Learning and navigating in hypertext: Navigational support by hierarchical menu or tag cloud?

Open Universiteit

www.ou.nl

Running head: DIFFERENCES IN HYPERTEXT NAVIGATION

Learning and Navigating in Hypertext: Navigational Support by Hierarchical Menu or

Tag Cloud?

ABSTRACT

As hypertext learning environments (HLE) are widely used in education, it is important to study and know the effects and consequences of its use. HLEs are non-linear which means that students have to develop ways of navigating through them. Thus, developing interfaces that facilitate and even guide navigation is important for learning. Research showed that successful learning in HLEs depends on both learner characteristics and HLE features. This study investigated an HLE navigation feature (navigational support with either a tag-cloud or conventional hierarchical menu), task complexity (fact-finding vs. information-gathering task) and a user characteristic (gender). Results show that neither navigational support nor gender is associated with differences in task performance. However, there are differences in information processing. Participants using a tag clouds looked longer at the navigational support and shorter at the overview pages. Combined with fewer revisits of webpages in the tag cloud condition, this indicates a more focused selection of pages. The deeper processing of information needed for the information-gathering task was reflected in fewer visits to, but longer viewing times of pages. As no differences in task performance were found, tag clouds seem to be as effective for performance as more traditional navigation structures for navigational support.

Keywords: hypertext learning; navigation behavior; gender differences; eye tracking

1. INTRODUCTION

In today's education, hypertext learning environments (HLEs) are widely used to enrich traditional education. Aleven, Stahl, Schworm, Fischer and Wallace (2003) describe a HLE as an on-demand help containing context specific hints, hyperlinked background material, textbooks and glossaries. Such HLEs are often used for inquiry-based learning, discovery learning, web quests, and so forth; that is in environments where the instruction does not necessarily follow a set order or structure. Due to its inherent nature, the content of these HLEs is presented in *non-linear ways* making it more difficult to navigate the contents than traditional linear text (Scheiter & Gerjets, 2007). A further complication is that HLEs make use of two different kinds of interfaces: the browser interface and the website (HLE) interface (Juvina & Van Oostendorp; 2006). While the browser interface / brand of browser hardly changes and the differences between browsers is minimal, the variation in website interfaces is enormous and often unique for each website. Thus, students have to develop *new ways of navigating* for each new HLE they encounter. For instructional designers, it is therefore important to design website interfaces for HLEs with ease of use in mind, and in such a way that navigation and learning are facilitated. In this study navigation in HLEs with tag clouds is compared with navigation with hierarchical menus. A tag cloud is a visual representation of keywords in the form of a cloud. While the hierarchical menu is standard, the novel tag cloud menu is upcoming. However, little is known about its use, yet. Both Anfinnsen, Ghinea and De Cesare (2011) and Voit, Andrews and Slany (2009) indicate the usefulness of tag clouds in browsing, but stress the need for more research. This study fills this research gap by comparing hierarchical menus with tag clouds. It can be expected that design choices have an influence on navigation behavior. As these influences are not known yet, the most commonly used design choices for tag clouds will be studied.

HLEs can be characterized as a database of pieces of information (often called 'nodes') where relations between the nodes are represented by explicit links (Conklin, 1987; Kim & Hirtle, 1995). In contrast to printed text where pieces of information are organized and made available in a fixed consecutive order, HLE users can jump from one node to another without following a predefined path. One often given advantage of the latter is that users are able to determine which links to follow and in which order (Conklin, 1987) and thus can adapt the amount and type of information they consume to their information needs. As the information nodes are modular, links to each node can be created from several different other nodes. This might increase the efficiency of the learning process (Dee-Lucas & Larkin, 1999) and learner motivation (Mobrand & Spyridakis, 2007). However, a negative consequence of this 'freedom' is that as a result of the less clear structure learners might get lost and consequently show less coherent reading order of text nodes and eventually create a 'fragile' network of knowledge offering "frail and casual webs of information that lead to the cultivation of similarly flimsy mental networks (the "Butterfly Defect)" (Salomon & Almog, 1998, p. 222). Likewise, users need to use additional cognitive processing capacity above the simple processing of the information because they have to decide where to go next and why. As more mental effort is needed to navigate through HLEs, the risk of overload of working memory will be increased as well. As Madrid, Oostendorp, and Puerta Melguizo (2009) concluded, a more coherent reading order of text nodes is correlated to reduced cognitive load. This indicates that design choices have an influence on cognitive load as well.

Previous research showed that the learning success from HLEs depends on learner characteristics (e.g., working memory capacity, expertise in a domain, ability to impose structure on information, gender) in conjunction with the particular features of different types of HLEs

(e.g., hyperlink structure, number of nodes), the complexity and type of task, and the type of navigational support, such as hierarchical or tag cloud (DeStefano & LeFevre, 2007; Puerta Melguizo, Vidya & Oostendorp, 2012). This study focuses on the type of navigational support and the task complexity, taking into account gender differences as an aspect of user characteristics. As little is known about the influence of design choices of tag clouds on navigation behavior, the most commonly used design choices for tag clouds will be studied.

2. BACKGROUND LITERATURE

2.1. Navigational support: Hierarchical or tag-cloud

Guiding learners while navigating through HLEs might reduce the cognitive burden of continuous decision-making, help them build a coherent mental structure of the information, and help them to use more coherent and efficient navigation behaviors thus facilitating learning. Not using such support might, in contrast, influence learning outcomes negatively (Minetou, Chen & Liu, 2008). However it is not clear if this is actually the case and if so, which type of navigational support should be used and how this support best can be designed.

Previous research on HLEs compared different types of graphical overviews (e.g., Amadieu et al, 2009; Bezdan, Kester & Kirschner, 2013) or different types of hierarchical menus (e.g., Leuthold, Schmutz, Bargas-Avila, Tuch & Opwis, 2011; Puerta Melguizo et al., 2012). Generally, one can conclude that a hierarchical organizer is slightly better for navigation compared to a network organizer (Amadieu et al., 2009), using dynamic menus hinders task performance (Bezdan et al., 2013), and it is not wise to severely restrict learner navigation paths (Bezdan et al., 2013). Furthermore, vertical menus are preferred over dynamic menus (Leuthold

et al., 2009). These studies show clearly that design choices have an influence on user

performance. It is therefore important to study the effect of these design choices.

