To Blend or Not to Blend? Perceptual Speed, Visual Memory and Aggregated Search

Lauren Turpin, Diane Kelly and Jaime Arguello
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
lwturpin@gmail.com, dianek@email.unc.edu, jarguell@email.unc.edu

ABSTRACT

While aggregated search interfaces that present vertical results to searchers are fairly common in today's search environments, little is known about how searchers' cognitive abilities impact how they use and evaluate these interfaces. This study evaluates the relationship between two cognitive abilities – perceptual speed and visual memory – and searchers' behaviors and interface preferences when using two aggregated search interfaces: one that blends vertical results into the search results (blended) and one that does not (non-blended). Cognitive tests were administered to sixteen participants who subsequently performed four search tasks using the two interfaces. Participants' search interactions were logged and after searching, they rated the usability, engagement and effectiveness of each interface, as well as made comparative evaluations. Results showed that participants with low perceptual speed spent significantly more time completing tasks when using the blended interface, while those with high perceptual speed spent roughly equivalent amounts of time completing tasks with the two interfaces. Those with low perceptual speed also rated both interfaces as significantly less usable along many measures, and were less satisfied with their searches. There were also main effects for interface: participants rated the non-blended interface significantly more usable than the blended interface.

Keywords

Aggregated search interfaces, search behavior, cognitive ability

1. INTRODUCTION

The goal of *aggregated search* is to combine results from multiple search engines in a single presentation. Most commercial search portals such as Bing and Google provide access to a wide range of specialized search engines called *verticals*. Different verticals focus on different types of media (images, news) or help users perform different types of tasks (shopping, local). Vertical results can always be accessed using tabs at the top of the search results page (SERP). However, in certain cases, depending on the query, the search portal may also decide to showcase a particular vertical by blending a few of its top results somewhere above, within, or below the first page of web results. The idea is to inform users that the underlying vertical might have relevant content. For example, in response to the query "tiger", Bing mixes results from the news, video, and image verticals along with web results. A different mix of results might be generated for a different query.

SIGIR '16, July 17-21, 2016, Pisa, Italy

ACM 978-1-4503-4069-4/16/07…$15.00
DOI: http://dx.doi.org/10.1145/2911451.2914809

Recently, there have been a growing number of studies on aggregated search interfaces [3, 4, 6, 12, 14]. Some of this work suggests individual differences might play a role in users' evaluations of aggregated search interfaces. Both Arguello et al. [4] and Bron et al. [6] compared blended and non-blended interfaces for aggregated search and found no clear preference. In Arguello et al.'s study, participants who preferred the non-blended display stated the blended display was confusing and harder to navigate, and the vertical results distracting. Participants who preferred the blended display described it as helpful, visually pleasing and easier to use. The authors concluded, "people who are more visual thinkers might prefer the blended display and people who are novice users and/or have attention difficulties might prefer the non-blended display" (p. 443). Bron, et al. studied graduate students conducting multi-session searches and found most searchers moved between the blended and non-blended displays, suggesting the usefulness of the display might depend on the task or stage. However, there was no clear switching pattern: some started with blended then switched to non-blended, while others did the opposite. The authors noted these variations were likely a result of individual differences.

One way that individuals differ is according to cognitive abilities. A person's cognitive abilities are their capacities to perform various mental activities such as problem solving, reasoning, remembering, and decision-making [8]. Examples of cognitive abilities include perceptual speed, associative memory, and visualization ability. Variations in some cognitive abilities might make it more challenging for certain people to interact with vertical search interfaces. For example, someone who has low visual memory might have a more difficult time navigating around, or filtering out, visual verticals such as images and videos. Hiding these vertical results from view until the searcher is ready to use them might lead to a better search experience.

In this work, we focus on two cognitive abilities, *perceptual speed* and *visual memory*. The basic research question we address is: *How do users' perceptual speed and visual memory abilities impact their search behaviors and interface preferences in the context of aggregated search?* Perceptual speed is a person's "speed in comparing figures or symbols, scanning to find figures or symbols, or carrying out other very simple tasks involving visual perception" [9, p. 123]. Several recent studies have shown that people with lower perceptual speed experience greater workload during search and interact at slower rates than those with higher perceptual speed [1, 5]. Another study has shown that those with lower perceptual speed achieved lower precision and recall [2]. Thus, we choose this ability in part because existing evidence shows it plays a role in search interaction. In the context of aggregated search, perceptual speed may come into play as a person scans the search results to identify relevant bits of information. The inclusion of verticals might make this process

more difficult for those with low perceptual speed since there are different types of information to parse.

