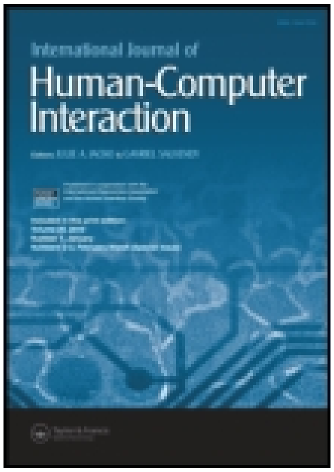


International Journal of Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hihc20

Students' Web Search Strategies With Different Task Types: An Eye-Tracking Study

Emine Şendurur[a] & Zahide Yildirim[b]
[a] Computer Education and Instructional Technology, Ondokuz Mayis University, Samsun, Turkey
[b] Computer Education and Instructional Technology, Middle East Technical University, Ankara, Turkey
Accepted author version posted online: 11 Sep 2014.Published online: 09 Jan 2015.

To cite this article: Emine Şendurur & Zahide Yildirim (2015) Students' Web Search Strategies With Different Task Types: An Eye-Tracking Study, International Journal of Human-Computer Interaction, 31:2, 101-111, DOI: 10.1080/10447318.2014.959105

To link to this article: http://dx.doi.org/10.1080/10447318.2014.959105

Intl. Journal of Human–Computer Interaction, 31: 101–111, 2015

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447318.2014.959105

Students' Web Search Strategies With Different Task Types: An Eye-Tracking Study

Emine Şendurur[1] and Zahide Yildirim[2]
[1]*Computer Education and Instructional Technology, Ondokuz Mayis University, Samsun, Turkey*
[2]*Computer Education and Instructional Technology, Middle East Technical University, Ankara, Turkey*

Common search engines may not be practical for students while searching their homework, no matter how easy or how hard the searched content. The purpose of this study is to investigate the whole search processes of seventh graders. 11 participants were assigned to three tasks that varied in difficulty. The findings demonstrated that search patterns can change according to types of tasks. Especially the number of page visits and correct hits, the way keywords used, and task completion success differed across task type. Although the statistical analysis did not generate significant relationships, qualitative analysis demonstrated that time spent, scan type, number of keywords, and reading styles differed for all tasks.

1. INTRODUCTION

Today, the Internet appeals to a wide range of users, including children. Children's use of web can be observed in various contexts, and using the web as an information source especially for school assignments can be considered as common (Large & Beheshti, 2005; Large, Beheshti, Nesset, & Bowler, 2004). Although this is the case, while engaging in searching tasks within the web environment, children are usually confronted with adult-specific interfaces, which might somehow confuse them. Interaction with the search engine interface and search results can be like navigating a maze, especially for novice searchers. This commonly happens in the utilization of such search engines as Google, because they are for general purpose searches. The technology behind such tools does not matter if the results do not appeal to or satisfy the user (Chau, Qin, Zhou, Tseng, & Chen, 2008), because a high level of user frustration translates to serious problems (Ceaparu, Lazar, Bessiere, Robinson, & Shneiderman, 2004).

Using a search engine might not be a simple process for a primary or secondary school student. No matter how advanced their technical skills, the related literature claims that they face certain difficulties while searching (Bar-Ilan & Belous, 2007). Even if meaningful keywords are entered, there can be irrelevant clicks, unnecessary content, long search durations, and so forth. Relevant keywords do not always generate all relevant results. Moreover, not all listed sites include the searched content; thus it is the user who filters the filtered list again and again. Users have to make a series of decisions in order to go to a relevant place on the web (Fang & Salvendy, 2000). Novice users generally do not make a plan before starting a search, and they visit links on the list without evaluating them critically (Quintana, Zhang, & Krajcik, 2005). If the searched topic is too specific, novice searchers can feel overwhelmed due to the expectation of finding ready-to-use information. Large and Beheshti (2000) exemplified this phenomenon through a study. They assigned 50 six-grade children a classroom project. Interviews with users and the final products indicated that the web search resulted in frustrations due to failed searches of specific information.

To eliminate unnecessary details and complexities of search engines, some alternative interfaces for children have been developed. Yahooligans was introduced as safe for kids in 1996, and now it is called Yahoo! Kids. Recently, Google has introduced KidRex, with an interface that is simple and free of ads. In addition, there are many other kid-specific interfaces using Google safe search filter, such as kids.net, KidzSearch, and quinturakids. Besides the search engines that provide a safe search environment for children, there are search engines that support visual search, such as AGA-Kids, Redz, Simploos, and Spacetime3d. This kind of search engine aims to reduce the time one must manage the generated results page. Despite these attempts, children are often alone during the search process, with no adult supervision. That's why there are other attempts to support a child's search process. Scaffolding the user through various tools such as scaffolded notepads (e.g., IdeaKeeper) or inquiry maps (e.g., Symphony) is one attempt. But although these tools have tried various approaches to help children in their search process, none provides scaffolding that helps students be cognitively active.

Therefore, a search tool that provides scaffolding for elementary school students during the web search was developed by the

Address correspondence to Emine Şendurur, Computer Education and Instructional Technologies, Ondokuz Mayis University, Egitim Fak. B-Blok, B-326 55139 Kurupelit, Samsun, Turkey. E-mail: eminesendurur@gmail.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

authors and named the Web-based Internet Search Scaffolding Tool (WISST; Sendurur, 2012). WISST was designed in the light of Quintana et al.'s (2005) framework for supporting metacognitive process of online inquiry through software-based scaffolding. In this framework, there are certain offerings for effective scaffolding that are applicable for both human and software scaffolding. This tool aims to scaffold users throughout web searching by emphasizing certain metacognitive skills improvement. According to Quintana et al., through an online inquiry, children experience certain cognitive phases. The inquiry starts with asking questions, then the search occurs, and then the evaluating and reading stages take place. Synthesizing is the final cognitive phase. The strategies applied in WISST can be summarized as follows:

- Asking a question: Provide driving questions; help to integrate results of multiple searches in one space.
- Searching: Encourage users to find rich resources; make search steps visible; help users to decide on keywords before searching; show the search history.
- Reading and Evaluating: Provide a prompted notepad; show users their goals; provide users with a list of evaluation criteria.
- Synthesizing: Encourage users to compare and contrast information across different resources; describe the criteria they should use; prompt users to reflect on different aspects of information.