In a hierarchical system, information is organized in a folder structure where folders can

contain subfolders and so further, which results in a tree-like structure. Recently, tag clouds have

come into use as a new tool for navigation. Examples of this tagging approach are Delicious®,

Pinboard® (favorite websites), Youtube® (video), Flickr® (photo's), Connotea® (science) or

Last.fm® (music). Few such studies exist (Trattner, Lin, Parra, et al, 2012) investigating how to

construct and visualize tag clouds. The same is true for comparing the use of tag clouds and more

traditional hierarchical menus (Civan, Jones, Klasnja & Bruce, 2009; Voit, Andrews & Slany,

2012).

Most participants reported that using tags is a better and more flexible way to search for

information (Civan et al., 2009). The greater flexibility and the descriptive characteristics of tag

clouds were appreciated. Research about filing and re-finding behavior with hierarchies and

tagging systems did not find significant differences between tags or folder hierarchies (Voit et al,

2012). However, fast performers required less time and fewer mouse clicks when using tags

while slow performers benefitted more from hierarchical folders. Feedback from the participants

indicated that the use of tags improved the subjective user experience, while requiring an

equivalent amount of time. However, these studies investigated only user behavior, but not the

cognitive and perceptual processing leading to this behavior. Whether the use of tags instead of

hierarchical menus has an effect on learning outcomes and its underlying processes is unknown,

but the findings of Civan et al. (2009) and Voit et al (2012) indicate that the use of tags can be a

viable alternative for navigational support in HLEs.

2.2. Task complexity

Experiments are frequently designed in such a way that the requested information can be found in one specific page of a certain website, although, in reality information is often scattered over various pages and websites (Puerta Melguizo et al., 2012). Hence, there is a distinction between 'fact-finding tasks', in which the information can be found in a specific place, and 'information-gathering tasks' where participants have to gather and combine information from different sources in order to find an answer. Information-gathering tasks are more difficult because collecting and integrating information from different sources requires that pieces of information be remembered while continuing the search task (Rouet, 2003). Consequently, cognitive load is higher in information-gathering tasks. As hierarchical menus present a more coherent reading order, they might be better suited for such complex tasks.

2.3. Individual characteristics: gender differences

Research found that boys are more active in browsing than girls (Large, Bahesti & Rahman, 2002; Roy & Chi, 2003): they looked at more pages, selected more hyperlinks and saved more information, while they spent less time on viewing pages. It can be expected that learning outcomes are affected negatively as spending less time viewing pages might hamper comprehension. Moreover, female students have been found to have more difficulties orienting and navigating (Ford, Miller & Moss, 2001). Female students also felt less in control and experienced feelings of getting lost more often. On the other hand, Hupfer and Detler (2006) found no evidence of gender differences in navigation behavior. Based on these gender differences we can only speculate which type of navigational support suits females better than males or vise versa. However, spatial ability research can gives us a direction.

Research found that when people talk about navigating through hypertext, they mostly use terms that are also used for spatial navigation (Kim & Hirtle, 1995; Maglio and Matlock, 1998; Hochmair & Luttich, 2006). When verbalizing our thoughts about navigating on the internet, phrases as "I *went to* this webpage", "I found that *at* Wikipedia" or "The *address* of this website" are often used. We also see this in the browser interface, which uses buttons labeled as "*home*", "*back*" or "*forward*". As navigation tasks require participants to move throughout cyberspace, spatial ability might be an important factor in determining success and/or difficulty of HLE navigation. In line with this argumentation, some research has shown a link between spatial ability and HLE navigation (Campagnoni & Ehrlich, 1989; Juvina & Oostendorp, 2006; Stanney & Salvendy, 1995). These studies showed that high spatial ability is connected to better and faster task performance on information retrieval tasks. This more efficient performance is characterized by visiting less non-relevant pages, using the back-button less often and lower experiences of feeling lost (Ahmed & Blustein, 2005). For gender differences in learning, memory and spatial ability see Andreano and Cahill (2009).

3. HYPOTHESES

The experiment presented in this article focuses on the effects of two types of navigational support, information tasks with differing complexity levels, and learner gender on navigation behavior through HLEs, visual processing of the navigation menu (as measured by eye tracking), and task performance.

With respect to *navigational support,* this study compares tag clouds as a navigational support system with the more commonly used hierarchical menus. Based on the review of the

literature in the previous section, it can be expected that participants using tag clouds will visit fewer pages because they will be able to locate the desired webpages more accurately (H1a). Consequently, based on the assumption of visiting fewer pages, it is likely that the use of a tag cloud will result in a faster task performance compared to the use of a hierarchical menu (H1b). Moreover, as a result of the higher information density of tag clouds, it is plausible that the participants in the tagging condition will need more viewing time for the navigation structure than participants using a hierarchical menu (H2a). This may neutralize the effect of visiting fewer pages (H2b). Furthermore, as tag clouds have a higher information density and are more cluttered, it could involve more time to take decisions (H2c). On the one hand, tag clouds give a more detailed description of the contents of the HLE which might help participants better locate the desired information. We therefore expect that participants in the tagging condition will carry out the tasks (H3a) in a better way. On the other hand, as Madrid, Oostendorp, and Puerta Melguizo (2009) concluded, a more coherent reading order of text nodes is correlated to reduced cognitive load. Thus, participants in the hierarchical condition will give better answers (H3b).

The *complexity of a task* can have an effect as well. A *gathering task* requires more mental effort and deeper processing of the texts because learners need to synthesize information from several pages into an answer. Whereas for the *fact-finding task*, learners only need to find the correct page with the desired information. Therefore we expect to see several differences between those two tasks. In the fact-finding task most time is probably spent on deciding which page to go to. We therefore expect to see more 'trial and error' behavior, which is characterized by relatively short page visits to more pages (H4). As locating the desired information is such an important part of the fact-finding task, we expect that more attention is given to the navigational support and in the gathering task the texts will receive more attention (H5).