Visual memory is "the ability to remember configuration, location and orientation of figural material" [9, p. 109]. Visualization ability, which is different from visual memory, has been studied frequently in search, although most studies have found little to no effects on visualization ability and search behavior and success. The one exception [5] found those with lower visualization ability issued fewer queries, made fewer SERP clicks and visited fewer URLs. Conversely, visual memory has not been investigated much in the context of search applications. In a study of email applications, it was found that people with low visual memory were slower carrying out certain tasks [10]. The authors observed that an interface that allows people with low visual memory to focus on the specific aspect of the interface that is most helpful to a particular task might be beneficial; this observation seems to have implications for aggregated search interfaces. When various verticals are blended into the SERP a person with low visual memory might have a more difficult time remembering the content of visual verticals and their locations on the SERP.

2. METHOD

A laboratory experiment was conducted using two aggregated search interfaces: one that blended vertical results into the Web results and one that only provided indirect access to the verticals through tabs. Figure 1 presents a screenshot of the blended interface. In both interfaces, vertical results could be accessed via tabs at the top and side of the interface. The blended interface also presented vertical results on the initial SERP, which were blocked in set positions for each vertical, if any were returned. They were blended in the same order: web results (1-3), images, video, web results (4-6), news, web results (7-10), and shopping The Bing API was used to generate Web results as well as results for the news, images and video verticals. The eBay API was used to generate results for the shopping vertical.

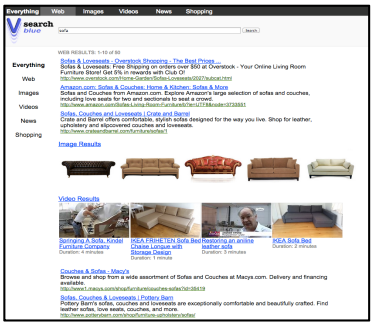

Figure 1. Screen shot of the blended search interface.

2.1 Participants

Participants were recruited through a mass email to the staff list at University of North Carolina. Sixteen participants were enrolled who had an average age of 36 (SD=11.81). Three participants were male and 13 female. Eleven participants had bachelor's degrees, three had master's degrees and two had doctorates. All participants had more than 10 years of computer experience. Participants were compensated with $15.00.

2.2 Cognitive Tests

Each participant completed two cognitive abilities tests from the Ekstrom Kit of Factor-Referenced Cognitive Test to assess perceptual speed and visual memory [9]. The *Finding A's* test was used to measure perceptual speed (PS) and the *Shape Memory* test was used to measure visual memory (VM). These are standardized tests that have been used in a large number of studies. Table 1 presents the descriptive statistics of the cognitive tests, which show our results were within one standard deviation of the reference score [9]. The median was used to divide participants equally into high/low PS and VM groups.

Table 1. Scores on cognitive abilities tests.

	Perceptual Speed	Visual Memory
Possible Range	0-200	-32 - +32
Mean (SD)	51.94 (10.41)	18 (8.69)
Median	51	20.5
Min, Max	34, 74	-2, 30
EKM Mean (SD) (Reference Scores)	Males: 47 (14.9) Females: 54 (14.9)	21.4 (4.3)

2.3 Search Tasks

Each participant completed four search tasks: two with each display. The order of the tasks and interfaces were counterbalanced, although participants always used one interface to complete two tasks before they moved to the next interface. The search tasks come from Kelly et al.'s [11] work; specifically, we used the four "Create" tasks. Create tasks require people to generate different things: a plan for constructing a soapbox derby car, an exercise program, a mascot for a sports team and a design for a living room. Kelly et al.'s [11] participants took about ten minutes completing these types of tasks, which is one reason we decided to use them. We also thought participants might need to use vertical results to address the tasks, or at least would issue queries that were likely to return vertical results. To limit the session to one hour, we gave participants a ten-minute task time limit; on average they spent about 6 minutes per task.

