How Readability Cues Affect Children's Navigation of Search Engine Result Pages

Christine Pinney
christinepinney@boisestate.edu
Boise State University
Boise, Idaho, USA

Benjamin Bettencourt
benjaminbettencourt@boisestate.edu
Boise State University
Boise, Idaho, USA

Jerry Fails
Boise State University
Boise, Idaho, USA
jerryfails@boisestate.edu

Casey Kennington
caseykennington@boisestate.edu
Boise State University
Boise, Idaho, USA

Katherine Landau Wright
katherinewright@boisestate.edu
Boise State University
Boise, Idaho, USA

Maria Soledad Pera
M.S.Pera@TUDelft.nl
TU Delft
Delft, Netherlands

ABSTRACT

Children often interact with search engines within a classroom context to complete assignments or discover new information. To successfully identify relevant resources among those presented on a search engine results page (SERP), users must first be able to comprehend the text included in SERP snippets. While this task may be straightforward for an adult user, children may encounter obstacles in terms of readability and comprehension when attempting to navigate a SERP. Previous research has demonstrated the positive impact of including visual cues on a SERP as relevance signals to guide children toward appropriate resources. In this work, we explore the effect of supplying visual cues related to readability and text difficulty on children's (ages 6-12) navigation of a SERP. Using quantitative data collected from user-interface interactions and qualitative data gathered from participant interviews, we analyze the impact of these visual cues on children's selection of results on a SERP when carrying out information discovery tasks.

CCS CONCEPTS

• **Social and professional topics** → **Children**; • **Information systems** → **Search interfaces**; *Web searching and information discovery*.

KEYWORDS

Web search; visual cues; text complexity; children

ACM Reference Format:
Christine Pinney, Benjamin Bettencourt, Jerry Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. 2024. How Readability Cues Affect Children's Navigation of Search Engine Result Pages. In *Interaction Design and Children (IDC '24), June 17–20, 2024, Delft, Netherlands.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3628516.3655818

This work is licensed under a Creative Commons Attribution International 4.0 License.

IDC '24, June 17–20, 2024, Delft, Netherlands

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0442-0/24/06
https://doi.org/10.1145/3628516.3655818

1 INTRODUCTION

Children in elementary classrooms are often tasked with assignments that require the use of search engines (**SE**) to retrieve information from online resources [5, 31]. For this, they prefer mainstream SE like Google or Bing over SE explicitly tailored to serve children [5, 11, 14, 18]. Mainstream SE, however, have been designed and optimized for adults and, consequently, may not address certain aspects of a Search Engine Results Page (**SERP**) that affect how children search for, retrieve and use information [4, 9–11, 22], including those related to text comprehension and readability (the ease with which a reader can understand the text's content, i.e., vocabulary and syntax).

Locating resources online requires that users scan and understand the text presented on a SERP, a process that includes reading and comprehending the snippets provided for each search result. A *SERP snippet* is a summary or explanatory text about (and extracted from) a result's contents [11]. While results on a SERP are ranked by their relevance to a given query, relevance algorithms employed by popular SE do not necessarily take into account the text difficulty or readability of these snippets when determining the most relevant results and subsequent rank order. This often causes the top-ranking positions on a SERP to be occupied by results whose snippets present text that is too complex for the child searcher to read and understand, regardless of the purpose of the online inquiry, i.e., for learning or leisure [4, 11]. In turn, this makes it more challenging for younger searchers — who are known to explore SERPs sequentially top-to-bottom [24] — to complete their information-seeking tasks because they often encounter incomprehensible or unreadable resources on the SERPs generated in response to their queries [9, 11, 13].

To assist children in finding information online, some researchers have proposed incorporating visual cues on a standard SERP to visually signal relevance; these cues help guide children to resources online that are most relevant to them [2, 24]. While previous studies [2, 24] underscore the value of visual cues in helping children better understand relevance and navigate ranked results on a SERP, there is a notable gap in research investigating the effect of visual cues on other factors that affect children's online search experiences. We posit that, when children are the primary stakeholders, text readability is a cornerstone factor in the context of web search. Thus, guiding SERP exploration toward resources children can read

and understand is a must — particularly within an educational context.

Inspired by the use of visual cues to help children identify relevant SERP results [24], we explore the effect of visual cues related to comprehension and readability on children's exploration of SERPs. Using the *four-pillar framework for designing and evaluating information retrieval systems for children* presented in [23], we define the pillars for this study as: children aged 6 to 12 make up the *user group*, web search for the classroom context serves as the *environment*, information discovery is the *task*, and visual cues signaling comprehension and readability are employed as the *strategy*.

We then pose the following research question (**RQ**): *On a SERP, how do visual cues related to comprehension and readability affect children's choices of which result(s) to pick?* To address this question, we conducted a user study (n=18) in which we asked child participants to conduct online inquiry assignments related to the primary school curriculum using different interfaces: a standard SERP and two SERPs extended with visual cues related to readability and comprehension based on (1) five hard words (algorithmically determined and aligned with educational development stages, §3.1.1), and (2) reading grade level (formulaically calculated, §3.1.2).

