How Do Active Reading Strategies Affect Learning Outcomes in Web Search?

Nirmal Roy, Manuel Valle Torre,
Ujwal Gadiraju, David Maxwell, Claudia Hauff

Delft University of Technology, The Netherlands
{n.roy, m.valletorre, u.k.gadiraju, d.m.maxwell, c.hauff}@tudelft.nl

Abstract. Prior work in education research has shown that various active reading strategies, notably highlighting and note-taking, benefit learning outcomes. Most of these findings are based on observational studies where learners learn from a *single* document. In a *Search as Learning (SAL)* context where learners have to iteratively scan and explore a large number of documents to address their learning objective, the effect of these active reading strategies is largely unexplored. To address this research gap, we carried out a crowd-sourced user study, and explored the effects of different highlighting and note-taking strategies on learning during a complex, learning-oriented search task. Out of five hypotheses derived from the education literature we could confirm three in the SAL context. Our findings have important design implications on aiding learning through search. Learners can benefit from search interfaces equipped with active reading tools—but some learning strategies employing these tools are more effective than others.[1]

1 Introduction and Prior Work

In the education literature, *active reading tools* such as highlighting and note-taking have been shown to improve learning outcomes in both low-level recall-oriented tasks [2,24,26], and high-level critical tasks [10]. These works also explore different strategies by which learners *use* these tools and their effects on learning outcomes [1,11,14,26]. However, in most of these works, learners are tasked to learn from a single document—often on paper. The effects of these strategies are unexplored in a *Search as Learning (SAL)* [5] context, where learners engage in an iterative exploration of the web, scanning and processing a number of documents with the goal of gaining knowledge pertaining to their learning objectives.

Previously, several *information organisational tools* have been developed for web search engines [3,8]. However, the effect that these tools have on learning has not been explicitly measured, nor do they study if participants employed different strategies while using these tools. Moreover, contemporary web search engines do not employ highlighting or note-taking tools—despite their benefits in

[1] This research has been supported by *DDS (Delft Data Science)* and *NWO* projects *SearchX* (639.022.722) and *Aspasia* (015.013.027).

Table 1. The five hypotheses and rationalisations used for this exploratory study.

Hypothesis		Rationale
H1	Learners who consider highlighting to be an important active reading strategy benefit less from it than learners who do not.	According to [26], learners who are *less* accustomed to highlighting put more effort into the act of highlighting and ultimately a better learning outcome is recorded for them.
H2	Learners directly copying considerable portions of their notes from documents they have viewed benefit less than participants who rephrase content in their own way.	Copying large portions of text reduces the attention of learners to critical details [1]. Rephrasing text while note-taking leads to a deeper processing and understanding of the said text while writing summaries [10].
H3	The number or amount of highlights by learners is *not* an indicator of learning outcomes.	Prior studies [12,17,26] have shown that the amount of highlights is not an indicator of learning outcomes.
H4	Learners who take wordier notes cover more facts in their essays.	Prior works [11,18] depict conflicting observations regarding wordy notes. For this study, we assume that wordier notes contain more facts [18].
H5	Trained highlighters and note-takers learn significantly more than their untrained counterparts.	[14] and [4] trained learners on effective highlighting and note-taking strategies respectively. They observed that the trained group of learners had significantly greater learning outcomes compared to control groups.

learning [10,26]. In order to address these shortcomings, we utilise data obtained from a crowd-sourced user study [21] to investigate how different highlighting and note-taking strategies (shown to be beneficial in learning outside of a SAL setup) affect learning outcomes during a complex, learning-oriented search task.

In this work we investigate whether five hypotheses (summarised in Table 1), inspired from the education literature, hold up in our SAL setup too.

2 Study Design

User Data, Topics and System In this work we make use of data collected during a user study conducted by Roy et al. [21]. The user study follows the setup by Moraes et al. [16], making use of the open source retrieval system, `SearchX` [20]. The standard interface, facilitated by the *Bing Search API*, provides a series of widgets, quality control features and generates fine-grained search logs, allowing us to capture a number of key behavioural measures. On top of the standard widgets of `SearchX`, we incorporate *highlighting* and *note-taking* tools, with a screenshot of the tools available in Figure 1 of Roy et al. [21]. In order to systematically evaluate the effect of active reading strategies (from our hypotheses) on learning, we consider four experimental conditions, namely:

- `CONTROL`: The standard `SearchX` search interface is provided *without* highlighting or note-taking tools.
- `NOTE`: In this condition, only the note-taking tool is enabled.
- `HIGH`: In this condition, only the highlighting tool is enabled.
- `HIGH+NOTE`: Both the highlighting and note-taking tools are enabled.

