

Published as: Ooms, K., De Maeyer, P., & Fack, V., 2015. Listen to the map user: Cognition, memory, and expertise. *Cartographic Journal*, vol. 52 (1), p. 3-19.

Listen to the Map User: Cognition, Memory, and Expertise

Kristien Ooms, Philippe De Maeyer and Veerle Fack
Ghent University, Ghent, Belgium
Email: kristien.ooms@ugent.be

Abstract: This paper aims at extending current research regarding map users' cognitive processes while working with screen maps. The described experiment investigates how (expert and novice) map users retrieve information from memory that was previously gathered from screen maps. A user study was conducted in which participants had to draw a map from memory. During this task, they were instructed to say out loud every thought that came into mind. Both user groups addressed the same general cognitive structures and processes to solve the task at hand. However, the experts' background knowledge facilitated the retrieval process and allowed them to derive extra information through deductive reasoning. The novices used more descriptive terms instead of naming the objects and could remember less, and less detailed map elements.

Keywords: cartography, user study, thinking aloud, cognition, memory

1. INTRODUCTION

As a consequence of the accessibility of the Internet, more and more novices started to use cartographic products on the web. However, because of the limited background knowledge and expertise in interpreting and working with these – often dynamic and interactive – screen maps, it might be difficult for these novice map users to process their content. What is more, maps found on the web nowadays are not only created by trained cartographers. With the increasing user involvements of Web 2.0, novice map users can contribute to the content of web maps (e.g. crowdsourcing and VGI) and can even – fairly easily – create cartographic products themselves (e.g. mashups, push pin maps, APIs). Since these 'new map makers' often did not receive any cartographic training, the created products might have issues regarding their design, resulting in a less effective communication process (Cartwright, 2012; Haklay, *et al.*, 2008; Turner, 2006).

At the beginning of the new century, a number of authors expressed their concerns regarding the cognitive limits of map users in the light of modern geovisualisation techniques (among others, MacEachren & Kraak, 2001; Slocum, *et al.*, 2001). An enormous amount of cartographic products have been released on the Internet without prior evaluation by representative users or without even considering these users' needs. As a consequence, almost no information regarding the influence of these screen based visualisations on the users' cognitive processes is available. A number of authors have already answered this need, especially in the case of (geographic) mobile applications (Haklay & Nivala, 2010) and navigation systems (Muenzer & Stahl, 2011; Schlender, *et al.*, 2000).

Recently, Montello (2009) and Fabrikant and Lobben (2009) pointed out that, although a rise in the research regarding the cognitive issues (and limits) of map users has been noticed, it is still a research challenge that needs attention. Detailed insights in the users' cognitive processes, while working with digital screen maps, are essential to be able to create effective digital maps (with their interactions, dynamic responses, and animations) in the future. However, one important issue that is often neglected in this is the influence of (cartographic) expertise on the users' cognitive processes.

The aim of the experiment described in this article is to investigate how map users retrieve information from memory that was previously gathered from screen maps. In order to use and apply this information later on, the users have to be able to retrieve this again from memory. Since both experts and novices in cartography nowadays make and use the screen maps that can be found on the web, it is important to study the difference between these target groups. Experts and novices have to be able to create (screen) maps that take into account the cognitive limits of their target users. An overview of the state of the art regarding research on human's cognitive and memory structures is given in the next sections.

1.1. Understanding the map users' cognitive structures

In order to understand how users store and retrieve information, a number of memory structures need to be explained: the *sensory stores*, the *working memory*, and the *long term memory*. The sensory stores receive the information from our senses, including vision. A part of this information is transferred to the working memory (WM), which was also called *short term memory* in the past. This working memory contains the information that is actively used, but has a limited capacity. Nevertheless, there is still some debate whether there is also a time limit on the information contained in the WM (Cowan, 2001; Ericsson & Simon, 1980; Miller, 1956). In order to be able to remember the information contained in the WM for a longer time span, it has to be transferred to the long term memory (LTM). The transmission of information from the WM to the LTM corresponds to the learning process and is accomplished through rehearsal. The LTM has a virtually unlimited capacity, but cannot be reported from directly. In order to actively use the information stored in the LTM, it has to be retrieved again: transmission from the LTM to the WM (Cowan, 2001; Matlin, 2002).

General knowledge regarding a certain item, event, or situation is coded into the LTM using *schemas*. These schemas describe the essential concepts or characteristics that define such an item, event, or situation in a hierarchical way. Furthermore, concepts in these schemas are interconnected (through *links* or *pointers*) with related concepts in other schemas. These schemas are used during the cognitive process of object recognition and assist in the selection of certain memories stored in the LTM (MacEachren, 1995; Matlin, 2002).

When these schemas contain knowledge regarding the spatial layout of certain features, it is often referred to as a *cognitive map* or *mental map* (Downs & Stea, 1997). Cognitive maps can be constructed from two different sources of knowledge. Primary knowledge is gathered when humans view or move through the environment itself. This is not the same information that is gathered when studying a map, which is a secondary source of information and has picture-like properties (Scholl, 1987). Kulhavy and Stock (1996) argued that these visual images of maps are described by features on the one hand – characteristics of point locations, needed to identify them and their visual variables as proposed by Bertin (1967) – and by structural information – the spatial framework and relations between the objects – on the other hand. They indicate that the structural information is a critical element of cognitive maps. Other authors consider the cognitive map to consist out of nodes (or landmarks) and paths, which form a

hierarchical structure in the associated schemas (e.g. Hirtle & Jonides, 1985; Huynh & Doherty, 2007). However, most research on the structure of human's cognitive maps was conducted in relation to urban environments when primary knowledge was gathered.

1.2. Influence of expertise

Miller (1956) stated that the WM can hold only seven 'chunks' of information, give or take two. More recent research found that this number is even less and corresponds to four chunks of information, give or take one (Cowan, 2001). These chunks can be defined as groups of information that have a strong internal relation, but weaker relations to other information. Chunking of information is closely linked to the structure of the schemas held in the LTM. Nevertheless, this limited number of chunks that can be contained in the WM does not say anything about the actual amount of information. Related information can be grouped together, forming only one larger chunk of information (corresponding to one 'slot' in the WM). Effective grouping of information can thus increase the amount of information stored in the WM (Gobet & Simon, 1998; Miller, 1956).

Gobet and Simon (1996, 1998) studied the cognitive processes and memory structures of expert chess players. They concluded that the better performance of these experts can, on the one hand, be explained by the higher amount of chunks that are stored in the LTM. These chunks of information are, on the other hand, also much larger: more information can be contained in a single chunk (Matlin, 2002). Because the experts have richer schemas stored in memory, they can structure this information better and more tightly in the chunks that are transferred to the WM. Similar results were found in other domains of expertise, such as bridging and computer programming (Gilhooly, *et al.*, 1998).

Based on theories of human cognition, Thorndyke and Stasz (1980) assumed that the descriptions stored in memory are based on existing knowledge and the available cognitive processes. This is closely linked with the findings of Hambrick and Engle (2002) that domain knowledge has a significant (positive) influence on a person's memory performance. They suggested that domain knowledge can serve as a retrieval structure that facilitates the retrieval process from memory. Kulhavy and Stock (1996) further distinguished between *general map knowledge* and *specific map knowledge*. They argued that experts and novices differ little in the amount of information that can be retrieved from memory when standard map information is considered. The specific map knowledge differs more between experts and novices and, consequently, influences the cognitive maps that are constructed when interpreting maps.

In their experiments Thorndyke and Stasz (1980) and Gilhooly, *et al.* (1988) investigated the difference between expert and novice map users in how they learn and recall the information presented on planimetric and contour maps. Besides the cognitive processes, they focussed on the different strategies that were addressed during learning and information retrieval. They found that experts did not perform better in recalling the planimetric maps and the non-contour features on the contour maps. However, they could recall (and draw) the contour information from the contour maps better. The authors concluded that experts retrieved both specialised and 'lay' or lower level schemas. Furthermore, the experts' lower level schemas were richer than these of the novices (although the same labels were used) and they could link these lower level schemas to the specialist schemas. These findings were also confirmed by Montello, *et al.* (1994), who used landscape scenes and topographic maps in their experiments.

More recently, Ooms, *et al.* (2012) found that expert map users are significantly more efficient in the performance of a visual search on a basic screen map design. The experts' higher efficiency in map interpretation was further confirmed when both expert and novices map users had to learn – store in memory – the content of topographic maps (Ooms, *et al.*, 2013). Nevertheless, the communication of

spatial information through maps is only successful if this information can be recalled and used later on. Still very little is known on the cognitive processes and strategies that are used to recall the information that was retrieved from these screen maps. This is important considering the spectacular rise of digital cartographic products available on the Web, which are also much more accessible to the novice map users. The aim of this paper is to study the cognitive processes and strategies involved in remembering (recalling) the information that was gathered previously from screen maps. The techniques used to study the cognitive processes during information retrieval are described in the next sections.

1.3. Accessing the users' cognitive processes

1.3.1 Thinking Aloud

Nielsen (1993, p.195) stated that thinking aloud could be “the single most valuable usability engineering method”. Thinking aloud is also a well-integrated method in psychological research. During a thinking aloud study, the participants are typically asked to say out loud every thought that comes into mind. In order to maintain objectivity, the participants should not interpret these thoughts, only verbalise them. As a consequence, these verbal protocols are direct and unfiltered as opposed to other methods such as retrospection and introspection (Nielsen, 1993; Trickett & Trafton, 2007; van Someren, *et al.*, 1994). Since humans can only report information contained in the WM (See Section 1.1), the verbal protocols obtained during concurrent verbalisation correspond to the content of the WM at that moment (Trickett & Trafton, 2007; van Someren, *et al.*, 1994).