Outcomes from this study show a clear preference by the participants for the "help" afforded by readability visual cues over the absence of help when no visual cues are present. Both quantitative and qualitative data collected suggest changes in children's SERP navigation habits during information-discovery tasks for the classroom context, as well as active engagement by children with the SERPs' visual cues and the information provided by them. This indicates that readability visual cues are helpful signals of text readability and have the potential to impact the SERP navigation habits of young searchers still learning to read and develop web search skills.

With this initial exploration, we focus our efforts on assessing whether the addition of visual cues related to comprehension and readability affects children's decisions of which SERP result(s) to examine. We do not directly gauge if children's web search experiences are positively or negatively affected by these visual cues in this iteration of our study. However, we anticipate that insights gained emerging from this research can inform future efforts aimed at enhancing children's web search experiences.

2 RELATED WORK

This work is rooted in two areas of research with relevant prior work, which we describe below.

Children's Information Retrieval and Search Engine Use. The design of effective strategies for improving children's online web search experiences has become an important area of research in an increasingly technological world. Besides exploring and using content filters to limit inappropriate content, studies have investigated several aspects that affect children's experiences when using SE, including children's query formulation and search behaviors, SE interface design, and relevance ranking of SERP results [3, 5, 13, 17, 27, 31, 37].

Algorithmic relevance is employed by SE to rank retrieved results in order of relevance to a given query [9]. While much of the existing research regarding the optimization and effects of relevance ranking for items on SERPs is focused on adult users, there have been more recent efforts towards child-centric relevance ranking [3, 5, 13, 27, 31]. Gwizdka et al. [17] investigated children's query analysis, click behavior, and thought processes when using a SE, finding that both younger and older children have misconceptions about the ranking of results on a SERP. Children often exhibit linear behavior when exploring a SERP or gravitate towards top-ranked results regardless of a result's relevance to their query or information needs [9, 37]. This work illustrates the need to provide better support for children in their development of search literacy and informed SERP navigation habits, where search literacy can be defined as the "ability to identify, locate and effectively use information" presented on SERPs [21]. Collins-Thompson et al. [13] explored the effect of incorporating both the estimated reading level difficulty of result snippets as well as the estimated reading level ability of a given user into SERP ranking algorithms; the authors assert that reading level features serve as valuable signals for relevance, suggesting that readability information could contribute to the development of improved SERP navigation habits.

Although approaches to alter rankings to prioritize education-friendly, child-suitable and readable results exist [3, 27, 31], there has been less focus on using this information to provide children with "hints" that could help guide their navigation towards results they can comprehend and that are relevant to their unique information needs. By providing visual cues that inform children of potential text readability, their navigation habits may shift as they become more informed searchers.

Readability Assessment & Text Difficulty Estimation. If a child is unable to read and comprehend the information presented on SERP snippets, then these results become useless. Consequently, comprehension and readability of SERP snippets become key factors influencing children's ability to effectively navigating SERPs. As demonstrated by Allen et al. [1], determining which strategy to use to estimate the level of difficulty of a text in the web search context is not an easy feat—different strategies result in different outcomes, which could affect the overall search experience. These strategies range from traditional formulas like Flesch-Kincaid [29] and Fry [16] that analyze shallow linguistic features like syllable counts, sentence length, and word complexity [26] to more complex methods involving language models and handcrafted linguistic features that take into account more complex lexical and semantic features of language [6, 8].

Regardless of the strategy employed, readability assessment methods are not without flaws–they are significantly affected by text length (performing worse on shorter texts) and often produce estimations that are several grade levels off of the intended grade level of the text [8, 20, 28]. Even more, the readability estimations are often used only behind the scenes within mechanisms that are unknown or not visible to a user, so users cannot directly benefit from these calculations. By providing visual cues that estimate the readability of the text on a SERP, these calculations may help inform children's decisions of which results best meet their information needs.

3 METHOD

For young SE users – more particularly, those still learning to read – navigating a SERP requires different comprehension strategies than those of their adult counterparts. Children's vocabularies (both in terms of knowing a word and being able to read/spell it) are not as developed or robust as most adults', so comprehension of online materials becomes a task of finding resources that are, first and foremost, readable. As is demonstrated in previous research [3, 13, 27, 31], leveraging in the retrieval and ranking process the estimations of the reading level of the indexed resources, and even the reading level of the searcher, can help provide children with more suitable materials. Still, while readability information may impact the ranking position of retrieved resources, this information is not revealed to the searcher.

In the rest of this section, we offer details of the user study we conducted to probe the degree to which, if any, visual clues related to readability and comprehension could impact the search habits of young searchers.

3.1 Cues

To help inform children's SERP result selection process, we proposed augmenting SERPs by including in each SERP snippet a visual cue related to readability. Specifically, we consider two types of readability cues: (1) the display of five challenging words, determined according to the identification of specific word patterns characteristic of more advanced reading levels [7], and (2) a simple, rounded readability score, determined by the Spache-Allen formula [1].