In line with prior works [15,22], learners are assessed based on a learning-oriented *critical task*. Two topics—*Genetically Modified Organisms* (**GMO**) and

Urban Water Cycle (**UWC**) inspired from Câmara et al. [7]—are used, and we ask learners to write a summary criticising and evaluating ideas from multiple perspectives [13]. In the data collected from the user study [21] (where highlighting and note-taking tools (*not* strategies) were examined over learning and search behaviour), we used: the text learners highlighted; the notes they have taken; the total time spent in taking notes; and their written essays. Depending on the experimental condition, learners had access to their saved documents (`CONTROL` and `NOTE`), their highlights together with the documents (`HIGH` and `HIGH+NOTE`) or their notes (`NOTE` and `HIGH+NOTE`) while writing the essays.

We collected data from $N = 115$ participants (referred to as *learners*) [21]; 71 of whom were assigned to the **GMO** topic, with the remaining 44 assigned to the **UWC** topic. In order to evaluate the learning outcomes from the essays, we employ two metrics inspired from Wilson and Wilson [25]. Specifically, we use **F-Fact**, which counts the number of individual facts present in the essays, and **T-Depth**, which rates the extent to which certain subtopics of the topics is covered in a summary essay, on a scale of 0-3 (from not covered at all, to covered with great focus). Both these measures were shown to be good indicators of learning. Three annotators (this paper's authors) split the 115 essays for manual annotation; 18 essays were analysed by all. They obtained a Pearson correlation of $0.78(p = 0.002)$ for **T-Depth** scores and a correlation of $0.76(p = 0.002)$ for **F-Fact** scores. We also calculated the Flesch-Kincaid[2] scores of the essays in order to assess their readability. A high score indicates that the essay is simple to read; a low score indicates a complicated text, best read by a graduate. After obtaining the essay scores, we operationalised our five hypotheses based on our collected data as follows:

H1: Learners were asked *Do you think highlighting is useful?* during the pre-questionnaire. This was an open question; we manually analysed their answers and divide them into *pro*, *unsure* and *anti* highlighters[3].

H2: We calculated how many terms from the learners' notes are taken verbatim from the documents they read. The more terms that overlapped, the more we assumed text was directly taken from the examined documents.

H3: We divided (median-split) learners into *heavy* and *light* highlighters based on two separate conditions: *(i)* the total number of highlighting actions; and *(ii)* the total number of words highlighted.

H4: We divided (median-split) learners into *heavy* and *light* note-takers based on the total number of words written in their note-taking tool.

H5: We make two assumptions to distinguish between trained and untrained highlighters and note-takers: *(i)* learners who frequently engaged in highlighting and note-taking prior to the study are considered to be trained (learners were asked the open question: *How often do you highlight and take notes while learning?* during the pre-questionnaire)[4]; and *(ii)* based on their

[2] We use `textstat` for computing the Flesch readability score.

[3] Pro - *A great extent*; Unsure -*It's a mild benefit to me*; Anti - *I don't think highlighting itself helps me all that much.*

[4] Trained - *Almost always if I see something very new to me*; Untrained - *Rarely*

Table 2. Mean (standard error) of learning metrics and metrics pertaining to active reading strategies across all participants in each condition. † indicates two-way ANOVA significance, while $^{\mathcal{C}}$,$^{\mathcal{H}}$,$^{\mathcal{N}}$,$^{\mathcal{B}}$ indicate post-hoc significance (TukeyHSD pairwise test, $p < 0.05$) increases vs. `CONTROL`, `HIGH`, `NOTE` and `HIGH+NOTE` respectively.

Measure		`CONTROL`	`HIGH`	`NOTE`	`HIGH+NOTE`
I	**#users**	32	29	29	25
II	**Session duration (minutes)**	$23m40s(1m51s)$	$28m19s(1m48s)$	$20m3s(1m15s)$	$29m17s(3m3s)$
III	***T-Depth* scores of essays**†	$1.2(0.1)^{\mathcal{H}}$	$1.6(0.1)^{\mathcal{C}}$	1.4(0.1)	1.5(0.1)
IV	***F-Fact* scores of essays**†	$14.6(1.8)^{\mathcal{N}}$	16.6(1.0)	$19.6(1.6)^{\mathcal{C}}$	15.9(1.6)
V	**Flesch scores of essays**†	32.2(7.0)	21.4(11.6)	$15.9(11.4)^{\mathcal{B}}$	$46.4(3.3)^{\mathcal{N}}$
VI	**#essay terms**	181.6(13.5)	200.8(15.9)	225.9(20.9)	193.0(17.6)
VII	**#highlight actions**	—	56.8(45.0)	—	54.9(48.4)
VIII	**#words highlighted**	—	1625.8(406.1)	—	1533.6(290.5)
IX	**Frac. essay terms in highlights**	—	0.4(0.0)	—	0.5(0.0)
X	**Overlap notes w/ documents**	—	—	10%(0.0)	10%(0.0)
XI	**#words in note-pad**	—	—	1000.1(460.0)	372.3(181.0)
XII	**Frac. essay terms in notes**†	—	—	$0.4(0.0)^{\mathcal{B}}$	$0.2(0.1)^{\mathcal{N}}$