Finally, as discussed in the previous section, from the gender perspective it can be expected that boys will visit more pages, view them for shorter periods of time than girls and will find the desired answers more quickly (H6). Hypotheses 4, 5, and 6 will be investigated in relation to the two types of organizers.

4. METHOD

4.1. Participants and Design

To study the influence of navigational support and gender in tasks of different complexity levels on navigation behavior and task performance in hypertext learning environments, a 2 x 2 factorial design with a within subject measurement was used with the factors *type of navigational support* (hierarchical menu vs. tag cloud) and *gender* (male vs. female) as between subject factors and *task complexity* (fact-finding task vs. gathering task) as a within subjects factor. The dependent variables were task performance and navigation behavior.

A total of 60 tenth grade students (M_{age} = 15.63 years, SD = 0.688) in the highest level of secondary education of in the south part of The Netherlands participated. 24 boys (M_{age} = 15.75, SD = 0.737) and 36 girls (M_{age} = 15.56, SD = 0.652) were involved. Participants were equally divided across the two conditions of navigational support (i.e., tag cloud vs. hierarchical menu). Participants had experience in using their school's HLE, but had no experience in using other HLEs. Participants had normal or corrected to normal vision.

4.2. Material and apparatus

4.2.1. Task. Two tasks on the topic of 'obesity' were constructed. This is a topic that students are familiar with, but that is not part of the regular school curriculum. The first task was a *fact-finding task* in which the students had to find the percentage of obese people in the Netherlands, which is 14 percent. The second task was an *information-gathering task* where students had to give three reasons why obesity is a bigger problem in lower social-economic classes of society than in higher ones, which are: (1) healthy foods are more expensive, (2) higher educated people have more knowledge about a healthy diet and (3) minorities (which are over-represented in the lower social-economic classes) consume more fastfood. Students located and stored relevant information while navigating through pages in either the hierarchical or the tagging hypertext environment. The students received three minutes to execute the simple fact-finding task and seven for the complex information-gathering task.

This study focuses on differences between two levels of task complexity. In order to describe the difference in complexity between the tasks, the prose task characteristics of Mosenthal (1998) were used. Task complexity can be judged along three dimensions: type of information requested, type of match and plausibility of distractors. The type of information requested relates to how concrete or how abstract the information is that must be identified. Questions about identifying an amount or percentage are considered to be easier to solve than questions requesting causes. Type of match relates to the similarity between the information in the question and the information in the text and the processes needed to connect those two. The cause-effect relationship requested in the *information-gathering task* is considered more difficult than locating a feature or characteristic as in the *fact-finding task*. The plausibility of distractors describes the degree in which other pieces of information have similar characteristics in relation

to the requested information. The more distractors and the closer they are located to the requested information in the text, the more difficult a task is.

Based on these three task characteristics, Mosenthal (1998) describes five levels of task difficulty. The *fact-finding task* can be assessed as a task of difficulty level one, whereas the *information-gathering task* is a task of difficulty level three.

4.2.2. Hypertext environments. Two computer based e-learning environments were set up on the topic of obesity, one for each condition. Both HLEs were built with the open source software Wordpress®. The environments for both conditions were almost identical; the only difference was the navigational structure. As the research presented in paragraph 2.1 shows, design choices affect the effectiveness and efficiency of users in their navigation behavior. As the influences of design choices of tag clouds are not known yet, it is appropriate to compare the most commonly used design choice for tag clouds with hierarchical menus. This results in a more ecological valid setting for the design of the tag clouds as well.

The navigational support was positioned at the left side of the screen, either as an hierarchical menu (Figure 1) or a tag cloud (Figure 2). The hierarchical menu showed the categories in a hierarchy with subcategories one level deep. In this study we decided for a static and vertical hierarchical menu. The tag cloud presented a collection of tags (i.e., labels). Relationships between the tags were not shown.

The environment consisted of two types of web pages: overview pages and text pages. A total of 32 text pages were created. These text pages were categorized in a hierarchical structure and a tag cloud by one of the researchers. The other researchers then reviewed this categorization. Clicking on a menu item in the hierarchical menu or on a tag in the tag cloud

provided the participant with an overview page in that specific category or labeled with that specific tag with a linear list of the corresponding page title for the selected category or tag. The participants could access the text pages by clicking on the titles at the overview pages. Each environment consisted of 32 text pages. The text pages in the environment were copied from existing Internet resources. A text page consisted of the following elements: title, information about the source, plain text without hyperlinks and sometimes a table and a button with which the page could be saved if appropriate for answering the question. Figure 1 presents an example.

<INSERT FIGURE 1>

<INSERT FIGURE 2>

4.2.3. Eye tracking equipment. Eye movements and logging data were recorded with a remote Tobii 1750 eye tracker with temporal resolution of 50 Hz (2003), which is integrated with a PC screen, and is operated with Studio software (see www.tobii.com) from the stimulus PC. The screen capture recording mode was used, so not only the eye movements, but the entire task performance process (including possible mouse and keyboard operations) was captured.

4.3. Measurements

Task performance was measured as the correctness of the answers to the questions in the tasks. Task performance on the fact-finding task was coded as correct or incorrect. For the information-gathering task, the number of correct reasons why obesity is a bigger problem in lower socio-economic classes of society was counted. The maximum number of reasons was three. *Time on task* was used as a control variable.

To gain insight in navigation behavior, logging and eye tracking data were recorded. While participants carried out the tasks, several actions (page id, start and end timestamp for each visited page) within the HLE were tracked, recorded and stored in a log file for each participant. Based on these log files, the following variables were calculated.

Time to first click. Time it took participants before they clicked on an item in the navigational support for the first time to visit an overview page. This included viewing the navigational support, making a choice and clicking on a navigation item with the mouse.

Visited pages. Total number of visited pages was logged. From these logs, the number of *uniquely visited* pages and the number of *revisited* pages were calculated.