The main components designed in WISST were *start, search, reading and evaluating,* and *end* modules, and use of the tool is restricted by user name and password. Although the components were designed as separate modules, they work in a meshed manner depending on the input of the user. The user has the freedom to jump from one component to another.

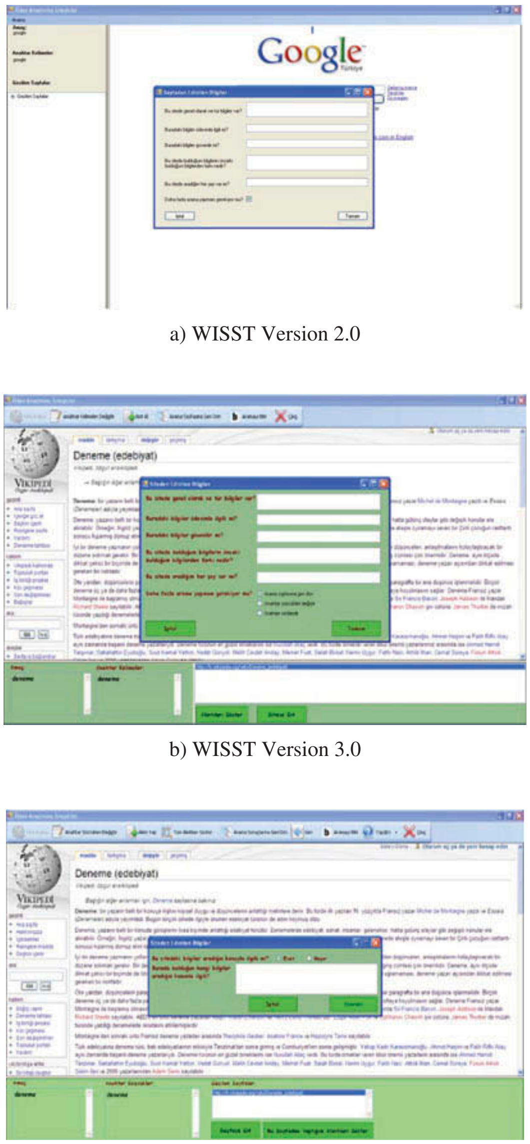
- Start Module: When the user logs in to the system for the first time, a help video provides the learner with strategies and steps for an effective search. It also describes how to cope with the integration of results from different sites. Moreover, to initiate building an accurate schema of searching, a whole search strategy from start to finish was exemplified in the help video.
- Search Module: To encourage students to expand their scope of search, the search steps are made visible on a part of the screen. In this module, learners are asked to write about the aim of their search and previous knowledge about the assigned topic and to decide on keywords. Users are not allowed to start the search without answering these questions. The aim, keywords, and visited sites are visible on the screen during the whole search. Users are allowed to either change their aims or keywords completely or edit partially whenever needed. In addition, they can revisit previously visited sites by just clicking on the site listed at the bottom of the screen. Those sites are listed according to the order of the visit.
- Reading and Evaluation Module: This module starts when the user clicks on a website's link. In this module, a prompted notepad is available at the middle top of the screen. It allows users to check the copied and pasted information. Users can edit the information or add their own interpretations. While leaving the page, users are encouraged to think about their reading performance by answering presented questions about relevancy and security of the information. Students were expected to answer all questions in regard to the sites they visited. In this module, users also have to decide on what to do next. They can choose to either end the whole search or continue the search. They can go back to the start module and continue the search with either new keywords or previous keywords.
- End Module: Learners are presented with the following questions when they intend to terminate the search session:

1. What is the topic of the search?
2. What have you learned from this search?
3. Which information you found was the most interesting? Why?

In answering these questions, they are expected either to think about the steps in the whole search or to compare information needed before the search to information gathered after the search. On this screen, cited information is also presented to users to help them remember. Figure 1 presents the screenshot of the tool's different versions, respectively. For this study, WISST Version 3.0 was used.

The development of WISST was based on iterative design. After five iterations, the tool has reached its ultimate appearance and functionality. In the *first iteration*, the prototype of WISST (WISST 1.0) was designed based on the created scenario. In the *second iteration,* the tool was designed with limited functions (WISST 2.0), and feedback from the potential users was gathered through focus group interviews in addition to experts' heuristic evaluations. In the *third iteration,* the actual tool was created (WISST 3.0). Its usability was checked with the help of eye-tracking technology. The *fourth iteration* included expert evaluations. After all necessary changes were completed, the final and *fifth iteration* included the control of functions (see Sendurur & Yildirim, 2010, 2012). In this study, data collected during the third iteration via eye-tracker were investigated in terms of general search patterns, search behavior, and task type.

