An Eye Tracking Study of Web Search by People With and Without Dyslexia

Srishti Palani
Design Lab, UC San Diego
srpalani@ucsd.edu

Adam Fourney
Microsoft Research, Redmond, WA
adamfo@microsoft.com

Shane Williams
Microsoft Research, Redmond, WA
shanewil@microsoft.com

Kevin Larson
Microsoft, Redmond, WA
kevlar@microsoft.com

Irina Spiridonova
Microsoft Research, Redmond, WA
irinasp@microsoft.com

Meredith Ringel Morris
Microsoft Research, Redmond, WA
merrie@microsoft.com

ABSTRACT

Web search is a key digital literacy skill that can be particularly challenging for people with dyslexia, a common learning disability that affects reading and spelling skills in about 15% of the English-speaking population. In this paper, we collected and analyzed eye-tracking, search log, and self-report data from 27 participants (14 with dyslexia) to confirm that searchers with dyslexia struggle with all stages of the search process and have markedly different gaze patterns and search behavior that reflect the strategies used and challenges faced. Based on these findings, we discuss design implications to improve the cognitive accessibility of web search.

CCS CONCEPTS

• **Information systems → Search interfaces**; • **Human-centered computing → Empirical studies in accessibility**.

KEYWORDS

Web search; Information Retrieval; Dyslexia; Eye tracking

ACM Reference Format:
Srishti Palani, Adam Fourney, Shane Williams, Kevin Larson, Irina Spiridonova, and Meredith Ringel Morris. 2020. An Eye Tracking Study of Web Search by People With and Without Dyslexia. In *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), July 25–30, 2020, Virtual Event, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3397271.3401103

1 INTRODUCTION

Dyslexia is a cognitive difference that impacts about 15% of English speakers [7] (incidence rates vary by language [30]). People with dyslexia tend to experience challenges in tasks involving reading, writing, spelling, and memory, despite having normal intelligence. These challenges often manifest in a slower reading rate and lower reading comprehension [35]. However, because dyslexia is a spectrum disorder, different people may experience different subsets and degrees of symptoms [7].

SIGIR '20, July 25–30, 2020, Virtual Event, China

ACM ISBN 978-1-4503-8016-4/20/07.
https://doi.org/10.1145/3397271.3401103

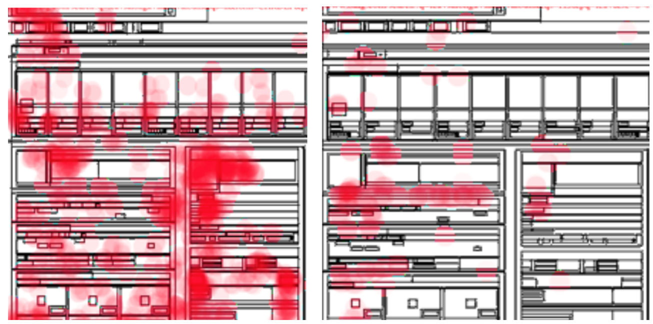

Figure 1: Example of a searcher with dyslexia exhibiting the Commitment fixation pattern (left), and a control group searcher exhibiting the F-shape fixation pattern (right) on a SERP. Each red dot is a fixation. DOM elements are outlined.

Relatively little is known about how dyslexia impacts web search, but several recent studies have begun to shed some light on this topic. Interviews with people with dyslexia suggest that search engine use - including query formulation, search result triage and information extraction from target webpages - is particularly challenging for this population [4, 19, 35, 37, 43]. Online experiments comparing search behaviors of people with and without dyslexia have identified some features of webpages, such as average line length and the ratio of images to text, that impact page readability for people with dyslexia [25].

In this paper, we build on this knowledge by conducting an eye tracking study to investigate the following research questions:
RQ1: Are there any differences between gaze patterns of searchers with and without dyslexia?
RQ2: What does the gaze data reveal about the challenges experienced and strategies employed by these two searcher groups?

To address these research questions, we conducted an eye tracking study with 27 participants (14 with dyslexia, 13 control). Each participant completed 6 informational search tasks using a modern, English-language, interactive search engine. Our analysis of eye tracking data, together with participants' self-reports about their experiences, extend findings from prior studies and contribute new insight into the differences in search behavior between people with and without dyslexia. Our findings validate prior self-report findings that Searchers with Dyslexia (SWD) struggle with all stages of the search process: query formulation, search results triage, and information extraction [37, 43]. Adding to prior work, we find SWD visually attend to pages in a markedly different fixation pattern than searchers without dyslexia (for e.g. see Figure 1). We conclude the paper by discussing the design implications of our findings to improve cognitive accessibility of search for people with dyslexia.

2 RELATED WORK

This research builds on prior work done to capture and understand search behavior, and on studies of searchers with dyslexia.

2.1 Methods for Understanding Web Search

Researchers in information retrieval and HCI have extensively studied user interaction with web search systems, particularly for complex informational [16] or exploratory [55] search tasks. Such tasks typically comprise three stages [3, 43]: *query formulation* (i.e., generating and refining search keywords), *search results triage* (i.e., determining which parts of the search engine results page – the SERP – are most relevant to the task at hand, and which link to open), and *information extraction* (i.e., gathering and making sense of the sought-after content). In this study, we gather data about this complete query-triage-extraction search pipeline for searchers with and without dyslexia.

Researchers have employed a variety of methods to study web search, including analyzing search engine and web browser logs (e.g., [32, 53, 54]), gathering self-report data from surveys, interviews, or diary studies of end-users (e.g., [42, 43]), and recruiting participants to perform controlled search tasks (e.g., [3, 25, 44]). As with any methodology, there are trade-offs: logs can provide in-situ data for a large set of users, but lack qualitative depth; self-report data may have gaps or inconsistencies with actual observed behavior; controlled, in-lab task performance may differ from natural search behavior in unanticipated ways, etc. Several researchers have begun to use eye tracking to understand which aspects of the SERP users attend to [22, 23, 36, 52]. Eye tracking allows us to log and track the amount of attention paid to specific parts of the pages and interactions at a granular level of space and time. For example, eye tracking studies have revealed that searchers usually fixate on SERPs and webpages in an F-shape pattern [45, 46]. These studies have also shown that searchers distribute their visual attention differently across organic and ad results on a SERP [24], and that one can determine a webpage's most salient parts by looking at the amount of visual attention paid to its different elements [17]. This paper builds on prior eye-tracking studies to understand how different searchers attend to different elements of SERPs and webpages. We complement our eye tracking data with participants' comments immediately after searching to gain qualitative insight into the meaning of the eye tracking results.

2.2 Dyslexia and Web Search

In this section, we discuss prior work on challenges faced by searchers with dyslexia (SWD) at each stage of the search process.