Time spent on overview pages and text pages. How long a participant stayed at each page. From these data, time spent on overview pages and time spent on text pages was calculated.

Eye tracking parameters were calculated for all pages in the HLE. All parameters were assigned to certain elements of these pages, so-called *areas of interest* (AOIs). A distinction was made between an AOI for the navigational support and an AOI for the content part of the pages. On the overview pages the list AOI covered the list of corresponding pages for the clicked tag or hierarchical category. On the text pages the content AOI covered the title and text of the page. All analyses were performed with Tobii Studio software version 2.2.4 (2007). Before further analyses were executed, the raw data was filtered with the Tobii ClearView fixation filter. Based on visual inspection of the data, a fixation definition of 30 pixels and 100 milliseconds was chosen (cf. Hegarty & Just, 1993; Loftus, 1981).

To analyze these AOIs, the total fixation duration was calculated. The fixation duration measures the sum of the durations for all fixations within an AOI or an AOI Group. If during the

recording the participant returns to the same media element then the new fixations on the AOI will be included in the calculations of the metric (Tobii, 2007).

4.4. Procedure

Prior to the experiment, participants received an introduction to the study. Also, all potential participants received a letter for their parents in which the study was explained. The experiment was conducted in individual sessions. Each session started with a detailed explanation of the procedure. For each participant, demographic data (i.e., age, gender) were collected. Next, they were introduced to the eye tracking equipment. The eye tracker was adjusted to the individual features of each participant by calibrating the system with a 9-point calibration. The maximum duration of the sessions was 50 minutes. Participants were randomly assigned to the two different conditions.

After a short instruction, participants first had to perform the fact-finding task and then the information-gathering task. Each task started with a question after which the participants searched in the hypertext environment at their own pace and in their own order. The maximum duration for the fact-finding task was three minutes and for the information-gathering task seven. When the participant thought they could answer the question, they could stop the recording themselves by pressing the spacebar. When the maximum duration was reached, the eye tracking software stopped the session. The whole procedure required maximally thirty minutes (including instruction for each task, calibration of the eye tracker for each task, make-up removal and a short break between the two tasks). For each participant, both tasks were recorded in one session.

5. RESULTS

The results of this study are presented in three parts: task performance, logged activities and eye movements. All relevant means and standard deviations are summarized in Table 1 (fact-finding task) and Table 2 (information gathering task). Two way factorial repeated measures ANOVAs were used to analyze the data. As part of the eye tracking data and the logging data were not normally distributed, thus, non-parametric statistical tests (Wilcoxon rank sum test and Kruskal-Wallis test for between subject factors; Wilcoxon signed rank test for the within subjects factor) were used to analyze these data when transformation did also not result in normally distributed data.

<INSERT TABLE 1>

<INSERT TABLE 2>

5.1. Task performance

Time on task was used as a control variable in this study. For the fact-finding task Wilcoxon rank sum tests revealed no significant differences across genders ($W = 352.5$, $p = .233$), or across type of organizer ($W = 427$, $p = .739$). A Kruskal-Wallis test with four groups (male-hierarchical, male-tagcloud, female-hierarchical and female-tagcloud) did also not show a significant difference between the groups ($H(3) = 2.706$, $p > .05$) on time on the fact-finding task. For the information-gathering task Wilcoxon rank sum tests revealed no significant differences across genders ($W = 443$, $p = .874$), or across type of organizer ($W = 429$, $p = .762$). A Kruskal-Wallis test with four groups did also not show a significant difference between the groups ($H(3) = 0.124$, $p > .05$) regarding time on the information-gathering task.

Task performance is related to hypotheses H1b, H3a, H3b and H6. The numbers of correct answers are summarized in Table 3. Logistic regression analysis revealed no statistical significant relationship between task performance in the fact-finding task and type of navigational support, gender or the interaction between type of navigational support and gender (see Table 4). For the gathering task, regression analysis revealed also no statistical significant relationship between task performance and type of navigational support, gender or the interaction between type of navigational support and gender (see Table 5).

<INSERT TABLE 3>

<INSERT TABLE 4>

<INSERT TABLE 5>

5.2. Differences in behavior

Behavioral differences of the participants in this study are divided in differences in durations, differences in page visits and differences in viewing behavior. All relevant means and standard deviations are summarized in Table 1 (fact-finding task) and Table 2 (information gathering task). Relevant test statistics are summarized in Table 6 (Analysis of Variance tests), 7 (Wilcoxon rank sum tests) and 8 (Kruskal-Wallis tests).

<INSERT TABLE 6>

<INSERT TABLE 7>

<INSERT TABLE 8>

Task complexity had the most prominent effect on behavior, closely followed by the type of navigational support. The more complex information-gathering task resulted in more attention

to the text. The tag cloud required more processing time and more effective selection of pages. Hardly any gender differences were found.

5.2.1. Differences in durations

Hypotheses H2c, H4 and H6 are related to differences in duration. *Time to make a decision* (hypothesis H2c) was measured by time to first click. The participants needed more time to make a decision in the information-gathering task compared to the fact-finding task ($F(1, 56) = 4.363$, $p = .041$, $\eta^2 = 0.072$). As the effect size is rather small, the effect of task complexity on the time to make a decision is not very distinct. No significant differences were found for type of navigational support, gender or an interaction between type of navigational support, gender and task complexity.

In the fact-finding task the participants spend significantly more time on the overview pages ($F(1, 56) = 52.507$, $p = .041$, $\eta^2 = 0.484$) and less time on the text pages ($F(1, 56) = 55.705$, $p < .001$, $\eta^2 = 0.499$) when compared to the information-gathering task. Task complexity has a medium effect size on page visit duration. No significant differences were found for type of navigational support, gender or an interaction between type of navigational support, gender and task complexity.