3.1.1 Challenging Words. Within the classroom, teachers aim to provide reading materials containing vocabularies with which students are approximately 95% familiar, resulting in around 5% of the words being unfamiliar or advanced [25]. In this way, students have the tools to understand and comprehend the majority of a piece of text but are still challenged to learn more difficult words with more complex linguistic patterns using contextual clues. To determine what those more challenging words may be and how to identify them, we turn to research in the field of education.

The Simple View of Reading posits that a child's reading ability is the product of their language comprehension and decoding skills [19], where decoding refers to a child's ability to recognize a word as a meaningful combination of its linguistic components [36]. In other words, the process of learning to read starts with the basic connection of spoken language to letter symbols on a page and progresses from simple word recognition (i.e., decoding) towards syntactic and semantic understanding of phrases and sentences to derive meaning from text (i.e., comprehension) [19]. Educational researchers have outlined the order in which readers typically learn the different word and spelling patterns, moving from straightforward letter-sound correspondences (e.g., adding the letter "e" to the end of "car" changes the vowel sound), to more complex patterns that contain information about both sound and meaning (e.g., the presence of "tion" at the end of a word indicates that it is a noun). For our purposes, we focus on the specific patterns identified by Bear et al. [7] in *Words Their Way*, where four stages of orthographic development are defined: Letter-Name Alphabetic, Within-Word Pattern, Syllables & Affixes, and Derivational Relations.

Words Their Way characterizes each development stage based upon the orthographic patterns learners in this stage have mastered, and which they still need to learn [7]. We describe these patterns below as they provide the foundation for the visual cues used to assist children in their comprehension of information presented on a SERP.

Letter-Name Alphabetic. In this primary stage, learners are just beginning to recognize the connection of speech sounds to written alphabetic symbols. Learners develop articulation skills by isolating individual and combination letter sounds, namely affricates (e.g., ch, dr), stop consonants (*e.g., b, d, g*), voiced/unvoiced pairs (e.g., *b* and *p*, *f* and *v*), and continuants (e.g., *l, m, r, s, v*). Further patterns include consonant digraphs (e.g., *sh* in "dish") and blends (e.g. *sp* in "gasp"), preconsonantal nasals (e.g., *mp* in "dump"), and short vowels following the pattern of consonant-vowel-consonant in monosyllabic (single-syllable) words (e.g., the *a* sound in "bat").

Within-Word Pattern. In this second stage, learners move to the more complex recognition of units of speech and patterns within words that distinguish one from the other: phonemes. Long vowel sounds that follow the form of consonant-vowel-vowel-consonant (e.g., the *ai* sound in "rain") and consonant-vowel-consonant ending in *e* (e.g., the *a* sound in "gate") are also explored within this stage. Learners additionally focus on increasingly complex consonant clusters (e.g., *dge* in "edge"), basic plural nouns, and irregular past tense verbs.

Syllables & Affixes. In this third stage, learners start to explore multisyllabic words, syllable junctures, and stress and accent patterns. Morphemes– meaningful units of language that cannot be further divided– become an important area of focus in this stage, with specific emphasis given to inflectional endings (e.g., *-ing* in "running" and *-ed* in "followed"), suffixes and prefixes (e.g., *-ness* in "darkness" and *-ify* in "classify"), and compound words.

Derivational Relations. In this final stage, special attention is given to more complicated affixes like advanced suffixes (e.g., *-tion* in "creation" and *-ible* in "audible") and assimilated/absorbed prefixes (e.g., *il-* in "illogical" and *op-* in "opposite"). These more complex linguistic patterns often signify differences in semantic meaning and can significantly affect the process of comprehension.

To identify the more challenging words within a snippet that are characteristic of the latter stages of orthographic development, we apply algorithmic functions designed and presented by Pinney et al. [30] to process each snippet and identify specific word patterns characteristic of the later development stages (i.e., Syllables & Affixes and Derivational Relations). These patterns include advanced suffixes, assimilated prefixes, compound words, and advanced inflectional endings. The presence of these linguistic patterns within a SERP snippet may indicate that the linked web page text is more sophisticated and requires a reading ability characteristic of children in these later stages of orthographic development, having implications for readability and comprehension; as such, visibly emphasizing the more challenging words within a snippet could aid children in deciding which results to click and which to avoid, resulting in changes to children's SERP navigation habits.