There are a few studies investigating the use of eye-tracking methodology as a way of analyzing web search behaviors (e.g., Cutrell & Guan, 2007; Dinet, Bastien, & Kitajima, 2010). The majority of these studies have focused on adult users or university students. In their study, Dinet et al. (2010) worked on students from fifth to 11th grades, but their study was based



a) WISST Version 2.0

b) WISST Version 3.0

c) WISST Version 4.0 (last version)

FIG. 1. Screenshots of different Web-based Internet Search Scaffolding Tool (WISST) versions.

upon the visual strategies associated with either typographical cues or topic familiarity. These studies fall short in explain the parameters of search behaviors in detail as well as the relationship between these behaviors and various task types. Therefore, the main focus of this study was to investigate children's search behaviors in detail through the use of an eye-tracking tool, particulary in relation to various task types. For this study, the focus is on the seventh graders (12–13 years old). There are two reasons for studying that age group. One is that the designed tool providing metacognitive scaffolds is targeted at those age groups due to their developmental characteristics. In the literature, metacognitive skills are associated with formal operational stage of Piaget, and at this stage children have the skills for metacognitive control. Before this stage, the children's egocentric characteristics can hinder metacognitive skill development (Berk, 2003; Kuhn, 1999; Veenman & Spaans, 2005; Veenman, Wilhelm & Beishuizen, 2004). Considering the development levels of children, inclusion of 12- or 13-year-old students would provide meaningful findings. The second reason is that secondary school students often perform web searches because of their school projects, homework, and so on. Sixth grade is a transition period for secondary school, but the students are less familiar with a web search than are seventh graders. Eighth graders have more experiences than others, but they have school attendance problems due to high school entrance exam preparations. Therefore, seventh graders were targeted in this study. The specific research questions guided this study are the following:

1. How do seventh-grade students' search patterns change across three task types, namely, ready-to-use (T1), easy-to-interpret (T2), and difficult-to-interpret (T3)?
2. How are students' scanning behaviors related to click accuracy and number of visits to various sites?
3. How are task types related to click accuracy, number of visited links, time spent on search, scan type, number of used keywords, and reading styles?
4. What, if any, leads to change in search behavior?

2. METHOD

2.1. Study Design

This study is an observational study. To test the usability of WISST 3.0, 11 seventh-grade students (12–13 years old) from three different public schools volunteered to participate in this study. Before the students participated, parental consent was obtained. Both students and parents were informed that during the sessions, students' eye movements would be tracked. The participants were used to searching the web for their homework in their regular courses and were familiar with search engines. They used the WISST 3.0 in the Human–Computer Interaction laboratory located at Middle East Technical University. Participants had no previous experience with eye-tracking technology. The detailed descriptives are listed in Table 1.

2.2. Apparatus

In the Human–Computer Interaction Laboratory, there was a PC with a 17-in. monitor set at 1024 × 768 resolutions. There was an adjustable chair for the user and a chair and a small table for the observer. All sessions were video recorded. Tobii 1750 Eye Tracker (Tobii Technology, 2008) with 50 Hz sampling rate was utilized to record eye movements while the user interacted with WISST 3.0. The eye tracker has 0.5° accuracy and .25° spatial resolution. Its freedom of head movement is about 30 × 16 × 20 cm (W × H × D) at 60 cm away from the tracker. Camera field of view is 21 × 16 × 20 cm at 60 cm distance. The Tobii 1750 allows binocular tracking.

TABLE 1
Descriptives

Descriptive	Frequency
Gender	
Female	6
Male	5
Age	
12	6
13	5
School	
Public School 1	4
Public School 2	4
Public School 3	3
Computer & Internet access at home	
Yes	11
No	0
Web search experience/Frequency	
Whenever needed . . .	8
If teacher assigns . . .	3
Preferred search engine	
Google	9
Others	2
Eye-tracker experience	
Yes	0
No	11

2.3. WISST

In addition to the eye tracker, the data were collected through WISST 3.0, which kept records of any search parameter including duration, used keywords, visited links, and number and duration of search sessions. For each user, the system kept the logs.

2.4. Procedures

After completing necessary permissions, each student was scheduled to come to the Human–Computer Interaction Lab. Students were expected to act in the same way that they search for a regular performance task at home or at the school computer lab; therefore, no training was given on how to use WISST. Three tasks were designed as representatives of regular performance work questions with the help of middle school science teacher. First, the teacher introduced the sample assignments to the researchers; then the researchers agreed on the three topics and prepared three questions with the guidance of the teacher. Meanwhile, the researchers categorized the questions according to three levels. Ready-to-use task refers to easy-to-find information with simple keywords through one or two clicks. It can be considered a kind of close-ended task, and the answer can be found by exploring the results page rather than visiting a page. Easy-to-interpret task is not as easy as the former one. It requires organizing keywords in a meaningful manner through the use of Boolean operators. The user might need to compare and contrast the found information if the keyword combination is not appropriate. The user might also need to visit more than one page to confirm the inclusion of all differences. Unlike the ready-to-use task, the hard-to-interpret task is neither accessible through simple keyword entry nor one click away. Like the easy-to-interpret task, it requires the use of meaningful combination of keywords with Boolean operators; however, the complete answer is not accessible directly, that is, the user needs to interpret and find the relations between the found information. In the literature, difficulty levels of tasks were distinguished as well defined or ill defined (Schacter, Chung, & Dorr, 1998), close-ended or open-ended (Tu, Shih, & Tsai, 2008), knowledge finding or argument (Tsai, Tsai, & Hwang, 2011), or fact-based or research tasks (Bilal, 2001). In our defined levels, ready-to-use is parallel to fact-based task; hard-to-interpret is parallel to argument task, but easy-to-interpret remains in the middle of the continuum. Before conducting the study, all tasks were performed many times by both researchers and the teacher to ensure that the tasks are consistent with the expectations.

Task 1 (ready-to-use): Search the Internet and find the name of native Australians.

Task 2 (easy-to-interpret): Search the Internet and find the differences between aerobic and anaerobic respiration.