In the past, some concerns regarding the *validity* and *reactivity* of the method have been expressed. The validity of the method, on the one hand, is related to how well the obtained verbal protocols reflect the user's thoughts (and cognitive processes). Ericsson (2006), among others, concluded that the validity of the verbal protocols is dependent on the time interval between the occurrence of the actual thought and its verbalisation. Consequently, concurrent verbalisations show the highest level of validity (and completeness). What is more, a considerable amount of the participants' cognitive load is already addressed to solve the assignment and consequently, there is no room left to interpret or change the thoughts before or during verbalisation (van Someren, *et al.*, 1994).

The reactivity of the method is, on the other hand, related to influences that the verbalisation may have on the user's thoughts (the sequence of the thoughts, for example) and the execution of the assignment. Ericsson and Simon (1980) found no influence due to the application of the thinking aloud method on the users' thoughts, in comparison to participants who did not verbalise their thoughts. However, the participants might need more time to complete the assignment when they have to verbalise their thoughts in addition to the assignment (Ericsson, 2006). As a consequence, thinking aloud studies should not be combined with time measurements.

An excellent overview of the methodology regarding the analyses of verbal protocols is provided by van Someren, *et al.* (1994) and Chi (1997). The structure of these analyses is depicted in Figure 1. Based on the task analysis and the existing psychological theories (such as described in Section 1.1 and Section 1.2), a psychological model can be constructed. This model depicts the different cognitive processes that are needed to complete the task, and the relation between these different processes. A set of standard codes can be linked to each of these cognitive processes, possibly on different levels of detail. The obtained (raw) verbal protocols (transcriptions) need to be segmented or divided into smaller units. A code can be assigned to each of these units, indicating which cognitive process takes place at that moment. The results of this latter process are the coded protocols, which list objective information

regarding the participants' cognitive processes while executing the assignment (Chi, 1997; van Elzaker, 2004; van Someren, *et al.*, 1994).

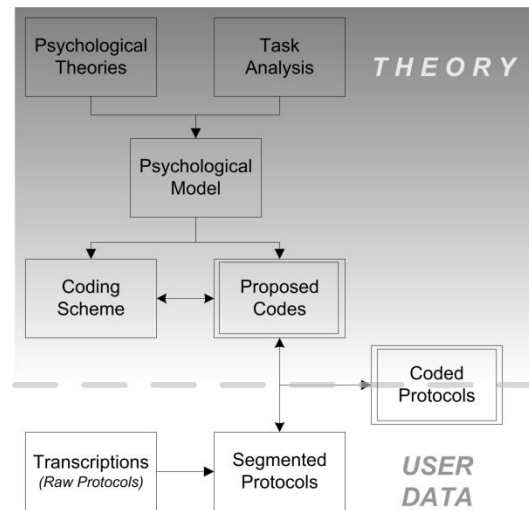


Figure 1: Analysing verbal protocols obtained from a thinking aloud study (after van Someren, *et al.*, 1994)

Suchan and Brewer (2000) considered qualitative methods in combination with content analysis suitable for research on map making and map use. One of the methods they proposed is the collection of verbal data through thinking aloud. The experiments by, for example, Thorndyke and Stasz (1980), Gilhooly, *et al.* (1988), and Montello, *et al.* (1994) (mentioned in Section 1.2) used (among others) thinking aloud to study differences in how expert and novice map users learn and recall information from paper maps (planimetric, contour, and topographic maps).

1.3.2 Sketch Maps

Huynh and Doherty (2007, p.286) defined sketch maps as “the extraction of information from a mental map through drawing”. On the most abstract level, Blaser (2000) defined sketches as a collection of line strokes that can be grouped together to form objects and these are in turn related to other objects. He distinguished between three main parts that build up a sketch map: sketched objects, spatial relations, and annotations. In his article, Blaser (2000) described the results of a survey that investigates which objects, spatial relations, and annotations are typically drawn while sketching a map.

Huynh, *et al.* (2008) and Huynh and Doherty (2007) focussed on the order with which the sketched objects were drawn. They distinguished between two elements that can be drawn - paths and landmarks – and concluded that paths are more frequently drawn at first; landmarks are more frequently drawn later on. However, they focussed on cognitive maps from urban environments (primary knowledge), not on information gathered from maps (secondary knowledge). Thorndyke and Stasz (1980), Gilhooly, *et al.* (1988), and Montello, *et al.* (1994) combined thinking aloud with sketch maps. They found that experts only performed better in the sketch maps when specific map knowledge had to be addressed: contours and topographic maps.

1.4. Research objective and question

Ooms, *et al.* (2012, 2013) studied the eye movements of (expert and novice) map users who interpreted – stored in memory – the content of screen maps with a simple and a complex content. This paper aims at extending this research by investigating the next step in the communication process: the retrieval of information from memory previously gathered from screen maps. The main research question related to this objective is to investigate whether experts and novices address different cognitive processes and/or use different strategies to recall the information. The set-up of the experiments is based on those conducted previously by Thorndyke and Stasz (1980), Gilhooly, *et al.* (1988), and Montello, *et al.* (1994) who studied these processes and strategies using printed planimetric and contour maps. The combination of two methodologies will be used during the experiment, in order to provide complementary data. Thinking aloud will be used to study the users' cognitive processes and strategies during information retrieval. However, the spatial and visual nature of the information that is recalled can be difficult to verbalise (Kulhavy & Stock, 1996). Therefore, the resulting sketch maps are also analysed regarding their content.

2. MATERIALS AND METHODS

2.1. Participants

Since the analyses of thinking aloud data are very labour intensive, the number of participants is preferably limited. However, a sufficient amount of users from each group (experts and novices in this case) is also indispensable to be able to detect differences between both. Nielsen (1994) recommended to select four (plus or minus one) participants. He based this finding on the ratio of usability problems that could be discovered with a certain number of participants. Tricket and Trafton (2007) concluded that it would be ideal to have more than five participants (preferably five to ten).

A total of 24 participants took part in the study described in this paper, with 12 experts and 12 novices. Both groups consisted out of six female and six male participants. It was chosen to work with 12 participants for each group (more than necessary) because some stimuli were only shown to half of them. The participants that took part in the expert group had at least a MSc. in Geography or Geomatics. Furthermore, they were employed at the Department of Geography (Ghent University) at the moment of the study and work with cartographic products on a daily basis. The novice map users did not receive any previous cartographic training and did not work with cartographic products on a daily basis. All participants' native language was Dutch, the language in which the study was conducted (explanation of the assignment, thinking aloud, etc.).

2.2. Tasks and stimuli

Four different maps were displayed on a screen to the participants during the experiment. They were instructed to remember the map that was currently displayed as best as possible in order to be able to draw it afterwards. They were told that they did not have to remember the location of each individual house, but certainly the main structuring elements on the map. The participants could decide for themselves how long they wanted to study each screen map, with a maximum of ten minutes. This latter time limit was introduced to keep the length of the study manageable, but this was rarely reached. Most participants ended this initial part of the trial after (on average) five minutes. During this initial part, the participants' eye movements were recorded. However, a detailed description of these eye movement recordings is beyond the scope of this article and will thus not be treated here (Ooms, *et al.*, 2013).

The stimuli that were used during the experiment are depicted in Figure 2. These maps were selected out of the Belgian topographic map series on 1 : 10 000, based on the criteria that they were not

overly complex, contained a number of main structuring elements, and did not contain any well-known areas or cities. Some extra deviations were introduced in these map images. Half of the participants saw map 3a, whereas the other half saw map 3b (with an equal share of experts and novices). These two stimuli are each other's mirrored equivalents (over its central vertical axis). The map displayed during the fourth trial was the mirrored version of the one displayed in the first trial, this time over its central horizontal axis. Finally, the colours of some objects in the second map – the water bodies and the village backgrounds – were adapted to obtain a non-conform result.