5.2.2. Differences in page visits

Hypotheses H1a, H2b, H4, H5 and H6 are related to differences in the number of page visits. Regarding the *number of total page visits*, the participants visited significantly more pages in the hierarchical condition compared to the tagging condition ($F(1, 56) = 4.309$, $p = .043$, $\eta^2 = 0.071$) and significantly more pages in the fact-finding task compared to the information-

gathering task ($F(1, 56) = 29.873$, $p < .001$, $\eta^2 = 0.348$). Moreover, a significant interaction

between type of navigational support and task complexity was found ($F(1, 56) = 10.164$,

$p = .002$, $\eta^2 = 0.154$). Pairwise comparisons with Bonferroni adjustment for this interaction

showed that especially participants in the hierarchical condition visited relatively more page

pages in the fact-finding task than in the information-gathering task or participants in the tagging

condition.

Regarding the *number of revisited pages*, participants in the hierarchical condition did

revisit significantly more pages than participants in the tagging condition for both the fact-

finding task ($W = 691$, $p < .001$, $r = -0.473$) and the information-gathering task ($W = 676$,

$p < .001$, $r = -0.431$). The participants revisited significantly fewer pages in the fact-finding task

than in the information-gathering task ($Z = 273$, $p < .001$, $r = -0.521$). A possible interaction

between type of navigational support and gender was tested with a Kruskal-Wallis test with four

groups, which did show a significant difference between groups for the fact-finding task

($H(3) = 13.830$, $p = .003$) as well as the information-gathering task ($H(3) = 13.108$, $p = .004$). In

the fact-finding task, female participants did revisit more pages in the hierarchical condition than

in the tagging condition or male participants in the tagging condition. In the information-

gathering task, male participants did revisit relatively more pages in the hierarchical condition

than in the tagging condition or female participants in the tagging condition.

5.2.3. Differences in viewing behavior

Hypotheses H2a, H2c, H5 and H6 are related to differences in viewing behavior. The

attention given to the navigational support was significantly higher in the tagging condition than

in the hierarchical condition ($F(1, 56) = 41.827$, $p < .001$, $\eta^2 = 0.428$). Furthermore, the

participants gave significantly more attention to the navigational support in the fact-finding task than in the information-gathering task ($F(1, 56) = 67.717$, $p < .001$, $\eta^2 = 0.547$). No significant differences were found for gender or an interaction between type of navigational support, gender and task complexity. Type of navigational support as well as task complexity has a medium effect on the attention given to the navigational support.

The *viewing time devoted to the main part of the overview pages* was significantly shorter in the tagging condition than in the hierarchical condition ($F(1, 56) = 21.111$, $p < .001$, $\eta^2 = 0.274$). Regarding task complexity, the participants looked significantly shorter on the main part of the overview pages in the fact-finding task compared to the information-gathering task ($F(1, 56) = 8.562$, $p = .005$, $\eta^2 = 0.133$). No significant differences were found for gender or an interaction between type of navigational support, gender and task complexity. Type of navigational support as well as task complexity has a medium effect on the time devoted to the main part of the overview pages.

Dedicated *viewing time for the main part of the text pages* was significantly shorter in the tagging condition than in the hierarchical condition in the information-gathering task ($W = 651$, $p = .003$, $r = -0.383$). No significant differences were found for gender or for type of navigational support and gender in the fact-finding task. Furthermore, participants did look significantly shorter on the main part of the text pages in the fact-finding task than in the information-gathering task ($Z = 307$, $p < .001$, $r = -0.577$). A possible interaction between type of navigational support and gender for the information-gathering task was tested with a Kruskal-Wallis test with four groups, which did also show a significant difference between groups ($H(3) = 9.991$, $p = .018$). The male participants in the hierarchical condition did look relatively longer on the main part of the text pages than the male participants in the tagging condition.

6. DISCUSSION

The main question of this study was whether it would make a difference using a hierarchical menu or tag cloud for navigation behavior and learning and how this is related to task difficulty (simple fact-finding vs. complex information-gathering) and learner characteristics (i.e., gender). Critical was whether task performance was affected or not. Based on the assumption that users of a tag cloud will visit fewer pages, we predicted that the use of a tag cloud would result in better task performance for participants with tag cloud navigation compared to hierarchical menu navigation (H1b, H3a and H3b). This expectation was not supported by the results. Although the relative frequency of correct answers was higher on the fact-finding task in the tagging condition, there were no significant differences between the two types of navigational support with either of the tasks. There were also no significant differences in task performance with regard to task difficulty or gender (H4 and H6). As there were also no differences in time on task, the results show neither negative nor positive effects of the use of a non-traditional type of navigational support (i.e., a tag cloud). One of the most intriguing findings concerning task performance is that, contrary to our expectations, users in the tagging condition did not need more time to make decisions (H2c). This could be due to the better descriptive value and better navigational flexibility of tag clouds (Civan et al., 2009) compared to a hierarchical menu.

Regarding navigation behavior, we predicted that participants who received a tag cloud would visit fewer pages compared to participants receiving a hierarchical menu (H1a) because they would be able to locate desired webpages more accurately (e.g., Voit, et al., 2012). What was found was that participants in the hierarchical condition visited relatively more pages in total

and also revisited relatively more pages than participants in the tag cloud condition. As there were no differences in time on task, it can be concluded that the use of tag clouds as navigational support leads to browsing behavior that is less active. Moreover the smaller number of revisited pages indicates more effective localization of the desired information. As a result of the higher information density of tag clouds, we expected that the participants in this condition would need more viewing time when using the navigational support (H2a and H2b). Consequently, it might be the case that they consider what to choose more deeply. As the results showed, the participants in the tagging condition viewed the tag cloud considerably longer compared to the participants who used the hierarchical menu. This was accompanied by shorter viewing times on the main part of the overview pages. In combination with the fewer page revisits, this indicates that the use of tag clouds may lead to more focused page selection and better processing of the navigational support compared to a hierarchical menu.

With respect to task complexity, we expected more active browsing behavior for the information-gathering task compared to the fact-finding task (H4). The results, however, showed partly the opposite. Participants visited relatively more (unique) pages in the fact-finding task, but at the same time revisited more pages in the information-gathering task. This latter effect could be because an information-gathering task requires more elaborate reading of the text in order to be able to carry it out than a fact-finding task. This is supported by the result obtained that participants spent relatively more time on the text pages in the information-gathering task (H5). Moreover, the longer decision time to the first click is also an indication that the information-gathering task required more elaborate reading. Finally, the higher frequency of page visits in the fact-finding task might point to a more trial-and-error type of navigation behavior.