2.4 Interface Evaluation

Participants evaluated the interfaces and their search experiences using three types of questionnaires: post-task, post-system and exit questionnaires. The post-task questionnaire asked participants to rate how difficult it was to find relevant documents, their ability to find relevant documents, the system's ability at retrieving relevant documents, how many relevant documents they thought they found and search success. The post-system questionnaire contained the 10-item System Usability Scale (SUS) [7] and four subscales of the User Engagement Scale (UES) [13]: (1) focused attention, (2) felt involvement, (3) perceived usability and (4) endurability. All items were evaluated with 5-point scales, where higher values indicated more of the construct being measured or greater agreement. For both the SUS and UES, we followed the analyses guidelines put forward by the authors [6, 13]. Participants' responses to the 10-item SUS were averaged to arrive at an overall SUS score. Participants' responses to each of the four UES subscales were averaged to obtain overall scores for each subscale. These averages were then averaged to arrive at an overall engagement score.

The post-system questionnaire also contained five items about search effectiveness and four about ease of use, including items about the usefulness of the information returned and the presentation method. These items were evaluated with 5-point agreement scales. Reliability analysis was performed to determine if we could average items. The Cronbach's alpha for the search effectiveness items was 0.861 and 0.681 for the ease of use items. These values support combining responses for analyses.

3. RESULTS

Unless otherwise reported, mixed ANOVAs were used to analyze the data, with cognitive ability (high, low) as a between-subjects factor and interface (non-blended, blended) as a within-subjects factor. There were no significant main effects for visual memory for any of the user experience or search behavior measures, and there were no significant interaction effects between interface and either of cognitive abilities except for time. Because of space, we only present results related to perceptual speed and interface, and the interaction for time.

Figures 2, 3 and 4 display the post-task and post-system measures according to perceptual speed and interface. There were significant main effects for perceptual speed on four of the five post-task measures (Figure 2). High-PS participants rated their search skills higher ($F(1, 16)=4.42$, $p=0.04$) and system performance higher ($F(1, 16)=8.75$, $p=0.005$). They also rated their searches as more successful ($F(1, 16)=4.41$, $p=0.04$) and believed they had found more of the relevant documents ($F(1, 16)=9.58$, $p=0.003$). There were several significant main effects for perceptual speed on participants' post-system evaluations (Figures 3 and 4). High-PS participants rated the ease of use (Figure 2) and perceived usability (Figure 3) of *both* systems higher than low-PS participants [($F(1, 16)=5.20$, $p=0.04$) and ($F(1, 16)=5.60$, $p=0.03$), respectively. There were also several significant main effects for interface. The non-blended interface received significantly greater SUS scores ($F(1, 16)=5.48$, $p=0.03$) and ease of use scores ($F(1, 16)=6.09$, $p=0.03$) than the blended interface (Figure 2).

Figure 5 displays the mean number of queries and clicks made by participants according to perceptual speed and interface, and the amount of time taken to complete tasks. There were no significant main effects for perceptual speed or interface on any measure, but there was a significant interaction effect between perceptual speed and interface for time: Low-PS participants spent longer completing tasks with the blended interface, while High-PS participants spent similar amounts of time completing tasks regardless of interface ($F(1, 16)=6.33$, $p=0.015$).

Participants' clicks were further analyzed to determine if they selected any vertical results or navigated to another SERP display. Table 2 shows these frequencies according to perceptual speed. Overall, participants with high perceptual speed clicked on more vertical results (n=16) than those with low perceptual speed (n=6). An analysis of the mean number of clicks was not significant. Participants also clicked on more verticals when they were using the blended interface, which is not surprising since the vertical results were integrated into the default SERP. An analysis of the means showed a significant difference [nonblended: 0.06 (SD=0.35); blended: 0.63 (SD=0.91); $F(1, 16)=11.15$, $p=0.001$]. The last three columns of the table show how many clicks originated from different types of SERPs, some of which were only accessible via tabs. *Everything SERP* was only available in the blended display; this was also the default SERP for this interface, while in the non-blended interface *Web SERP* was the default. Table 2 shows that High-PS and Low-PS participants made similar numbers of clicks from the *Everything SERP*, but High-PS participants navigated more often to the vertical SERPs (images, videos, news or shopping) (7 vs. 0 clicks). Participants made more clicks on the *Web SERP* when using the non-blended display, which is not surprising since this was the default SERP. Of potential interest is that while there were 10 cases where High-PS participants made the effort to select the *Web SERP* over the default *Everything SERP* when using the blended interface, there were no instances of Low-PS participants doing this.