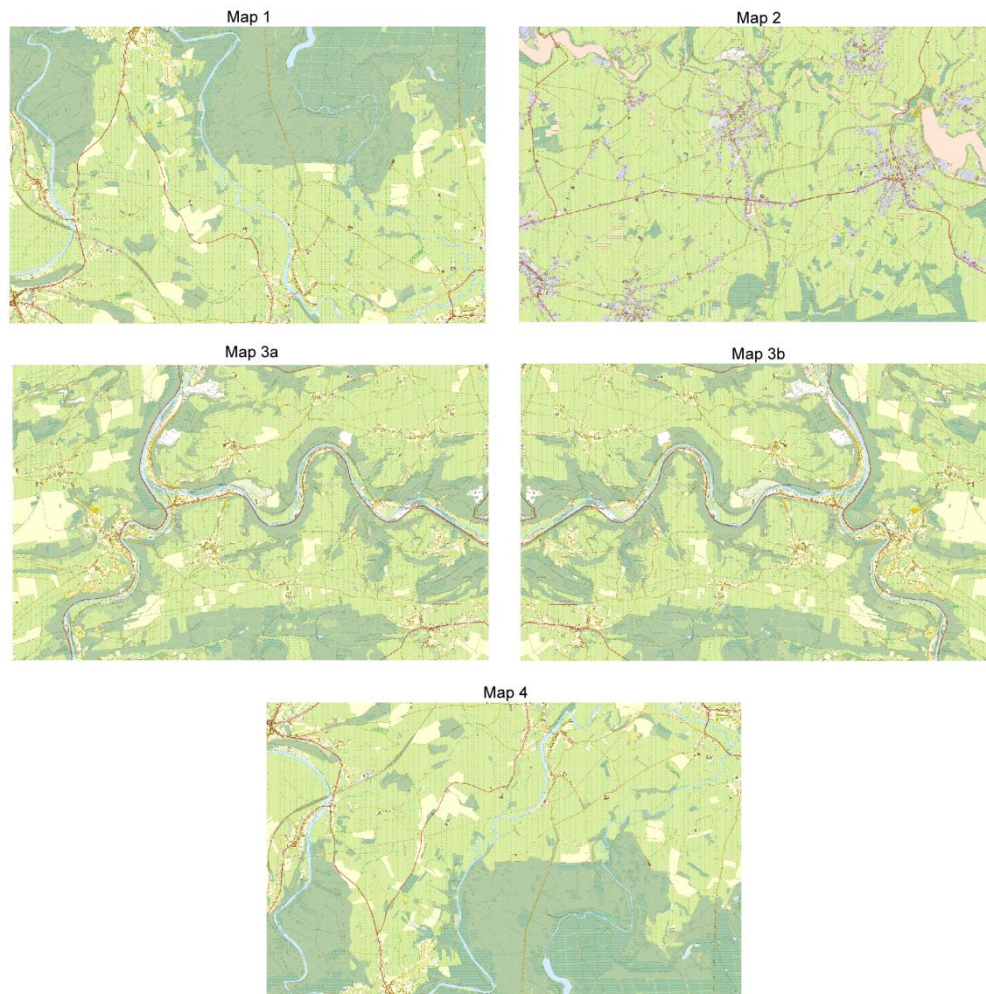


Figure 2: The stimuli presented on screen during the initial part of the trial

After ending the display of the screen map, the participants were asked to take place at another table where they could draw the map which they had just studied. They were instructed to say out loud every thought that came into mind, even if it did not had to do anything with the experiment itself. No time limit was set on this drawing assignment to avoid any pressure on the participants that could influence their behaviour. The absence of a time limit is also important as the participants were asked to execute two concurrent tasks: drawing and verbalising their thoughts. These verbalisations influenced the drawing times, and these measurements were – as a consequence – not used in the analyses.

On the table, participants could find a pile of blank A3 pages on which they could draw the map. Twelve colour pencils, a sharpener, and an eraser were placed at the upper side of the A3 pages. The colour pencils were arranged in the same order before the start of each experiment, but were not touched by the experiment moderator between the trials. On the right side of the A3 pages, a coloured post-it was placed on which the participant's code was written (e.g. ef_p3: expert, female, third participant). The colour of the post-it corresponded to the trial number. This trial number could thus easily be derived from the recorded videos (see Section 2.3). The participants were informed that they could start drawing (and verbalising their thoughts) once this post-it was in place.

In total, the participants had to complete four trials of remembering and drawing a map. The assignment was read out loud to each participant, and the same text was also displayed on the screen before the start of the first trial. At this point the participant could ask any question if the assignment was not clear. These instructions also informed the participants that they had to fill in a post study questionnaire after the completion of the fourth trial. This questionnaire was necessary to check whether any of the participants were familiar with the displayed regions, to register personal characteristics, and to receive feedback.

This task of 'remembering information previously gathered from a map' is often executed in real life. This holds especially true when considering topographic maps: users study the content of a map to get a general impression of the region and the relative positions of the objects in it. This information is added to the map user's cognitive map of the region (using a secondary source of information). Later on, this information can be used in a specific situation: in-situ orientation, planning a route, linking the cognitive map to other information, explaining the structure of the region to somebody else, etc. This use (and thus recalling) of the information, previously gathered from a (topographic) map, is an essential part of the communication process: failure in recalling this information suggests a problem in the communication process.

2.3. Apparatus and recordings

The experiment was conducted in the Eye Tracking Labo of the Department of Experimental Psychology (Ghent University). This laboratory is standard equipped with an *EyeLink1000* (SR Research) to record participants' eye movements while studying stimuli on a screen. The stimuli (maps in this case) were depicted on a 21 inch monitor. Furthermore, two camera's and a headset were used to record the essential auditory and visual components related to a thinking aloud study. The participants had to put on a headset (with a microphone) before the start of the second phase of each trial: drawing the map. This headset was connected to a laptop whose webcam recorded the front view of each participant: face and A3 paper on which the map was sketched. This was an *Acer Crystal Eye Webcam* that captures HD videos, including the audio recorded with the headset.

A second HD camera (*Sony HandyCam HDR-CS115E*) was placed on a tripod and captured the drawing process on the A3 page from a top view. The coloured post-it placed next to each A3 page was visible in all recorded videos. Furthermore, the code assigned to each participant was written on the post-it and could thus be read from the top view video. The synchronisation and analysis of the recordings from the two cameras (webcam and HandyCam) was done in the software ELAN (*EUDICO Linguistic Annotator*, developed by the Technical Group of the Max Planck Institute for Psycholinguistics). These analyses are described in more detail in Section 2.4.

2.4. Methodology: analysing thinking aloud data

The recorded videos were analysed based on the methodology described by, among others, Chi (1997), van Elzakker (2004), and van Someren, *et al.* (1994). This methodology was also briefly described in Section 1.3 and illustrated in Figure 1. A number of subsequent steps are necessary to obtain the final coded protocols in an objective way. First, the obtained videos have to be transcribed, which will form the raw (verbal) protocols. Second, the obtained verbal utterances have to be divided into smaller units (to which the codes will be assigned). Next, a suitable coding scheme has to be developed, based on a psychological model. Finally, these codes will be assigned to all segmented protocols. These steps are described in more detail in the next paragraphs.

Chi (1997) described a number of issues that have to be considered when choosing the ‘size’ of these units. She concluded that there is a trade-off between the segmentation granularity and the information that can be retrieved. Related to the main research question of this article (how do different types of map users mentally store and retrieve the visual information presented to them) two segmentation sizes were selected: sentences and words. These two levels were chosen to be able to check the validity of the data, on the one hand, and to be able to obtain complementary (more detailed) information on the other hand. The sentence segmentation corresponds to the structure of the users’ full thought.

Based on current psychological theories (see Section 1.1 and Section 1.2), the cognitive processes that would be addressed during the execution of this task were identified, including their interrelations. On the highest (general) level, independent of the actual assignment, this task could be modelled as depicted in Figure 3. The participant first orientates himself and his surroundings (what material is available, how do I begin, etc.). Next, he starts to execute the assignment and at a certain moment he evaluates his temporary result. If the participant finds that this result is acceptable, the assignment is completed. Otherwise, the participant will correct the result by adding or adapting elements, until he will evaluate this (new) result again. The ‘Execute Assignment’ and ‘Evaluate Result’ sub-parts of this model is described in more detail in the next paragraphs.

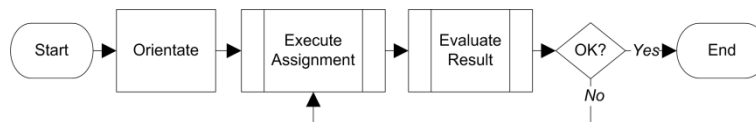


Figure 3: Highest (general) level of the psychological model.

Table 1: Proposed codes related to the map level (Level 1)

<i>psych. model</i>	<i>category</i>	<i>code</i>	<i>description</i>
Orientate	orientate	or1	<i>before the start of the task execution, gather thoughts, check tools, assignment, ...</i>
Execute Assignment	execute	ex1	<i>drawing, talking, explaining, correcting items, ...</i>
Evaluate Result	evaluate	ev1	<i>check if everything is on the map (count objects, check size, relative positions,...); not the correction of individual items when drawing them</i>
OK?	evaluate	ev1	

Figure 4 depicts the psychological model related to the actual assignment execution. At this level, links with the participants’ cognitive processes are made, including the relations between these processes. First, the participant consults his WM whether this contains information (map elements) that still needs to

be drawn. If so, these elements will be drawn. If not, the participant will have to consult his LTM (through pointers to schemas) in order to retrieve this information: LTM to WM. Information has to be transferred to the WM to be able to draw it. Next, the item that was just drawn will be evaluated for correctness and possibly adapted. If the participant finds the drawn item acceptable, he will consult his WM again. This process continues until the participant evaluates the total image that was drawn hitherto.

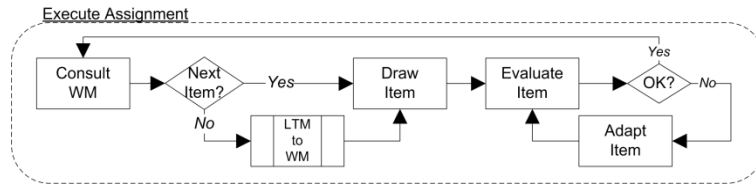


Figure 4: Psychological model of the assignment execution

Table 2: Proposed codes related to the item level (Level 2)

<i>psych. model</i>	<i>category</i>	<i>code</i>	<i>description</i>
Consult WM	gather thoughts	gt2	<i>participant is clearly thinking, gathering thoughts, finding links between items, consult previous knowledge, similarity between objects</i>
Next Item	gather thoughts	gt2	
LTM to WM	gather thoughts	gt2	
Draw Item	draw	dw2	<i>items are draw on paper, colour pencil was taken, ...</i>
Evaluate Item	evaluate	ev2	<i>check correctness of item that was drawn, or items related to it</i>
OK?	evaluate	ev2	
Adapt Item	correct	cr2	<i>correct items which were evaluated (drawing, erasing)</i>

During the evaluation process (see Figure 5), the image of the sketched map will be ‘captured’ through the participant’s (visual) sensory store and transferred to the WM. The memories of the original screen map are stored in the LTM and have to be retrieved again. Next, the image of the sketched map is compared with the memories of the original image, based on shapes, locations, completeness, colours, distributions, etc.