education level—learners having a bachelor's, master's or a doctorate degree are considered to be trained.

3 Results and Discussion

The basic learner statistics for each condition are shown in Table 2. We observe that `HIGH` learners cover significantly more subtopics in their essays (**T-Depth**, **III**), whereas `NOTE` learners write significantly more facts than their `CONTROL` counterparts (**F-Fact**, **IV**). Essays written by `NOTE` learners were also significantly more complex to read compared to `HIGH+NOTE` learners (**Flesch**, **V**). Incorporating both highlighting and note-taking tools does not lead to a significant improvement in learning outcomes.

H1: We did not observe a significant difference (Table 3) for Flesch scores (**V**) and F-Fact (**III**) between the three groups of highlighters belonging to `HIGH` and `HIGH+NOTE` when compared to the three groups of `CONTROL`. However, we observed significant differences for T-Depth ($F(2, 77) = 6.44, p = 0.002$). Post-hoc tests revealed that unsure highlighters belonging to both `HIGH` and `HIGH+NOTE` cover significantly more subtopics in their essays than their `CONTROL` counterparts. Anti-highlighters belonging to `HIGH` show better learning outcomes compared to anti-highlighters belonging to `CONTROL`, whereas pro-highlighters belonging to `HIGH` and `HIGH+NOTE` gain no benefits. This is in line with the findings of [26] and shows evidence *for* our hypothesis. This might be attributed to the fact that learners who are not sure about the benefits of highlighting put more effort in the act of highlighting itself. This also indicates that highlighting makes some learners process text in a way different from how they normally would, which eventually leads to a better understanding of the text.

Table 3. H1: Learners are divided into *pro-highlighters*, *unsure* or *anti-highlighters*. † indicates two-way ANOVA significance, while $^{\mathcal{C}}$,$^{\mathcal{H}}$,$^{\mathcal{B}}$ indicate post-hoc significance (TukeyHSD pairwise test, $\mathbf{p < 0.05}$) with Holm-Bonferroni correction.

	CONTROL			HIGH			HIGH+NOTE		
	Pro	**Unsure**	**Anti**	**Pro**	**Unsure**	**Anti**	**Pro**	**Unsure**	**Anti**
I #users	9	13	10	13	11	5	11	7	7
II #words highli.	—	—	—	1529.8 (333.1)	1944.6 (1018.2)	1174.2 (126.6)	1703.0 (319.0)	1826.7 (790.1)	974.1 (490.3)
III F-Fact	13.1(1.9)	16.3(3.9)	13.6(3)	17.1(1.3)	14.6(1.5)	19.6(3.6)	16.2(2.4)	17.9(3.7)	13.6(2.5)
IV T-Depth†	1.2(0.2)	1.2(0.1)$^{\mathcal{H},\mathcal{B}}$	1.2(0.1)$^{\mathcal{H}}$	1.4(0.1)	1.6(0.1)$^{\mathcal{C}}$	2.3(0.2)$^{\mathcal{C}}$	1.2(0.2)	1.7(0.1)$^{\mathcal{C}}$	1.8(0.3)
V Flesch	35.7(7.7)	25.9(12.1)	37.3(15.4)	8.0(18.4)	27.3(21.7)	43.3(3.1)	48.9(5.2)	41.7(2.9)	47.2(8.7)

H2: From Table 2, we find that notes of learners from both NOTE and HIGH+NOTE on average have 10% overlap with the documents they read (row **X**). Hence, when we combine all note-takers, we see that those who have more than 10% of their notes overlapped with the viewed documents, covered significantly more facts (F-Facts) than whose notes overlapped less than 10% ($t(38) = 2.04, p = 0.04$), which shows evidence *against* our hypothesis. However, the former explored less subtopics and wrote more complex essays (although not significantly) than the latter. This shows that although copying considerable portions of text into notes might not be beneficial for certain aspects of essay writing like topical coverage, they can be useful when the essays require more factual information.