Task 3 (hard-to-interpret): Search the Internet and find how steam affects global warming.

Students were assigned to complete three tasks ranking from easy to hard (from Task 1 to Task 3), and the tasks were given one at a time in order. All three tasks were expected to be completed without any time limit. Users were free to try any method they needed, that is, there were no limits for the trials. Students were required to write their answers on paper for each task.

Each search session was recorded, and gathered data were visualized with the help of Tobii Studio software. ClearView software (Tobii Technology, 2008) enabled a fine calibration adjustment. At the beginning of the sessions, the calibration was checked for each participant. While looking for the research question answers, the researchers first decided and defined the appropriate areas of interest (AOI). Figure 2 includes the AOI defined for WISST by the researcher. The upper part consisting of menu elements were labeled TOOLBAR. The center of the interface was named MAIN because this is the main part of the software where they search and read. The bottom part was divided into two: DISPLAYBAR; HISTORYBAR. DISPLAYBAR shows user inputs including goal of the search and keywords, and HISTORYBAR includes visited sites and related buttons. Such metrics as fixation counts/durations/numbers/sequence, gazing time, total visit duration, and scan path were analyzed through the guidance of AOI. MAIN AOI was the busiest area in terms of fixations. Reading and scanning styles were decided within those boundaries. HISTORYBAR and DISPLAY AOI were used to compare

FIG. 2. Defined Areas of Interest of the Web-based Internet Search Scaffolding Tool.

if the logs of visited sites and keywords were consistent with the kept logs.

Eye-tracking data were divided into segments manually. New search, end search, and keyword change segments were defined within separate intervals where the user starts seeking for new search, end search, and keyword change buttons, that is, when the gazing starts on that AOI and ends when the user clicks the buttons. The commonality of these buttons is that they all located within TOOLBAR AOI. To generate specific metrics, sometimes scenes were generated from appropriate segments.

Quantitative data, consisting of click accuracy, number of visits, time spent, and number of keywords, were stored by WISST logs and were analyzed using SPSS 15.0 program. The click accuracy as a dichotomous variable refers to the relevance of the visited pages. The number of visits variable refers to the number of web pages visited during a search process. The time-spent variable means how much time spent on scanning, on reading, or on specific AOIs. The number of keywords refers to the entered keywords for a web search task. The data gathered from search parameters were used to explore such descriptive data as percentages. In addition, those data had the benefit of revealing relations between variables through the use of either Spearman's rho or Pearson correlation methods. Eye-tracker data such as gaze plots were used to explore the scan types. For example, the sequence of fixations was used to define the scan types of users with the combination of WISST logs. First, all gaze plots were reviewed to find common points; second, themes were written; third, connections between themes were defined; and fourth, themes were grouped under meaningful titles that are the scan types and reading styles. Last, two other collegues reviewed the process.

3. RESULTS

Next is a description of data collected from the participants about the first research question. Then the second and third questions are addressed in relation to search parameters. Search parameters were investigated in two different parts. In the first part, students' scan types were focused, and the relations to both click accuracy and number of visited links were analyzed. In the second part, the relations between search parameters consisting of click accuracy, number of visited links, time spent on search, scan type, number of used keywords, and reading styles were explored with the focus of task types. The session records of WISST and skimming data of eye tracker were used to explore both qualitative and quantitative parameters. Last, the findings for the fourth question were presented.

3.1. Search Patterns Accross Tasks

To explore the search patterns of participants, the skimming data gathered through eye-tracker and the data gathered through WISST session recordings were used. The gaze data within MAIN-AOI were used to demonstrate the fixations during the search process. Because the flow of search process is the key issue to specify the search patterns of users, the scan path metric was used, that is, the sequence of fixations within the defined area of interest (MAIN-AOI) were taken as gaze plots to make scan paths interpretable. Figure 3 includes examples of gaze plots on which fixations represents via dots varying in size.

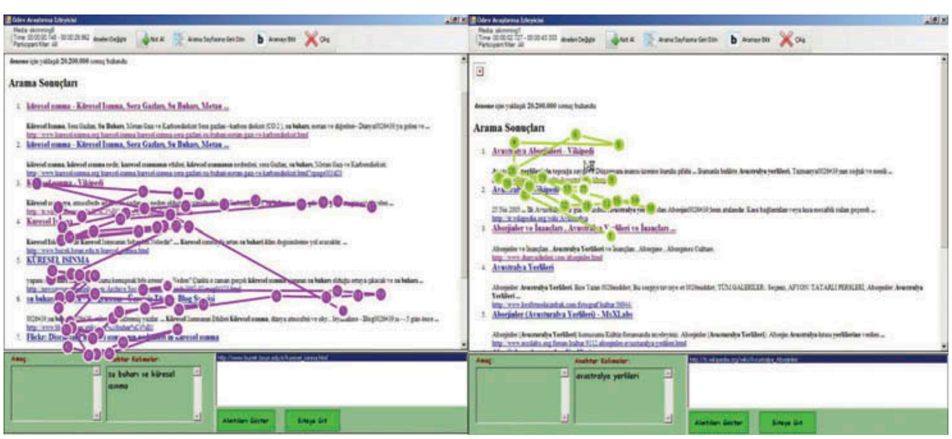


FIG. 3. Gaze plots of two different users.

