying magma. This contact metamorphism results in a rock that is altered and re-crystallized. Any pre-existing type of rock can be modified by the processes of metamorphism.

Figure A.14: 7: Oceanic depths and divisions - Presentation without cues

Oceanographers divide the ocean into regions depending on physical and biological conditions of these areas. The pelagic zone includes all open ocean regions, and can be divided into further regions categorized by depth and light abundance. The photic zone covers the oceans from surface level to 200 metres down. This is the region where photosynthesis can occur and therefore is the most biodiverse. Since plants require photosynthesis, life found deeper than this must either rely on material sinking from above or find another energy source; hydrothermal vents are the primary option in what is known as the aphotic zone (depths exceeding 200 m).

The pelagic part of the photic zone is known as the epipelagic. The pelagic part of the aphotic zone can be further divided into regions that succeed each other vertically according to temperature. The mesopelagic is the uppermost region. Its lowermost boundary is at a thermocline of 12 řC, which, in the tropics generally lies at 700Ű1,000 metres. Next is the bathypelagic lying between 10-4 řC, typically between 700Ű1,000 metres and 2,000Ű4,000 metres. Next is the abyssalpelagic which lies along the abyssal plain, its lower boundary lies at about 6,000 metres. The last zone includes the deep trenches, and is known as the hadalpelagic. This lies between 6,000Ű11,000 metres and is the deepest oceanic zone.

Corresponding to the pelagic aphotic zones there are also benthic aphotic zones which run along the continental slope down to the ocean floor. The bathyal zone covers the continental slope down to about 4,000 metres. The abyssal zone covers the abyssal plains between 4,000 and 6,000 m. Lastly, the hadal zone corresponds to the hadalpelagic zone which is found in the oceanic trenches. The pelagic zone can also be split into two subregions, the neritic zone and the oceanic zone. The neritic encompasses the water mass directly above the continental shelves, while the oceanic zone includes all the completely open water. In contrast, the littoral zone covers the region between low and high tide and represents the transitional area between marine and terrestrial conditions.

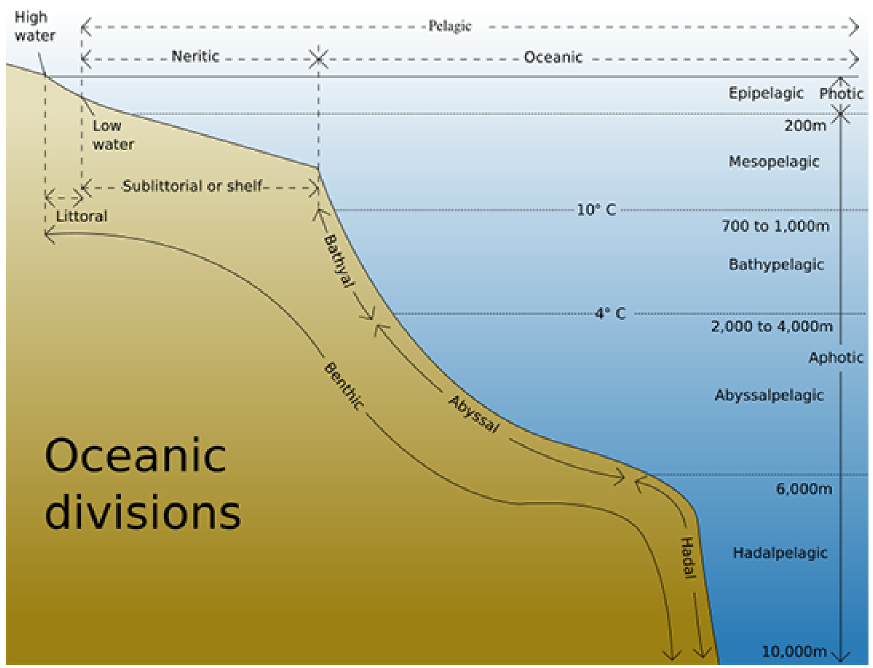


Figure A.15: 7: Oceanic depths and divisions - On the left the text with simple cues and on the right the text with explicit cues

Oceanographers divide the ocean into regions depending on physical and biological conditions of these areas. The pelagic zone includes all open ocean regions, and can be divided into further regions categorized by depth and light abundance. The photic zone covers the oceans from surface level to 200 metres down. This is the region where photosynthesis can occur and therefore is the most biodiverse. Since plants require photosynthesis, life found deeper than this must either rely on material sinking from above or find another energy source; hydrothermal vents are the primary option in what is known as the aphotic zone (depths exceeding 200 m).

The pelagic part of the photic zone is known as the epipelagic. The pelagic part of the aphotic zone can be further divided into regions that succeed each other vertically according to temperature, see the image. The mesopelagic is the uppermost region. Its lowermost boundary is at a thermocline of 12 řC, which, in the tropics generally lies at 700Ű1,000 metres. Next is the bathypelagic lying between 10-4 řC, typically between 700Ű1,000 metres and 2,000Ű4,000 metres. Next is the abyssalpelagic which lies along the abyssal plain, its lower boundary lies at about 6,000 metres. The last zone includes the deep trenches, and is known as the hadalpelagic. This lies between 6,000Ű11,000 metres and is the deepest oceanic zone.

Corresponding to the pelagic aphotic zones there are also benthic aphotic zones which run along the continental slope down to the ocean floor (see the image). The bathyal zone covers the continental slope down to about 4,000 metres. The abyssal zone covers the abyssal plains between 4,000 and 6,000 m. Lastly, the hadal zone corresponds to the hadalpelagic zone which is found in the oceanic trenches. The pelagic zone can also be split into two subregions, the neritic zone and the oceanic zone (see the image). The neritic encompasses the water mass directly above the continental shelves, while the oceanic zone includes all the completely open water. In contrast, the littoral zone covers the region between low and high tide and represents the transitional area between marine and terrestrial conditions.

Oceanographers divide the ocean into regions depending on physical and biological conditions of these areas. The pelagic zone includes all open ocean regions, and can be divided into further regions categorized by depth and light abundance. The photic zone covers the oceans from surface level to 200 metres down. This is the region where photosynthesis can occur and therefore is the most biodiverse. Since plants require photosynthesis, life found deeper than this must either rely on material sinking from above or find another energy source; hydrothermal vents are the primary option in what is known as the aphotic zone (depths exceeding 200 m).