H3: Again from Table 2, we observe no significant difference between learners of HIGH and HIGH+NOTE when comparing learning metrics, the number of highlight actions (**VII**) and words highlighted (**VIII**). Following this, dividing learners into *heavy* and *light* highlighters, we see from Table 4 the amount of highlighting is not an indicator of learning since there is no significant difference between *heavy* and *light* highlighters (**I**, **II**), thereby providing evidence *for* our hypothesis. This indicates that the act of highlighting alone does not benefit learning—it has to be coupled with a deeper cognitive processing of the text.

H4: NOTE learners cover significantly more facts in their essays compared to their CONTROL counterparts (**IV**), cover significantly more essay terms in their notes (**XI**), and write more complex essays (**V**) than their HIGH+NOTE counterparts (Table 2). Furthermore, albeit not significantly, NOTE learners write wordier notes (**XI**) compared to HIGH+NOTE learners (Table 2). This shows evidence *for* our hypothesis that wordy notes benefit learners in our given task. Table 4 further corroborates our hypothesis where we see that learners who take wordier notes (*heavy* note-takers) cover significantly more facts in their essays, and write significantly more complex essays (**III**). This indicates that taking wordy notes and having access to them while writing their essays help learners to cover more factual information.

H5: When we divide learners based on their prior highlighting experience, we observe a significant difference for T-Depth (Table 5)—untrained highlighters

Table 4. H3, H4: Learners are divided into two groups (*heavy* and *light*) based on the median values for each active reading strategy. The learning metrics are computed separately for each group. The significant differences obtained from TukeyHSD pairwise test are highlighted in **bold**.

	F-Fact		**T-Depth**		**Flesch Score**s	
	Heavy	Light	Heavy	Light	Heavy	Light
I. **#Highlight Actions**	15.9(1.2)	16.6(1.4)	1.5(0.1)	1.6(0.1)	32.4(7.1)	33.5(11.3)
II. **#Highlighted Words**	17.0(1.3)	15.5(1.3)	1.4(0.1)	1.7(0.1)	26.4(9.5)	39.5(9.2)
III. **#Words in Note-pad**	**20.0**(**1.8**)	**15.7**(**1.4**)	1.4(0.1)	1.5(0.1)	**11.6**(**12.0**)	**48.4**(**2.8**)

Table 5. H5: Participants are divided into two groups (*trained* and *non-trained*) based on their self reported highlighting and note-taking frequency and also based on their education level. The learning metrics are computed separately for each group. The significant differences obtained from TukeyHSD pairwise tests are highlighted in **bold**.

	F-Fact		**T-Depth**		**Flesch Score**s	
	Trained	Non-trained	Trained	Non-trained	Trained	Non-trained
I. **Prior highlighting frequency**	16.8(1.3)	15.8(1.3)	**1.4**(**0.1**)	**1.7**(**0.1**)	28.8(12.2)	36.6(6.5)
II. **Highlighter Education Level**	16.4(1.2)	15.7(1.5)	1.7(0.1)	1.5(0.1)	36.6(8.8)	27.4(10.8)
III. **Prior note-taking frequency**	18.9(1.5)	16.6(1.8)	1.6(0.1)	1.3(0.1)	28.7(8.8)	31.8(10.3)
IV. **Note-taker education level**	19.5(1.6)	15.7(1.7)	1.5(0.1)	1.4(0.1)	23.8(10.8)	36.9(6.4)

cover more subtopics in their essays (**I**). Prior note-taking experience does not benefit learners. We also do not see any significant learning difference between trained and untrained highlighters/note-takers when we divide them based on their education level. These results show evidence *against* our hypothesis that being trained in highlighting and note-taking benefits learners. This indicates that if learners are prevented from learning using strategies they employ, the cost of prevention does not outweigh the benefits of using a highlighting or a note-taking tool. Although these results do not follow the observations from [4,14], it needs to be considered that in those studies, the experimental groups of learners were trained specifically about efficient highlighting and note-taking strategies.

Contributions and Conclusions In our work we investigated the extent to which five findings (i.e. our hypotheses) from the education literature [2,4,14,26] hold up in a SAL context. We confirmed three of those hypotheses, and showed that while engaging in complex learning-oriented search tasks on the web, the acts of highlighting and note-taking themselves may not benefit learners. Rather, it is *how* these tools change the way the learners scan and processes text that is more important for learning while searching. The observations from this work has design implications for search interfaces, where we must consider incorporating active reading tools within web search engines. For future work, we will build on existing literature that looks into search behaviours as proxies for learning [9,6,16,19,23]. This can be done by analysing if active reading strategies can also be used to predict learning outcomes.