To identify the presence of the advanced linguistic patterns in a given snippet and then determine corresponding "hard words" to be displayed, we rely on several functions. These functions utilize syllable counts, regular expressions, parts of speech, and IPA

(International Phonetic Alphabet) translations. IPA provides a standardized representation of written text to represent oral language and speech sounds.[1] For example, to identify the presence of advanced suffixes (e.g., adopt*able*, deposit*ory*, crea*tion*), the function first determines whether or not a given word contains two or more syllables, as words containing advanced suffixes are rarely monosyllabic. The next step limits the application of this function to words with noun and adjective parts of speech; advanced suffixes are not characteristic of other parts of speech. If a word is still a candidate for this linguistic pattern after these steps, it is scanned for specific advanced suffixes using regular expressions of IPA translations. The application of IPA translations over regular alphabetic text allows the function to account for advanced suffixes that vary in spelling. For instance, the words *musician*, *evaluation*, and *commission* employ the same advanced suffix "shun" but do so with different spelling; IPA translation represents all three spellings with a singular notation: [ʃən].

These steps limit the need for text processing to only words that pre-qualify, resulting in more efficient calculations and quicker display of content on the interface. Functions are applied sequentially in descending order of development stages; patterns characteristic of the Derivational Relations stage are identified first, and the patterns characteristic of the Syllables & Affixes stage are identified second. Once the amount of "hard words" reaches a maximum of five words or all patterns have been applied, the list of words is displayed within the corresponding result box from which the snippet was extracted (Figure 1b). The number of words calculated and displayed varies across queries and retrieved results (averaging at about 1.396 words), but we limit the maximum number to be displayed to five so as not to overwhelmingly alter the appearance of the SERP.

3.1.2 Spache-Allen Readability Score. The Spache-Allen readability score, first introduced in [1], expands the original Spache formula for readability estimation [34, 35]. This updated Spache-like formula extends the original vocabulary list which consists of terms deemed "simple" (i.e., familiar to everyday writing) by explicitly accounting for more contemporary terms with which children are familiar.

To determine the readability score (i.e., grade level) of a SERP snippet S, we use the Spache-Allen formula in Equation 1. Our choice of this formula over other popular ones (e.g., Flesch-Kincaid [15]) is based on the fact that it is more effective when determining the readability of web texts targeting younger audiences, and its performance is comparable to other popular counterparts when estimating the complexity level of snippets (i.e., short texts) [1].

$$Spache\text{-}Allen(S) = (0.141 \times w_S/s_S) + (0.086 * dif(S)) + 0.839 \quad (1)$$

where w_S and s_S are the number of words and sentences in S, respectively. $dif(S)$ captures the percentage of difficult words in S, where a word is deemed difficult if it does not appear in the extended "simple" term list, which consists of 65,669 unique terms that children learn through instruction and/or are exposed to online, in addition to the original Spache's term list.

[1]International Phonetic Association website: https://www.internationalphoneticassociation.org

For each snippet, the predicted Spache-Allen sore is rounded to signify a discrete grade level and displayed within the result box of the corresponding SERP snippet to visualize the reading grade level of the text (Figure 1c).

3.2 Search Topics

To simulate an information-discovery process that can take place within a classroom context, searchers in our study are asked to use a tool that mimics a familiar SE to complete online quests that cover three educational topics: (1) maps, (2) ecosystems, and (3) banks. These topics were chosen as they align with the Common Core curriculum for children in primary school (2nd, 3rd, and 4th grade, respectively) and are thereby familiar to the participants.[2] The order of presentation of these topics is randomized as the educational question is not what is being investigated, but rather the process through which we analyze the effect of readability visual cues on children's SERP navigation habits.

3.3 Interface Design

The user interface for our user study is derived from a standard SERP. In particular, we consider three interface versions that are the experimental condition to see if readability cues influence how children navigate a SERP. The three versions provide different, additional information for each search result related to comprehension and readability: (1) **Standard SERP**, where no additional information (see Figure 1a), (2) **Word SERP**, where up to five challenging words found within a result snippet (as specified in Section 3.1.1, Figure 1b), or (3) **Grade SERP**, where a Spache-Allen readability score (as stipulated in Section 3.1.2, Figure 1c).

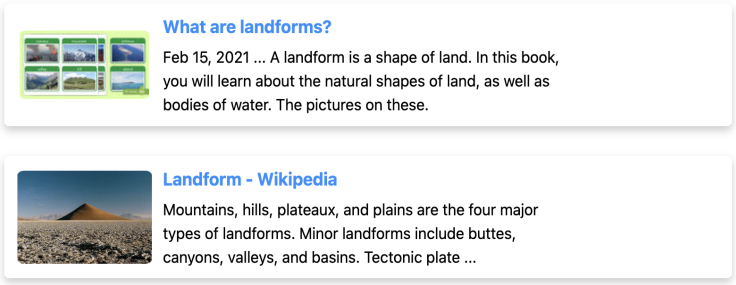
The interface version displaying the five challenging words does so under the label "Hard Words" rather than "Challenging Words" as the word 'challenging' can be a difficult word for some in the age range – particularly the younger children – and "hard" is more easily recognizable across the full range of ages for this study (6-12). The interface version displaying the Spache-Allen readability score does so under the label "Grade Level" as children are familiar with the connection between grade level and readability.