The pelagic part of the photic zone is known as the epipelagic. The pelagic part of the aphotic zone can be further divided into regions that succeed each other vertically according to temperature. Please look at the vertical regional division starting with the epipelagic region in the photic zone and the regions below it in the aphotic zone. The mesopelagic is the uppermost region. Its lowermost boundary is at a thermocline of 12 řC, which, in the tropics generally lies at 700Ű1,000 metres. Next is the bathypelagic lying between 10-4 řC, typically between 700Ű1,000 metres and 2,000Ű4,000 metres. Next is the abyssalpelagic which lies along the abyssal plain, its lower boundary lies at about 6,000 metres. The last zone includes the deep trenches, and is known as the hadalpelagic. This lies between 6,000Ű11,000 metres and is the deepest oceanic zone.

Corresponding to the pelagic aphotic zones there are also benthic aphotic zones which run along the continental slope down to the ocean floor. Notice the benthic zones on continental slope on the left side of the image. The bathyal zone covers the continental slope down to about 4,000 metres. The abyssal zone covers the abyssal plains between 4,000 and 6,000 m. Lastly, the hadal zone corresponds to the hadalpelagic zone which is found in the oceanic trenches. The pelagic zone can also be split into two subregions, the neritic zone and the oceanic zone, which can be seen in the top horizontal axis in the image. The neritic encompasses the water mass directly above the continental shelves, while the oceanic zone includes all the completely open water. In contrast, the littoral zone covers the region between low and high tide and represents the transitional area between marine and terrestrial conditions.

Figure A.16: 8: The thermohaline circulation - Presentation without cues

The term thermohaline circulation refers to the part of the large-scale ocean circulation that is driven by global density gradients created by surface heat and freshwater fluxes. Cold water, in general, is denser than warm water. Likewise, water with a high salinity is denser than water that contains less salt. Surface ocean currents are primarily driven by winds. Deep ocean currents, on the other hand, are mainly a result of density differences.

The thermohaline circulation links major surface and deep water currents in the Atlantic, Indian, Pacific, and Southern Oceans. Wind-driven surface currents (such as the Gulf Stream) head polewards from the equatorial Atlantic Ocean, cooling all the while and eventually sinking at high latitudes forming North Atlantic Deep Water. This dense water then flows into the ocean basins and flows very slowly into the deep abyssal plains of the Atlantic. This high latitude cooling and the low latitude heating drives the movement of the deep water in a polar southward flow. The deep water flows through the Antarctic Ocean Basin around South Africa where it is split into two routes: one into the Indian Ocean and one past Australia into the Pacific.

At the Indian Ocean, some of the cold and salty water from the Atlantic causes a vertical exchange of dense, sinking water with lighter water above. In the Pacific Ocean, the rest of the cold and salty water from the Atlantic undergoes Haline forcing and slowly becomes warmer and fresher. The out-flowing undersea of cold and salty water makes the sea level of the Atlantic slightly lower than the Pacific and salinity or halinity of water at the Atlantic higher than the Pacific. This generates a large but slow flow of warmer and fresher upper ocean water from the tropical Pacific to the Indian Ocean through the Indonesian Archipelago to replace the cold and salty Antarctic Bottom Water. This warmer, fresher water from the Pacific flows up through the South Atlantic to Greenland, where it cools off and undergoes evaporative cooling and sinks to the ocean floor, providing a continuous thermohaline circulation.

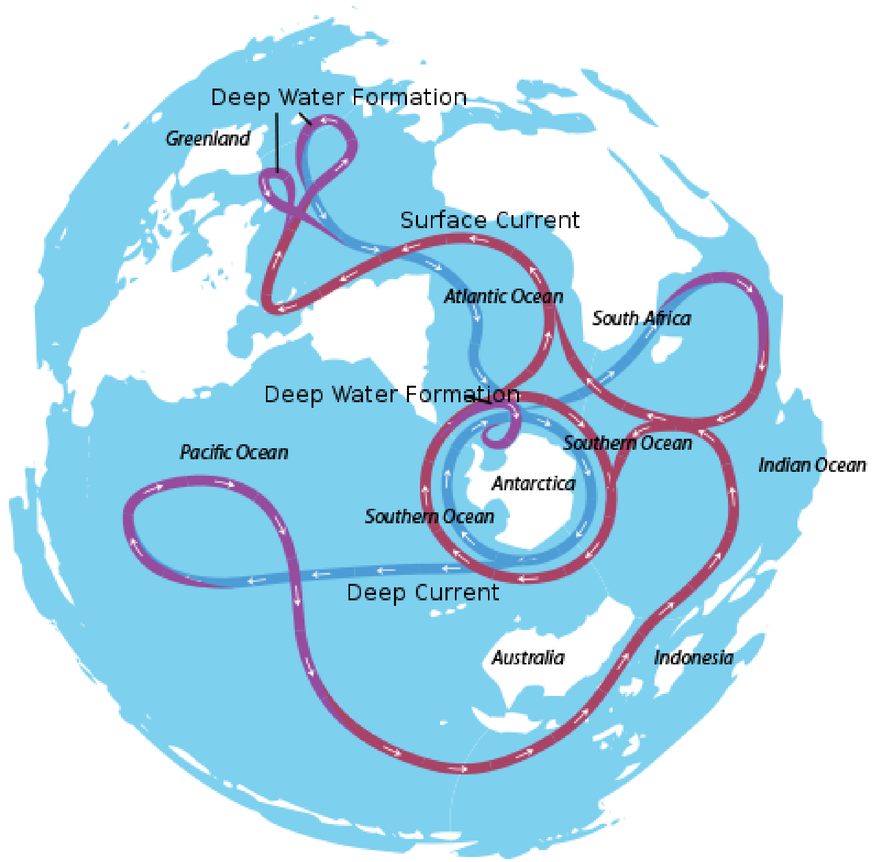


Figure A.17: 8: The thermohaline circulation - On the left the text with simple cues and on the right the text with explicit cues

The term thermohaline circulation refers to the part of the large-scale ocean circulation that is driven by global density gradients created by surface heat and freshwater fluxes. Cold water, in general, is denser than warm water. Likewise, water with a high salinity is denser than water that contains less salt. Surface ocean currents are primarily driven by winds. Deep ocean currents, on the other hand, are mainly a result of density differences.