2.2.1 Query Formulation: Since spelling and query formulation are closely related, this stage of search has been reported to be challenging for those with dyslexia. From their interview study, Morris et al. [43] reported SWD have trouble spelling words at the phonetic level. They also found SWD report a heavy reliance on voice input and on autocomplete when forming queries. In 2015, Berget and Sandness [13] compared search logs of 21 students with and 21 without dyslexia who had formulated Norwegian queries for a Norwegian academic library system with no query-formulation aids. They found that SWD took longer to search, possibly because of spelling errors in their queries. Consequently, they also issued more queries overall. In their 2016 qualitative study, Cole et al. [19] confirmed these findings for English queries in a system without any query formulation aids. They reported that SWD found choosing keywords, spelling, and forming/refining complex queries using Boolean search strategies of AND and OR operators to be more challenging than a control group. However, in 2016, Berget and Sandness [14] found that when SWD used Google, a modern interactive search engine with query formulation aids, for formulating Norwegian queries, there were no differences in query formulation behavior. This suggests query formulation aids could help those with dyslexia. Through our eye-tracking study, we validate these self-report results for a modern interactive English-language search engine with query-formulation aids.

2.2.2 Search Results Triage: In 2010, MacFarlane et al. [37] conducted a search log analysis of people with dyslexia interacting with an information retrieval interface (Okapi), and found SWD read fewer documents on average, had fewer search interactions, and took more time to complete searching. Concluding their article, they state the need for a more granular approach to understand search behavior differences between searchers with and without dyslexia than search log analysis.

Furthermore, prior work suggests a strong correlation between dyslexia and lower phonological working memory (i.e., the ability to hold words in short-term memory) [10, 38]. Work done to investigate the effects of working memory on search results triage echoes MacFarlane et al.'s findings. In 2019, a search log study of participants with low and high working memory found that those with lower working memory take more time to first click on the SERP, open fewer links, and take more time between events [18].

2.2.3 Information Extraction: In 2007, Al-Wabil et al.'s [4] interview study observing searchers with dyslexia navigate multiple pages within a website found navigational trails (such as site maps, back and forward buttons) and menus helped SWD locate where they were within a site. Similarly, in 2008, an eye tracking study by Al-Wabil et al. [5], observed 7 participants (2 with dyslexia and 5 controls) navigate webpages within 6 websites to extract information. They found SWD took longer to complete the tasks, had more fixations on the page, looked at the page for longer, and their scan paths were markedly different than the control group. A follow-up analysis reported SWD changed scan direction more on the webpage than the control group [39]. They suggest this is an indication of short term memory problems and adds to prior research on backtracking [4]. However, the small number of participants (2 SWD) makes it hard to extrapolate definitive quantitative differences from these results.

In 2018 Fourney et al. [25] found SWD's relevance ratings of webpages were highly correlated with their readability scores. They identified several visual and textual features such as line length, number of headings, and ratio of images to text, that impact SWDs' readability and relevance judgements of webpages. In 2019, Li et al. [35] found using the "Reader View" mode in Firefox web browser, which simplifies a page's visual structure, improved reading speed for SWD without reducing comprehension.

Our study builds on these prior efforts by capturing eye gaze patterns of 27 participants (14 SWD, 13 control) using a modern, English-language, interactive search engine during the complete query-triage-extraction search pipeline.

3 METHOD

Our primary research question asks if there are differences between gaze patterns and search behavior of searchers with and without dyslexia. To answer this question we designed a study that collects search log, eye tracking, and self-report data through all three stages of search: formulating queries, triaging search results, and extracting information from the resultant pages.

3.1 Participants

We recruited participants using paid social media ads targeted towards residents of our (anonmyized, US-based) metropolitan area, who followed the *#dyslexia* hashtag or any of a set of organizations related to dyslexia, including a local special-education secondary school. As gratuity for participating in the hour-long study, participants received a $50 Amazon.com gift card. The study took place over a three-week period in the summer of 2019.

We recruited 32 participants, half of whom had a medical diagnosis of dyslexia. However, eye tracking data could not be reliably collected for 5 participants: in 4 cases this was due to a malfunction of the instrumentation, and in 1 case tracking failed because the participant had an eye condition (amblyopia). These 5 participants are removed from further consideration, leaving 27 total participants: 14 searchers with dyslexia (SWD), and 13 searchers in the control group.

Participants' ages ranged from 13-72 years ($\overline{x}$ =27.3, σ =17.3). 12 participants were between the ages 13-17 (and received parental permission to participate), and 15 participants were 18 or older. Participants with and without dyslexia were well-balanced across age groups with 7 teenagers with dyslexia and 6 without, and 7 adults with dyslexia and 7 without. One adult reported having attention deficit hyperactivity disorder (ADHD) in addition to dyslexia. ADHD, a neurological difference that is characterized by difficulty focusing, is a common comorbidity that occurs with dyslexia [26]. Of the 27 participants, 20 reported using web search engines multiple times on the average day, 5 reported using search at least once per day, and 2 reported searching multiple times per week. When asked to rate their search expertise, 8 participants self-rated as expert searchers, 17 as intermediate, and 2 as novice. Adult participants had a diverse set of occupations, including: program managers, designers, engineers, analysts, a school superintendent, and a human-resources director. Teen participants were all enrolled in secondary school.

9 participants were male and 18 were female. In the general population, the gender ratio for dyslexia is typically near parity, or skewed slightly towards males [29]. Within our sample, 5 participants with dyslexia identified as male, while 9 identified as female. In the control group, 4 participants identified as male and 9 identified as female. This gender skew likely reflects a limitation of our recruitment methods – women are more likely to use social media, follow mailing lists [27], and be special education teachers [2], so may have been more likely to see our ads. Lastly, all participants were fluent or native English speakers.

3.2 Apparatus

We conducted the study on a 17" LCD monitor, positioned at desk-level approximately 28" from the participant's eyes. This distance and head position were weakly enforced by seating participants in a four-legged chair, and then marking the location of each chair leg on the floor with tape. We used the monitor's native resolution of 1920 × 1200 pixels, and accepted the operating system's default display scaling of 150%. With this display setting, three or four search results were visible on the screen before scrolling. Participants could scroll pages freely, but could not adjust the scaling or browser zoom factors. Together, these controls ensured a consistent visual experience for all participants.

We used the Tobii 4c eye tracker [http://www.tobii.se/] upgraded with the Tobii Pro SDK to record each participant's gaze patterns as they interacted with the search engine and other webpages. The eye tracker sampled the position of the participant's gaze at a rate of 90Hz. From this data stream we identified fixations, which we define as a sequence of gaze points where: (1) the sequence accounts for at least 100 milliseconds of time, and (2) all gaze points fall within 15 pixels of a common centroid. A custom Google Chrome browser extension then mapped each centroid, in real-time, to an HTML element on the webpage that the participant was visiting. We logged both information about the fixation (location, duration) and information about the HTML element (e.g., the font size, element type, etc.). Finally, the browser was further instrumented to record page navigation events and search queries. To ensure a degree of ecological validity, all queries were issued to a major commercial web search engine (anonymized).

