The role of pre-existing highlights in reader–text interactions and outcomes

Samuel Dodson
University of British Columbia
dodsons@mail.ubc.ca

Luanne Freund
University of British Columbia
luanne.freund@ubc.ca

Rick Kopak
University of British Columbia
r.kopak@ubc.ca

ABSTRACT

Many digital information environments enable sharing of readers' highlights and other annotations, despite the lack of clear evidence of the effects on interaction behaviours and outcomes. We report on an experimental user study (n=15) of the impact of pre-existing highlights of varying quality on the digital reading process and outcomes of participants with different cognitive styles. We found that highlight quality affects surface level comprehension, but not deeper understanding. Participants were able to assess highlight quality and expressed different approaches to highlighting that influenced their interpretation of pre-existing highlights. Results regarding the impact of cognitive style were inconclusive.

CCS CONCEPTS

• **Information systems → Collaborative and social computing systems and tools**; **Users and interactive retrieval**; • **Applied computing → Collaborative learning**;

KEYWORDS

annotation, comprehension, highlighting, reading, social reading

ACM Reference Format:
Samuel Dodson, Luanne Freund, and Rick Kopak. 2018. The role of pre-existing highlights in reader–text interactions and outcomes. In *JCDL '18: The 18th ACM/IEEE Joint Conference on Digital Libraries, June 3–7, 2018, Fort Worth, TX, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3197026.3197066

1 INTRODUCTION

Annotation is widely considered to be an effective method to support active and effective reading [19], and highlighting, the most commonly used form of annotation [14], is available as a social, or shared, feature in many reading and learning platforms [7]. Within the digital library research community, there is more than two decades of research on annotation, and with much of that work focused on the conceptual and technical implementation [e.g., 1, 12].

Much less is known about the effects of shared and pre-existing highlighting on reading and learning. There is an assumption that sharing annotations can be useful for learning, as indicated by the development of social annotation systems for this purpose [e.g., 8], but the evidence is far from conclusive. It is unclear, for example, how shared highlights affect reading processes, particularly across different groups of readers. A recent study of the effects of pre-existing highlights on reading outcomes [3] found that highlights did not improve comprehension and may be a disadvantageous feature of the text. To better understand this phenomenon and to inform the design of social reading systems and digital libraries, this study addresses the following research questions: i) What is the impact of variable quality highlights (i.e., appropriate and inappropriate) on comprehension? ii) How are such highlights used and perceived by readers with different cognitive styles? iii) What factors affect the use and usefulness of highlights for readers?

2 PRIOR WORK

Given the central role of reading in learning, and the impact of reading comprehension on post-secondary education outcomes, it is essential to find ways to support effective reader–text interactions [17]. There has been a trend to increase readers' interactions with texts through active reading, where readers continuously evaluate the meaning of the text and its use for their task [21], or in deep reading, defined as reading with the aim of comprehension and long term retention [16]. A related goal is to facilitate social and collaborative reading and learning, through systems that enable multiple reader–text interactions, often through shared annotation capabilities, whether within private reading groups or public digital reading systems, such as Hypothesis (https://web.hypothes.is).

Though highlighting is one of the most commonly used reading strategies [14], existing studies of highlighting are inconclusive as to its benefits. Dunlosky et al. [4] carried out a review of a wide range of learning techniques employed by students, and concluded that highlighting, while widely practiced by students, was not found to be effective in improving outcomes. They suggest that these findings are limited, in that prior studies have not controlled for the amount of text highlighted or the quality of the highlights, nor has research focused on what the conditions for effective highlighting are. The majority of studies Dunlosky et al. reviewed were print based, and do not address the concerns raised by digital reading practices and the greater challenges in attention and deep reading.

The strongest effects of highlighting seem to be those for directing attention to particular words and for recall of facts, rather than for comprehending concepts. Fowler and Barker [5] found that pre-existing highlights can have a positive effect on recall. Similarly, Chi, Gumbrecht, and Hong [2] found that readers were able to answer reading questions faster and more accurately using a digital text with pre-existing highlights than without. Silvers and Kreiner [18] studied the effects of appropriate and inappropriate

JCDL '18, June 3–7, 2018, Fort Worth, TX, USA

ACM ISBN 978-1-4503-5178-2/18/06…$15.00
https://doi.org/10.1145/3197026.3197066

pre-existing highlighting on comprehension and found that comprehension levels were not improved by appropriate highlights but were significantly lower for inappropriate highlighting. Similar effects were found by [3] in a digital reading environment.