The thermohaline circulation links major surface and deep water currents in the Atlantic, Indian, Pacific, and Southern Oceans (see the image). Wind-driven surface currents (such as the Gulf Stream) head polewards from the equatorial Atlantic Ocean, cooling all the while and eventually sinking at high latitudes forming North Atlantic Deep Water. This dense water then flows into the ocean basins and flows very slowly into the deep abyssal plains of the Atlantic. This high latitude cooling and the low latitude heating drives the movement of the deep water in a polar southward flow. The deep water flows through the Antarctic Ocean Basin around South Africa where it is split into two routes: one into the Indian Ocean and one past Australia into the Pacific, see the image.

At the Indian Ocean, some of the cold and salty water from the Atlantic causes a vertical exchange of dense, sinking water with lighter water above. In the Pacific Ocean, the rest of the cold and salty water from the Atlantic undergoes Haline forcing and slowly becomes warmer and fresher. The out-flowing undersea of cold and salty water makes the sea level of the Atlantic slightly lower than the Pacific and salinity or halinity of water at the Atlantic higher than the Pacific. This generates a large but slow flow of warmer and fresher upper ocean water from the tropical Pacific to the Indian Ocean through the Indonesian Archipelago to replace the cold and salty Antarctic Bottom Water, see the image. This warmer, fresher water from the Pacific flows up through the South Atlantic to Greenland, where it cools off and undergoes evaporative cooling and sinks to the ocean floor, providing a continuous thermohaline circulation.

The term thermohaline circulation refers to the part of the large-scale ocean circulation that is driven by global density gradients created by surface heat and freshwater fluxes. Cold water, in general, is denser than warm water. Likewise, water with a high salinity is denser than water that contains less salt. Surface ocean currents are primarily driven by winds. Deep ocean currents, on the other hand, are mainly a result of density differences.

The thermohaline circulation links major surface and deep water currents in the Atlantic, Indian, Pacific, and Southern Oceans. Please look at the image how the thermohaline circulation drives through the oceans. Wind-driven surface currents (such as the Gulf Stream) head polewards from the equatorial Atlantic Ocean, cooling all the while and eventually sinking at high latitudes forming North Atlantic Deep Water. This dense water then flows into the ocean basins and flows very slowly into the deep abyssal plains of the Atlantic. This high latitude cooling and the low latitude heating drives the movement of the deep water in a polar southward flow. The deep water flows through the Antarctic Ocean Basin around South Africa where it is split into two routes: one into the Indian Ocean and one past Australia into the Pacific. Notice the blue stream coming through the Atlantic Ocean and then going to the Indian- and Pacific Ocean.

At the Indian Ocean, some of the cold and salty water from the Atlantic causes a vertical exchange of dense, sinking water with lighter water above. In the Pacific Ocean, the rest of the cold and salty water from the Atlantic undergoes Haline forcing and slowly becomes warmer and fresher. The out-flowing undersea of cold and salty water makes the sea level of the Atlantic slightly lower than the Pacific and salinity or halinity of water at the Atlantic higher than the Pacific. This generates a large but slow flow of warmer and fresher upper ocean water from the tropical Pacific to the Indian Ocean through the Indonesian Archipelago to replace the cold and salty Antarctic Bottom Water. See in the image how the Pacific stream mixes with the Indian stream and then above Antarctica mixes with the Antarctic Bottom Water. This warmer, fresher water from the Pacific flows up through the South Atlantic to Greenland, where it cools off and undergoes evaporative cooling and sinks to the ocean floor, providing a continuous thermohaline circulation.

Figure A.18: 9: The beer brewing process - Presentation without cues

The process of making beer is known as brewing. The aim is to convert grain starches to sugar, extract the sugar with water, and then ferment it with yeast to produce the alcoholic beverage. The brewing starts with malting which is a 3 step process where the barley is made ready by releasing the starches. Malting begins with steeping (immersing barley in water) for 45 hours. As the grain imbibes water its volume increases and a white root sheath (a chit) breaks through the husk. The chitted barley is then removed from the steep. In the 2nd step, germination, the grain is spread out in a room for 5 days. Germination causes the secretion of enzymes that convert the starch into sugars and the proteins into amino acids. The grain is then called green malt. In the third step, kilning, green malt is dried which stops enzyme activity but leaves 50% in an active state.

For efficient extraction of water the malt must be milled. The aim is to retain the husk relatively intact while breaking up the brittle starch into particles. The milled malt (grist) is mashed thus converting the starches into sugars that can be fermented. The malt is mixed together with water in a vessel to create a cereal mash. The leftover sugar rich water is then strained through the bottom of the cereal mash in a process known as lautering. At this point the liquid is called wort. After separation, the wort is transferred to a kettle for boiling, to stop enzyme activity, and to obtain the bitter taste of the added hops. The kettle boil sterilizes the wort, evaporates undesirable aromas, and precipitates insoluble proteins (known as trub).

At the end of the boil the wort settles in a vessel called a "whirlpool" by separating out the solid particles. The wort is then cooled in a heat exchanger. Then the cooled wort is moved into a fermentation tank where yeast is added. During the fermenting process the sugars turn into alcohol, carbon dioxide, and other components. During maturation, residual sugars ferment generating carbon dioxide which is vented and purges undesirable compounds. Filtration gives beer a natural polish and colour and stabilizes the flavour. Most beers that are packed in bottles, metal cans or metal kegs are pasteurized by heating.