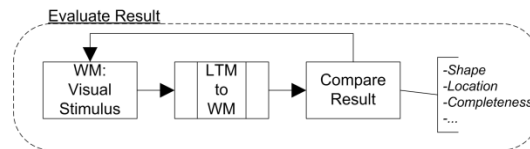


Figure 5: Psychological model of the evaluation process

Figure 6 models the information retrieval process, from LTM to WM, in more detail. In order to be able to retrieve chunks of information from the LTM, this information has to be ‘activated’ to create the so called Long Term Working Memory (LTWM) (Cowan, 2001). This activation and retrieval process makes use of pointers to the schemas (or links between these schemas), stored in the LTM. These pointers work as triggers to activate and subsequently retrieve the correct (chunks of) information from the LTM. If the correct, usable, information is retrieved, the chunk of information is placed in a slot of the WM. As mentioned before, the size of these chunks can vary, containing, for example, the name of an object, its shape, location, related background information, etc.

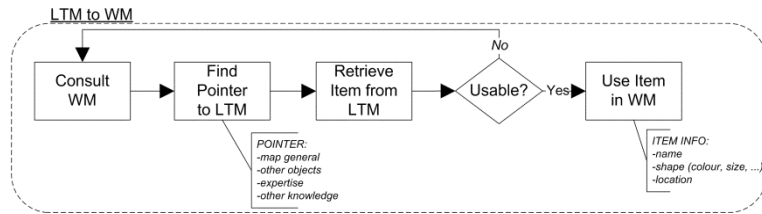


Figure 6 Psychological model of the information retrieval process (from LTM to WM)

In Table 3, the proposed codes related to the participants' level of confidence are listed. The assignment of these codes was based on the level of confidence that the participant expresses during the completion of the task. These 'expressions of confidence' can take two forms. First, the participant can say that he is (not) sure: 'I am not sure', 'I suspect', 'I guess', etc. Second, the level of confidence can be derived from the actual verbal utterances on the videos: intonations, pauses between words, etc. Both types are combined to assign a level of confidence (confident, not confident, or neutral) to each segmented protocol (full thought).

Table 3: Proposed codes related to the participants' level of confidence (Level 3)

<i>confidence</i>	<i>category</i>	<i>code</i>	<i>description</i>
Confident	confident	cf3	<i>participant is sure that it is correct</i>
Neutral	neutral	nt3	<i>level of confidence is not expressed</i>
Not Confident	not confident	nc3	<i>participant is not sure that it is correct</i>

The codes in Table 4 capture the actual actions of the participants during the task. While thinking out loud (talking), the participant was also doing something else to complete the assignment: drawing, checking the drawing, erasing a part of the drawing, taking a pencil, etc. Since these actions occurred concurrent with the verbalisations, they had to be coded separately.

Table 4: Proposed codes related to the participants' actions (Level 4)

<i>action</i>	<i>category</i>	<i>code</i>	<i>description</i>
Draw	draw	dw4	<i>participant draws an item</i>
Take Pencil	take pencil	tp4	<i>participant takes a pencil</i>
Correct	correct	cr4	<i>participant corrects item that was drawn previously</i>
Erase	erase	er4	<i>participant uses the eraser</i>
Fill Colour	fill colour	fc4	<i>participant is filling up items with colour</i>
Checking	checking	ck4	<i>participant checks drawing</i>
Talking	talking	tk4	<i>participant is only talking</i>

In contrast to previous tables, the codes listed in Table 5 were not assigned to the sentence segmentations. Multiple codes would have to be assigned to a single segmented protocol in this case. These codes were used to categorise the aggregated word segmentations. The number of occurrences of each word was counted, and a code was assigned to a subset of them. The categories listed in Table 5 are based on how the visual information is encoded in the users' memory (see Figure 5 and Figure 6). As mentioned in Section 1.1, visual images of maps can be described by concepts or features (their shape,

shape, location, colour, etc.) and structural information (spatial framework, distances, directions, relations, etc.) (Kulhavy & Stock, 1996; Thorndyke & Stasz, 1980).

Table 5: Proposed codes related to the word segmentation (Level 5)

<i>word categories</i>	<i>category</i>	<i>code</i>	<i>description</i>
Colour	colour	cl5	<i>object is described by colour</i>
Geometry	geometry	gm5	<i>object is described by its geometry (size, shape,...)</i>
Name Object	name object	no5	<i>object is recognised and named (forest, city,...)</i>
Location	location	lc5	<i>object's location is described (left, right, above,...)</i>
Strange Item	strange item	si5	<i>strange item is mentioned in the description of the object (turtle, ghost, Africa,...)</i>

To obtain the coded protocols, the segmented protocols were linked to the proposed codes in the scheme. The resulting coded protocols (of the sentence segmentation) have four codes assigned to each protocol: map level, item level, level of confidence, and action level. This process of assigning codes to protocols was accomplished using the software ELAN. In order to maintain objectivity during the coding process, a subset of the data was coded by three independent coders and their results were compared with these of the main coder. A correspondence of 75.5% between the main coder and the three independent coders was reached, which was considered acceptable.

The coded protocols derived from the word segmentation were obtained differently. The codes listed in Table 5 were used to categorise a subset of the (aggregated) list with words and their number of occurrences. The results that are obtained from the analyses of both segmentations are described in Section 3.

3. RESULTS

3.1. Word segmentation

The initial word segmentation (2,043 different words) was aggregated by grouping words that have the same meaning, such as for example forest and woodland or buildings and houses. This aggregation operation resulted in a list of 1,189 words. Next to each word in the aggregated result, its total number of occurrences was listed, separately for each user group. However, these aggregated data showed considerable differences in the total word count between expert and novice participants, with 21,332 versus 13,747 respectively. This difference could be explained by the fact that no fixed time limit was set on the completion of this part of the assignment. Experts took generally longer to complete it. In order to be able to compare the use of wording between experts and novices in an objective way, a normalised set of values had to be used. These normalised values were calculated by setting the total word count of each group equal to 1,000. The occurrence of each word thus corresponded to its ratio in comparison to the total word count of 1,000 (in ‰).

The total number of occurrences for each word (sum of experts and novices) were ordered from largest to smallest. The word with the highest total number of occurrences is ‘een’ (‘a’ in English), with a ratio of 97.42‰ and 89.44‰ for the expert and novice participants respectively. These ratios were compared between the expert and novice group for each word individually: $(\text{exp}\% - \text{nov}\%) / (\text{exp}\% + \text{nov}\%)$. The result is a value between -1 (only the novices used the word) and 1 (only the experts used the word). In the case of the word ‘een’ this values is 0.043, which indicates that the experts used it only slightly more than the novices.

Next, these test values (TVs) were classified to indicate the level of difference in frequency of word use between expert and novice users. The first class contains the test value (and thus the related words) that show very little difference between the expert and novice users $[0.00;0.25]$. In the second class, this test value is between 0.25 and 0.50, indicating a more pronounced difference. The third class contains words that have an even larger deviating frequency of use between both user groups, with a test value in the interval $]0.50;1.0[$. Finally, the last class contains the words that were only used by one group of users. These categories are thus based on the absolute test values. However, the sign of these test values indicates which user group used the word more frequently: a negative result indicates that the novices used the word more frequently; a positive result indicates that the experts used the word more frequently.

Figure 7 depicts an overview of the occurrence of these test values for expert and novice users across the first half of the ordered word list (first 50% or 595 words). The first group on the left side of the graph depicts 10% of the words that are most frequently used; the second group corresponds to the next 10% of words in the list, etc. The left bar in each group visualises all positive test values (experts use them most frequently); the right bar all negative test values (novices use them most frequently). The sum of both graphs always corresponds to 119 words (10% of the list with words). The shades of gray indicate the amount of difference with the other user group, based on the TVs' classes.

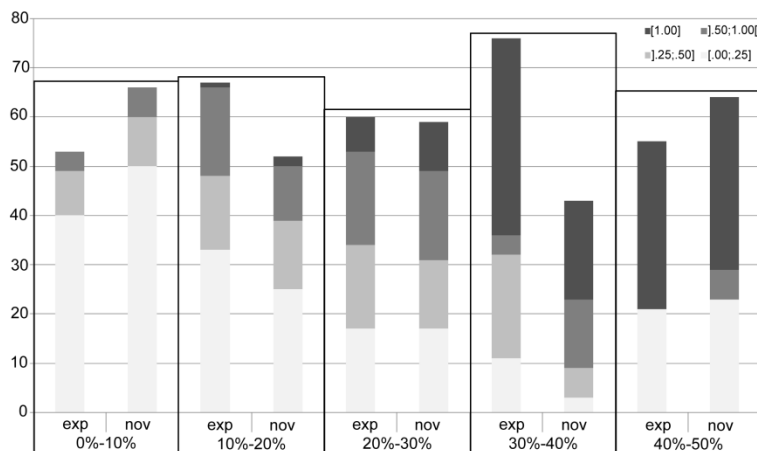


Figure 7: Overview of the ordered word segmentation (first 50% of words) and their differences in word frequencies between expert and novice users

A trend can be noticed in the bar chart in Figure 7. The most occurring words (first 20 to 30%) show a very high similarity between the experts and novices: very high light grey bars. The height of these light grey bars (associated with a very small difference of less than 0.25) decrease systematically until the group of 30-40% of the data is reached. The height of the bars associated with the second category (TV between 0.25 and 0.50) is relatively stable. However, the bar heights of the highest categories (TV between 0.50 to 1) show an increase over the first three groups. Furthermore, the highest category (which indicates that only one user group used a certain word) shows a considerable increase over all groups. It can thus be concluded that the words that are most frequently used are very similar between expert and novice users. However, a diverging trend is noticed between experts and novices regarding words that are less frequently used. The less frequently used words are thus influenced by the users' level of expertise.