Reading depends on bottom-up and top-down cognitive processes [22]. Integrating a theory of comprehension into our study offers insight into the cognitive processes that influence how readers create and restructure mental representations of the text. Theories of comprehension that use an interaction of bottom-up and top-down processes, such as the Construction-Integration (CI) Model [9], "are currently considered the best frameworks for understanding individual differences in reading comprehension" [10]. Kintsch [9], for example, distinguishes between how readers generate representations of a text at different levels: the microstructure and macrostructure. The CI model shows that a reader can establish a microstructure representation of a text that is sufficient to answer some questions about the text without comprehending the deeper meaning of the text through the macrostructure.

Messick [13] defines cognitive style as the "strategies determining a person's typical modes of perceiving, remembering, thinking, and problem solving." One of the most studied cognitive styles is Field Dependence-Independence (FDI) [23]. Depending on their degree of FDI, individuals process, encode, and recall information differently. Field Dependents (FDs) use external cues when interacting with information, whereas Field Independents (FIs) rely on internal cues [24]. When reading, for example, FDs are more likely to use the pre-existing structure of a text than FIs, who tend ignore external cues. Wolfe [25] states, "annotations provide a critical scaffolding that can support students' critical thinking and argumentation activities." We posit that FDI may play a role in shaping how individuals perceive and use highlights.

3 METHODS

We conducted a within-subjects study in which each of 15 participants, undergraduate students recruited using a listserv, read two articles with pre-existing highlights. The study employed eye-tracking, questionnaires, and interviews for data collection. Participants were asked to read two articles from *Scientific American* [15, 20] containing highlights created by the authors. Their instructions asked them to read the articles as if they were assigned for an upcoming class, and noted that the highlights were left by previous readers. Articles were selected so as to be interesting to undergraduate students across disciplines, and texts of approximately 3,000 words were used, longer in comparison to those used by [18]. Participants read the texts using a twenty-four inch liquid-crystal display on a desktop computer equipped with a Tobii Pro X2-60 eye-tracker.

Articles were presented in two conditions, which we termed appropriate (APP) and inappropriate (INAPP) highlighting. In keeping with Kintsch's CI model of comprehension [9], APP highlights marked key concepts and themes important to understanding the overall gist or message of the article, and INAPP highlights marked content that was secondary or peripheral. To create these conditions, the three authors read the texts independently, highlighting appropriate content first, and on a second pass, highlighting inappropriate content. The three sets of highlights were aggregated and spans of text that were highlighted by at least two of the authors were selected to represent each condition. In the final versions of the articles, highlighting emphasized no more than fifteen percent of the text, in accordance with guidelines based on previous findings [11]. The instruments and highlights were identical to those used in a previous experiment [3], however, we eliminated the control condition (no highlighting) as our purpose was to compare the more specific effects of APP and INAPP highlights on reading.

The paper version of the Group Embedded Figures Test (GEFT) was administered to participants at the beginning of the study to identify their cognitive style. Participants were classified as FD or FI using the median split used in [3]. The response variables were participants' comprehension scores and eye fixations while reading the articles. Comprehension scores were determined using Term–Definition Matching, where participants were asked to match a term with one of nine possible definitions, and a modified Cloze test, where participants were asked to fill-in the blanks of a paragraph summary of each article. The quiz was designed to measure both shallow (surface level recall) and deep (conceptual understanding and gist) levels of comprehension, in accordance with the CI model.

Participants were given five minutes to read each article and seven minutes to complete the quiz, and they were informed of these contstraints in the instructions. Through pilot testing, we found this was enough time for participants to skim the articles while also encouraging use of the highlights. At the end of each session we asked participants to rate their level of prior topical knowledge, topical interest, and perceptions of the highlights on a 5-point scale. After completing the reading tasks, a 10 minute semi-structured interview was carried out to learn more about participants' task-related experiences, and their reading and highlighting practices more generally. Participants received a $20 honorarium.

4 RESULTS

Quantitative results are summarized first in three sections, followed by findings from the interview data. The assumptions necessary to run parametric tests were met, and an alpha value of .05 was used.