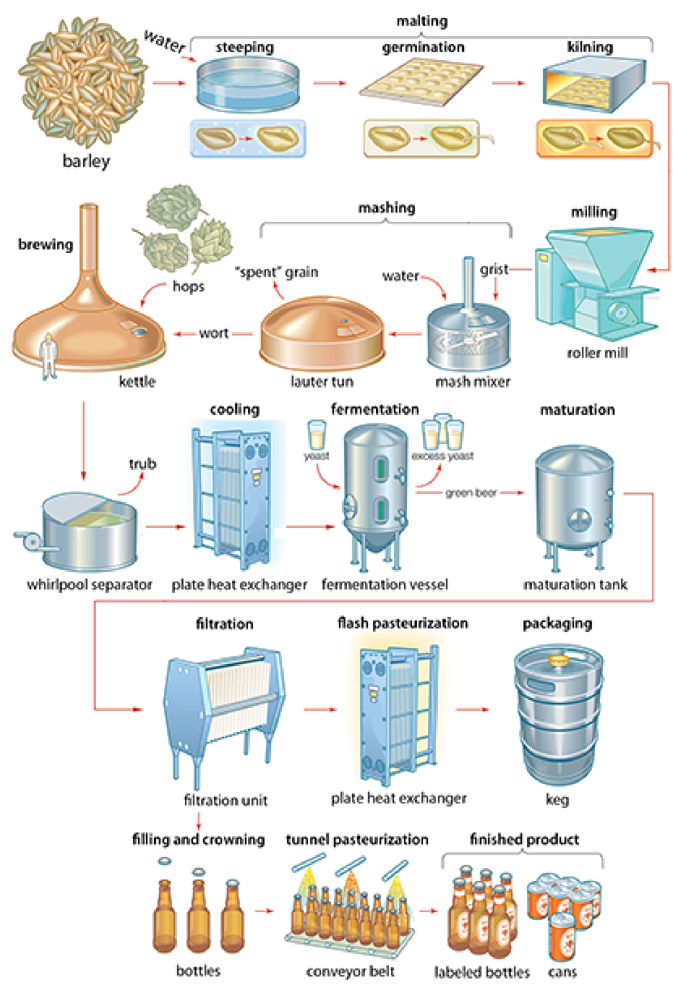


Figure A.19: 9: The beer brewing process - On the left the text with simple cues and on the right the text with explicit cues

The process of making beer is known as brewing. The aim is to convert grain starches to sugar, extract the sugar with water, and then ferment it with yeast to produce the alcoholic beverage. The brewing starts with malting. Malting is a 3 step process where the barley is made ready by releasing the starches, see the image. Malting begins with steeping (immersing barley in water) for 45 hours. As the grain imbibes water its volume increases and a white root sheath (a chit) breaks through the husk. The chitted barley is then removed from the steep. In the 2nd step, germination, the grain is spread out in a room for 5 days. Germination causes the secretion of enzymes that convert the starch into sugars and the proteins into amino acids. The grain is then called green malt. In the third step, kilning, green malt is dried which stops enzyme activity but leaves 50% in an active state; see the image.

For efficient extraction of water the malt must be milled. The aim is to retain the husk relatively intact while breaking up the brittle starch into particles. The milled malt (grist) is mashed thus converting the starches into sugars that can be fermented. The malt is mixed together with water in a vessel to create a cereal mash. The leftover sugar rich water is then strained through the bottom of the cereal mash in a process known as lautering. At this point the liquid is called wort. After separation, the wort is transferred to a kettle for boiling, to stop enzyme activity, and to obtain the bitter taste of the added hops. (See the image.) The kettle boil sterilizes the wort, evaporates undesirable aromas, and precipitates insoluble proteins (known as trub).

At the end of the boil the wort settles in a vessel called a "whirlpool" by separating out the solid particles. The wort is then cooled in a heat exchanger. Then the cooled wort is moved into a fermentation tank where yeast is added. During the fermenting process the sugars turn into alcohol, carbon dioxide, and other components. During maturation, residual sugars ferment generating carbon dioxide which is vented and purges undesirable compounds. Filtration gives beer a natural polish and colour and stabilizes the flavour. Most beers that are packed in bottles, metal cans or metal kegs are pasteurized by heating.

The process of making beer is known as brewing. The aim is to convert grain starches to sugar, extract the sugar with water, and then ferment it with yeast to produce the alcoholic beverage. The brewing starts with malting. Malting is a 3 step process where the barley is made ready by releasing the starches. In the top of the image the malting steps are grouped together. Malting begins with steeping (immersing barley in water) for 45 hours. As the grain imbibes water its volume increases and a white root sheath (a chit) breaks through the husk. The chitted barley is then removed from the steep. In the 2nd step, germination, the grain is spread out in a room for 5 days. Germination causes the secretion of enzymes that convert the starch into sugars and the proteins into amino acids. The grain is then called green malt. In the third step, kilning, green malt is dried which stops enzyme activity but leaves 50% in an active state.

Please follow in the image how milling and mashing changes the green malt. For efficient extraction of water the malt must be milled. The aim is to retain the husk relatively intact while breaking up the brittle starch into particles. The milled malt (grist) is mashed thus converting the starches into sugars that can be fermented. The malt is mixed together with water in a vessel to create a cereal mash. The leftover sugar rich water is then strained through the bottom of the cereal mash in a process known as lautering. At this point the liquid is called wort. After separation, the wort is transferred to a kettle for boiling, to stop enzyme activity, and to obtain the bitter taste of the added hops. Notice in the image how the wort continues as the product through the brewing process. The kettle boil sterilizes the wort, evaporates undesirable aromas, and precipitates insoluble proteins (known as trub).

At the end of the boil the wort settles in a vessel called a "whirlpool" by separating out the solid particles. The wort is then cooled in a heat exchanger. Then the cooled wort is moved into a fermentation tank where yeast is added. During the fermenting process the sugars turn into alcohol, carbon dioxide, and other components. During maturation, residual sugars ferment generating carbon dioxide which is vented and purges undesirable compounds. Filtration gives beer a natural polish and colour and stabilizes the flavour. Most beers that are packed in bottles, metal cans or metal kegs are pasteurized by heating.

Figure A.20: 10: The malaria parasite and its replication - Presentation without cues

The malarial parasite, Plasmodium, is a very small, single-cell blood organism, or 'protozoan'. It lives as a parasite in other organisms, namely man and mosquito. The parasite is the cause of the tropical disease malaria. The parasite's primary (definitive) hosts and transmission vectors are female mosquitoes of the Anopheles genus, while humans and other vertebrates are secondary hosts. A mosquito becomes infected when it takes a blood meal from an infected human. Once ingested, the parasite gametocytes taken up in the blood will further differentiate into male or female gametes and then fuse in the mosquito gut. This produces an ookinete that penetrates the gut lining and produces an oocyst in the gut wall. When the oocyst ruptures, it releases sporozoites that migrate through the mosquito's body to the salivary glands, where they are then ready to infect a new human host.