TABLE 2
Scan Types of Participants

Scan Type	%
One shot	6.9
Forward linear	13.8
Backward linear	17.2
Mixed	27.6
Nonlinear	34.5

Scan types were determined with the combination of gaze plots and WISST records of visited links. The categorization was first determined by one of the researchers, and then two other colleagues categorized the data in the same way. In other words, there was a 97% agreement across observers. Finally, all observers agreed on five types of scanning: (a) one shot, (b) forward linear, (c) backward linear, (d) mixed, and (e) nonlinear. *One-shot scan* refers to those visiting only one web page, not judging the relatedness of the content. Moreover, even if they could not find the information they looked for, they did not proceed with the search. The second and third scan types were *linear* with different orientations; that is, whereas forward linear scan types start with high-ranked pages and continue to the lower ranks (1, 3, 4, 6 . . .), backward linear types first start with low-ranked pages and then continue to higher ranks (10, 7, 5, 3 . . .). *Mixed* types demonstrated a different pattern, including both characteristics of forward and backward linear scanning at one search session. Besides those types, there was another scan type named *nonlinear*, of which percentage is the highest. The distribution of percentages was summarized in Table 2.

Search patterns for Task 1 (ready-to-use). The findings showed that only four users tried the same keywords once. The keyword characteristics ranged from one word to a complete question. The most frequently used keywords were "native Australians." For this task, the word "Australia" and the complete question were used equally as keywords, but their frequencies were fewer than the others. Five out of seven users who entered "native Australians" as the keywords completed the task correctly. One of the common points in their search was that they also tried two other keywords ("Australia," "name of the native Australians").

Another common action was a visit to Wikipedia, but the rank of the visit varied across users. For instance, User7, who failed to answer correctly, used the complete question format besides the "native Australians" keyword. This user spent a total of 6 min deciding on which sites to visit, that is, to explore the results page. She only visited Wikipedia and read the content for 2 min, then gave up the task.

Another failing participant was User3. Similar to User7, she spent 10 min on a results page and 2 min reading the website contents. Although she clicked all related content, she failed to determine the necessary keywords.

Among all, six users were able to write the correct answer. They spent 2 to 10 min investigating the results page and 2 to 15 min within websites. They visited Wikipedia at least once. Except for User10, all users visited more than one website and changed keywords at least once. Their *number of website visits* was two to four. One common characteristic for those who completed the task successfully was that they were not distracted by visuals or ads and they scanned the titles before focusing on the linear reading of the related content.

Others who failed the task spent too much time deciding what link to click on the results page. The least amount of time spent was 6 min, and the most was 18 min. In line with decision making on the links, the *reading time* ranged from 2 to 25 min. They were easily distracted by ads especially. Their *scan type* is quite linear. Their incorrect *clicks* are more than the correct ones. Two of the users insisted on entering Wikipedia more than once, but due to its irrelevance, they stuck with the table of contents on the page. These two users went through nine website visits.

Search Patterns for Task 2. All users completed this task correctly. The most frequently used format as *keyword* was the question statement. Four students used it more than once and did not try any other keyword. Except for User4, they were very selective, and all focused on the content. A common pattern among these users was that they scanned the results and the pages in a linear fashion. Five users used the keywords in a divided form as aerobic and anaerobic respiration. They performed three to nine *visits*, which was more than others. They also spent much more time than others on their visits. Three of these users visited Wikipedia first. Their *click accuracy* was better than question searchers, who used the question statement as the keyword. Two of these users plus two other users tried to combine the words "aerobic" and "anaerobic" within one search session. Two users who used only this combination completed *scanning* the results within about 5 min and *reading* within about 3 min, which were less than all others. Their *scan type* was nonlinear, but they were observed focused on the content.

Search Patterns for Task 3. Except for User11, all users referred to the question format for this task in any part of the search. All users visited Wikipedia at least once. For most of the users, this site was both the first and the last visited site. Results *scan time* ranged from 4 to 17 min. Investigation of the *visited sites* ranged from 4 to 35 min. Except for User11, users *click accuracy* was about 50%. Only User5 failed to complete the task. All users were observed *reading and scanning* linearly with a close focus on the titles and the content.

3.2. Relations of Scan Types With Click Accuracy and Visits

After a series of examinations on gaze plots and WISST logs together, the researchers decided on the five scan types (one shot, forward linear, backward linear, mixed, and nonlinear). Other parameters (click accuracy and number of visits) were

TABLE 3
Correlations Between Number of Visits and Click Accuracy

	Click Accuracy
No. of visits	
Pearson correlation	.605**
Sig. (two-tailed)	.000
N	33

**Correlation is significant at the .01 level (two-tailed).

TABLE 4
Click Accuracy/Number of Visits Ratio Across Scan Types

Scan Type	Click Accuracy/No. of Visits (%)
One shot	65
Forward linear	27
Backward linear	44
Mixed	86
Nonlinear	85

analyzed and coded with the help of WISST recordings and gaze records for each task type (11 × 3). Click accuracy was assessed based on the relatedness of the entered keyword for search. One hundred percent agreement was reported between the observers. Number of visits was counted from WISST logs. Pearson correlation was run to see if there was any relation between these two variables. It was reported that there was a relatively strong correlation between click accuracy and number of visits. As the number of visited sites increased, the tendency to find out related links also increased ($r = .61$; see Table 3).

Forward linear scanning type can be interpreted as the least efficient scan method among others, because click accuracy per visit is lower than others. On the other hand, either mixed or nonlinear scanning seems the most efficient styles, because 86% or 85% of visited sites are related with the searched concept (see Table 4).

3.3. Task Types' Relations With Click Accuracy, Visits, Time Spent, Scan Type, Keywords, and Reading Styles

Data collected from WISST records and gaze records of eye tracker were used to explore click accuracy, number of visited links, time spent, and keywords used. In addition, the gaze data were used to decide on both scan type and reading style through the utilization of fixations.