Both the hard words and Spache-Allen score calculations are carried out on-the-fly using the SERP snippets that result from children's unique queries; the information is then displayed within each SERP result in a similar manner (same font and colors to not draw attention away from the result information itself and thus avoid introducing unfair bias between the experimental conditions). To aid in the algorithmic identification of challenging word patterns (§3.1.1), we utilize NLTK's syllable tokenizer and POS tagger [12].

3.4 Procedure

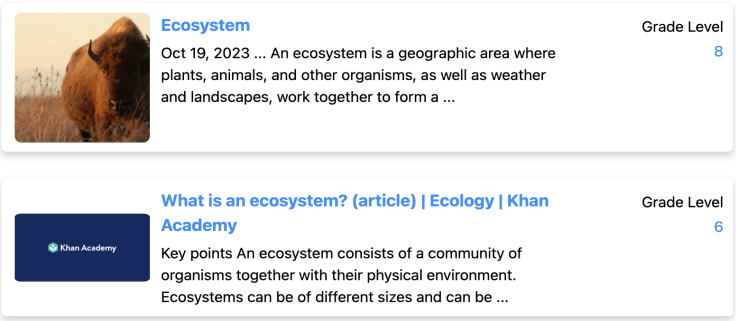
Study participants were children between the ages of 6 and 12 years old. We selected this age group because children within this range are actively learning to read and develop their vocabularies and decoding skills and, thus, may have the most to gain from readability visual cues. Thirty minutes were allotted for each participant.

[2]The Common Core curriculum is an educational initiative that outlines what children in grades K-12 in the US should know by the end of each grade in terms of language arts and mathematics. Common Core State Standards Initiative website: https://www.loc.gov/item/lcwaN0010852/



(a) Interface Version: Standard SERP.

(b) Interface Version: Word SERP.



(c) Interface Version: Grade SERP.

Figure 1: Sample interfaces participants engage with during the information-seeking process.

Each participant carries out three search "quests," containing three tasks. The search quests were administered via Zoom with links provided for each interface version. Each search quest pairs one interface version with one of the aforementioned educational topics presented in randomized order, allowing participant exposure to all three interface versions. Topics were explored sequentially, with background information presented to help children frame their queries before engaging with the interface. For example, when introducing the topic of maps, the following script was read aloud: "Now we're going to learn about how to read a map. To start, let's talk about physical features on a map. Maps can show us where landforms are located." Participants were then given the prompt: "Use the search bar to find out what landforms are and some types of landforms." Participants were allowed several minutes to shout out any answers they found appropriate, or state that they had not found any information before moving on to the next prompt. For each topic, two prompts were verbal instructions (as with the previous example) and one was a simple question (i.e., "What map symbol is used to show capital cities?").

Upon completion of the search quests, participants were given an exit interview. Employing the This-or-That evaluation method for children's experiences [33], each participant was asked to pick a most favorite and a least favorite version following a brief review of the different interfaces using screenshots (i.e., "Out of all three versions, which would you pick as your favorite/least-favorite?"). Using the Fun Toolkit survey method, participants were asked where they would place each version on the Smileyometer scale [32]. For each decision, participants were asked to supply verbal justification for their choice, providing further insight into the reasons for their selection.

3.5 Experiments

To measure the effect of readability visual cues on children's SERP navigation, we operationalize search behavior using logged user-interface interaction events and snippet readability calculations. This quantitative data allows us to explore similarities and differences in children's navigation of SERPs across all three interface versions.

(1) *Interface Events:* Average SERP clicks, average SERP position of first result clicked, average result hovers (when a user hovers the mouse over a result) per query, average SERP position of first result hovered, and average time hovering any result. Both click and hovering events allow us to probe some known navigation habits of the user group under study, i.e., preference for top results [9, 37].
(2) *Readability-Related Information:* Average number of hard words of the first result clicked and first result hovered, and average Spache-Allen readability score of the first result clicked and first result hovered. This allows for the analysis of children's engagement with and employment of the information provided by the visual cues.

From the collected exit interviews we analyze children's overall search experiences with the visual cues. This data provides insight into children's engagement with the readability visual cues, as well as their general preferences and interpretation of the information presented with each visual cue.

Paired t-tests were conducted where appropriate to analyze the statistical significance of the reported results.

4 RESULTS & DISCUSSION

Following the protocol discussed in Section 3, we collected data from 18 children (all native English speakers, residing within the US); 10 participants were male; 8 female. The age groups 6, 7, and 12 each had two participants, age groups 8 and 9 had four participants, age group 10 had three participants, and age group 11 had one participant.

From the collected data, we computed the signals described in Section 3.5; this information is summarized in Table 1. The Word SERP and Grade SERP had the highest average numbers of clicks, and the Grade SERP had the greatest depth of the position of the first result clicked. The average readability score of the first result clicked and the first result hovered was lower with both the Standard and Grade SERP when compared to the Word SERP, with the Grade SERP having the lowest readability score of the first hovered result. The Word SERP and Grade SERP had lower average amounts of words displayed on the first result clicked (1.4 and 0.7) when compared with the Standard SERP (3.7). The position of the result where a participant first hovered had the greatest depth with the Word SERP, with both the Standard and Grade SERP having lesser position depths regarding this measure. Hover duration was most significant with the Grade SERP, with both the Standard SERP and Word SERP obtaining more equal and lesser hover duration.