3.3 Procedure

As part of the recruitment process, participants answered a brief screening and demographics questionnaire that collected information about age, gender, occupation, language fluency, education, dyslexia diagnosis, dyslexia-related training/tutoring received, and search engine use (search frequency, preferred sites, and any additional devices or apps used to support searching). As noted above, we excluded participants who were not fluent in English, who were under the age of 13, and who were not located in our general metropolitan area. These exclusion criteria, together with the entire study procedure, were approved by our organization's IRB.

When each participant arrived at the research lab, the experimenter reviewed the study procedure with the participant, then calibrated the eye tracker to the participant using a 20-point calibration process. The study then proceeded as follows: Participants were asked to complete 6 search tasks, with a maximum of 10 minutes to complete each task, for a total study time of approximately one hour (see Table 1). For each task, the experimenter read a backstory out loud to the participant that outlined a particular information need that the participant needed to address in order to successfully complete that task. This procedure ensured that reading difficulties did not impact participants' understanding of the tasks, and also avoided giving participants clues regarding the spelling of search queries.

Participants indicated that they were done with a search task by pressing a button on the keyboard that was marked with a red

Phase	Task Complexity	Domain	Search Task	Query
1	Remember	Science	You recently watched a show on the Discovery Channel, about fish that can live so deep in the ocean that they're in darkness most of the time. This made you more curious about the deepest point in the ocean. What is the name of the deepest point in the ocean bed?	*Participant Formulates*
1	Understand	Entertainment	You are an avid tennis fan and are excited that your favorite player has finished the Wimbledon tournament undefeated. When the ATP (Association of Tennis Professionals) rankings come out, however, you notice that several other players are ranked higher than your favorite player. How are the ATP rankings determined?	*Participant Formulates*
2	Remember	Entertainment	You recently attended a concert and heard an artist called Lea Salonga. You really enjoyed the artist and want to purchase their latest album. What is the name of their latest (full-length) album?	"lea salonga latest album"
2	Understand	Health	Your cousin wants to join a fencing club sports team. Most of your relatives are supportive of the idea, but you think the sport is dangerous and are worried. You want to know what the potential risks they could face while fencing are.	"risks of fencing"
3	Remember	Commerce	Your family is considering buying a house in Orange County, North Carolina, but first want to check the current county property tax rate. What is the current property tax in Orange County, North Carolina?	"property tax rate for orange county nc "
3	Understand	Commerce	You have noticed that some coffee shops in your neighborhood advertise that they only sell 'fair trade' coffee. In order to decide whether to support these coffee shops you want to understand what the label 'fair trade' really entails. What are three requirements for coffee to be labeled as fair trade?	"fair trade coffee label"

Table 1: Search Tasks

sticker. They then dictated their answer out loud to the experimenter. The experimenter then administered an oral version of the NASA-TLX (task load index), which was modified to allow participants to respond on a 5-point (rather than 21-point) scale [18, 28]. This was also done orally to reduce reading strain while reporting the task load scores. Participants also reported their prior domain knowledge using a 3-point scale ranging from 1 ("*I could have answered this without searching for any information at all*") to 3 ("*I did not have any prior knowledge, and had to search for all information.*"). Finally, participants described any challenges they faced during the task, and reported any strategies they used to overcome those challenges.

3.3.1 Search Phases. The 6 search tasks (see Table 1) were divided into 3 phases, with each phase having 2 tasks to check if searcher behavior remained consistent across task differences. Tasks were presented in a randomized order. Each phase placed additional variables under experimental control so as to more effectively probe different stages of web search. These phases are as follows:

- *Phase 1:* Participants were presented with the landing page of the search engine, and the query box was left empty (See Supplementary Materials for example https://tinyurl.com/sigir2020supplementarymaterial). After hearing the description of a search task, participants were free to formulate any query, and could reformulate as often as desired. Likewise, participants were free to open as many search results as necessary. This phase was the most ecologically valid, and allowed us to collect differences at the query formulation stage of search between SWD and the control group. However, the ensuing diversity of queries, search results, and page-clicks rendered it difficult to compare the populations in later stages of web search. We address this by introducing additional experimental controls in phases 2 and 3, below.
- *Phase 2:* After hearing the description of a search task, participants were presented with a pre-populated search results page for a fixed query that could not be changed. From here, participants were again free to open as many search results as was necessary to complete the current task. Since all participants viewed the same SERP, we can more effectively measure differences in the results-triage stage of search. However, participants may open different pages, or may open no pages at all (i.e., completing the task using search results' snippets), thus rendering it difficult to compare populations in the final information extraction phase of search.
- *Phase 3:* Similar to Phase 2, participants were presented with a pre-populated SERP for a fixed query that could not be changed, but were additionally asked to open at least three webpages from the search results. This overlap of many participants looking at particular webpages allows us to compare information extraction behavior from webpages.

3.3.2 Task Types. All 6 tasks (2 per phase) were informational search tasks (i.e., tasks with an intent to acquire information present in one or more webpages), as opposed to transactional or navigational tasks [16]. To observe if searcher behavior remained consistent across tasks, we ensured each of the 3 phases had 2 tasks. Each phase's tasks occupied a different level of cognitive complexity, as defined in Kelly et al.'s framework [33]. Specifically, one task occupied the *Remember* level of complexity, and required participants to retrieve at most one piece of relevant knowledge. The other task occupied the *Understand* level of complexity, and required constructing meaning from multiple sources or documents. Of the two, *Understand* tasks are reported to be harder [18, 56]. When developing tasks for this study, we chose a set that covered diverse domains to reduce effects of prior domain knowledge [20].

3.4 Measures

To observe and analyze differences in search and gaze patterns across SWD and control searchers, we measured the following:

3.4.1 Self-Reported Task Load: Each task's mental, temporal, performance, effort, and stress level self-rated by participants on a 5 point Likert-type post-task questionnaire (a simplified NASA TLX)

[18, 28]. Higher TLX scores indicate more challenging tasks, and an overall task load is computed by summing the responses to all five questions, yielding a score between 5 and 25.