Malaria in humans develops via two phases: an exoerythrocytic and an erythrocytic phase. The exoerythrocytic phase involves infection of the hepatic system, or liver, whereas the erythrocytic phase involves infection of the erythrocytes, or red blood cells. When an infected mosquito pierces a person's skin to take a blood meal, sporozoites in the mosquito's saliva enter the bloodstream and migrate to the liver. The sporozoites infect liver cells and multiply asexually and asymptomatically for a period of 6Ű15 days. The parasite matures to a schizont containing many merozoites in it, which, following rupture of their host cells, escape into the blood and infect red blood cells, thus beginning the erythrocytic stage of the life cycle.

Within the red blood cells the merozoite grows to an immature trophozoite. Mature trophozoites develop into the schizont stage where the parasite divides several times to produce new merozoites. These break out the red blood cells and travel within the bloodstream to invade new red blood cells. Some merozoites turn into male and female gametocytes. If a mosquito pierces the skin of an infected person, it potentially picks up the gametocytes within the blood, thus completing the cycle.

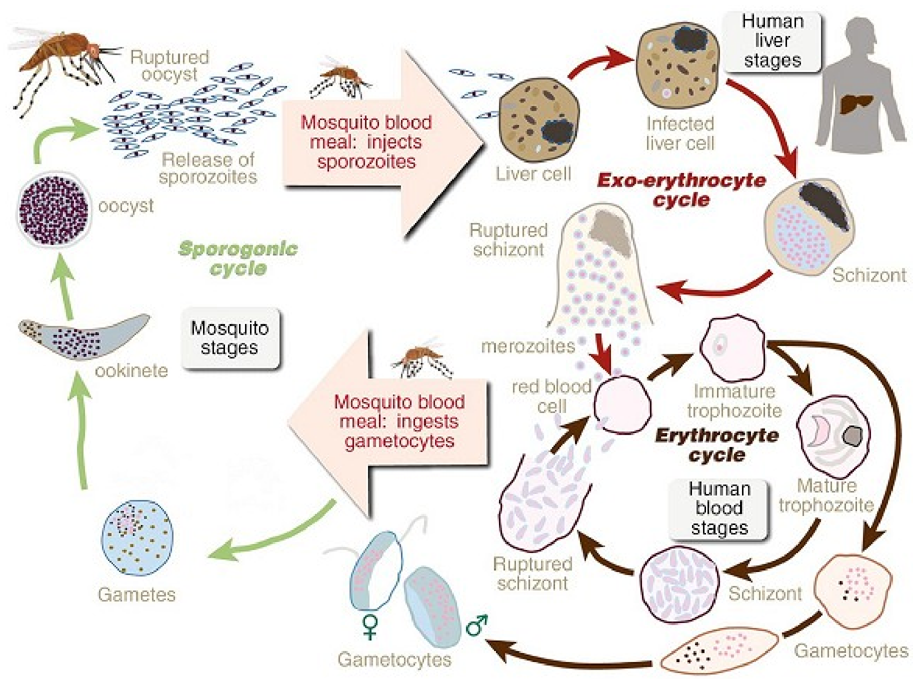


Figure A.21: 10: The malaria parasite and its replication - On the left the text with simple cues and on the right the text with explicit cues

The malarial parasite, Plasmodium, is a very small, single-cell blood organism, or 'protozoan'. It lives as a parasite in other organisms, namely man and mosquito. The parasite is the cause of the tropical disease malaria. The parasite's primary (definitive) hosts and transmission vectors are female mosquitoes of the Anopheles genus, while humans and other vertebrates are secondary hosts. A mosquito becomes infected when it takes a blood meal from an infected human, see the image. Once ingested, the parasite gametocytes taken up in the blood will further differentiate into male or female gametes and then fuse in the mosquito gut. This produces an ookinete that penetrates the gut lining and produces an oocyst in the gut wall. When the oocyst ruptures, it releases sporozoites that migrate through the mosquito's body to the salivary glands, where they are then ready to infect a new human host.

Malaria in humans develops via two phases: an exo-erythrocytic and an erythrocytic phase. The exo-erythrocytic phase involves infection of the hepatic system, or liver, whereas the erythrocytic phase involves infection of the erythrocytes, or red blood cells; see the image. When an infected mosquito pierces a person's skin to take a blood meal, sporozoites in the mosquito's saliva enter the bloodstream and migrate to the liver. The sporozoites infect liver cells and multiply asexually and asymptomatically for a period of 6Ű15 days. The parasite matures to a schizont containing many merozoites in it, which, following rupture of their host cells, escape into the blood and infect red blood cells, thus beginning the erythrocytic stage of the life cycle, see the image.

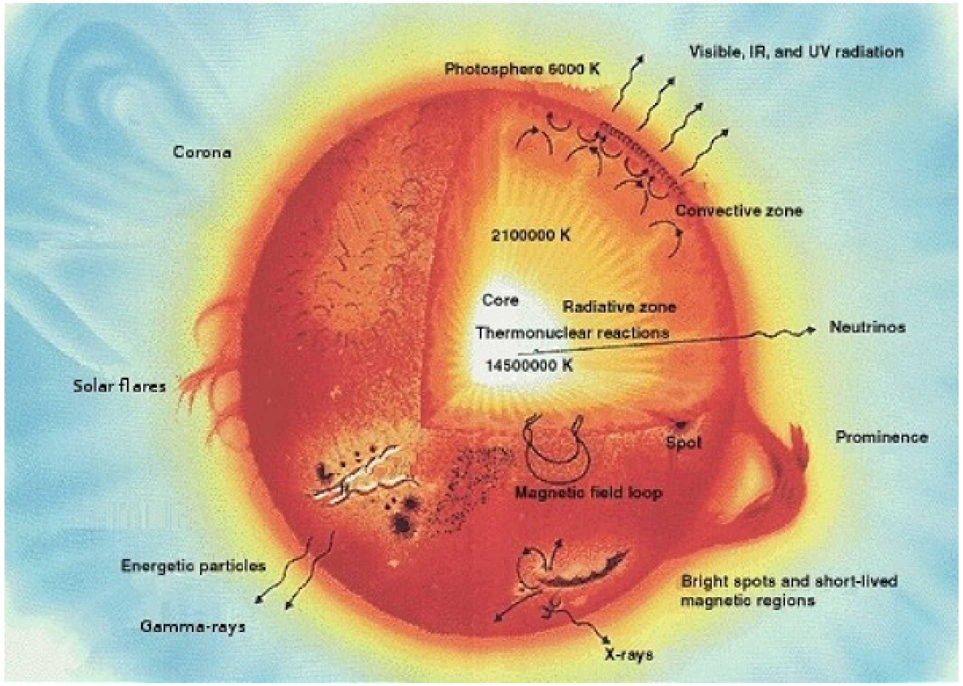
Within the red blood cells the merozoite grows to a immature trophozoite. Mature trophozoites develop into the schizont stage where the parasite divides several times to produce new merozoites. These break out the red blood cells and travel within the bloodstream to invade new red blood cells. Some merozoites turn into male and female gametocytes. If a mosquito pierces the skin of an infected person, it potentially picks up the gametocytes within the blood, thus completing the cycle.