Although most of the results did not reveal differences, some minor differences between the genders were found. These findings were contradictory to the expectations based on earlier research (Large, Bahesti & Rahman, 2002; Roy & Chi, 2003). While both genders needed fewer revisits when given a tag cloud, males profited more from tag clouds than females. Thus, the use of tag clouds leads to less active browsing behavior, in particular for males. On the other hand, males in the tagging condition viewed the texts for shorter periods of time, which might be an indication of more active browsing behavior.

Since students are most used to receiving and using hierarchical structures in their learning, navigating with the use of a hierarchical menu is less demanding for them than navigating with less familiar navigation structures such as the tag clouds used in this study. Although they had to carry out two tasks, this was probably not enough time to get used to using tag clouds for navigation. As such tags are valued for there descriptive value and are perceived to be slightly more favorable than hierarchical structures (Civan et al., 2009), when the use of tags becomes more common, users might be able to use them more efficiently.

Finally, nearly every design decision influences behavior. It is therefore likely that our design decisions have influenced behavior. However, our design decisions ware based on common practice in constructing hierarchical menu structures and tag clouds.

7. DIRECTIONS FOR FUTURE RESEARCH

The most important conclusion is that tag clouds can be considered as a viable option for navigational support. As little is known about the influence of design choices of tag clouds on navigation behavior, this study is therefore a first step in studying the use of tag clouds as navigational support in HLEs. However, more research is needed concerning the different design

choices of tag clouds. Studying variations in the design of the navigational support will further develop the understanding of the use of tag clouds. For example, varying the level of detail or varying the visual presentation will give more insight into the influence of specific design decisions on behavior and performance. A more thorough understanding will also result in sound recommendations for the design of HLEs.

In this study two task were used, a fact-finding task and a more complex information-gathering task. It has to be noted, though, that while the information-gathering task was more complex than the fact-finding task, neither task was really very complex. Another future research direction is to look at other, more complex tasks and maybe also other task types. An example of a more complex task is an information-comparison task. In such a task the learner not only has to find and gather information, but also has to determine relevant dimensions on which to compare the information and then actually carry out the comparison. A step further is an information-evaluation task where after the comparison decisions need to be made as to value, use for a certain purpose, et cetera. Examining navigation behavior for other, more complex, tasks is necessary to build a more coherent understanding of different types of navigational support. Moreover, future research should make use of other task topics as well.

Another aspect for future research is the size of the HLE's. In our study, the HLE was limited to 32 pages. In reality, HLE's can contain more and sometimes almost endless amounts of content (e.g., Wikipedia®). It would be interesting to see whether our conclusions are applicable in large scale HLE's. In this study we choose for a specific lay out, but other layouts might induce other effects (Lohman, Ziegler & Tetzlaff, 2009).

Furthermore, in this study we choose gender as the individual characteristic to investigate. In future studies it would be interesting to see whether other individual

characteristics (e.g., working memory capacity, expertise in a domain, ability to impose structure on information) do have an influence.

8. CONCLUDING REMARKS

Navigating in hypertext learning environments is a complex process; one that needs to be understood much better by instructional designers and educational practitioners (e.g., teachers). The choice of a menu style and item can have far reaching implications. Choosing the wrong type for the task at hand or for a specific type of learner may, for example, slow a student down because (s)he selects a wrong menu item or hyperlink. As a result the reading and information processing is interrupted, which in turn might hamper the learning process. The results of this study show that learning outcomes can benefit from using a novel tag cloud as navigational support as compared to more traditional navigation structures. Tag clouds can unfold their full potential, in particular for large HLEs, making large amounts of information easier accessible. Moreover, the descriptive nature of tag clouds could enhance the information processing as well.

REFERENCES

Ahmed, I., & Blustein, J. (2006). Influence of spatial ability in navigation: using look-ahead breadcrumbs on The Web. *International Journal of Web Based Communities*, *2*, 183-196.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research, 73*, 277–320. doi: 10.3102/00346543073003277

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, *19*, 376-386. doi: 10.1016/j.learninstruc.2009.02.005

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory*, *16*, 248-266. doi: 10.1101/lm.918309

Anfinnsen, S., Ghinea, G., & De Cesare, S. (2011). Web 2.0 and folksonomies in a library context. *International Journal of Information Management*, *31*(1), 63-70. doi: 10.1016/j.ijinfomgt.2010.05.006

Bezdan, E., Kester, L., & Kirschner, P. A. (2013). The influence of node sequence and extraneous load induced by graphical overviews on hypertext learning. *Computers in Human Behavior*, *29*, 870-880. doi: 10.1016/j.chb.2012.12.016

Campagnoni, F. R., & Ehrlich, K. (1989). Information retrieval using a hypertext-based help system. *ACM Transactions on Information Systems (TOIS)*, *7*, 271-291. doi: 10.1145/65943.65948

Civan, A., Jones, W., Klasnja, P., & Bruce, H. (2009). Better to organize personal information by folders or by tags?: The devil is in the details. *Proceedings of the American Society for Information Science and Technology*, *45*, 1–13. doi: 10.1002/meet.2008.1450450214

Conklin, J. (1987). Hypertext: an introduction and survey. *IEEE Computer, 20*, 17-41. doi: 10.1109/MC.1987.1663693

Dee-Lucas, D., & Larkin, J. H. (1999). Hypertext segmentation and goal compatibility: Effects on study strategies and learning. *Journal of Educational Multimedia and Hypermedia*, *8*, 279-313.