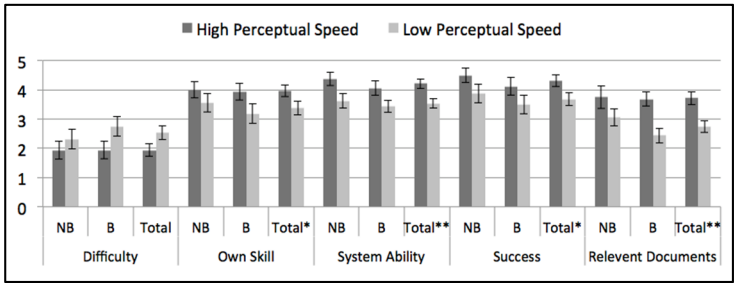

Figure 2. Post-task evaluations according to perceptual speed and interface (NB=Non-blended; B=Blended; *p<0.05; **p<0.01).

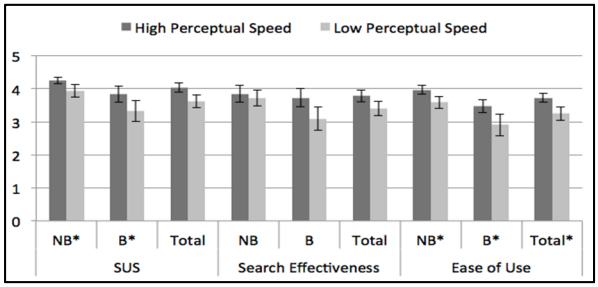

Figure 3. Post-system evaluations according to perceptual speed and interface (NB=Non-blended; B=Blended; *p<0.05).

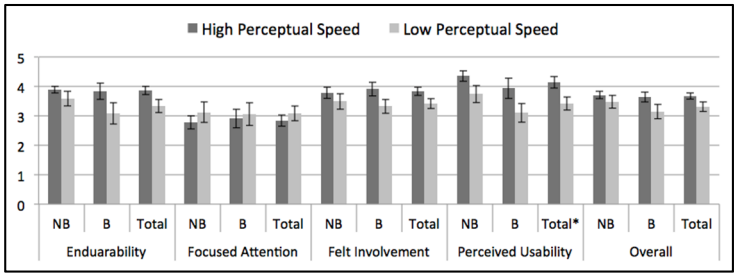

Figure 4. User Engagement post-system evaluations according to perceptual speed and interface (NB=Non-blended; B=Blended; *p<0.05).

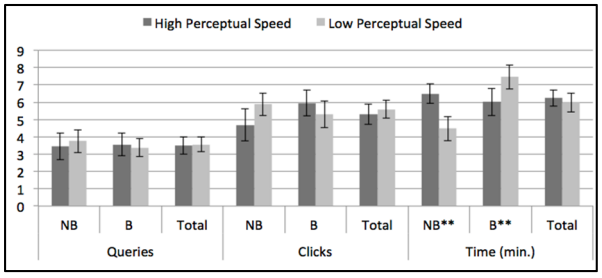

Figure 5. Search behaviors according to perceptual speed and interface (NB=Non-blended; B=Blended; **p<0.01).

Table 2. Number of clicks on vertical results and SERPs according to perceptual speed and interface.

		Vertical Results	Everything SERP	Web SERP	Vertical SERP
HPS	NB	2	-	73	2
	B	14	80	10	5
	Total	16	80	83	7
LPS	NB	0	-	94	0
	B	6	85	0	0
	Total	6	85	94	0
Total	NB	2	-	167	2
	B	20	165	10	5

4. DISCUSSION AND CONCLUSION

In this paper, we investigated the impact of two cognitive abilities, perceptual speed and visual memory, on searchers' behaviors and interface evaluations when using two aggregated search interfaces: one that blended vertical results into the search results and one that did not. Overall, we found no effects for visual memory ability, but a number of effects for perceptual speed, which is consistent with other studies (although none of these past studies have focused on aggregated search).