Looking at the Word SERP, the average number of clicks (0.34) was higher than the Standard SERP (0.28). The average number of "hard words" calculated for first clicked and first hovered results was lower than the Standard SERP, i.e., 1.4 (clicked)/1.7 (hovered) compared with 3.7/2.1 on the Standard SERP. These results suggest that when participants had access to the "hard words" for each result, they opted for results that had fewer hard words displayed, implying that participants may have interpreted the presence of more hard words as an indication of greater text complexity. Furthermore, the average position of the first hovered result with the Word SERP was higher (lower on the page) than both other SERP versions, i.e., 1.052 compared with 1.037 (Standard SERP) and 1.029 (Grade SERP). This suggests that participants navigated further down the page in a nonlinear fashion before focusing on a specific result, indicating more active engagement in the SERP navigation process.

Looking at the Grade SERP, the average number of clicks (0.38) was higher than both the Standard SERP (0.28) and Word SERP (0.34), and the depth of the first clicked result's position was greatest, i.e., 1.57 compared with 1.38 (Standard SERP) and 1.29 (Word SERP). The average grade level of the first clicked result was equal to the Standard SERP but lower than the Word SERP, i.e., 6.33 compared with 6.33 (Standard SERP) and 8.0 (Word SERP). The average grade level of the first hovered result was lower than with both other SERP versions, i.e., 6.1 compared with 6.6 (Standard SERP) and 6.8 (Word SERP). These findings indicate that participants were willing to go past the first SERP result towards results with perhaps more appropriate grade levels before making any decisions on which result to click, suggesting a higher level of engagement with this SERP version in terms of exploring the result page for readable resources. We posit that, when considering the target reading levels of SERP snippets and click interactions, participants seem to be best served by the Grade SERP version.

Results from qualitative data indicate that most participants prefer the presence of readability visual cues (Word and Grade SERP) over the absence of readability visual cues (Standard SERP). The data from the exit interviews reveals that 14 (78%) participants identified a SERP version that supplied additional information (i.e., Word SERP or Grade SERP) as their favorite (two participants declined to identify a favorite). When asked why they made this decision, 8 participants implied that the visual cue gave them some sort of insight in terms of readability. Regarding the Word SERP, one participant (age 12) mentioned that having the hard words available "helped [her] decide if [she] would understand it or not." Another participant (age 9) mentioned that she enjoys reading challenges, so the hard words helped her decide how challenging the text would be. When referring to the Grade SERP, one participant (age 7) said that it "lets you know what grade you need to be in to read it." Another participant (age 10) pointed out that it was "helpful because if [he] was in fifth grade, [he] probably wouldn't click on [a result] with [grade level] 10 because it would be too hard." These findings suggest that children leverage the presence of readability visual cues and explore SERPs in their quest for results that are potentially more readable.

Of the 13 (72%) participants that chose a least favorite SERP version (5 participants did not choose a least favorite SERP version), 9 participants chose the Standard SERP. One participant (age 12) mentioned that this SERP version "didn't really tell you anything" about the text, and two other participants (ages 9 and 10) made reference to the absence of any "help" with this SERP version. Two participants chose the Word SERP as their least favorite, but the justifications for their decisions imply that this SERP version still

	Standard SERP	Word SERP	Grade SERP
# of SERP clicks	0.28	0.34	0.38
First click position	1.38	1.29	1.57
First hover position	1.037	1.052	1.029
Hover duration (in seconds)	8.28	9.13	13.71
First click readability score	6.33	8.0	6.33
First click # of hard words	3.7	1.4	0.7
First hover readability score	6.6	6.8	6.1
First hover # of hard words	2.1	1.7	1.0
Average Smileyometer points	3.04	3.58	3.91

Table 1: Overview of results inferred from logged searcher-system interactions and exit interviews, grouped by interface type.

provided insight in terms of readability: one participant (age 8) explained that he chose this SERP version as his least favorite because "hard words are hard!", and the other (age 7) stated that "hard words means that [the text] is hard." Even though these two participants indicated that this SERP version was their least favorite, it is still apparent that Word SERP provided information concerning readability, affecting their decisions on what resource to further explore.

Comparing the values from the Smileyometer for each SERP version (Awful, Not Very Good, Good, Really Good, Brilliant), the Grade SERP received the most "Brilliant" labels with a total of five. The Word SERP obtained the most "Really Good" labels totaling at six. The only "Awful" label was applied to the Standard SERP. Converting the Smileyometer to a 5-point scale (where 1 corresponds with "Awful" and 5 corresponds with "Brilliant") and averaging the scores for each SERP version, the Standard SERP received 2.91 points, the Word SERP received 3.38 points, and the Grade SERP received 3.76 points, making it evident that children preferred having the readability visual cues. We used a paired t-test analysis to compare the various conditions. The pairing of Standard and Grade SERP had a p-value of 0.02 indicating a statistically significant difference. The other pairs were above the 0.05 threshold (0.13 for the Standard SERP and Word SERP pair, and 0.23 for the Word and Grade SERP pair).