The malarial parasite, Plasmodium, is a very small, single-cell blood organism, or 'protozoan'. It lives as a parasite in other organisms, namely man and mosquito. The parasite is the cause of the tropical disease malaria. The parasite's primary (definitive) hosts and transmission vectors are female mosquitoes of the Anopheles genus, while humans and other vertebrates are secondary hosts. A mosquito becomes infected when it takes a blood meal from an infected human. Please follow the mosquito infection stage starting with the ingesting of gametocytes. Once ingested, the parasite gametocytes taken up in the blood will further differentiate into male or female gametes and then fuse in the mosquito gut. This produces an ookinete that penetrates the gut lining and produces an oocyst in the gut wall. When the oocyst ruptures, it releases sporozoites that migrate through the mosquito's body to the salivary glands, where they are then ready to infect a new human host.

Malaria in humans develops via two phases: an exo-erythrocytic and an erythrocytic phase. The exo-erythrocytic phase involves infection of the hepatic system, or liver, whereas the erythrocytic phase involves infection of the erythrocytes, or red blood cells. Please follow both human infection stages on the right side of the image. When an infected mosquito pierces a person's skin to take a blood meal, sporozoites in the mosquito's saliva enter the bloodstream and migrate to the liver. The sporozoites infect liver cells and multiply asexually and asymptomatically for a period of 6Ű15 days. The parasite matures to a schizont containing many merozoites in it, which, following rupture of their host cells, escape into the blood and infect red blood cells, thus beginning the erythrocytic stage of the life cycle. Notice in the middle right of the image the merozoites moving from the exo-erythrocytic phase to the red blood cells in the erythrocyte phase.

Within the red blood cells the merozoite grows to a immature trophozoite. Mature trophozoites develop into the schizont stage where the parasite divides several times to produce new merozoites. These break out the red blood cells and travel within the bloodstream to invade new red blood cells. Some merozoites turn into male and female gametocytes. If a mosquito pierces the skin of an infected person, it potentially picks up the gametocytes within the blood, thus completing the cycle.

Figure A.22: 11: An anatomy of the Sun - Presentation without cues



The innermost layer of the Sun is the core and it generates the Sun's energy. Gravitational pressures compress and heat the material in the core to 13.6 million kelvins which keeps it in a gaseous state. Fusion reactions produce energy in the form of gamma rays and neutrinos. The energy passes from the core into the cooler radiative zone (4 mil. K.). Here the energy moves randomly from atom to atom, with some of the energy moving toward the surface.

As energy moves out of the radiative zone, it enters the convective zone. Here the atoms do not pass the energy from particle to particle; the atoms themselves move, carrying the heat with them. The hotter material near the radiative zone rises to the cooler surface of the convective zone. As it reaches the top of the convective zone, it cools, sinks and reheats creating huge circulating cells. The photosphere (Şsphere of lightŤ) is the ŞsurfaceŤ of the Sun. It has temperatures that reach about 5800 K. and is the layer that releases most of the light. The surface has continuously changing dark regions or sunspots. The spots are dark because they are cooler than the surrounding gas. The Sun's magnetic field loops oppose convection in the convective layer and stop the flow of energy to the surface. This results in the cooler and darker spots.

The corona is a thin outer layer that is seen during a solar eclipse and emits energy at many different wave lengths. Loops and arches of matter, called solar flares and prominences, are often seen extending along lines of the Sun's magnetic field. Solar flares sometimes reach the size of the Earth and their temperature can reach tens of millions of kelvins. The extreme heat produces x rays that create light when they hit the corona. Prominences are generally less violent and cooler than solar flares.

Figure A.23: 11: An anatomy of the Sun - On the left the text with simple cues and on the right the text with explicit cues

The innermost layer of the Sun is the core and it generates the Sun's energy. Gravitational pressures compress and heat the material in the core to 13.6 million kelvins which keeps it in a gaseous state. Fusion reactions produce energy in the form of gamma rays and neutrinos, see the image. The energy passes from the core into the cooler radiative zone (4 mil. K.). Here the energy moves randomly from atom to atom, with some of the energy moving toward the surface.

As energy moves out of the radiative zone, it enters the convective zone. Here the atoms do not pass the energy from particle to particle; the atoms themselves move, carrying the heat with them. The hotter material near the radiative zone rises to the cooler surface of the convective zone. As it reaches the top of the convective zone, it cools, sinks and reheats creating huge circulating cells, see the image. The photosphere (Şsphere of lightŤ) is the ŞsurfaceŤ of the Sun. It has temperatures that reach about 5800 K. and is the layer that releases most of the light. The surface has continuously changing dark regions or sunspots. The spots are dark because they are cooler than the surrounding gas. The Sun's magnetic field loops oppose convection in the convective layer and stop the flow of energy to the surface. This results in the cooler and darker spots, see the image.

The corona is a thin outer layer that is seen during a solar eclipse and emits energy at many different wave lengths. Loops and arches of matter, called solar flares and prominences, are often seen extending along lines of the Sun's magnetic field. Solar flares sometimes reach the size of the Earth and their temperature can reach tens of millions of kelvins. The extreme heat produces x rays that create light when they hit the corona. Prominences are generally less violent and cooler than solar flares.