4.1 Comprehension Outcomes

Across all participants, scores were significantly higher in the APP than the INAPP condition for the Term–Definition Matching questions (Table 1). While the mean comprehension was also higher in APP for the Cloze questions, the difference was not significant. Within FDI groups, there were no significant differences between APP and INAPP comprehension scores, perhaps due to the smaller group sizes. A t-test showed no difference between FD and FI participants (t=-0.620, df=9.873, p=.550), indicating that the impact of the APP and INAPP conditions was similar across these groups.

4.2 Use and Perceptions of Highlights

A summary of the eye-tracking fixation data (Table 2) shows a higher mean time spent looking at highlights in the APP condition compared to INAPP. This difference is significant for the FD participants ($t(5)$=3.921, p=.011), but not for FIs ($t(8)$=2.180, p=.061). While mean times spent looking at highlights are higher for FDs than FIs in both conditions, t-tests indicate that differences are not significant: APP (t(6.801)=1.665, p=.141); INAPP (t(12.861)=0.353,

Table 1: Term–Definition Matching and Cloze score by FDI and condition.

		Term–Definition Matching			Cloze		
FDI	n	APP M (SD)	INAPP M (SD)	APP–INAPP	APP M (SD)	INAPP M (SD)	APP–INAPP
FD	6	6.33 (2.42)	4.50 (2.07)	*t(5)*=1.611, *p*=.168	3.50 (2.17)	3.83 (2.79)	*t(5)*=-0.415, *p*=.695
FI	9	7.22 (1.48)	5.56 (2.65)	*t(8)*=1.826, *p*=.105	4.11 (2.93)	3.56 (3.17)	*t(8)*=0.342, *p*=.741
all	15	6.87 (1.88)	5.13 (2.42)	*t(14)*=2.525, *p*=.024	3.87 (2.59)	3.67 (2.92)	*t(14)*=0.199, *p*=.845

p=.730). So, while there is some suggestion that FDs relied more heavily on the highlights, the impact of the conditions on both groups was similar in terms of their fixation times.

Table 2: Mean fixation duration (seconds) on highlights by FDI and condition.

FDI	n	APP M (SD)	INAPP M (SD)	Both
FD	6	117.18 (23.50)	88.00 (11.71)	102.59 (23.36)
FI	9	99.84 (12.13)	85.44 (16.34)	92.64 (15.81)
all	15	106.78 (18.94)	86.46 (14.25)	96.62 (19.44)

Results indicate that, overall, highlights did not attract a disproportionate amount of attention in comparison to non-highlighted text. Given that the Areas of Interest marking highlighted texts for the eye-tracking analysis constituted about a third of the content for each condition, the baseline fixation time would be 100 seconds of the total reading time of 300 seconds (5 minutes). Only the FDs in the APP condition exceeded this baseline (Table 2).

Participants' perceptions of the usefulness of highlights are summarized in Table 3. Overall, participants rated the usefulness of highlights significantly higher in the APP condition than the INAPP. These differences are not significant within the FDI groups.

Table 3: Mean usefulness scores of highlights by FDI and condition (5-point scale).

FDI	n	APP M (SD)	INAPP M (SD)	APP–INAPP
FD	5*	3.60 (0.89)	3.20 (0.84)	*t(4)*=1.000, *p*=.374
FI	9	3.89 (0.60)	3.22 (0.97)	*t(8)*=2.000, *p*=.081
all	14	3.79 (0.70)	3.21 (0.89)	*t(13)*=2.280, *p*=.040

*Data missing from one participant.

4.3 Role of Prior Knowledge and Interest

To investigate the association between comprehension and self-report measures we calculated nonparametric correlations. Results show significant positive correlations between interest and Cloze test scores in both the APP $r_s(12) = .564$, $p=.036$ and INAPP $r_s(12)=.713$, $p=.004$ conditions, with the strongest correlation observed in the latter. Prior knowledge and comprehension scores were not significantly correlated. This suggests that participants' level of interest in the content of the articles is a factor in their ability to grasp their overall meaning, or gist, as assessed through the Cloze test. This was particularly true in the INAPP condition.

4.4 Findings From Interviews

The interview data was coded inductively using QDA Miner. Responses provide further insights on uses and perceptions of highlights and factors that affect use. Two different approaches to highlighting emerged. One group, including many FIs, viewed highlights as a means to gain overall understanding: "the passages were too long, so what I did was to only read the highlights because they helped me get the main idea of the articles" [P12]. This was based on the idea that highlights convey "the gist of the paragraph" [P4].