3.4.2 Search Log Measures: From the search logs we measured: **(i) Task completion time**: total time (in milliseconds) to complete each search task. **(ii) Number of queries issued per task**. **(iii) Length of query**: number of terms in a query. **(iv) Number of spelling errors in query**: including number of phonetic and typographic spelling errors, calculated using Damerau–Levenshtein edit distance [9] which is the minimum number of insertions, deletions, or substitutions of a single character required to change one query term into the correct term. **(v) Number of SERPs and webpages visited. (vi) Number of returns to each SERP and webpage**

3.4.3 Eye-tracking Measures: To observe what searchers looked at we divided the SERPs and webpages into several Areas of Interest (AOIs outlined in red in Supplementary Materials https://tinyurl.com/sigir2020supplementarymaterial). We measured the following across the AOIs from eye tracker data: **(i) Attention to an AOI**: is measured in terms of fixation duration. This is the total time that someone fixated on the AOI. **(ii) Number of search results viewed**: items on the search results list on the SERP with at least one fixation. **(iii) Fixation Patterns**: To observe how participants spatially distribute their visual attention, we looked at a visualization of their fixations on the page to detect any patterns. Previous work [45, 47] has established that searchers examine webpages using the following patterns (see Figure 2):

- *F-shape Pattern:* The F-shape Pattern is similar to the shape of its namesake, the letter F. Text on the left and towards the top of the page is read more than text on the right or towards the bottom of the page. [45]
- *Spotted Pattern:* The spotted scanning pattern involves fixating on specific words or parts of the page. Searchers choose these spots either because of visual salience (like bold or large text), or because word shapes resemble those of specific words they are looking for to answer the task. [47]
- *Layer-Cake Pattern:* Searchers fixate mostly on the page's headings and subheadings. There are few other fixations on the text in between — that is, until searchers locate the heading or part of the page that they are interested in. At that point, there are many fixations in that particular part. It is the most effective scanning method. [47]
- *Commitment Pattern:* Searchers fixate on all, or most, content words in the body of the page. This pattern demonstrates traditional reading, not scanning, and is the most effective for reading comprehension. [47]

We randomly selected 20% of all 167 SERPs and 20% of all 476 webpages visited for two raters to independently label using the above four category definitions. If a fixation pattern did not fit any of the four categories, then these were labelled as "NA". The raters reached an agreement of 0.78 Cohen's Kappa. The rest were classified by one of the coders.

4 RESULTS

In this section we report the findings of the study, beginning with general metrics of task success, and then discuss findings specific

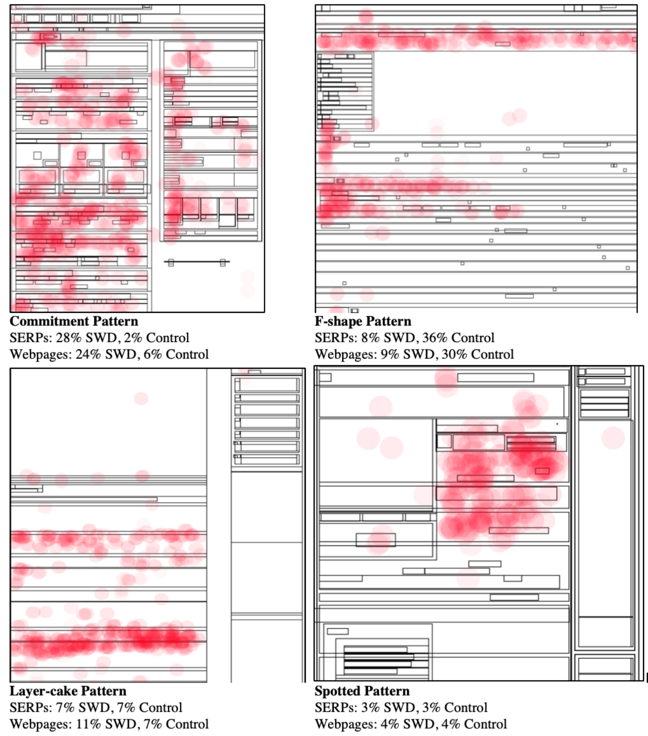

Figure 2: Classification of Fixation patterns on SERPs and webpages. AOIs outlined in black, each red dot is a fixation.

Search Task	SWD	Control	p	t_{26}
Phase 1 Remember	254,617 (229894)	198,875 (160033)	0.47	0.74
Phase 1 Understand	170,614 (165556)	107,444 (90555)	0.23	1.24
Phase 2 Remember	231,544 (120374)	220,911 (227397)	0.88	0.15
Phase 2 Understand	222,430 (110391)	196,066 (162980)	0.63	0.49
Phase 3 Remember	200,870 (72105)	144,639 (87571)	0.08	1.81
Phase 3 Understand	343,042 (140112)	283,348 (165509)	0.32	1.01

Table 2: Mean(σ) task completion time, in milliseconds.

to each study phase, sequentially focusing on query formulation, search results triage, and information extraction from webpages.

4.1 General Results

All 27 participants successfully completed each of the 6 search tasks within the time limit of 10 minutes per task. Prior research has shown that people with dyslexia are slower when reading and composing text [35, 38, 39], so we hypothesized that SWD would take longer to complete search tasks than the control group. Though the control group was slightly faster for all tasks, no differences were significant at the $\alpha = 0.05$ level (See Table 1).

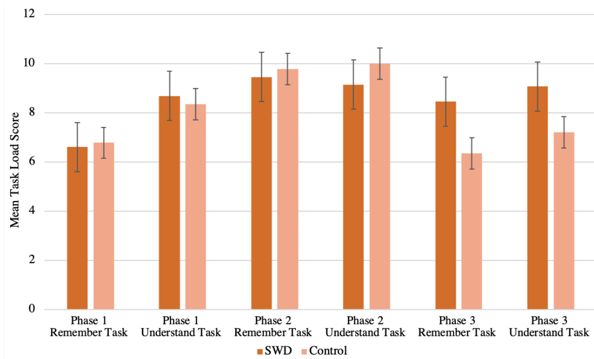

Figure 3: Average task load scores for all 6 tasks, spanning the 3 phases of the experiment. The minimum possible task load score is 5, while the maximum possible score is 25.

We also hypothesized that SWD would report exerting more cognitive effort when completing tasks, given the documented challenges that SWD experience when reading, writing, and remembering information [10, 38] together with their tendency to express lower levels of self-esteem and confidence when performing such tasks [6, 31, 43]. We were not able to accept this hypothesis for any task in the first two phases of the experiment, but found indications that SWD may have experienced a higher task load in the third phase, which required participants to read more webpages (Figure 3). Specifically, SWD reported higher task load for both tasks, but the difference was significant only for the *Remember* task ($p = 0.03$, $t = 2.27$). Given the median task load across all participants and tasks was 7 – only 2 points higher than the minimum score of 5 – it is possible that our scale was not sensitive enough at the lower end to detect population differences. Furthermore, in 2019, Choi, Capra and Arguello found that differences in working memory did not affect participants' post-task perceptions about workload during search tasks. Since dyslexia is strongly correlated with lower phonological working memory this result is consistent with their findings [18]. Also, when asked to report their prior domain knowledge for each task, most participants reported that they *"did not have any prior knowledge and had to search for all information"* ($\overline{x} = 2.62$, $\sigma = 0.41$). Therefore, we do not include it in our analysis of search behavior.

Finally, we note that in contrast to prior work [33, 56], we did not find *Remember* tasks to be systematically easier, or harder, than *Understand* tasks – there were no significant differences in task completion time, nor were there significant differences in reported task load. For the remainder of this paper, we therefore do not differentiate between task complexity types, and instead treat the pairs of tasks as independent trials within each phase. Next, we present phase-specific findings.

4.2 Phase 1 Results: Query Formulation

In Phase 1 of the study, participants completed a pair of search tasks by composing queries, reviewing SERPs, and optionally retrieving relevant documents. In this phase, we are best-positioned to compare differences in query formulation; other comparisons are obscured by differences in the results and webpages seen by searchers.