Click accuracy and number of visits. The results of Spearman's rho analysis showed that there was a significant correlation between click accuracy and task type ($r = .399$). On the other hand, no correlations were detected among the number of visits and task type ($r = .143$). Comparing the three tasks, the efficiency in terms of number of visits and correct hits was

TABLE 5
Correlations Between Number of Visits and Click Accuracy in Different Task Types

	Task Type
Click accuracy	
Correlation coefficient	.399*
Sig. (two-tailed)	.021
N	33
No. of visits	
Correlation coefficient	.143
Sig. (two-tailed)	.427
N	33

*Correlation is significant at the .05 level (two-tailed).

72% for Task 1, 78% for Task 2, and 87% for Task 3. Table 5 summarizes the correlations.

Time spent. When the task completion durations were compared, it was observed that they were very close to each other ($M_{task1} = 433$ s; $M_{task2} = 485$ s; $M_{task3} = 475$ s). Among users, 51% were below the average task completion duration for the first task. For Task 2, the percentage was about 59%, and for Task 3 it was 55%.

Scan type and number of used keywords. No significant correlations were found between five scan types and three task types. In addition, the number of used keywords was also found to be unrelated to task types.

Reading style. Three types of reading styles can be identified within collected data: (a) distracted, (b) linear, and (c) skimming readers. Most of the participants showed different patterns of reading during the search periods. Their styles changed depending on the tasks, keywords, and visited pages. At one session, more than one style was observed for each participant. "Distracted reading style" took place whenever users came up with advertisements or any other irrelevant visuals. This style was coupled with unfocused keywords, and then irrelevant page visits occurred, which sometimes caused students to be engaged in the table of contents of Wikipedia. "Linear reading style" emerged whenever the keywords were very specific and focused. It was observed that participants who started with linear reading stopped, scrolled, and skimmed if they saw unrelated content. The last reading style, "skimming over," generally benefitted from titles, colors, familiar keywords, or sentences. The users with this style read one or two sentences at most, and these were generally from the first paragraph. It was observed that if the content is consistent with the entered keyword, then the style turns out to be linear reading.

For Task 1, five users were unfocused readers, and they were easily distracted by advertisements, unrelated visuals, and pop-up windows. This reading style was less common for Task 2, which was more relatively structured and clearly defined than the other two tasks. The number of linear readers was the highest in Task 2, but the readers were relatively slower in completeing the task than readers in the other styles. Reading strategy that involved skimming before deep reading was used during each task, but the percentage of use was lower than the other strategies. The results of the correlation run did not generate significant correlations between reading style and task type.

3.4. Conditions Leading to Changes in Search Process

Findings revealed that participants changed their manner of searching in three ways: (a) after irrelevant page visits, (b) after relevant but poor page content, and (c) after a prescan of a results page. The first condition occurred generally after a couple of incorrect website visits. This is valid especially for Tasks 1 and 2, which were relatively easier than Task 3. The second condition was observed after the correct but unsatisfactory visits. In Task 3, more than half of the users frequently changed keywords due to poor visits. For the last condition, users preferred to try more new keywords for Task 3 than others as a result of irrelevant or insufficient results.

While changing the search behavior, the strategies were also revised parallel to that change. For Task 1, almost all users started with simple keywords such as "native Australians." Only one user directly used the question as the keywords. When they needed to search more, users used either the same keywords or very similar keywords. For Task 2, five users preformed the search in two steps—aerobic respiration and unaerobic respiration. Four users directly input the question instead of a generation of keywords. Only two users used simple keywords. For Task 3, six users started with simple keywords and then continued with a complex combinations of keywords, except one user who tried a question sentence and then went back to simple keywords. The rest of the users simply entered a question in the keyword field. However, three of five students decided to convert the keywords into a simpler form, whereas two users tried to find the results through the exact question sentence.

Overall, most of the users started with simple keywords for a simple task. For the more difficult task, they used division strategy to manage the search process, and they divided the contents into manageable parts to search. However, the ones who treated the second question as a keyword also continued to use the same strategy for the next harder task. The users, who preferred simple keywords for the most difficult task, felt a need to arrange the keywords in different and preferably more complex combinations. On the other hand, a few users changed their question strategy after a couple of trials into simple keyword forms.

4. DISCUSSION AND CONCLUSION

This study took place in a human–computer interaction laboratory equipped with an eye-tracker machine and video recorders. The data were collected from both WISST and eye tracker. Eleven participants were assigned to three web search tasks and were observed throughout the search process by the researchers. Because the tasks differed in terms of difficulty and style (ready-to-use, easy-to-interpret, and hard-to-interpret), it was expected that the patterns of search for each task would differ as well. The results indicated that there were some differences between search patterns in terms of task types. For example, as the tasks became more complex, users performed more efficiently in terms of selection of link visits. The efficiency calculated with number of visits and correct hits was 72% for Task 1, 78% for Task 2, and 87% for Task 3. In Task 3, the users had to think deeper to figure out the main point, to find exact keywords, to discriminate the related content, and to interpret the found information in order to complete the ill-structured task successfully. The initiation and the flow of the search process depended on the users' understanding of the information they are requested (Lucas & Topi, 2004). Although there were not too many failures of Task 3, the whole question statement was frequently treated as a keyword combination (Tu et al., 2008), which was very common when users did not know where to start.