When studying the list of 119 words that were most frequently used (first 10%), it was noticed that these were all very general, commonly used words: 'tekenen', 'ook', 'dan', 'in', 'links', 'huis',

'kant', etc. (in English: 'drawing', 'also', 'than', 'in', 'left', 'house', 'side'). Words that were more frequently used by the novice users ($TV > 0.50$) seem to correspond to general descriptive terms such as 'green', 'middle', 'red', 'line', 'area'. In the second group of the ordered word list, the distinction in the use of specific wording is starting to show. The words that were used (considerably) more frequently ($TV > 0.50$) by the novices correspond again to general descriptive terms, including colours: 'orange', 'part', 'surface', 'both', 'beige', etc. The words that were more frequently used by the experts seem to reflect their level of expertise: 'north', 'field', 'core', 'south', 'meander', etc.

This initial trend is also visible in the subsequent groups in the ordered list. Experts seem to be able to name the objects using the exact terms, such as 'farm', 'drainage-basin', 'cluster', 'industry', 'agglomeration', 'quarry', 'electricity wire', 'marshland', etc. The novices used more general words to describe the shape of the objects, including comparisons with other objects that seem strange in this context, such as 'turtle', 'ghost' or 'Africa'. Furthermore, novices seem to use relative object locations more frequently ('top right', 'below left', 'inside', 'straight ahead', etc.). This is in contrast with the experts who used wordings like 'north', or 'southwest' to describe locations. These differences in word use to describe the map content are more pronounced near and in the second half of the ordered list of words.

In Figure 8, the word segmentations are grouped based on the codes presented in Table 5: the expression of colour, geometry, locations, the name of objects, and strange objects. Likewise as in Figure 7, a comparison is made regarding the frequency with which expert and novice participants used the words in these thematic categories. The calculation and classification of these test values (TVs), to analyse which words are more frequently used by experts or novices, is done in the same way as with the ordered word list. The two most striking differences in Figure 8 are to be found in the categories 'geometry' and 'object'. The novice users show a tendency to describe objects based on their geometry: 'line', 'surface', 'thin', 'curve', 'bloc', etc. This is also found in the verbalisations of the experts, but less pronounced.

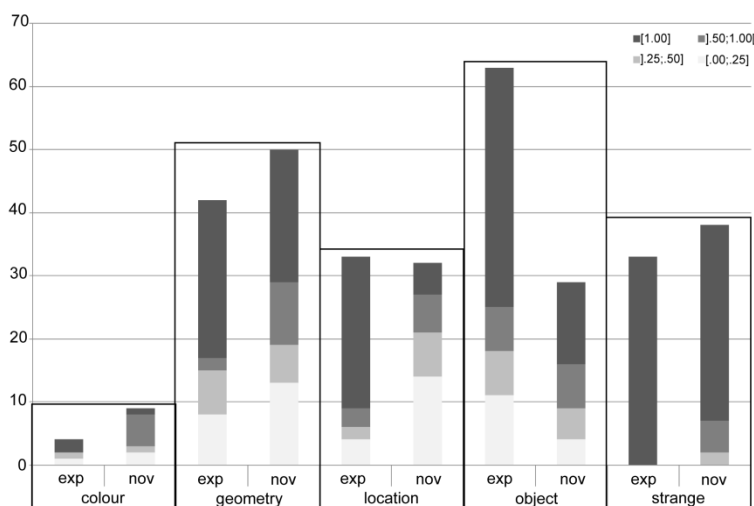


Figure 8: Overview of the word segmentation, based on a thematic categorisation

The experts tended to use the actual name of the object instead of giving the description of its shape. This is reflected in the two bars related to the thematic category 'object'. The experts' bar is much higher, with a large section depicted in dark grey colours. This indicates that experts used these words much more frequently than novices. Even when studying the actual words that the novices used more frequently than the experts, it should be noted that these are less specific or detailed objects. The novices, on the one hand, used for example 'plain', 'crossroad', 'ground', 'parcel', 'residential area', 'field', 'hill', 'sand', 'dirt track', etc. The experts, on the other hand, named more detailed objects such as 'drainage-basin', 'meander', 'bridge', 'farm', 'quarry', 'coniferous', 'electric wire', 'marshland', 'hamlet', 'firebreak', etc.

Only two colours were used considerable more by the expert participants: purple and pink. These colours correspond to less obvious objects in the map, such as the electrical wires. Furthermore, the large portion of dark grey in the experts' bar related to location can be explained by the use of compass directions such as 'north', 'northwest' or 'west'. These were almost never used by novice users. Finally, the thematic groups that contains the 'strange words' is more pronounced with the novice users, which can also be explained by the fact that they tried to describe the shape of the objects instead of naming them. They used 'strange objects' with a similar shape to describe this. The words 'Asia', 'Africa', and 'turtle' were for example used by multiple persons to describe the shape of different forest patches. Experts also used these comparisons with other objects to describe the shape of objects, but less frequently. Strange words only used by the expert group were for example 'candlestick' and 'catapult' to describe the shape of the main roads.

3.2. Sentence segmentation: coded protocols

Besides the word segmentations, the verbal protocols were also segmented based on the criteria of 'full thought'. The full thoughts can be marked both by linguistic elements (sentences) and actions (longer pauses). This coarser segmentation granularity allows taking the contexts of the individual words into account. Furthermore, one segment (a full thought) can more easily be linked to a specific cognitive process. A psychological model and coding scheme was proposed in a previous section (see Section 2.4). Each segment was linked to a cognitive process using this coding scheme, resulting in coded protocols. When analysing these coded protocols, the time ratio of each segment (or code) in comparison to the whole trial duration, was taken into account. This time ratio was summed for all corresponding codes, taking into account the four different stimuli and two user groups. The results of these coded protocols are presented in Figure 9. The total sum of all bar heights from one user group thus corresponds to 1 within each graph. For example for map 1 the experts' results for each separate code in Level 1 are $ev1=.219$, $ex1=.754$, and $or1=.027$. The summed time ratios of these codes should thus cover the total trial duration, which is equal to one.

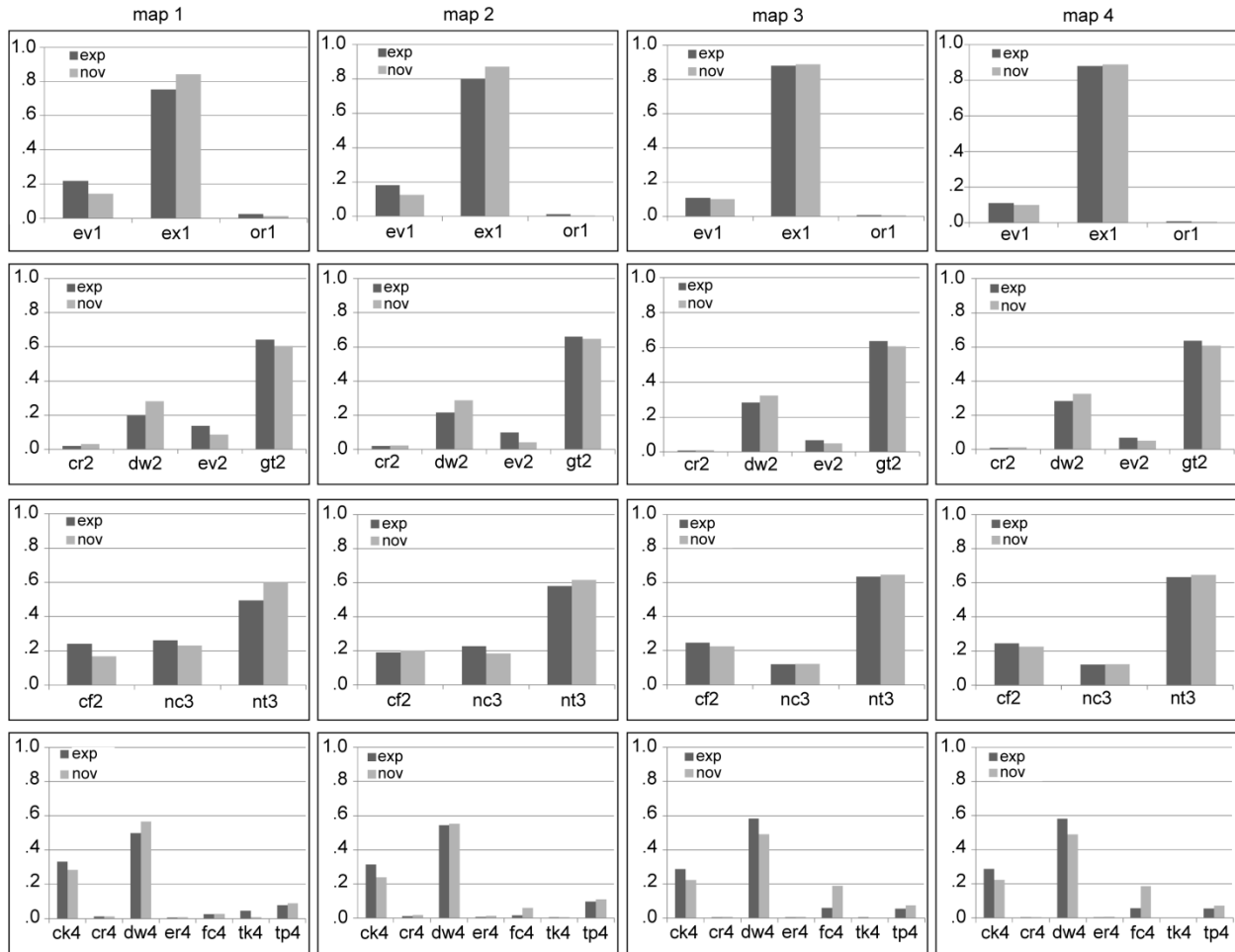


Figure 9: Overview of the segmented protocols (full thought) per level and map

A first glance at these graphs shows that the general trend – the time ratios for each code or cognitive process – is similar for both user groups; no strong deviations can be noticed. What is more, these general trends are repeated over the four different stimuli (or maps) for each level. This means that the ratio of the different cognitive processes that were addressed to solve the problem at hand (remembering and drawing the maps) are similar for both user groups and independent of the actual map content (different topographic maps in this case). At the first level, the shortest time interval is associated with the orientation process, then the evaluation of the results, and then the execution of the task. This latter process took considerably more time than the others (about 80%).