We expected SWD would formulate more queries than the control group, as mentioned in previous studies [11, 19, 37]. Our findings are consistent with this hypothesis: SWD formulated an average of 2.27 ($\sigma = 1.37$) queries per task, versus 1.54 ($\sigma = 1.07$) queries per task in the control population ($p = 0.03$, $t_{52} = 2.20$).

Likewise, we expected SWD to make more spelling errors when composing queries, as reported in [11, 37, 38]. With this measure, we are able to strongly confirm our hypothesis: SWD made an average of 0.67 errors per query ($\sigma = 0.69$), versus 0.04 errors per query ($\sigma = 0.13$) within the control group. This difference is highly significant ($p < 0.01$, $t_{52} = 3.26$). To further explore what type of errors these were, we manually labelled the errors into: *phonetic error* (an error at the phoneme level like "calqulation" instead of "calculation" or "dipest" instead of "deepest") and *typographic error* (error because of a typing mistake, identified if characters typed are close together on the keyboard like "tennois" instead of "tennis"). Since those with dyslexia have been shown to have trouble spelling because of challenges with phonological decoding [18, 38], we expected SWDs to make more phonetic errors than the control group. Our analysis confirmed that SWDs make more phonetic errors ($\overline{x}$=0.35, σ=0.49) than the control group ($\overline{x}$=0.07, σ=0.26, p=0.01).

However, as reported in Table 3, we did not find significant differences in average query length, time spent fixating on the query input box, or number of typographic errors. We suspect that autocomplete and other query suggestions may serve as a normalizing factor for query length, but further research is necessary to confirm this hypothesis. Overall, from these results we confirm that SWD struggle with query formulation - issuing more queries and making more errors per query, especially phonetically.

Measure	SWD	Control	p	t_{52}
# of queries	2.27 (1.37)	1.54 (1.07)	0.03	2.20
Errors per query (edit distance)	0.67 (0.69)	0.04 (0.13)	< 0.01	3.26
Phonetic Errors per query (edit distance)	0.35 (0.49)	0.07 (0.26)	0.01	2.82
Typographic Errors per query (edit distance)	0.79 (1.40)	0.07 (0.27)	0.07	1.90
Terms per query	6.12 (2.53)	6.04 (2.92)	0.91	0.11
Attention to Search Box (ms)	788 (1800)	744 (1811)	0.93	0.01

Table 3: Mean(σ) Query-formulation behavior by searchers with and without dyslexia during Phase 1. Significant differences at the $\alpha = 0.05$ level are highlighted in grey.

4.3 Phase 2 Results: Search Results Triage

In Phase 2 of the study, participants were presented with a preformulated query-SERP pair per task. They were required to complete tasks by inspecting results on the SERP and optionally opening linked webpages. Since all participants viewed the same SERPs, we can compare SERP triage behavior between groups.

We begin by reporting the fixation patterns observed when inspecting the SERPs, and then examine search behavior across specific AOI types. We found that SWD were most likely to employ the Commitment pattern, and exhibited this strategy for 63.89% of search result pages (vs. 0.05% in the control group). Conversely,

Measure	Phase 2			
	SWD	**Control**	***p***	t_{49}
# returns to the same SERP	4.33 (3.16)	2.35 (1.37)	0.00	-0.74
# of webpages opened	2.77 (1.33)	2.59 (1.49)	0.44	-0.78
# results viewed on SERP	7.67 (4.93)	3.20 (2.91)	0.04	-0.80
Attention to Instant Answers (ms)	13,340 (13979)	7,261 (6394)	0.04	-2.07
Attention to Ad results (ms)	5,187 (8161)	837 (1713)	0.01	-2.79
Attention to Organic results (ms)	20,120 (46631)	4,8059 (20724)	0.01	-2.88
Attention to Result titles (ms)	22,339 (19244)	9,507 (8767)	0.00	-3.18
Attention to Result snippets (ms)	24,898 (28977)	7,692 (7360)	0.00	-2.95
Attention to Related Searches (ms)	22,853 (20437)	12,499 (10020)	0.02	-2.38
Attention to Right Rail (ms)	2,506 (3747)	1,607 (2389)	0.32	-1.04
Attention to Images (ms)	4,032 (5094)	4,592 (9270)	0.80	0.26
Attention to Bold (ms)	3,966 (5139)	4,555 (9112)	0.78	0.27

Table 4: Mean (σ) gaze patterns of searchers with and without dyslexia on SERPs in Phase 2. Significant differences at $\alpha = 0.05$ level are highlighted in grey.

people in the control group were more likely to employ the F-shape pattern, exhibiting this strategy for 71% of search results (vs. 13.89% in the SWD group). Other scan patterns were rarely employed (Figure 2). The Fisher's exact test for differences in proportions finds these Commitment and F-shape differences to be highly statistically significant ($p < 0.01$ in both cases). In the context of web search, the Commitment pattern corresponds with a careful and systematic examination of the document, whereas the F-shape pattern indicates quick assessment of relevance (e.g., based on titles), followed by more focused attention on one or more promising snippets [45, 47]. This result helps explain prior findings reporting SWD struggle with search results triage [37, 43].

To further quantify the differences in search results triage patterns, we count the number of search results visually inspected by participants. We find that SWD inspected an average of 7.67 results ($\sigma = 4.93$), which is significantly more than the average of 3.20 results ($\sigma = 2.91$) inspected by people in the control group ($p = 0.04$, $t = 0.80$). Since without scrolling the SERP shows only 3-4 results as per our experimental setup, this result suggests that SWD scroll further down the SERP than the control group to less relevant results [1]. They might be scrolling further to find things that are more readable to them, as suggested by [25, 43]. On the other hand, this might be to find more information to add to what they've already found, as suggested by P9, an SWD: *"Even if the answer's given right there at the top, I will continue to look on the page to find more information to support it before I finish searching."* Concern with finding additional support for an answer may reflect a lack of confidence, as suggested by prior literature [6, 31, 43].

Additionally, we found that SWD returned to look at the same SERP more times, on average 4.33 ($\sigma = 2.35$) compared with 2.30 returns by the control group ($\sigma = 2.00$, $p = 0.00$). Since SWD do not open significantly more pages than the control group (See Table 3), this suggests that they return to examine the same page more closely or more of it. SWD might re-visit and re-read information because dyslexia is associated with having a shorter working memory [7, 18, 38]. Furthermore, SWD spent more time examining organic search results, instant answers, and advertised results. We also expected SWD to pay more visual attention to bold fonts and images than the control group, because prior work relates these to higher readability scores and SWD prefer more readable documents [25, 35, 41]; however, we could not accept this hypothesis because we did not find any significant differences at the $\alpha = 0.05$ level. Nonetheless, feedback from participants in the post-task questionnaire suggests that they rely on these features. For example, P4, an SWD, describes the importance of images: *"I clicked on this (link) because the image showed me it referred to the fencing I was looking for"*. P5, also an SWD, says about bold text: *"the answer is right there in bold, so I think it's right"*. Therefore, it is possible that both groups depend equally on these features, or that differences arise in saccadic patterns rather than fixations.