The investigation of time spent on a web search indicated that it was not related to task type. Despite 35-min page visits in Task 3, there were no significant correlation between time spent and task difficulty. These findings might be due to the other parameters such as the density of the searched content, familiarity with the content, previous experiences, distracters on visited page, and so forth. Moreover, relatedness of the keyword might have affected the results because there are many possible combinations and there is no direct answer of the question, that is, the user needs to interpret the found information. Although there was no statistically significant relationship between task type and time spent, it was observed that longer task completions were more frequent in Task 3, which was parallel to the Lorigo et al. (2006) findings. Unlike other tasks, this task led users to focus more on the content and required much more cognitive effort to distinguish between relevant and irrelevant information. This might be related to the complexity of the content. On the other hand, task completion durations were observed very close to each other. It was not an expected result due to the varying difficulties. The result might be different if more than 11 users participated in the study, because in our sample the variance is considerably high and the extreme cases can influence the mean scores. The participants had no previous experience with the eye-tracking machine, and being in a laboratory environment might have distracted them in the beginning of the time for Task 1, the simplest one. The reading fluency/speed, comprehension skills, speed of motor response, and many other factors might help to explain these close values of task completion. This issue needs to be studied in details with the control of other variables. For example, there are certain psychometrics such as cognitive styles or search experiences (Kim, 2001) affecting the quality of interaction for web pages (Van Schaik & Ling, 2005), and these can also have effects on task completion duration.

Other search parameters including scan type, number of keywords, and reading style were not correlated to task type. However, when the scanpaths were explored, it can be concluded that users generally applied linear scanning (both backward and forward) at Task 1, nonlinear scanning at Task 2, and mixed scanning (combination of backward and forward) at Task 3. There have been findings about the tendency to scan linearly (Cutrell & Guan, 2007; Joachims, Granka, Pan, Hembrooke, & Gay, 2005; Klöckner, Wirschum, & Jameson, 2004), but the current study implies that task type can play a role on scan type strategy. While comparing navigational and informational search tasks at different ranks, Guan and Cuttrell (2007) reported that the low ranks could bring about longer and less successful search results, especially for informational tasks. The similar pattern occurred in this study, that is, task type might have caused the emergence of different search behaviors. In addition, the user can apply more than one strategy during the whole search process, because searching the web, which requires decisions to be made to access relevant information, does not always generate intended results (Fang & Salvendy, 2000); thus, if the entered keyword is not focused enough, then the strategy could depend on the relatedness of the generated results page.

"Number of keywords" parameter was found statistically not related to task type, but the way keywords were used was different. For example, simple keywords were used widely both in Task 1 and Task 3, but "question statement" keywords were more common in Task 3. On the other hand, Task 2 brought about the divided usage of the whole statement, that is, most of the users performed two-step search sessions using keywords one by one. It can be inferred that because of the complexity of tasks, the user might not be certain where to start, which can lead to the use of whole question as a keyword because of the complexity and indirect access of the answer (Marchionini, 1989).

Reading style and task type were also found statistically not related, and this might be because of the utilization of more than one reading style at one session. Reading style types did not exactly fall into one task; instead, each task included more than one style. Among three reading styles (distracted, linear, and skimming), distracted style was observed least for Task 2, which was relatively more structured than the other two tasks. An explanation for that situation could be the consistency between the keywords and the listed content. Distracters causing slower reading can be listed as irrelevant keywords and irrelevant visuals, especially the pop-up ads and moving objects. It was also observed that users who applied skimming as a reading style tried to find a familiar keyword or sentence. At this point, the user decided on the relatedness of the content and then either continued to read in a linear manner or went back to the results page. This condition can refer to a kind of strategy change. There were other cases bringing about some strategy changes during the search process. For example, when there was an inconsistency between the entered keyword and listed pages, the user needed to change the keywords after a couple of unrelated page visits. This was espeacially observed for Task 1 and Task 2. On the other hand, during Task 3 engagement, users generally changed strategies in two situations: The first one occurred after visiting pages with related but poor content, and the second one occurred after a prescan of the results page. This is an interesting finding, because it was expected that users could visit more relevant pages for the first two tasks; these are more basically stated in question format, and therefore it could be easier to find related keywords. Nevertheless, no matter how pointed the keywords are, the users rarely conducted a prescan of the results page, which in turn caused unrelated page visits and then strategy changes. Lorigo et al. (2006) also reported that participants did not spend much time on the search results page but did on web documents, which was obvious for most of cases in this study.

During three task types, there were both successes and failures. The ones who failed to finish the task or wrote wrong answer generally visited Wikipedia at least once. This might be because of this site's nature of organized content format bringing about students' frequent visits. Even when the keyword was correct and the page visit was relevant, some users failed. This could be explained by scan types and reading styles, because failing searchers were observed spending a considerable amount of time deciding on what to click on the results page (Guan & Cutrell, 2007). Successful searchers, whose characteristics were very similar to Tabatabai and Shore's (2005) description of successful web search, exhibited less distracted reading and continued reading after skimming the content regardless of the task type. Users who achieved the relevant content through appropriate keywords were observed to be more focused on the task. This might be related to the number of accurate clicks. In other words, unrelated visited sites might disturb the attention; thus for further cognition, the user might become exhausted and overloaded. Moreover, students' knowledge about the searched content might have also affected the success of achieving related information (Brandt, 1997).

This study might be considered as one of the pioneers regarding the web search pattern analysis through the use of both a web search scaffolding tool and eye-tracking methodology. The findings of this study can contribute to studies trying to define and explain the usability of specific search engines. One of the interesting findings of the study is about the frequent use of Wikipedia. Because of the well-defined structure of this page, the web designers as well as instructional material designers can benefit from the tips about the page. This study may also contribute to the definition and model of successful versus unsuccessful web searchers. Finally, this study can provide some guidance for those who plan to create new search engines specific to children.