The innermost layer of the Sun is the core and it generates the Sun's energy. Gravitational pressures compress and heat the material in the core to 13.6 million kelvins which keeps it in a gaseous state. Fusion reactions produce energy in the form of gamma rays and neutrinos. Notice the core and excited neutrinos and gamma-rays in the image. The energy passes from the core into the cooler radiative zone (4 mil. K.). Here the energy moves randomly from atom to atom, with some of the energy moving toward the surface.

As energy moves out of the radiative zone, it enters the convective zone. Here the atoms do not pass the energy from particle to particle; the atoms themselves move, carrying the heat with them. The hotter material near the radiative zone rises to the cooler surface of the convective zone. As it reaches the top of the convective zone, it cools, sinks and reheats creating huge circulating cells. Notice in the image the circulating energy cells near the surface of the Sun. The photosphere (Şsphere of lightŤ) is the ŞsurfaceŤ of the Sun. It has temperatures that reach about 5800 K. and is the layer that releases most of the light. The surface has continuously changing dark regions or sunspots. The spots are dark because they are cooler than the surrounding gas. The Sun's magnetic field loops oppose convection in the convective layer and stop the flow of energy to the surface. This results in the cooler and darker spots. See in the image the dark spots and the magnetic field loops flowing outwards.

The corona is a thin outer layer that is seen during a solar eclipse and emits energy at many different wave lengths. Loops and arches of matter, called solar flares and prominences, are often seen extending along lines of the Sun's magnetic field. Solar flares sometimes reach the size of the Earth and their temperature can reach tens of millions of kelvins. The extreme heat produces x rays that create light when they hit the corona. Prominences are generally less violent and cooler than solar flares.

Figure A.24: 12: The working of the fluorescence microscope - Presentation without cues

In fluorescence microscopy, the sample you want to study is itself the light source. It is based on the phenomenon that certain materials emit energy detectable as visible light when irradiated with the light of a specific wavelength. Most biological molecules do not emit fluorescence on their own, so they must be linked with fluorescent molecules. The basic task of the fluorescence microscope is to permit excitation light to irradiate the specimen and then to separate the much weaker emitted fluorescent light from the brighter excitation light. Thus, only the light from the specimen reaches the eye. The resulting fluorescing areas shine brightly with sufficient contrast to permit detection.

A fluorescence microscope uses a mercury or xenon lamp as light source to produce ultraviolet light. To view the fluorescence, filters are needed to isolate and cut-off the excitation and emission wavelengths of a fluorochrome. The excitation filter (EF), after the light source, only lets through radiation with the desired wavelength that matches the fluorescing material. The light then hits a dichromatic mirror. The dichromatic mirror reflects the ultraviolet light up to the specimen. The radiation collides with the atoms in the specimen and electrons are excited to a higher energy level. When they relax to a lower level, they emit light. The condenser and objective lens collect the fluorescent-wavelength light produced.

This fluorescent light then passes through the dichromatic mirror. The mirror's special reflective properties allow it to separate the two light paths. Each dichromatic mirror has a set wavelength value, called the transition wavelength value, which is the wavelength of 50% transmission. The mirror reflects wavelengths of light below the transition value and transmits the wavelengths above. This property accounts for the name given to this mirror (dichromatic - two color). The dichromatic mirror and the filters are mounted on an filter block. After the mirror, the emission filter (BF) separates the emitted light from the much brighter excitation light. The filtered light then makes it to the eyepiece.

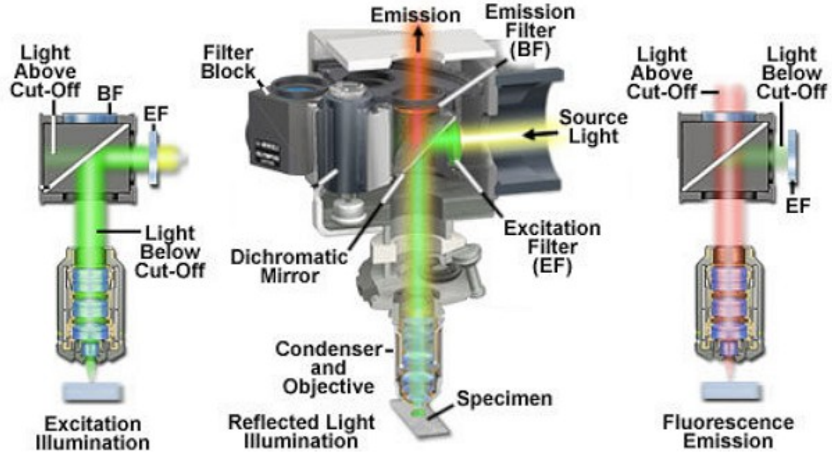


Figure A.25: 12: The working of the fluorescence microscope - On the left the text with simple cues and on the right the text with explicit cues

In fluorescence microscopy, the sample you want to study is itself the light source. It is based on the phenomenon that certain materials emit energy detectable as visible light when irradiated with the light of a specific wavelength. Most biological molecules do not emit fluorescence on their own, so they must be linked with fluorescent molecules. The basic task of the fluorescence microscope is to permit excitation light to irradiate the specimen and then to separate the much weaker emitted fluorescent light from the brighter excitation light. Thus, only the light from the specimen reaches the eye. The resulting fluorescing areas shine brightly with sufficient contrast to permit detection.

A fluorescence microscope uses a mercury or xenon lamp as light source to produce ultraviolet light. To view the fluorescence, filters are needed to isolate and cut-off the excitation and emission wavelengths of a fluorochrome. The excitation filter (EF), after the light source, only lets through radiation with the desired wavelength that matches the fluorescing material, see the image. The light then hits a dichromatic mirror. The dichromatic mirror reflects the ultraviolet light up to the specimen. The radiation collides with the atoms in the specimen and electrons are excited to a higher energy level. When they relax to a lower level, they emit light. The condenser and objective lens collect the fluorescent-wavelength light produced; see the image.

This fluorescent light then passes through the dichromatic mirror. The mirror's special reflective properties allow it to separate the two light paths. Each dichromatic mirror has a set wavelength value, called the transition wavelength value, which is the wavelength of 50% transmission. The mirror reflects wavelengths of light below the transition value and transmits the wavelengths above. This property accounts for the name given to this mirror (dichromatic - two color), see the image. The dichromatic mirror and the filters are mounted on an filter block. After the mirror, the emission filter (BF) separates the emitted light from the much brighter excitation light. The filtered light then makes it to the eyepiece.