When observing participants work through search quests, there was a noticeable difference in digital literacy across the age range of 6 to 12 years old. In the exit interviews, children on the younger end of this range (i.e., 6-8 years old) often pointed out that they did not readily notice the different visual cues across SERP versions, while older children (i.e., 9-12 years old) had much more to say about the effect of the different cues. Younger children spent more time forming and typing queries than analyzing the results on a SERP. Across the age range, most children's first strategy for finding information was to scan the snippets of each result; if the answer was available in the snippet, they often did not end up clicking on any results. This observation further motivates the need for age-appropriate, readable snippets on a SERP.

From the reported results we infer that children generally exhibited changes in SERP navigation habits when using the Word and Grade SERP compared with the Standard SERP. We surmise that both the Word and Grade SERP provided visual information that was interpreted by participants as an indicator of readability and thus resulted in children steering their navigation towards potentially more readable resources, whether that be resources with fewer "hard words" displayed or resources with lower grade levels. This suggests that children find visual cues related to text readability useful when navigating SERP information within the classroom context, indicating that readability signals could provide children with the tools to more effectively locate resources online with information that they can read, comprehend, and utilize. By visualizing readability information rather than strictly incorporating it into relevance and ranking calculations, children become more involved in the process of navigating towards resources that may be able to meet their unique needs in terms of readability.

5 CONCLUDING REMARKS

With this work, we have advanced knowledge of the feasibility of using visual clues to better support children turning to SE to conduct online inquiries related to the classroom. We explored the use of readability-based clues that can serve as hints that young searchers can rely on to identify SERP snippets that better align with their reading and comprehension skills.

Findings emerging from the analysis of quantitative and qualitative data collected as a result of a user study involving 18 children indicate that children's SERP navigation habits can change when readability visual cues are provided. Initial results suggest that children navigate towards more readable resources when supplied with readability information via these visual cues; children demonstrate a preference for interfaces that provide clues regarding readability. A limitation of this work is that it does not measure the effect of readability visual cues on children's comprehension of the clicked results' content nor the extent to which such cues guide children towards reliable (or unreliable) resources and it only takes into account the readability of the snippet.

In the future, we plan to take into account the readability level of the resources included in SERPs rather than the result snippets, as searchers will have to go through their content to identify the information needed to complete their search. Also of interest is verifying if the overall search experience can meaningfully impact the accurate completion of inquiry assignments related to the classroom context, i.e., whether readability-related clues prompt searchers to resources they can comprehend and in turn ease the task of distilling information from resources, which is known to be a problem among young searchers [22]. The four-pillar framework

utilized within this work allows for future work focused on extending the user group to other individuals for whom readability may be an important factor in navigating a SERP (e.g., second language learners, aspiring readers, etc.).

6 SELECTION AND PARTICIPATION OF CHILDREN

Participants were recruited via social media. The Institutional Review Board (IRB) approved recruiting practices and experimental design. The study was conducted via Zoom with a parent/guardian of the child present. The purpose of the study was explained to participants and their parents. Parents signed consent forms to allow their children to participate, and children assented to participating.

ACKNOWLEDGMENTS

We thank the children who participated in this study. This work was partially funded by NSF Award #1763649.

REFERENCES

[1] Garrett Allen, Ashlee Milton, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. 2022. Supercalifragilisticexpialidocious: Why Using the "Right" Readability Formula in Children's Web Search Matters. In *European Conference on Information Retrieval*. Springer, 3–18.

[2] Garrett Allen, Benjamin L Peterson, Dhanush Kumar Ratakonda, Mostofa Najmus Sakib, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. 2021. Engage!: co-designing search engine result pages to foster interactions. In *Interaction Design and Children*. 583–587.

[3] Garrett Allen, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. 2023. A Multi-Perspective Learning to Rank Approach to Support Children's Information Seeking in the Classroom. *arXiv preprint arXiv:2308.15265* (2023).

[4] Oghenemaro Anuyah, Ashlee Milton, Michael Green, and Maria Soledad Pera. 2020. An empirical analysis of search engines' response to web search queries associated with the classroom setting. *Aslib Journal of Information Management* 72, 1 (2020), 88–111.

[5] Ion Madrazo Azpiazu, Nevena Dragovic, Maria Soledad Pera, and Jerry Alan Fails. 2017. Online searching and learning: YUM and other search tools for children and teachers. *Information Retrieval Journal* 20 (2017), 524–545.

[6] Ion Madrazo Azpiazu and Maria Soledad Pera. 2019. Multiattentive recurrent neural network architecture for multilingual readability assessment. *Transactions of the Association for Computational Linguistics* 7 (2019), 421–436.