Each of these comparisons is detailed in Table 4. Overall, from these results we confirm that SWD struggle with triage of search results: they inspect more results, more carefully, and take longer in this phase than searchers in the control group.

4.4 Phase 3 Results: Information Extraction

In Phase 3 of the study, participants were presented with a pre-formulated query-SERP pair per task. They were required to complete tasks by inspecting results on the given SERP and opening at least 3 linked webpages. From the 3 webpages selected by each participant, there were only 2 webpages (1 in each task) that all 27 participants all viewed. Therefore, we chose to focus our analysis on those 2 webpages. Phase 3's main focus was to studying behaviors on retrieved webpages. While there was an overlap in the SERPs viewed by participants in this phase too, we cannot analyze and report on the SERP triage behavior for this phase because triage behavior could have been affected by our requirement to open at least 3 webpages when one or none might have sufficed.

To analyze information extraction behavior on webpages, we identified one page for each task that all participants visited. For the webpage seen by all in the Phase 3 *Remember* Task [1], SWD usually exhibit a Commitment pattern, with 87.5% of SWD exhibiting this strategy (vs 12.5% of the control group). Conversely, participants in the control group were more likely to employ the F-shape pattern, with 77.78% of the control group exhibiting this strategy (vs 23% of SWD). Other scan patterns were rarely employed. The Fisher's exact test for differences in proportions finds these Commitment and F-shape differences to be highly significant ($p < 0.01$). On the webpage in the Phase 3 *Understand* Task [2], SWD were, again, most likely to employ the Commitment pattern, with 74.89% of them

[1] http://www.tax-rates.org/north_carolina/orange_county_property_tax
[2] https://theexoticbean.com/blog/fair-trade-coffee/what-the-fair-trade-coffee-label-really-means/

exhibiting this strategy (vs 20% of the control group). Conversely, participants in the control group were more likely to employ the F-shape pattern, with 75% of the control group exhibiting this strategy (vs 25% of SWD). Again, other scan patterns were rarely employed. The Fisher's exact test for differences in proportions finds these to be highly significant ($p < 0.01$).

These results reiterate the fixation patterns seen in Phase 2 for inspecting SERPs. Since the Commitment pattern signifies reading each part of the page carefully rather than skimming, this result indicates that SWD may be employing less efficient page-scanning strategies during information extraction. Furthermore, these results confirm that SWD exhibit different gaze patterns than the control group, and this may indicate that SWD struggle with information extraction, as intimated by the NASA TLX responses reported earlier.

Results from the post-task questionnaire enrich our understanding of the challenges that SWD face when extracting information from webpages, as well as some of the strategies that SWD employed at this stage of the search process. Specifically, participants referred to the usefulness of images, menus, tables, and lists; they also reported a desire to encounter fewer ads, less italicized text, and more bold and large text. Referring to menus, P4, an SWD, said, "*I always go to the table of contents on Wikipedia to direct me*", and P23, an SWD, said, *"I couldn't find what I was looking for so I went to the drop down menu and selected 'sort by' to filter"*. P17, also an SWD, said about lists, *"I just want to see a list of the answers in front of me"*. About textual features, P18 from the control group said, *"the bold print led me to where I want to go."* They also said, *"well, it's got a box around it and it's bold so I think that helped."* P12, an SWD, said, *"I found what I was looking for because it was big and bold"*. These qualitative results are consistent with prior self-report results [25, 39, 43]. Conversely, though each of these properties was mentioned by participants, we did not find any significant quantitative differences in how SWD and the control group fixated on these features (Table 5). It is possible that both groups depend equally on these features, or that differences arise in saccadic patterns rather than fixations.

Measure	Phase 3 Remember Task[1]			
	SWD	**Control**	p	t_{24}
Attention to Headings (ms)	1,651 (1800)	5,043 (11368)	0.29	1.06
Attention to Images (ms)	348 (655)	1,783 (4769)	0.29	1.07
Attention to Italics (ms)	3,095 (2277)	3,751 (6866)	0.74	0.33
Attention to List (ms)	1,019 (1962)	1,878 (4302)	0.52	0.66
Attention to Paragraphs (ms)	47,507 (38938)	50,260 (42306)	0.86	0.17
Attention to Table (ms)	1,964 (4189)	1,000 (2076)	0.43	-0.74
	Phase 3 Understand Task[2]			
	SWD	**Control**	p	t_{24}
Attention to Headings (ms)	1,736 (1923)	1,769 (1768)	0.82	0.05
Attention to Images (ms)	9,857 (11141)	8,215 (8214)	0.68	-0.42
Attention to Italics (ms)	126 (308)	10 (35)	0.21	-1.29
Attention to List (ms)	825 (1425)	435 (295)	0.36	-0.93
Attention to Paragraphs (ms)	71,366 (41384)	56,597 (47659)	0.42	-0.83

Table 5: Mean (σ) fixation duration of searchers with and without dyslexia attending to page elements on a webpage encountered in Phase 3's Understand Task

5 DISCUSSION

This paper is the first to present findings from gaze data collected from English-speaking searchers with and without dyslexia using a modern interactive search engine during the complete query-triage-extraction process of search. The gaze patterns and search behavior demonstrate that searchers with dyslexia have more spelling errors (particularly phonetic errors) during query formulation, look further down the SERP, and use less-efficient page-scanning patterns than searchers in the control group. In this section we outline several design implications of our findings, discuss the study's limitations, and suggest opportunities for future work.

5.1 Design Implications

5.1.1 Query Formulation: During query formulation, SWD issued more queries and made more errors per query (particularly more phonetic errors as opposed to typographic errors) than the control group. To help with phonetic spelling errors, autocomplete could optionally be configured to more aggressively correct spelling errors and/or to incorporate dyslexia-specific updates to spellcheck algorithms, such as proposed by Rello et al. in the context of word processing [48]. Furthermore, to generally help with formulating their information need, autocomplete could suggest not only popularly searched queries, but also make sure that suggestions are semantically diverse to cover more possibilities. Also, since SWD issued more queries than the control group to find more relevant, readable, and trustworthy results, the browser could show searchers rich, multi-modal, instant answers as they type in the query. This could help searchers find relevant information quicker, make the result more glanceable, and help them evaluate and refine the query as they're typing. Furthermore, since instant answers are usually sourced from Wikipedia, a source deemed trustworthy by SWD [34], these answers will reduce the challenge to triage trustworthiness by SWD and non-SWD alike [40].