In fluorescence microscopy, the sample you want to study is itself the light source. It is based on the phenomenon that certain materials emit energy detectable as visible light when irradiated with the light of a specific wavelength. Most biological molecules do not emit fluorescence on their own, so they must be linked with fluorescent molecules. The basic task of the fluorescence microscope is to permit excitation light to irradiate the specimen and then to separate the much weaker emitted fluorescent light from the brighter excitation light. Thus, only the light from the specimen reaches the eye. The resulting fluorescing areas shine brightly with sufficient contrast to permit detection.

A fluorescence microscope uses a mercury or xenon lamp as light source to produce ultraviolet light. To view the fluorescence, filters are needed to isolate and cut-off the excitation and emission wavelengths of a fluorochrome. The excitation filter (EF), after the light source, only lets through radiation with the desired wavelength that matches the fluorescing material. View the 1st and 2nd image to see the path of the light through the EF filter. The light then hits a dichromatic mirror. The dichromatic mirror reflects the ultraviolet light up to the specimen. The radiation collides with the atoms in the specimen and electrons are excited to a higher energy level. When they relax to a lower level, they emit light. The condenser and objective lens collect the fluorescent-wavelength light produced. Notice their location in the microscope seen in the image.

This fluorescent light then passes through the dichromatic mirror. The mirror's special reflective properties allow it to separate the two light paths. Each dichromatic mirror has a set wavelength value, called the transition wavelength value, which is the wavelength of 50% transmission. The mirror reflects wavelengths of light below the transition value and transmits the wavelengths above. This property accounts for the name given to this mirror (dichromatic - two color). In the images its effects are visualized by the paths of the fluorescent light (in red) and the excitation light (in green). The dichromatic mirror and the filters are mounted on an filter block. After the mirror, the emission filter (BF) separates the emitted light from the much brighter excitation light. The filtered light then makes it to the eyepiece.

A.3 Presentation questions

In this section are presented all presentation questions. The first answer in the list below is the correct answer to the question.

- Example: The structure and development of hurricanes
  - *Question:* Hurricane surface winds ..?
    1. Circulate in the cyclonic wind pattern
    2. Push rainbands higher
    3. Suck falling air down
    4. Rush down through the eye to the surface
- 1: The Tuareg and Tuareg conflicts
  - *Question:* Where are the most Tuareg confederations located?
    1. Niger
    2. Burkina Faso
    3. Mali
    4. Algeria
- 2: African migration routes to western Europe
  - *Question:* To which destination do migrants go who wish to avoid the central saharan routes?
    1. Canary islands
    2. Lampedusa
    3. Spain
    4. Ceuta and Melilla
- 3: The mid-ocean ridge system
  - *Question:* How does slab-pull work?
    1. The weight of the underlying plate pulls itself underneath an overlying plate
    2. The weight of the overlying plate pushes the underlying plate further down
    3. The gravitational pull from the moon slides the overlying plate on the underlying plate
    4. Magma flowing up from the trench pulls the underlying plate deeper under the overlying plate
- 4: Replication of the influenza virus
  - *Question:* After the influenza virus enters the cell it disassembles in the?
    1. In the cytoplasm
    2. In the nucleus
    3. In the viral partical

4. In the ribosomes

- 5: The various beef cuts from a cow
  - *Question:* Which part of the cow is called the 'Round'?
    1. The hind leg
    2. The lower neck and upper shoulder
    3. The lower half
    4. The upper belly
- 6: The rock cycle
  - *Question:* How does igneous rock form?
    1. It forms from cooled magma
    2. It is fused together from eroded material
    3. It forms through pressure
    4. It forms after rock heats up
- 7: Oceanic depths and divisions
  - *Question:* How many regions does the Aphotic zone have?
    1. 4
    2. 5
    3. 3
    4. 2
- 8: The thermohaline circulation
  - *Question:* The thermohaline circulation sinks to deep water in the?
    1. Atlantic ocean
    2. Pacific ocean
    3. Indian ocean
    4. Antarctic ocean
- 9: The beer brewing process
  - *Question:* Steeping is part of?
    1. The malting process
    2. The mashing process
    3. Fermentation process
    4. Milling process
- 10: The malaria parasite and its replication

- *Question:* After a mosquito takes a human blood meal the injected sporozoites migrate to the human...?

  1. Liver
  2. Blood cells
  3. Gut
  4. Schizont

- 11: An anatomy of the Sun

  - *Question:* How do the moving cells in the convection zone form?

    1. Hot material rises to the top of the convection zone where it cools and sinks back down to reheat
    2. Energy moving randomly from atom to atom causes large cells of moving energy
    3. Magnetic field loops force material and energy to move in large streams
    4. Cooler sunspots cause a difference in temperature which causes the energy to move as cells

- 12: The working of the fluorescence microscope

  - *Question:* The function of a dichromatic mirror is to?

    1. Separate the fluorescent light from the excitation light
    2. To block the emission fluorescence
    3. To mirror the image of the specimen
    4. Mix the fluorescent light from the excitation light

Appendix B

Participant data

Table B.1: Gender per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Male	6 (32%)	7 (35%)	5 (24%)	18 (30%)
Female	13 (68%)	13 (65%)	16 (76%)	42 (70%)

Table B.2: Average age per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mean	20.74	20.95	21.29	21
Standard deviation	1.45	1.43	2.39	1.81
Range	18-24	19-24	18-27	18-27

Table B.3: Language per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Dutch	10 (53%)	14 (70%)	8 (38%)	32 (53%)
German	9 (47%)	6 (30%)	12 (57%)	27 (45%)
Other	0	0	1 (5%)	1 (2%)

Table B.4: Reported English level per condition

	No cues (n=19)	Simple cues (n=20)	Explicit cues (n=21)	Total (N=60)
Mother tongue	0	0	0	0
Proficient/Excellent	5 (26%)	1 (5%)	1 (5%)	7 (12%)
Advanced	9 (47%)	12 (60%)	15 (71%)	36 (77%)
Intermediate	4 (21%)	6 (30%)	5 (24%)	15 (25%)
Elementary	1 (5%)	0	0	1 (2%)
Beginner	0	1 (5%)	0	1 (2%)