[7] Donald R. Bear, Marcia Invernizzi, Shane Templeton, and Francine Johnston. 2020. *Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction*. "Pearson Education, Inc.".

[8] Rebekah George Benjamin. 2012. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review* 24 (2012), 63–88.

[9] Dania Bilal. 2012. Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Y ahoo!, B ing, Y ahoo! K ids, and ask K ids. *Journal of the American Society for Information Science and Technology* 63, 9 (2012), 1879–1896.

[10] Dania Bilal and Meredith K Boehm. 2013. Towards new methodologies for assessing relevance of information retrieval from web search engines on children ' s queries. 3 (March) (2013), 93–100.

[11] Dania Bilal and Li-Min Huang. 2019. Readability and word complexity of SERPs snippets and web pages on children's search queries: Google vs Bing. *Aslib Journal of Information Management* 71, 2 (2019), 241–259.

[12] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[13] 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*. 403–412.

[14] Judith H Danovitch. 2019. Growing up with Google: How children's understanding and use of internet-based devices relates to cognitive development. *Hum. Behav. Emerg. Technol.* 1, 2 (April 2019), 81–90.

[15] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[16] Edward Fry. 1968. A Readability Formula That Saves Time. *Journal of Reading* 11, 7 (1968), 513–578.

[17] Jacek Gwizdka and Dania Bilal. 2017. Analysis of Children's Queries and Click Behavior on Ranked Results and Their Thought Processes in Google. In *the 2017 Conference*. unknown, 377–380.

[18] Jacek Gwizdka and Dania Bilal. 2017. Analysis of children's queries and click behavior on ranked results and their thought processes in google search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. 377–380.

[19] Wesley A Hoover and Philip B Gough. 1980. The simple view of reading. *Reading and Writing: An Interdisciplinary Journal* (1980).

[20] Tapas Kanungo and David Orr. 2009. Predicting the readability of short web summaries. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. 202–211.

[21] Ioannis Karatassis. 2017. WebSAIL: Computer-based Methods for Enhancing Web Search Literacy. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 403–405.

[22] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 2021. Right way, right time: Towards a better comprehension of young students' needs when looking for relevant search results. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 256–261.

[23] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 2022. Have a Clue! The Effect of Visual Cues on Children's Search Behavior in the Classroom. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 310–314.

[24] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 2022. Have a clue! the effect of visual cues on children's search behavior in the classroom. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 310–314.

[25] Lauren Leslie and JoAnne Schudt Caldwell. 2021. *Qualitative Reading Inventory*. "Pearson".

[26] Ion Madrazo Azpiazu and Maria Soledad Pera. 2020. Is cross-lingual readability assessment possible? *Journal of the Association for Information Science and Technology* 71, 6 (2020), 644–656.

[27] Ashlee Milton, Oghenemaro Anuya, Lawrence Spear, Katherine Landau Wright, and Maria Soledad Pera. 2020. A ranking strategy to promote resources supporting the classroom environment. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE, 121–128.

[28] Shingo Nahatame. 2021. Text readability and processing effort in second language reading: A computational and eye-tracking investigation. *Language learning* 71, 4 (2021), 1004–1043.

[29] J Peter Kincaid, Robert P Fishburne, Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. (1975).

[30] Christine Pinney, Casey Kennington, Maria Soledad Pera, Katherine Landau Wright, and Jerry Alan Fails. 2024. Incorporating Word-level Phonemic Decoding into Readability Assessment. In *Proceedings of the LREC-COLING Joint Conference*. European Language Resources Association, Torino, Italy.

[31] R Rajalakshmi, Hans Tiwari, Jay Patel, R Rameshkannan, and R Karthik. 2020. Bidirectional GRU-based attention model for kid-specific URL classification. In *Deep Learning Techniques and Optimization Strategies in Big Data Analytics*. IGI Global, 78–90.

[32] Janet C Read and Stuart MacFarlane. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 conference on Interaction design and children* (Tampere, Finland) *(IDC '06)*. Association for Computing Machinery, New York, NY, USA, 81–88.

[33] Gavin Sim and Matthew Horton. 2012. Investigating children's opinions of games: Fun Toolkit vs. This or That. In *Proceedings of the 11th International Conference on Interaction Design and Children* (Bremen, Germany) *(IDC '12)*. Association for Computing Machinery, New York, NY, USA, 70–77.

[34] George Spache. 1953. A new readability formula for primary-grade reading materials. *The Elementary School Journal* 53, 7 (1953), 410–413.

[35] George D Spache. 1968. Good reading for poor readers. (1968).

[36] Gail Tompkins. 2001. *Literacy for the 21st Century: A Balanced Approach 7th Edition*. Pearson.

[37] Nicholas Vanderschantz and Annika Hinze. 2021. Children's query formulation and search result exploration. *International Journal on Digital Libraries* 22, 4 (2021), 385–410.