Because the number of participants and tasks were limited, the findings of this study might not be generalized. The design of the study could be replicated at different settings with an increased number of participants. Although the study is indirectly concerned with participants' reading styles, any measurement of reading/comprehension abilities was applied before the task assignment. This is one of the limitations of this study and further studies needed to understand the impact of this ability. The cases occurred in a laboratory setting with a PC; thus the results could change if the context changes. For example, mobile platforms to observe search behavior can be compared to desktop search behavior. The Google search engine was used for this study, but further research could try many other search engines to distinguish between the advantages and disadvantages of each engine. Moreover, further versions of WISST can be developed through the results of this study. Observing adults and elderly peoples' search behavior can also move this research a step forward. In addition to these issues, elaborating on why users frequently visited Wikipedia can also be a challenge for researchers.

REFERENCES

Bar-Ilan, J., & Belous, Y. (2007). Children as architects of web directories: An exploratory study. *Journal of the American Society for Information Science and Technology*, *58*, 895–907.

Berk, L. E. (2003). *Child development* (6th ed.). Boston, MA: Allyn and Bacon.

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine: II. Cognitive and physical behaviors on research tasks. *Journal of the American Society for Information Science and Technology*, *52*, 118–136.

Brandt, S. D. (1997). Constructivism: Teaching for understanding on the Internet. *Communications of the ACM*, *40*, 112–116.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., & Shneiderman, B. (2004). Determining causes and severity of end-user frustration. *International Journal of Human–Computer Interaction*, *17*, 333–356.

Chau, M., Qin, J., Zhou, Y., Tseng, C., & Chen, H. (2008). Spidersrus: Creating specialized search engines in multiple languages. *Decision Support Systems*, *45*, 621–640.

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in web search. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 407–416.

Dinet, J., Bastien, J. M. C., & Kitajima, M. (2010). What, where and how are young people looking for in a search engine results page?: Impact of typographical cues and prior domain knowledge. *Proceedings of IHM'10 Conference Internationale Francophonesur I'Interaction Homme-Machine*, 105–112.

Fang, X., & Salvendy, G. (2000). Keyword comparison: A user-centered feature for improving web search tools. *International Journal of Human–Computer Studies*, *52*, 915–931.

Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 417–420.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *Proceedings of ACM Conference on Research and Development on Information Retrieval (SIGIR)*, 154–171.

Kim, K.-S. (2001). Implications of user characteristics in information seeking on the World Wide Web. *International Journal of Human-Computer Interaction*, *13*, 323–340.

Klöckner, K., Wirschum, N., & Jameson, A. (2004). Depth- and breadth-first processing of search result lists. In *Extended Abstracts of ACM Conference on Computer-Human Interaction* (p. 1539). New York, NY: ACM Press.

Kuhn, D., (1999). A development model of critical thinking. *Educational Researcher*, *28*(2), 16–26.

Large, A., & Beheshti, J. (2000). The web as a classroom resource: Reactions from the users. *Journal of the American Society for Information Science*, *51*, 1069–1080.

Large, A., & Beheshti, J. (2005). Interface design, web portals and children. *Library Trends*, *54*, 318–342

Large, A., Beheshti, J., Nesset, V., & Bowler, L. (2004). Designing Web portals in intergenerational teams: Two prototype portals for elementary school students. *Journal of the American Society for Information Science and Technology*, *55*, 1140–1154

Lorigo, L. Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on serach and evaluation behavior using Google. *Information Processing and Management*, *42*, 1123–1131.

Lucas, W., & Topi, H. (2004). Training for Web search: Will it get you in shape? *Journal of the American Society for Information Science and Technology*, *55*, 1183–1198.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of American Society for Information Science*, *40*, 54–66.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*, 235–244.

Schacter, J., Chung, G. K., & Dorr, A. (1998). Children's Internet searching on complex problems: Performance and process analysis. *Journal of the American Society for Information Science*, *49*, 840–849.

Sendurur, E. (2012). *Effects of a web-based Internet search scaffolding tool on metacognitive skills improvement of students with different goal orientations* (Unpublished doctoral dissertation). Middle East Technical University, Ankara, Turkey.

Sendurur, E., & Yildirim, Z. (2010). Formative evaluation of web-based internet search scaffolding tool. In M. Simonson (Ed.), *33rd Annual Proceedings of Association for Educational Communications and Technology* (Vol. 2, pp. 277–281). Available at: http://aect.site-ym.com/?page=ConvProceedings (Accessed November 21, 2014)

Sendurur, E., & Yildirim, Z. (2012). *Iterative design of a web-based Internet search scaffolding tool: Putting a theoretical framework into practice*. Paper presented at The European Conference on Educational Research, September 18–21, Cadiz, Spain.

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research*, *27*, 222–248.

Tobii Technology. (2008). *Tobii Studio 1.2 user manual*. Danderyd, Sweden: Tobii Technology AB.

Tsai, P-S., Tsai, C-C., & Hwang, G-J. (2011). The correlates of Taiwan teachers' epistemological beliefs concerning Internet environments, online search strategies, and search outcomes. *The Internet and Higher Education*, *14*, 54–63.

Tu, Y.-W., Shih, M., & Tsai, C.-C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers & Education*, *51*, 1142–1153.

Van Schaik, P., & Ling, J. (2005). Five psychometric scales for online measurement of the quality of human-computer interaction in web sites. *International Journal of Human–Computer Interaction*, *18*, 309–322.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, *15*, 159–176.

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*, 89–109.

ABOUT THE AUTHORS

Emine Şendurur is an assistant professor in the Computer Education and Instructional Tehcnology Department at Ondokuz Mayis University. She completed her Ph.D. at Middle East Technical University. Her main research interest includes

human–computer interaction and web search skills. She teaches undergraduate and graduate courses about human–computer interaction and instructional design issues.

Zahide Yildirim is a professor in the Computer Education and Instructional Technology Department at Middle East Technical University. Her research interests include use of technology as cognitive tools, distributed learning environments, instructional message design, and human performance technology. She teaches courses on instructional design and theoretical bases for technology integration into education topics.