Perceptual speed (PS) had a significant effect on many of our user experience measures. High-PS participants rated the perceived usability and ease of use of both systems higher than Low-PS participants. These participants also rated their search skills higher, system performance higher, their searches as more successful and believed they had found a greater number of relevant documents. Overall, these results indicate that High-PS participants had more positive search experiences than Low-PS participants regardless of interface.

We also found a main effect for interface on usability and easy use. The non-blended interface received significantly higher usability and ease of use scores than the blended interface. An examination of participants' clicks showed that they did not select many vertical results. Thus, the difference in usability and easy or use measures might be because the verticals were not useful and therefore an unwelcome (and unusable) distraction on the blended interface. In this study, we controlled the type and placement of verticals and this finding might not apply to vertical displays that are generated dynamically in response to queries. Although participants could change the display of verticals by selecting *Web SERP*, there were few instances where this occurred and no instances of Low-PS participants clicking on *Web SERP*. Initially, we thought that Low-PS participants might be more inclined to do this so as to reduce the complexity of the perceptual space, but it might be that these participants did not notice or realize this was possible; indeed, Low-PS participants never changed the default SERP. Low-PS participants also spent significantly longer completing tasks with the blended interface, which might have contributed to its lower usability scores.

Overall, our results call into question the idea of one-size-fits-all search interfaces, especially when they involve aggregating different types of search results. Our results, combined with those of other researchers', highlight the importance of perceptual speed in search. Current search modalities place great demands on the perceptual system and future research might focus on developing interfaces that improve the search experience for people with Low-PS. Projecting even further into the future, researchers might also begin to consider the role of other cognitive abilities, such as those that emphasize auditory and memory skills, on voice-based search interactions.

5. ACKNOWLEDGEMENTS

This work was supported in part by NSF grant IIS-1451668. Any opinions, findings, conclusions, and recommendations expressed in this paper are the authors and do not necessarily reflect those of the sponsor.

6. REFERENCES

[1] Al-Maskari, A. & Sanderson, M. (2011). The effect of user characteristics on search effectiveness in information retrieval. *IP&M, 47*, 719-729.

[2] Allen, B. (1994). Perceptual speed, learning and information retrieval performance. *Proc. of SIGIR*, 71-80.

[3] Arguello, J. & Capra, R. (2012) The effect of aggregated search coherence on search behavior. *Proc. of SIGIR*, 1293-1302.

[4] Arguello, J., Wu, W.-C., Kelly, D., & Edwards, A. (2012). Task complexity, vertical display and user interaction in aggregated search. *Proc. of SIGIR*, 435-444.

[5] Brennan, K., Kelly, D. & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. *Proc. of IIiX*, 165-174.

[6] Bron, M., van Gorp, J., Nack, F., Baltussen, L.B. & de Rijke, M. (2013). Aggregated search interface preferences in multi-session search tasks. *Proc. of SIGIR*, 123-132.

[7] Brooke, J. (1996). SUS: A "quick and dirty" usability scale. *Usability Evaluation in Industry*. Taylor & Francis.

[8] Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-analytic studies*. Cambridge University Press.

[9] Ekstrom, R.B., French, J.W., Harman, H.H. and Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests.* Educational Testing Service, Princeton, NJ.

[10] Gwizdka, J. & Chignell, M. (2004). Individual differences and task-based user interface evaluation: A case study of pending tasks in email. *Interacting with Computers,* 769-797.

[11] Kelly, D., Arguello, J., Edwards, A. & Wu, W.-C. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. *Proc. of ICTIR*, 101-110.

[12] Liu, Z., Liu, Y., Zhou, K., Zhang, M. & Ma, S. (2015). Influence of vertical result in Web search examination. *Proc. of SIGIR*, 193-202.

[13] O'Brien, H.L. & Toms, E.G. (2010). The development and evaluation of a survey to measure user engagement. *JASIST, 61*(1), 50-69.

[14] Sushmita, S., Joho, H., & Lalmas, M. (2009). A task-based evaluation of an aggregated search interface. *Proc. of SPIRE*, 322–333.