5.1.2 Search Results Triage: When triaging search results, SWD came back to look at the same SERP more times and inspected more results. They examined the SERP exhaustively using the Commitment fixation pattern, and scrolled far below the control group to less relevant results to possibly find more readable and trustworthy results. Search engines and their interfaces must make it easier to find more readable results and assess trustworthiness quickly. These findings bolster the need to redesign the search algorithm to re-rank results using readability and trustworthiness more heavily than it currently does [21, 51] and add to previous work [25, 35] that informs what factors (e.g., amount of images, headings, ads) lead to more readable SERPs and webpages. To make it easier to triage trustworthiness and readability of a result, we could crowdsource trustworthiness and readability ratings for each result and present it to searchers as a visualization (like in [50]) or an icon (like in [12]). Additionally, since we now know that SWD examine more of the SERP, in a more committed manner (focusing on organic and advertised search results, instant answers, etc.) we can spatially organize SERP elements accordingly to prioritize more readable, relevant, and trustworthy content. For example, regulatory bodies who wish to ensure equal access to content by people with dyslexia may need to pass regulations limiting layout options for ads, pushing more readable, relevant, and credible content to the top-left.

5.1.3 Information Extraction: When extracting information from webpages, SWD took longer and reported it to be harder than the control group. We also found that SWD examine webpages in a more committed manner than the control group. This could inform how we design webpages and spatially organize relevant and important information. For example, by providing summaries of each webpage at the top that can be used to help navigate the website and to expand only more relevant parts so that SWD don't have to committedly read the entire page. Furthermore, the browser interface and plugins could increase readability by automatically enlarging text as proposed by Bigham in [15], and to mitigate memory challenges faced by SWD preserve highlighted and zoomed sections. The browser could enable this information to be extracted easily by either taking a screenshot or allowing easy export of these sections into a notepad.

While we propose several design implications for search engines, browsers, SERPs, and webpages based on our findings, implementing these design ideas and verifying their effectiveness for searchers with dyslexia is left to future work. While our investigation focused on web search, many of these design suggestions may also enhance the usability and inclusivity of other information systems (e.g., e-books). We also need to do more research to test if these findings can help inform better search tools for those with other causes for reading differences such as English language learners and children who are learning to read.

5.2 Limitations and Future Work

This study has limitations as it tries to balance between ecological validity and the need to impose experimental control. Here, we discuss their potential impact, how we tried to address the limitations, and propose future work. First, we did not allow searchers to use any assistive technology, or change any of the default display settings of the browser. This controlled the visual interface all participants experienced, and allowed us to more easily compare behaviors across different participants. However, this altered the user experience for at least two people; one participant reported using a more aggressive spelling auto-correct technology on their personal computer and another reported difficulty because the study prevented them from zooming in.

Instructing participants to open at least 3 webpages from the SERP in phase 3 was another limitation of the study design. This potentially impacted how they triaged the search results and extracted information. We tried to reduce these effects by allowing them to choose which 3 links they wanted to open. This freedom in choosing the links resulted in having only two webpages common between all participants. If there had been more overlap in the visited webpages in phase 3, we may have had more statistical power in analysing differences in visual attention paid to elements on webpages.

We recruited only searchers fluent in English, above the age of 13, and within a major metropolitan area using ads targeted towards dyslexia and dyslexia-related organizations. Recruiting via social media biased us to a population with a certain level of technical literacy. Furthermore, self-selection bias may also impact our findings - maybe the people who responded to our ads had more severe dyslexia-related challenges, or had friends or family who had dyslexia and therefore followed dyslexia-related organizations. In future studies we could employ different recruitment and sampling methods to reduce these biases.

While we ensured that all SWD had a medical diagnosis of dyslexia, we recognize that dyslexia is a spectrum disorder, and there could be individual differences that do not generalize to other searchers with dyslexia. In future studies, when quick diagnostic tests of dyslexia (for example, [49]) become widely or clinically accepted or available in English language [8], we can control for individual differences and better model related behaviors and strategies to different points along the spectrum. Future research is required to overcome these limitations, build out and test suggested design implications of these findings.

6 CONCLUSION

By conducting an eye-tracking study observing 27 participants (14 with dyslexia, 13 control) use a widely-used modern English-language interactive search engine, we confirmed that searchers with dyslexia struggle with all stages of the search process: query formulation, search results triage, and information extraction. Furthermore, we established that searchers with and without dyslexia have noticeably different ways of visually attending to SERPs and webpages, including to individual page elements and to the page overall. Moreover, we learned that many of the challenges and strategies reported by participants in self-report studies manifest in the eye tracking data. We conclude by reflecting on our findings to propose design implications for improving utility and cognitive accessibility of search systems for people with dyslexia.

REFERENCES

[1] [n. d.]. https://web.archive.org/web/20130702063520/https://www.google.com/competition/howgooglesearchworks.html
[2] [n. d.]. Special education teachers. https://datausa.io/profile/soc/252050/#demographics
[3] Mikhail Ageev, Qi Guo, Dmitry Lagun, and Eugene Agichtein. 2011. Find it if you can: a game for modeling different types of web search success using interaction data. In *Proceedings of SIGIR 2011*. ACM.
[4] Areej Al-Wabil, Panayiotis Zaphiris, and Stephanie Wilson. 2007. Web navigation for individuals with dyslexia: an exploratory study. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 593–602.
[5] Areej Al-Wabil, Panayiotis Zaphiris, and Stephanie Wilson. 2008. Examining visual attention of dyslexics on web navigation structures with eye tracking. In *2008 International Conference on Innovations in Information Technology*. IEEE, 717–721.
[6] Neil Alexander-Passe. 2006. How dyslexic teenagers cope: an investigation of self-esteem, coping and depression. *Dyslexia* 12, 4 (2006), 256–275.
[7] International Dyslexia Association. [n. d.]. Dyslexia Basics. https://dyslexiaida.org/dyslexia-basics/.
[8] International Dyslexia Association. [n. d.]. Testing and Evaluation. https://dyslexiaida.org/testing-and-evaluation/.
[9] Gregory V Bard. 2007. Spelling-error tolerant, order-independent pass-phrases via the Damerau-Levenshtein string-edit distance metric. In *Proceedings of the*

fifth Australasian symposium on ACSW frontiers-Volume 68. Citeseer, 117–124.

[10] Harald Beneventi, Finn Egil Tønnessen, and Lars Ersland. 2009. Dyslexic children show short-term memory deficits in phonological storage and serial rehearsal: an fMRI study. *International Journal of Neuroscience* 119, 11 (2009), 2017–2043.

[11] Gerd Berget. 2016. Search and find? An accessibility study of dyslexia and information retrieval. (2016).

[12] Gerd Berget and Frode Eika Sandnes. 2015. The effect of dyslexia on searching visual and textual content: are icons really useful?. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 616–625.

[13] Gerd Berget and Frode Eika Sandnes. 2015. Searching databases without query-building aids: implications for dyslexic users. *Information Research: An International Electronic Journal* 20, 4 (2015), n4.

[14] Gerd Berget and Frode Eika Sandnes. 2016. Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of G oogle. *Journal of the Association for Information Science and Technology* 67, 10 (2016), 2320–2328.

[15] Jeffrey P Bigham. 2014. Making the web easier to see with opportunistic accessibility improvement. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 117–122.