Another group, primarily FDs, expected highlights to signal information or facts of significance. P11 noted, "I found the ones that were highlighting a certain word or something…like definitions, I found those really helpful." P5 clearly preferred such highlights: "in the first article…the highlights…skipped over the useful details that I would have personally highlighted and they went over the main statement of the topic being discussed." This preference may have been driven by the task (i.e. reading to prepare for a test), as P15 noted, "the more that [highlights] pertained to technical terminology as opposed to whatever the overall theme…the more I found it useful to answer the questions."

A theme that arose in the interviews is the tension between the value of highlights for focusing attention and the negative potential for distraction. P15 expressed this well: "[I] found it problematic because it would try and grab your attention to something that I would try and think, 'oh, that should be relevant. I should pay attention to this,' but it's in the middle of something that A) has no context — you don't have any frame of reference to try and understand that and B) it kind of disorients the natural flow of your reading, because you get drawn to this."

Participants identified intrinsic and extrinsic factors that affected highlight use. Chief among the former is the type of content highlighted, notably the distinction between thematic and fact-based highlighting. Other factors included location and coverage. Issues arose with highlights of partial sentences or those situated mid paragraph, due to a lack of context. Two extrinsic factors were noted: time, and prior knowledge. Many participants expressed that the usefulness of highlights increased under time pressure, when they "wouldn't have enough time to read the entire thing word for word" [P4]. P3 noted that highlights "really helped guide my attention to where, probably, the bulk of the information for the quiz was going to be." However, some were disappointed when they used this technique and the answers were not among the highlighted text. A small number of participants indicated that when they are more familiar with the content and vocabulary of a text, they would have less use for highlights as a guiding or structuring feature.

5 DISCUSSION & CONCLUSION

Results clearly indicate that there are negative effects of exposing readers to poor quality highlights in this type of time limited reading task. This adds evidence to earlier findings [18]. Comprehension scores were lower in the INAPP condition, although only the surface level test was affected. Given that the APP highlights were designed to focus attention on the main themes and concepts, we expected to see an effect on conceptual understanding as well, but this was not the case. Readers with higher levels of interest were better able to reach a deeper conceptual understanding, perhaps because the interest enabled them to focus their attention even in the face of inappropriate highlights. These results provide support for the conclusion in [4], that highlights serve better as signaling devices than as scaffolding for text comprehension.

We found that most participants were aware of the quality variation, and some clearly attended less to the INAPP highlights. This suggests that post-secondary readers have the capacity to recognize and ignore less useful highlights that they may encounter in social reading systems. The interview data suggests that ignoring such highlights was not always easy, and that the effort needed to assess and ignore distracting highlights might subtract from the effort available to engage in deep reading of challenging texts.

We expected highlights would have more value for FDs, who typically benefit from external structure. However, results show no clear differences between FD and FI readers. FDs did spend more time viewing highlights in the APP conditions than in the INAPP condition, while FIs did not spend any more time in the highlights than the rest of the text, perhaps because they had little need of them. This brings into question the hypothesis posed in [3] that FIs perform badly when faced with inappropriate highlights, because they spend valuable time trying to make sense of them. Interestingly, our initial analysis of the interview data suggests that FDs and FIs may have used highlights differently, with FDs using them predominantly to focus on factual information and FIs to gain a general conceptual understanding.

Clearly time is a key factor in the use of highlights, but more research needs to be done to determine the extent to which highlights can serve to increase reading efficiency, as suggested in [6]. A central finding of this work, is that for social highlighting to be effective, systems should support diverse approaches to highlighting. While the thematic and factual approached identified here are not new [e.g., 12], results suggest that there may be both personal and task-based reasons to adopt one or the other. We can conclude that in social reading systems, including digital libraries, there can be benefit in creating guidelines for shared highlighting, to maximize potential benefits.

Highlight quality was shown to affect surface level text comprehension, although readers were able to assess the quality of highlights and to adjust their reliance upon them accordingly. Results indicate that readers approach highlighting differently. As such, there is danger in adopting a one size fits all approach in which shared highlights are aggregated and displayed to readers. A task-based approach that adds information to highlights and allows readers to know why a highlight was created, by whom, or for what task, may be valuable, but the negative potential of distraction needs to be addressed. We fully acknowledge the limitations of this work, which is small-scale, lab-based and therefore non-naturalistic. The time constraint on readers was also quite severe and results may not generalize to more relaxed or lengthy reading tasks. Future work will move towards organic studies of social reading and learning systems in use in university courses to validate and extend these findings.