On the second level, the test persons spent most time gathering their thoughts (around 60%). This corresponds to the consultancy of their WM to check whether suitable information is already present. If not, this information is retrieved from the LTM. Next, about 20% of the time was spent on the drawing task. On this level (Level 2), the code *dw2* was noted down when the participant was not verbalizing his thought, only drawing or taking a pencil. This pause in the verbalisations during the thinking aloud assignment can be explained by the interference of two tasks in the WM. The drawing task consumes a lot of the cognitive load available in the WM, and as a consequence the participant could not verbalise his thoughts anymore. However, the interference between the drawing and thinking aloud task was not always as strong that the participant stopped talking every time he drew something. The graphs related to

Level 4 (which list all actions, independent of the verbalisations) show that the participants were drawing about 50-60% of the time and taking a pencil during about 10% of the time, which is much more than the 20% that was registered on Level 2. The participants spent about 5-10% of their time evaluating the objects which they had just drawn, and only 1-2% on the correction of these evaluated objects.

While drawing, the participants could express their level of confidence by using specific words (I think, I guess, I am sure, etc.) or by intonations and hesitations. Most of the time, no clear expression of confidence was detected (*nt3*). The expressions of confidence versus not confidence seem to be in balance for all four stimuli (10-20% of the time). The actions that the participants performed during the verbalisations were recorded on Level 4. As mentioned before, the participants spent most of their time drawing, which is followed by checking (*ck4*). This checking corresponds to evaluating the map (is everything on it, do I miss something, counting the number of objects on the map, etc.) and evaluating individual objects (is the shape correct, is the location correct, etc.). The smallest intervals are related to the correction of items (*cr4*), erasing items (*er4*; which is a part of correcting on the second level, *cr2*), filling the objects' interior with colours (*fc4*), and talking without any other operation (*tk4*).

A more detailed look on these graphs allows identifying smaller variations in the cognitive behaviour of expert and novice map users. The experts, for example, spent more time evaluating the result, whereas the novices dedicated more time to the execution of the task. These differences between expert and novice users tend to become smaller over each trial. On the second level, the experts seem to devote more time on gathering their thoughts (*gt2*) and evaluating their objects (*ev2*) than the novices, who spent more time drawing (*dw2*). Furthermore, during the first trial expert participants expressed their confidence (both higher and lower) more than novice participants. However, during the last two trials, the expressed level of confidence is very similar. The experts' tendency to check and evaluate their results more than the novices, (see Level 1 and Level 2) is also reflected in the participants' actions (*ck4*). Nevertheless, the bar heights related to code *dw4* (drawing) show a variation between the expert and novice users. In the first trial, the novices used considerably more of the trial time to draw, which is nearly equal between both user groups for the second trial. For the third and the fourth trial, however, the experts spent more time drawing.

3.3. Sentence segmentation: contexts

While analysing the segmented protocols based on the full thought, it came apparent that experts also used their knowledge to place objects on a logical position, although they might not even remember that object. This deductive reasoning enabled them to solve the assignment more efficiently. For example, if experts drew a village they knew it should be connected to at least one road and then they tried to find a logical location for this road. If there was a valley with a river in it, the experts found it logical that there is also a road and a railway located next to this river. Furthermore, experts tried to figure out the reason behind the structure of the map: the direction of the rivers' flow, elevations, physical properties of the soil, etc.

The novice users did not have the same geographical background knowledge, and were less tempted to use deduction to place an object. They would guess, place an object on a certain location because that part of the map was still rather empty, but they did not give a specific (valid) reason for doing so. As mentioned before, novice users tended to describe the objects instead of naming them. However, when the novices did name an object, it was noticed during the analyses, that they often used a wrong name. They, for example, said that they were drawing a railway, whereas this was in fact not the case. This 'railway' was presented by a thick red line on the screen map, which actually corresponded to a

major road. The novices also talked about ‘sandy areas’, whereas these were in fact fields (depicted in yellow).

Some of the colours in the second map were adapted in order to study the influence of this deviating colour use on the participants’ cognitive processes. The colour of the water bodies was changed from cyan to light orange and the background of the villages from light yellow to purple. Most of the experts said that the second stimulus had objects with a natural shape, similar like a meander (top left object) or a drainage-basin (top right object). Some said that it would have been a river if it was blue, but considering its colour might be some kind of quarry. Most of the novices just described it as a capricious beige spot, looking like a worm. Two of them mentioned that it could have been a river if it was depicted in blue. No influence was found regarding the deviating colour use for the village backgrounds.

The mirror operation between the first and last stimuli was also mentioned by several participants at the beginning of the fourth trial. The participants said that they recognised the objects and suspected that the fourth map was a mirrored version of the first one. However, several participants also indicated that this did not facilitate the task: they recognised the objects but they were on a different position. They said that it was difficult to distinguish between the objects in the first and fourth map.

3.4. Score on drawn maps

The final maps that were drawn by the participants were compared with the original screen maps. Figure 10 shows four examples of how the fourth map stimulus was drawn by some participants. Based on how well these maps were drawn, a score was assigned to them. The criteria ‘how well’ was mainly based on the question, ‘is the object present on the map?’. The aesthetics of the sketched objects or drawing skill of the participants were not taken into account. A similar scoring system as used by Kulhavy and Stock (1996) and Thorndyke and Stasz (1980) was applied to obtain a score for each map. If the object was present and drawn on the correct location a score of one point was assigned to it. If the object was present, but drawn on a considerably wrong location, only half a point was assigned to it.

The objects that were drawn on the four maps contained villages, rivers, lakes, farms, roads, fields, forests, railroads, etc. In map 1 and map 4, 34 objects were checked, 27 in map 2, and 32 in map 3. These fit in the 19 categories of objects that are typically drawn on sketch maps, as identified by Blaser (2000). These 19 categories are aggregated into five major themes: hydrography, land cover, settlements, roads, and other. The class ‘other’ contains some detailed objects such as electrical wires and islands in the river. The obtained scores were recalculated to percentages (%). These scores are presented in Figure 11, on the left side for each theme (all maps) and on the right side for each map independently.



Figure 10: Examples of some maps sketched by the participants (fourth stimulus)

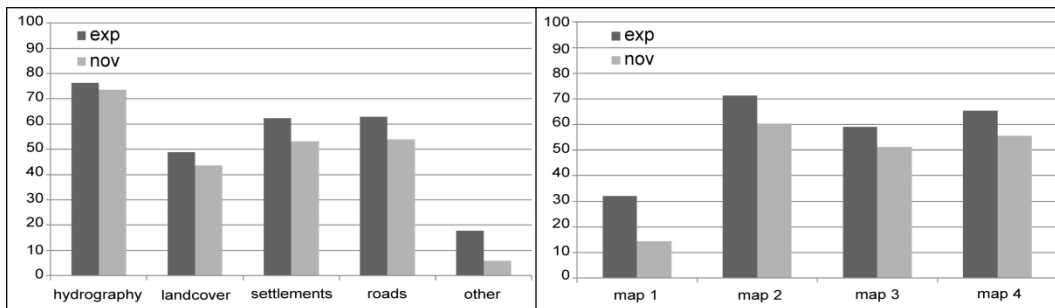


Figure 11: Overview of the scores (%) on the drawn maps, per theme (left) and per map (right)

From these graphs it can be derived that the expert participants score considerably better on all categories than the novices. The most pronounced difference can be found in the theme ‘other’, which contains less obvious objects on the map. Both user groups obtain the best scores in the theme hydrography. This seems to correspond to the findings regarding the order with which the objects were drawn (see Section 3.5). These hydrographic objects were drawn more frequently in the beginning of the assignment, resulting in more accurate results. Next are the settlements and the roads. These roads were also frequently drawn in the beginning of the assignment.

The lowest overall score was obtained for the first map, both by the experts and novices. This could be explained by the fact that this was the first time they had to execute this assignment. Most participants seemed to have underestimated the assignment and did not study the map good or long

enough. The second map, however, corresponds to the highest score: the participants seemed to have studied the stimulus better than the initial one. The fourth map shows again an increase of the score in comparison to the third map. However, this higher score is not as pronounced as could be expected. This fourth map is the mirrored equivalent of the first map, and contains consequently the same objects. Since the participants saw these objects for the second time (however mirrored) it could be expected that they would have remembered this in more detail. As can be derived from the bar heights, the participants' performance is not significantly better. During this trial they seemed to be confused by the mirrored appearance of these objects.

3.5. Order of drawing

The order with which the different structuring elements on the map were drawn can give insights in how these elements were stored in the users' memory. Elements that are systematically among the first objects to be drawn are more accessible in the LTM; they can be retrieved more easily. By studying patterns in the order with which objects were drawn, it can be derived which map elements are considered more important, and are retrieved more easily. In their article, Huynh *et al.* (2008) also investigated the order with which elements are recalled while drawing a sketch map. They argued that this sequence gives insights in the hierarchical structure of the users' cognitive map: a basic framework in which lower level elements are subsequently placed. This hierarchical structure is closely linked with the perceived importance of objects by the user (Huynh & Doherty, 2007; Huynh, *et al.* 2008).

For each map, eight categories of map elements were identified. These categories contain structuring map elements and are an extension of the four categories used to present the map scores: electrical wire (other), field (landcover), forest (landcover), settlement, hydrography, meadow (landcover), railroad (other), and road. A 'score' was assigned to each category based on the order with which the first element of that category was drawn: 100 for the first element, 50 for the second, 25 for the third, 5 for the fourth, 4 for the fifth, and so on. If a certain element was not drawn, no score was assigned. The graphs visualising the experts and novices their results are depicted in Figure 12, for each map separately. The three categories that are associated with the highest scores are the forests, roads and hydrography. These form the main structuring elements on the maps.