DeStefano, D., & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*, 1616-1641. doi: 10.1016/j.chb.2005.08.012

Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society for Information Science and Technology*, *52*, 1049–1066. doi: 10.1002/asi.1165

Hochmair, H. H., & Luttich, K. (2006). An Analysis of the Navigation Metaphor – and Why It Works for the World Wide Web. *Spatial Cognition and Computation*, *6*, 235-278. doi: 10.1207/s15427633scc0603_3

Hupfer, M., & Detlor, B. (2006). Gender and Web information seeking: A self-concept orientation model. *Journal of the American Society for Information Science and Technology*, *57*, 1105–1115. doi: 10.1002/asi.20379

Juvina, I., & Oostendorp, H. V. (2006). Individual differences and behavioral metrics involved in modeling web navigation. *Universal Access in the Information Society*, *4*, 258–269. doi: 10.1007/s10209-005-0007-7

Kim, H., & Hirtle, S. C. (1995). Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology*, *14*(4), 239-250. doi: 10.1080/01449299508914637

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*, 1-10. doi: 10.1016/S0959-4752(01)00014-7

Lohmann, S., Ziegler, J., & Tetzlaff, L. (2009). Comparison of tag cloud layouts: Task-related performance and visual exploration. In *Human-Computer Interaction–INTERACT 2009* (pp. 392-404). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-03655-2_43

Madrid, R. I., Van Oostendorp, H., & Puerta Melguizo, M. C. (2009). The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. *Computers in Human Behavior*, *25*, 66-75. doi: 10.1016/j.chb.2008.06.005

Maglio, P. P., & Matlock, T. (1998). Metaphors we surf the web by. In *Workshop on Personalized and Social Navigation in Information Space*.

Minetou, C., Chen, S., & Liu, X. (2008). Investigation of the use of navigation tools in web-based learning: A data mining approach. *International Journal of Human-Computer Interaction, 24*, 48-67. doi: 10.1080/10447310701771480

Mobrand, K. A., & Spyridakis, J. H. (2007). Explicitness of local navigational links: comprehension, perceptions of use, and browsing behavior. *Journal of Information Science*, *33*, 41-61. doi: 10.1177/0165551506068144

Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal, 35*, 269-307. doi: 10.3102/00028312035002269

Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative web searching

behavior: an elementary school study. *Information Processing & Management*, *38*, 427–

443. doi: 10.1016/S0306-4573(01)00034-6

Leuthold, S., Schmutz, P., Bargas-Avila, J. A., Tuch, A. N., & Opwis, K. (2011). Vertical versus

dynamic menus on the world wide web: Eye tracking study measuring the influence of

menu design and task complexity on user performance and subjective preference.

Computers in Human Behavior, *27*, 459-472. doi: 10.1016/j.chb.2010.09.009

Liu, Z., & Huang, X. (2008). Gender differences in the online reading environment. *Journal of

Documentation*, *64*, 616–626. doi: 10.1108/00220410810884101

Puerta Melguizo, M. C., Vidya, U., & Van Oostendorp, H. (2012). Seeking information online:

the influence of menu type, navigation path complexity and spatial ability on information

gathering tasks. *Behaviour & Information Technology*, *31*, 59-70. doi:

10.1080/0144929X.2011.602425

Rouet, J. F. (2003). What was I looking for? The influence of task specificity and prior

knowledge on students search strategies in hypertext. *Interacting with Computers, 15*, 409–

428. doi: 10.1016/S0953-5438(02)00064-4

Roy, M., & Chi, M. T. H. (2003). Gender Differences in Patterns of Searching the Web. *Journal

of Educational Computing Research*, *29*, 335–348. doi: 10.2190/7BR8-VXA0-07A7-

8AVN

Salomon, G., & Almog, T. (1998). Educational Psychology and Technology: A Matter of

Reciprocal Relations. *Teachers College Record, 100*, 222-241.

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational

Psychology Review*, *19*, 285-307. doi: 10.1007/s10648-007-9046-3

Schoenfeld, R., Lehmann, W., & Leplow, B. (2010). Effects of age and sex in mental rotation and spatial learning from virtual environments. *Journal of Individual Differences, 31*, 78-82. doi: 10.1027/1614-0001/a000014

Stanney, K. M., & Salvendy, G. (1995). Information visualization; assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics*, *38*, 1184-1198. doi: 10.1080/00140139508925181

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*, 257-285. doi: 10.1207/s15516709cog1202_4

Tobii 1750 [Apparatus]. (2003). Danderyd, Sweden: Tobii Technology AB.

Tobii Studio [Software]. (2007). Danderyd, Sweden: Tobii Technology AB.

Toms, E. G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., ... & Macnutt, A. (2008). Task effects on interactive search: The query factor. In *Focused access to XML documents* (pp. 359-372). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-85902-4_31

Trattner, C., Lin, Y. L., Parra, D., Yue, Z., Real, W., & Brusilovsky, P. (2012). Evaluating tag-based information access in image collections. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 113-122). ACM. doi: 10.1145/2309996.2310016

Voit, K., Andrews, K., & Slany, W. (2009). Why personal information management (PIM) technologies are not widespread. In *PIM Workshop, ASIS&T* (Vol. 2009, pp. 60-64).

Voit, K., Andrews, K., & Slany, W. (2012). Tagging might not be slower than filing in folders. *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 2063. doi: 10.1145/2212776.2223753

Wildemuth, B. M., & Freund, L. (2012). Assigning search tasks designed to elicit exploratory

      search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and*

      Information Retrieval (p. 4). ACM. doi: 10.1145/2391224.2391228

Table 1

Means and Standard Deviations for Eye Tracking and Logging Data for Task 1

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Control variable						
Time on task	137.33 (56.31)	131.67 (47.78)	133.94 (50.49)	142.46 (51.10)	122.50 (51.49)	130.48 (51.41)
Logging						
Time to first click	7.78 (2.31)	7.18 (2.19)	7.42 (2.22)	7.34 (2.35)	7.19 (1.84)	7.25 (2.02)
Time on overview pages	77.74 (52.64)	63.40 (38.96)	69.14 (44.63)	65.15 (34.26)	48.29 (34.19)	55.03 (34.66)
Time on text pages	43.06 (29.96)	51.35 (30.18)	48.04 (29.85)	55.05 (24.86)	53.27 (24.61)	53.98 (24.30)
Visited pages	12.17 (6.09)	11.78 (5.06)	11.93 (5.39)	9.00 (3.98)	9.17 (6.81)	9.10 (5.76)
Revisited pages	3.75 (3.84)	3.06 (2.31)	3.33 (2.97)	0.83 (1.27)	1.28 (2.40)	1.10 (2.01)
Eye tracking						
Fixation duration navigation	33.35 (16.72)	31.99 (17.34)	32.53 (16.82)	54.98 (28.21)	44.70 (26.48)	48.81 (27.19)
Fixation duration overview pages	55.32 (37.73)	45.28 (26.08)	49.30 (31.04)	38.36 (19.69)	28.04 (16.80)	32.17 (18.41)
Fixation duration text pages	34.40 (24.05)	37.68 (23.66)	36.37 (23.46)	34.57 (13.56)	37.56 (19.13)	36.36 (16.93)