[16] Andrei Broder. 2002. A taxonomy of web search. *SIGIR FORUM* 36, 2 (2002), 3–10.

[17] Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. 2009. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 21–30.

[18] Bogeum Choi, Robert Capra, and Jaime Arguello. 2019. The Effects of Working Memory during Search Tasks of Varying Complexity. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, 261–265.

[19] Lynne Cole, Andrew MacFarlane, and George Buchanan. 2016. Does dyslexia present barriers to information literacy in an online environment? A pilot study. *Library and Information Research* 40, 123 (2016), 24–46.

[20] Michael J Cole, Xiangmin Zhang, Chang Liu, Nicholas J Belkin, and Jacek Gwizdka. 2011. Knowledge effects on document selection in search results pages. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 1219–1220.

[21] Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian De La Chica, and David Sontag. 2011. Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 403–412.

[22] Edward Cutrell and Zhiwei Guan. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 407–416.

[23] Soussan Djamasbi, Marisa Siegel, and Tom Tullis. 2011. Visual hierarchy and viewing behavior: An eye tracking study. In *International Conference on Human-Computer Interaction*. Springer, 331–340.

[24] Susan T Dumais, Georg Buscher, and Edward Cutrell. 2010. Individual differences in gaze patterns for web search. In *Proceedings of the third symposium on Information interaction in context*. ACM, 185–194.

[25] Adam Fourney, Meredith Ringel Morris, Abdullah Ali, and Laura Vonessen. 2018. Assessing the readability of web search results for searchers with dyslexia. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 1069–1072.

[26] Eva Germanò, Antonella Gagliano, and Paolo Curatolo. 2010. Comorbidity of ADHD and dyslexia. *Developmental neuropsychology* 35, 5 (2010), 475–493.

[27] S Greenwood, A Perrin, and M Duggan. [n. d.]. Social Media Update 2016. Pew Research Center for Internet and Technology. November 11, 2016.

[28] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[29] Jesse L Hawke, Richard K Olson, Erik G Willcut, Sally J Wadsworth, and John C DeFries. 2009. Gender ratios for reading difficulties. *Dyslexia* 15, 3 (2009), 239–242.

[30] F. Hoeft, P. McCardle, and K. Pugh. 2015. The Myths and Truths of Dyslexia in Different Writing Systems. International Dyslexia Association blog.

[31] Neil Humphrey and Patricia M Mullins. 2002. Self-concept and self-esteem in developmental dyslexia. *Journal of Research in Special Educational Needs* 2, 2 (2002), no–no.

[32] Maryam Kamvar, Melanie Kellar, Rajan Patel, and Ya Xu. 2009. Computers and Iphones and Mobile Phones, Oh My!: A Logs-based Comparison of Search Users on Different Devices. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*. ACM, New York, NY, USA, 801–810. https://doi.org/10.1145/1526709.1526817

[33] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*. ACM, 101–110.

[34] Birgit Kvikne and Gerd Berget. 2018. When trustworthy information becomes inaccessible: The search behaviour of users with dyslexia in an online encyclopedia. (2018).

[35] Qisheng Li, Meredith Ringel Morris, Adam Fourney, Kevin Larson, and Katharina Reinecke. 2019. The Impact of Web Browser Reader Views on Reading Speed and User Experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 524.

[36] Lori Lorigo, Maya Haridasan, Hrönn Brynjarsdóttir, Ling Xia, Thorsten Joachims, Geri Gay, Laura Granka, Fabio Pellacini, and Bing Pan. 2008. Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology* 59, 7 (2008), 1041–1052.

[37] Andrew MacFarlane, Areej Al-Wabil, Chloe Ruth Marshall, A Albrair, Susan A Jones, and Panayiotis Zaphiris. 2010. The effect of dyslexia on information retrieval: A pilot study. *Journal of Documentation* 66, 3 (2010), 307–326.

[38] Andrew MacFarlane, A Albrair, CR Marshall, and George Buchanan. 2012. Phonological working memory impacts on information searching: an investigation of dyslexia. In *Proceedings of the 4th Information Interaction in Context Symposium*. ACM, 27–34.

[39] Andrew MacFarlane, George Buchanan, Areej Al-Wabil, Gennady Andrienko, and Natalia Andrienko. 2017. Visual analysis of dyslexia on search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 285–288.

[40] Matt McGee, Matt McGee, Matt McGee, and Matt McGee. 2010. Google Instant Search: The Complete User's Guide. https://searchengineland.com/google-instant-complete-users-guide-50136

[41] Aliaksei Miniukovich, Michele Scaltritti, Simone Sulpizio, and Antonella De Angeli. 2019. Guideline-Based Evaluation of Web Readability. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 508.

[42] Meredith Ringel Morris. 2013. Collaborative Search Revisited. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1181–1192. https://doi.org/10.1145/2441776.2441910

[43] Meredith Ringel Morris, Adam Fourney, Abdullah Ali, and Laura Vonessen. 2018. Understanding the needs of searchers with dyslexia. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 35.

[44] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An Interface for Collaborative Web Search. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1294211.1294215

[45] Jakob Nielsen and Jakob Nielsen. 2006. F-Shaped Pattern For Reading Web Content (original eyetracking research). https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/

[46] Kara Pernice and Kara Pernice. 2017. Exhaustive Review or "I Can't Believe It's Not There" Phenomenon. https://www.nngroup.com/articles/exhaustive-review-eyetracking/

[47] Kara Pernice and Kara Pernice. 2019. Text Scanning Patterns: Eyetracking Evidence. https://www.nngroup.com/articles/text-scanning-patterns-eyetracking/

[48] Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*. Springer, 203–219.

[49] Luz Rello, Miguel Ballesteros, Abdullah Ali, Miquel Serra, Daniela Alarcón Sánchez, and Jeffrey P Bigham. 2016. Dytective: Diagnosing risk of dyslexia with a game. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and ..., 89–96.

[50] Julia Schwarz and Meredith Ringel Morris. 2011. Augmenting web pages and search results to support credibility assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1245–1254.

[51] Milad Shokouhi, Ryen W White, Paul Bennett, and Filip Radlinski. 2013. Fighting search engine amnesia: Reranking repeated results. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 273–282.

[52] Sav Shrestha and Justin W Owens. 2008. Eye movement patterns on single and dual-column web pages. *Usability News* 10, 1 (2008), 0–6.

[53] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael A. S. Potts. 2007. Information Re-retrieval: Repeat Queries in Yahoo's Logs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07)*. ACM, New York, NY, USA, 151–158. https://doi.org/10.1145/1277741.1277770

[54] Ryen W. White and Dan Morris. 2007. Investigating the Querying and Browsing Behavior of Advanced Search Engine Users. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07)*. ACM, New York, NY, USA, 255–262. https://doi.org/10.1145/1277741.1277787

[55] Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan and Claypool Publishers.

[56] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. 2012. Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th Information Interaction in Context Symposium*. ACM, 254–257.