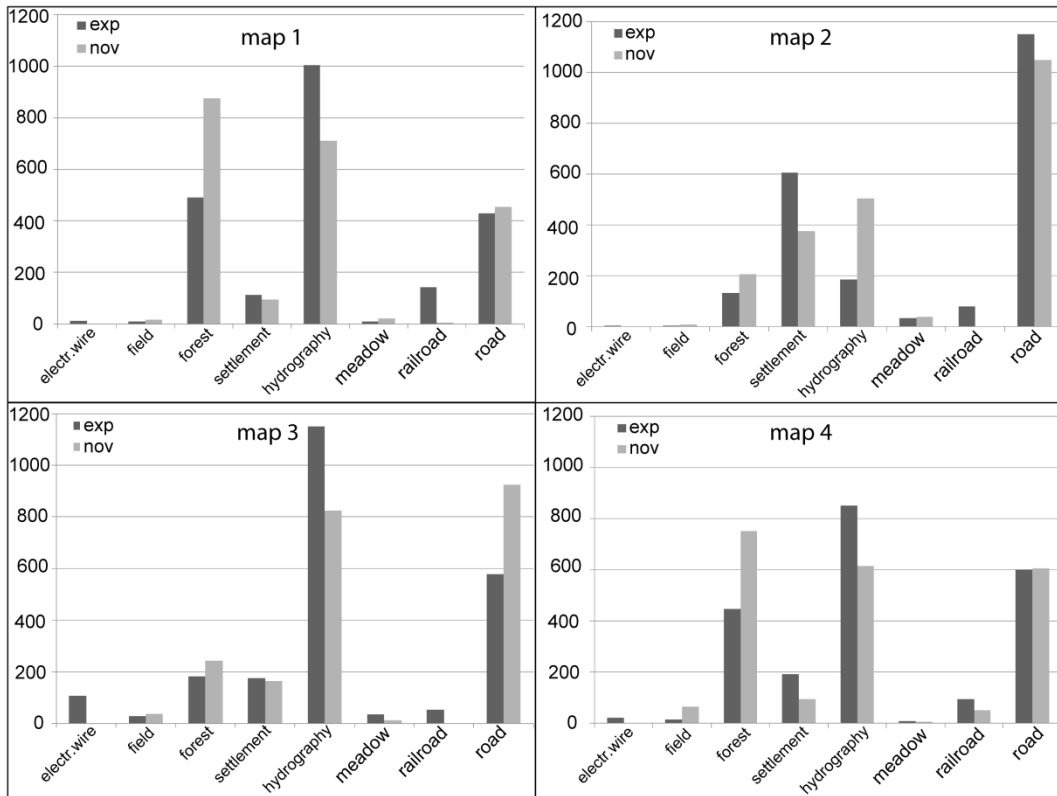


Figure 12: The order with which the objects were drawn for each map

In the first map, subjects first drew the forests, the hydrography, and the roads; however a different order for the experts and novices can be noticed. Most novices drew first the forest, than the hydrographic elements, and then the roads, whereas the experts mainly chose to draw the hydrographic elements at first. Less difference can be detected between the draw order of the roads and forests for this user group. The same structure of these bars (their relative heights) is also present in the graph related to map 4. This last stimulus was the mirrored version of the first one, which means that the same elements were present. However, the bar that represents the score assigned to the hydrography is not as high, especially not for the expert groups. The novices' bar related to the forests is also much lower. Nevertheless, both user groups tended to draw the roads earlier than in the first map. Finally, the detailed objects also received a higher 'score' in the fourth map. This indicates that more participants remembered these objects when the map object (although mirrored) were displayed for the second time.

In the second map, the roads were drawn most frequently as initial objects. This is followed by habitation (experts) or hydrography (novices). The colour of the hydrographic objects was adapted in this map, which could explain the lower score in comparison to the other maps. In the third map, the experts drew the hydrographic elements most frequently as the initial objects, whereas the novices focused more on the roads. All other objects received a much lower score. In all maps, the objects 'railroad' and 'electrical wire' were never drawn as first element. However, from the graphs it could be derived that more experts drew these element than novices.

3.6. Questionnaire results

After the participants had completed the four trials, they had to fill in a questionnaire. In this questionnaire, they had to report personal characteristics such as age, gender, level of education, and so on. They also had to indicate if they recognised any of the regions that were displayed during the study, as this could bias the results. None of the 24 participants indicated that they recognised one of the regions. They were also asked if they worked with Belgian topographic maps before, how often, and in what context. On this question, the experts noted down that they did use this type of map before during their studies and sometimes still do for their work. Most novices had not worked with these maps before, or not frequently. Furthermore, the participants also had to indicate if they noticed the deviating colour use on the second stimulus and that the fourth stimulus was a mirrored version of the first one. On this latter question, most participants (22 out of 24) said that they did notice it. However, few participants seemed to have noticed the adapted colour use on the second stimulus: 6 out of 24, from which four were experts.

Finally, the participants also had to indicate how well they thought they had drawn each of the maps on a five point scale (ranging from 'very bad' to 'very well'). A score was assigned to each option (-2|-1|0|1|2) and these were averaged for each map, separately for the two user groups. These average scores of confidence are presented in Figure 13. No clear trend can be noticed between the expert and novice users. Furthermore, the level of confidence regarding 'how well' the maps were drawn does not agree between the two user groups.

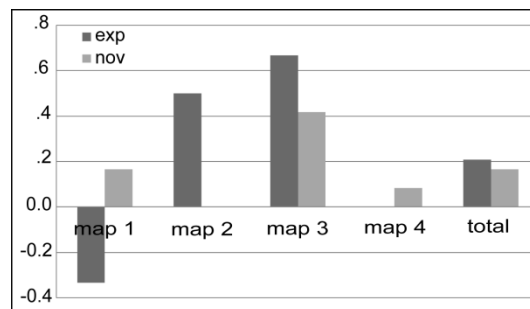


Figure 13: Result of the questionnaire regarding the confidence of the result (drawn map)

The experts are most confident regarding the second and the third map, whereas the novices seem to appreciate their results for the first and last map more. Only the experts have an overall negative score for one map, which is the first one. This was the initial map and the experts seem to be most displeased with these first drawn maps. The score of their fourth map is equal to zero, indicating that they were not pleased with the result, but also not displeased. The novices gave a better score for their fourth map in comparison with the experts, although this is also not considered very good. Both user groups seem to be most confident regarding the result of the third map, which is more pronounced in the expert group. This, however, is also not reflected in the actual map scores (see Figure 11). The novices are most displeased with their results regarding the second map.

4. DISCUSSION

The thinking aloud method was used to study the map users' cognitive processes during a drawing task. The participants had to retrieve information regarding a screen map they had just studied from their LTM. The verbalisations and derived segmented protocols can be linked to the different cognitive processes that were addressed during the drawing assignment. These protocols give direct, unfiltered insights in the

information contained in the WM, which had to be retrieved from the LTM. Especially the difference between expert and novice map users was considered in this.

The analyses of the segmented protocols (full thoughts and word counts) indicate that the structure of the users' cognitive processes during this task is very similar. What is more, both user groups first draw certain main structuring elements on the maps, a reference frame, and then the detailed objects. This reference frame is generally based on the location of the main rivers or roads in the map. In their experiments, Huynh and Doherty (2007) and Huynh, *et al.* (2008) obtained similar results. They argued that the sequence of drawing (and thus the recalling from memory) corresponds to the hierarchical structures of the schemas stored in the LTM. They found that their participants tended to draw first the main linear structures and only later on the landmarks. They concluded that these cognitive maps are thus hybrid (consisting out of both paths and nodes) and hierarchically structured.

However, the experts' results (the maps that were drawn) contain more, and more detailed objects than the novices. This better map score can, on the one hand, be explained by a higher efficiency in the interpretation of maps (Ooms, *et al.*, 2012; 2013). These authors analysed experts and novices their eye movements while studying a screen map with a very basic design (Ooms, *et al.*, 2012) and topographic screen maps (Ooms, *et al.*, 2013). They concluded that the experts can interpret a larger part of the map in the same amount of time as novices. As a consequence, experts can store more detailed information regarding these maps. Furthermore, the experts seem less distracted by deviating elements on the map, such as strange colours or mirrored objects.

The analysed protocols also indicate, on the other hand, that experts can retrieve the stored information regarding a map they had studied more efficiently. They know the name of specific objects on the map, which can serve as a pointer to schemas in the LTM that contain more extensive information regarding these objects. These pointers facilitate information retrieval from the LTM: the object itself and all related background knowledge. This information comes in the form of a tight group that can form one chunk of information in the WM. Furthermore, this background information can be used in deductive reasoning so that other information (for example, the logical position of linked objects) can be derived from it. This means that experts need to store less information in the LTM because they can retrieve it from knowledge that was already stored.

The background knowledge of the experts is not only related to map use and cartographic syntax, but also regarding geography. In order to be able to interpret a map correctly, the cartographic expertise is of most importance. However, when storing and retrieving information this geographical background knowledge has also a major impact in the users' cognitive processes: structures in the landscape (past situations, evolutions), physical properties of certain sub soils (vegetation that can grow on it, locations of quarries), economical characteristics (typical locations of industry, structure of road networks), etc. The experts that took part in the experiment have at least a Master degree in Geography or Geomatics and have considerable knowledge regarding these geographical issues, besides the cartographic expertise. While investigating the Knowledge-Is-Power hypothesis, Hambrick and Engle (2002) also demonstrated this facilitative effect that domain knowledge can have on the performance of the users' memory.