REFERENCES

[1] Maristella Agosti and Nicola Ferro. 2007. A formal model of annotations of digital content. *ACM Transactions on Information Systems* 26, 1 (2007), 244–255.

[2] Ed H Chi, Michelle Gumbrecht, and Lichan Hong. 2007. Visual foraging of highlighted text: an eye-tracking study. In *Human-Computer Interaction: HCI Intelligent Multimodal Interaction Environments*. Springer, 589–598.

[3] Samuel Dodson, Luanne Freund, and Rick Kopak. 2017. Do Highlights Affect Comprehension?: Lessons from a User Study. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, New York, NY, 381–384.

[4] John Dunlosky, Katherine A Rawson, Elizabeth J Marsh, Mitchell J Nathan, and Daniel T Willingham. 2013. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest* 14, 1 (2013), 4–58.

[5] Robert L Fowler and Anne S Barker. 1974. Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology* 59, 3 (1974), 358–364.

[6] Luanne Freund, Rick Kopak, and Heather O'Brien. 2016. The effects of textual environment on reading comprehension: implications for searching as learning. *Journal of Information Science* 42, 1 (2016), 79–93.

[7] Anis Kalboussi, Omar Mazhoud, and Ahmed Hadj Kacem. 2016. Functionalities provided by annotation systems for learners in educational context: an overview. *International Journal of Emerging Technologies in Learning* 11, 2 (2016).

[8] Nancy Kaplan and Yoram Chisik. 2005. In the company of readers: the digital library book as "practiced place". In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, New York, NY, 235–243.

[9] Walter Kintsch. 1998. *Comprehension: a paradigm for cognition*. Cambridge University Press, Cambridge, United Kingdom.

[10] Jacqueline Leighton and Mark Gierl. 2011. *The learning sciences in educational assessment: the role of cognitive models*. Cambridge University Press, Cambridge, United Kingdom.

[11] Robert F Lorch, Jr. 1989. Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review* 1, 3 (1989), 209–234.

[12] Catherine C Marshall. 2009. *Reading and writing the electronic book*. Morgan & Claypool Publishers, San Rafael, CA.

[13] Samuel Messick. 1976. *Individuality in learning*. Jossey-Bass Publishers, San Francisco, CA.

[14] Ilia A Ovsiannikov, Michael A Arbib, and Thomas H McNeill. 1999. Annotation technology. *International Journal of Human-Computer Studies* 50, 4 (1999), 329–362.

[15] Matthieu Ricard, Antoine Lutz, and Richard J Davidson. 2014. Mind of the meditator. *Scientific American* 311, 5 (2014), 38–45.

[16] Judith C Roberts and Keith A Roberts. 2008. Deep reading, cost/benefit, and the construction of meaning: Enhancing reading comprehension and deep learning in sociology courses. *Teaching Sociology* 36, 2 (2008), 125–140.

[17] Jean-Francois Rouet. 2006. *The skills of document use: from text comprehension to web-based learning*. Psychology Press, Hove, United Kingdom.

[18] Vicki L Silvers and David S Kreiner. 1997. The effects of pre-existing inappropriate highlighting on reading comprehension. *Literacy Research and Instruction* 36, 3 (1997), 217–223.

[19] Michele L Simpson and Sherrie L Nist. 1990. Textbook annotation: an effective and efficient study strategy for college students. *Journal of Reading* 34, 2 (1990), 122–129.

[20] Keith C Summa and Fred W Turek. 2015. The clocks within us. *Scientific American* 312, 2 (2015), 50–55.

[21] Craig Tashman and W Keith Edwards. 2011. Active reading and its discontents: the situations, problems and ideas of readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 2927–2936.

[22] Ludo Verhoeven and Charles Perfetti. 2008. Advances in text comprehension: model, process and development. *Applied Cognitive Psychology* 22, 3 (2008), 293–301.

[23] Herman A Witkin, R B Dyk, H F Fattuson, Donald R Goodenough, and S A Karp. 1962. *Psychological differentiation: studies of development*. Wiley, New York, NY.

[24] Herman A Witkin and Donald R Goodenough. 1977. Field dependence revisited. *ETS Research Bulletin Series* 1977, 2 (1977), i–53.

[25] Joanna Wolfe. 2008. Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies. *International Journal of Computer-Supported Collaborative Learning* 3 (2008), 141–164.