References

1. Bauer, A., Koedinger, K.: Pasting and encoding: Note-taking in online courses. In: Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06). pp. 789–793. IEEE (2006)
2. Ben-Yehudah, G., Eshet-Alkalai, Y.: The contribution of text-highlighting to comprehension: A comparison of print and digital reading. J. Edu. Multimedia & Hypermedia **27**(2), 153–178 (2018)
3. Bharat, K.: Searchpad: Explicit capture of search context to support web search. Computer Networks **33**(1-6), 493–501 (2000)
4. Boyle, J.R.: Thinking strategically to record notes in content classes. American Secondary Education pp. 51–66 (2011)
5. Collins-Thompson, K., Hansen, P., Hauff, C.: Search as learning (dagstuhl seminar 17092). In: Dagstuhl reports. vol. 7 (2017)
6. Collins-Thompson, K., Rieh, S.Y., Haynes, C.C., Syed, R.: Assessing learning outcomes in web search: A comparison of tasks and query strategies. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. pp. 163–172. ACM (2016)
7. Câmara, A., Roy, N., Maxwell, D., Hauff, C.: Searching to learn with instructional scaffolding. In: Proc. 6th ACM CHIIR (2021)
8. Donato, D., Bonchi, F., Chi, T., Maarek, Y.: Do you want to take notes? identifying research missions in yahoo! search pad. In: Proc. 19th WWW. pp. 321–330 (2010)
9. Eickhoff, C., Teevan, J., White, R., Dumais, S.: Lessons from the journey: a query log analysis of within-session learning. In: Proc. 7th ACM WSDM. pp. 223–232 (2014)
10. Hagen, Å.M., Braasch, J.L., Bråten, I.: Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. Journal of Research in Reading **37**(S1), S141–S157 (2014)
11. Howe, M.J.: Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. The Journal of Educational Research **64**(2), 61–63 (1970)
12. Lauterman, T., Ackerman, R.: Overcoming screen inferiority in learning and calibration. Computers in Human Behavior **35**, 455–463 (2014)
13. Lee, H., Lee, J., Makara, K., Fishman, B.J., Hong, Y.: Does higher education foster critical and creative learners? an exploration of two universities in south korea and the usa. Higher Education Research & Development **34**(1), 131–146 (2015)
14. Leutner, D., Leopold, C., den Elzen-Rump, V.: Self-regulated learning with a text-highlighting strategy. Zeitschrift für Psychologie/Journal of Psychology **215**(3), 174–182 (2007)
15. Liu, H., Liu, C., Belkin, N.: Investigation of users' knowledge change process in learning-related search tasks. Proc. ASIS&T **56**(1), 166–175 (2019)
16. Moraes, F., Putra, S.R., Hauff, C.: Contrasting search as a learning activity with instructor-designed learning. In: CIKM 2018. pp. 167–176. ACM (2018)
17. Norman, E., Furnes, B.: The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? Computers in Human Behavior **54**, 301–309 (2016)
18. Nye, P.A., Crooks, T.J., Powley, M., Tripp, G.: Student note-taking related to university examination performance. Higher Education **13**(1), 85–97 (1984)
19. Pardi, G., von Hoyer, J., Holtz, P., Kammerer, Y.: The role of cognitive abilities and time spent on texts and videos in a multimodal searching as learning task. In: Proc. 5th ACM CHIIR. pp. 378–382 (2020)

20. Putra, S.R., Grashoff, K., Moraes, F., Hauff, C.: On the development of a collaborative search system. In: DESIRES. pp. 76–82 (2018)
21. Roy, N., Valle, M., Gadiraju, U., Maxwell, D., Hauff, C.: Searching to learn with instructional scaffolding. In: Proc. 6th ACM CHIIR (2021)
22. Song, X., Liu, C., Liu, H.: Characterizing and exploring users' task completion process at different stages in learning related tasks. Proceedings of the Association for Information Science and Technology **55**(1), 460–469 (2018)
23. Syed, R., Collins-Thompson, K.: Exploring document retrieval features associated with improved short-and long-term vocabulary learning outcomes. In: Proc. 3rd ACM CHIIR. pp. 191–200 (2018)
24. Wang, S., Unal, D., Walker, E.: Minddot: Supporting effective cognitive behaviors in concept map-based learning environments. In: Proc. 38th ACM CHI. pp. 1–14 (2019)
25. Wilson, M., Wilson, M.: A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. JASIST **64**(2), 291–306 (2013)
26. Yue, C.L., Storm, B.C., Kornell, N., Bjork, E.L.: Highlighting and its relation to distributed study and students' metacognitive beliefs. Educational Psychology Review **27**(1), 69–78 (2015)