The novices only store and retrieve the information regarding the objects they saw, most of the times without linking it to other information: 'a red line going from top to bottom over the map'; 'in the upper part of the map there was a large dark green spot'; 'there were several yellow spots distributed across the map'; etc. This information is thus stored separately from other information in the LTM, which makes it more difficult to retrieve it. The information that can be retrieved is also less: a description of the shape, the colour and its position (possibly relative to another object on the map). Due to the lack of

additional information in the WM, the novices cannot use deductive reasoning to retrieve the position of objects that they did not remember. The novices, for example, need to remember that there was a thick red line (road) and its position in order to be able to draw it, whereas the experts know (based on deduction) that there has to be a road between two large villages.

Thorndyke and Stasz (1980) and Gilhooly, *et al.* (1988) did not find any difference in the number of items that experts and novice could draw from memory when considering planimetric maps or de non contour objects of contour maps. They found that the novices spent more time retrieving names, whereas the experts retrieved more information through both lower level and specialised schemas. The stimuli used in the current study were topographic maps on which no names and contour lines could be found. However, the content of the map was rather complex, with many different object types. This complex content could explain why a difference was found in the number of objects that were drawn on the sketch maps by expert and novice participants.

5. CONCLUSION AND FUTURE WORK

It can thus be concluded that the experts can retrieve the information from the LTM more efficiently, which also results in the retrieval of more information. Additional background information regarding the objects depicted on the map is also retrieved and placed in the WM. This additional information allows the expert users to use deductive reasoning to derive extra information regarding the map objects. As a consequence, the experts do not have to store this 'extra information', in contrast to the novices who cannot use deductive reasoning simply because they do not own as much additional information regarding the objects. Since they cannot derive the information, the novices have to store it as such, which consumes more space in the WM. The additional background information that experts have is contained in the schemas and thus in the (larger) chunks of information that are transferred to the WM.

These findings correspond to what was described by Gobet and Simon (1996, 1998) regarding the expertise in chess, and which was also found in other domains (among others, bridging and computer programming) (Gilhooly, *et al.*, 1998). Experts can access more schemas in the LTM based on the higher amount of background knowledge. This information can be grouped to form larger tight chunks of information that can be retrieved and placed in the WM. Hambrick and Engle (2002) argued that domain knowledge can function as a retrieval structure which enables the retrieval of specific domain-related information.

These insights in the map users' cognitive structures, and the differences due to expertise, are an essential contribution to obtain a better understanding of the actual end user of the fast evolving digital cartographic products. Especially the cognitive limits of the novice map users and the differences with the experts are crucial to be able to keep creating effective cartographic products in the future. Nevertheless, these products are often characterised by possibilities for user interactions and the display of animations. In this (initial) study, only static screen maps are considered, which should show the same information as paper maps. However, the limited dimensions and resolution of the screen could have a profound impact on how the information is perceived. Future studies, considering user interactions and animations, should thus not be compared with the results obtained with (static) paper maps, but with the static digital maps. These different types of interactions and animations could have a major influence on the map users' cognitive processes (and their processing capabilities) and should thus be taken into account in future research.

ACKNOWLEDGEMENTS

We would like to thank the Belgian national mapping agency (NGI, 'Nationaal Geografisch Instituut') to make a number of digital maps from the Belgian topographic map series on 1 : 10 000 available to us, and thus making it possible to incorporate them in the user study.

REFERENCES

- Bertin, J. (1967). **La sémiologie graphique**. Mouton-Gauthier-Villars, Paris.
- Blaser, A. D. (2002). 'A study of people's sketching habits in GIS'. **Spatial Cognition and Computation**, 2(4), pp. 393-419.
- Cartwright, W. (2012). 'Neocartography: Opportunities, issues, and prospects'. **South African Journal of Geomatics**, 1(1), pp. 14-31.
- Chi, M. T. H. (1997). 'Quantifying qualitative analyses of verbal data: A practical guide'. **Journal of the Learning Sciences**, 6(3), pp. 271-315.
- Cowan, N. (2001). 'The magical number 4 in short-term memory: A reconsideration of mental storage capacity'. **Behavioral and Brain Sciences**, 24(1), pp. 87-114.
- Downs, R. M., and Stea, D. (1997). **Maps in minds. Reflection on cognitive mapping**. Harper & Row, New York.
- Ericsson, K. A. (2006). 'Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks'. In: K. A. Ericsson, N. Charness, P. J. Feltovich and R. Hoffman (Eds.), **The Cambridge handbook of expertise and expert performance**. Cambridge University Press, Cambridge; New York.
- Ericsson, K. A., and Simon, H. A. (1980). 'Verbal reports as data'. **Psychological Review**, 87(3), pp. 215-251.
- Fabrikant, S. I., and Lobben, A. (2009). 'Introduction: Cognitive issues in geographic information visualization'. **Cartographica**, 44(3), pp. 139-143.
- Gobet, F., and Simon, H. A. (1996). 'Templates in chess memory: A mechanism for recalling several boards'. **Cognitive Psychology**, 31(1), pp. 1-40.
- Gobet, F., and Simon, H. A. (1998). 'Expert chess memory: Revisiting the chunking hypothesis'. **Memory**, 6(3), pp. 225-255.
- Gilhooly, K. J., Wood, M., Kinnear, P. R., and Green, C. (1988). 'Skill in map reading and memory for maps'. **Quarterly Journal of Experimental Psychology: Human Experimental Psychology**, 40(1), pp. 87-107.
- Hambrick, D. Z., and Engle, R. W. (2002). 'Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is power hypothesis'. **Cognitive Psychology**, 44, pp. 33-387.
- Haklay, M., and Nivala, A. (2010). 'User-centred design'. In: M. Haklay (Ed.), **Interacting with geospatial technologies**. Wiley-Blackwell, West Sussex.
- Haklay, M., Singleton, A., and Parker, C. (2008). 'Web mapping 2.0: The neogeography of the GeoWeb'. **Geography Compass**, 2(6), pp. 2011-2039.
- Hirtle, S. C., and Jonides, J. (1985). 'Evidence of hierarchies in cognitive maps'. **Memory & Cognition**, 13(3), pp. 208-217.
- Huynh, N. T., and Doherty, S. T. (2007). 'Digital sketch map drawing as an instrument to collect data about spatial cognition'. **Cartographica**, 42(4), pp. 285-296.
- Huynh, N. T., Hall, C. B., and Smith, W. W. (2008). 'Interpreting urban space through cognitive map sketching and sequence analysis'. **The Canadian Geographer**, 52(2), pp. 222-240.
- Kulhavy, R. W., and Stock, W. A. (1996). 'How cognitive maps are learned and remembered'. **Annals of the Association of American Geographers**, 86(1), pp. 123-145.
- MacEachren, A. M., and Kraak, M. (2001). 'Research challenges in geovisualization'. **Cartography and Geographic Information Science**, 28(1), pp. 3-12.
- Matlin, M. W. (2002). **Cognition** (5th ed.). Harcourt College Publishers, Fort Worth.

- Miller, A. M. (1956). 'The magical number seven, plus or minus two: Some limits on our capacity for processing information'. **The Psychological Review**, 63(2), pp. 81-97.
- Montello, D. R., Sullivan, C. N., and Pick, H. L. (1994). 'Recall memory for topographic maps and natural terrain: Effects of experience and task performance'. **Cartographica**, 31(3), pp. 18-36.
- Montello, D. R. (2009). 'Cognitive research in GIScience: Recent achievements and future prospects'. **Geography Compass**, 3(5), pp. 1824-1840.
- Muenzer, S., and Stahl, C. (2011). 'Learning routes from visualizations for indoor wayfinding: Presentation modes and individual differences'. **Spatial Cognition and Computation**, 11(4), pp. 281-312.
- Nielsen, J. (1993). **Usability engineering**. Morgan Kaufmann San Francisco.
- Nielsen, J. (1994). 'Estimating the number of subjects needed for a thinking aloud test'. **International Journal of Human-Computer Studies**, 41(3), pp. 385-397.
- Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., and Witlox, F. (2012). 'Interpreting maps through the eyes of expert and novice users'. **International Journal of Geographical Information Science**, 26(10), pp. 1773-1788.
- Ooms, K., De Maeyer, P., and Fack, V. (2013). 'Understanding expert and novice map users: Reading and interpretation'. **Cartography and Geographic Information Science**. (*under revision*)
- Schlender, D., Peters, O. H., and Wienhofer, M. (2000). 'The effects of maps and textual information on navigation in a desktop virtual environment'. **Spatial Cognition and Computation**, 2(4), pp. 432-433.
- Scholl, M. J. (1987). 'Cognitive Maps as Orienting Schemata'. **Journal of Experimental Psychology: Learning, memory, and Cognition**, 13(4), pp. 615-628.
- Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrman, S., *et al.* (2001). 'Cognitive and usability issues in geovisualization'. **Cartography and Geographic Information Science**, 28(1), pp. 61-75.
- Suchan, T. A., and Brewer, C. A. (2000). 'Qualitative Methods for Research on Mapmaking and Map Use'. **The Professional Geographer**, 52(1), pp. 145-154.
- Thorndyke, P. W., and Stasz, C. (1980). 'Individual differences in procedures for knowledge acquisition from maps'. **Cognitive Psychology**, 12(1), pp. 137-175.
- Trickett, S. B., and Trafton, J. G. (2007). 'A primer on verbal protocol analysis'. In: D. Schmorow, J. Cohn and D. Nicholson (Eds.), **Handbook of virtual environment training**. Praeger Security International, Westport, CT.
- Turner, A. J. (2006). **Introduction to neogeography**. O'Reilly Media, Sebastopol.
- van Elzakker, C. (2004). **The use of maps in the exploration of geographic data**. Labor Grafimedia b.v., Utrecht.
- van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C. (1994). **The think aloud method. A practical guide to modeling cognitive process**. Harcourt Brace & Company, London.