Table 2

Means and Standard Deviations for Eye Tracking and Logging Data for Task 2

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Control variable						
Time on task	298.28 (100.39)	302.71 (112.21)	300.94 (105.87)	297.99 (99.40)	312.06 (112.31)	306.43 (105.79)
Logging						
Time to first click	8.42 (1.94)	7.69 (2.06)	7.98 (2.01)	7.79 (1.16)	8.35 (3.06)	8.12 (2.46)
Time on overview pages	76.39 (38.58)	72.26 (43.51)	73.91 (40.97)	67.40 (31.91)	86.82 (65.84)	79.05 (54.96)
Time on text pages	206.93 (71.62)	216.98 (79.97)	212.96 (80.92)	212.20 (78.92)	204.18 (84.34)	207.39 (80.92)
Visited pages	19.50 (7.88)	18.67 (9.56)	19.00 (8.79)	17.17 (4.53)	20.22 (11.33)	19.00 (9.24)
Revisited pages	7.42 (3.42)	6.89 (5.93)	7.10 (5.01)	3.25 (2.38)	4.61 (3.42)	4.07 (3.07)
Eye tracking						
Fixation duration navigation	37.84 (19.52)	41.20 (19.64)	39.85 (19.32)	66.24 (29.02)	76.40 (48.65)	72.33 (41.62)
Fixation duration overview pages	67.01 (35.09)	57.47 (32.98)	61.29 (33.57)	47.06 (20.80)	52.80 (36.18)	50.50 (30.65)
Fixation duration text pages	181.92 (63.73)	185.02 (65.96)	183.78 (63.98)	168.05 (70.21)	157.49 (63.43)	161.72 (65.24)

Table 3

Task performance

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Fact-finding task						
Correct answer	5	8	13	7	11	18
Information-gathering task						
One correct reason	2	4	6	1	3	4
Two correct reasons	6	7	13	7	8	15
Three correct reasons	4	7	11	4	7	11

Table 4

Logistic regression fact-finding task: Variables in the Equation

	B	S.E.	Sign.	95% CI for odds ratio		
				Lower	Odds ratio	Upper
Constant	-.22	.47	.64	.30	.80	2.03
Gender	-.11	.75	.88	.20	.89	3.92
Navigation	.68	.68	.32	.53	1.96	7.69
Gender by Navigation	-.00	1.07	1.00	.12	1.00	8.35

Note: R^2 = .021 (Hosmer and Lemeshow), 0.28 (Cox and Snell), 0.37 (Nagelkerke).

Model χ^2 (3) = 1.72, p > .05

Table 5

Regression information-gathering task for predicting task performance

Variable	B	SE(B)	β
Navigation	0.06	0.24	.23
Gender	0.00	0.27	.00
Navigation x gender	0.03	0.38	.07
ΔR^2		.05	

*p < .05. **p < .01.

Table 6

Analysis of Variance (ANOVA) between navigational support, task complexity and gender

	df	F	η^2	p
Time to first click				
- navigation support	1	0.045	0.001	> .05
- task complexity	1	4.363	0.072	.041
- gender	1	0.233	0.004	> .05
- error	56			
Time spent on overview pages				
- navigation support	1	1.367	0.024	> .05
- task complexity	1	52.507	0.484	.041
- gender	1	0.967	0.017	> .05
- error	56			
Time spent on text pages				
- navigation support	1	2.101	0.036	> .05
- task complexity	1	55.705	0.499	< .001
- gender	1	1.101	0.019	> .05
- error	56			
Number of visited pages				
- navigation support	1	4.309	0.071	.043
- task complexity	1	29.873	0.348	< .001
- gender	1	1.101	0.019	> .05
- interaction: navigation support x task complexity	1	10.164	0.154	.040
- error	56			
Fixation duration navigation				
- navigation support	1	41.827	0.428	< .001
- task complexity	1	67.717	0.547	< .001

- gender	1	0.012	0.000	> .05
- error	56			
Fixation duration overview pages				
- navigation support	1	21.111	0.274	< .001
- task complexity	1	8.562	0.133	.005
- gender	1	0.706	0.012	> .05
- error	56			

Table 7

Wilcoxon rank sum tests for navigational support and gender

	W	*r*	*p*
Number of revisited pages (fact-finding task)			
- navigation support	691	-0.473	< .001
- gender	471	-0.076	> .05
Number of revisited pages (information-gathering task)			
- navigation support	676	-0.431	< .001
- gender	385	-0.092	> .05
Fixation duration main part text pages (fact-finding task)			
- navigation support	540	-0.205	> .05
- gender	346	-0.146	> .05
Fixation duration main part text pages (information-gathering task)			
- navigation support	651	-0.383	.003
- gender	419	-0.024	> .05

Table 8

Kruskal-Wallis tests for interaction between navigational support and gender

	df	H	*p*
Number of revisited pages			
- fact-finding task	3	13.380	.003
- information-gathering task	3	13.108	.004
Fixation duration main part text pages			
- fact-finding task	3	5.445	> .05
- information-gathering task	3	9.991	.018

FIGURE CAPTIONS

Figure 1: HLE with hierarchical navigation

Figure 2: HLE with tag cloud navigation