

Do highlights affect comprehension? Lessons from a user study

Samuel Dodson
iSchool
University of British Columbia
Vancouver, BC, Canada
dodsons@mail.ubc.ca

Luanne Freund
iSchool
University of British Columbia
Vancouver, BC, Canada
luanne.freund@ubc.ca

Rick Kopak
iSchool
University of British Columbia
Vancouver, BC, Canada
r.kopak@ubc.ca

ABSTRACT

A largely unquestioned assumption in social reading is that publicly shared annotations improve reading outcomes. In this study, we explore the specific assumption that relevant and irrelevant passive highlighting affects comprehension. Participants were divided by cognitive style based on their degree of Field Dependence-Independence [19]. We found that irrelevant highlights had significant negative effects on reading comprehension for Field Independents (FIs), but not Field Dependents (FDs). This is a surprising result because FDs typically rely on external cues to structure and help process information, whereas FIs use internal cues. This suggests that highlighting cues information but does not structure it.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability;

KEYWORDS

annotation, comprehension, highlighting, reading, social reading

ACM Reference Format:

Samuel Dodson, Luanne Freund, and Rick Kopak. 2017. Do highlights affect comprehension? Lessons from a user study. In *CHIIR '17: 2017 Conference on Human Information Interaction & Retrieval, March 7–11, 2017, Oslo, Norway*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3020165.3022158>

1 INTRODUCTION

Highlighting is a common form of text annotation widely used to facilitate reading, especially when reading to learn. Given the connection between reading, highlighting, and learning, it was not surprising that Marshall's [11] seminal study of annotations found that most second-hand textbooks examined were thoroughly highlighted. The practice of highlighting has carried over from print to digital reading environments, most of which offer tools for highlighting. Features such as Popular Highlights on the Amazon Kindle go even further by storing and aggregating highlights and displaying them to readers as a layer of social information complementary to the original text.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '17, March 7–11, 2017, Oslo, Norway

© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-4677-1/17/03...\$15.00

<https://doi.org/10.1145/3020165.3022158>

The underlying assumption of such systems is that both active (made by the reader during the act of reading) and passive (made by previous readers) highlights have a positive effect on reading outcomes, such as increased comprehension, retention, and engagement. However, very little research exists to support this assumption, particularly in the digital context. This paper reports on a study of passive, or social highlights, and investigates their effect on comprehension, an important outcome of human information interaction and a key component of learning.

2 PREVIOUS WORK

Marshall [11] identified multiple uses of highlights: as signals for future use; as place marks for a specific passage to be referenced later; and, to focus attention on the text, especially when a passage is difficult to read. Of these, the first is most applicable to passive highlights, which signal to readers the most important or interesting sections of the text. There is some evidence that highlights that cue key passages increase recall regardless of whether they were actively created by the reader or passively encountered [5, 9, 10]. However, one of the concerns with passive highlights, is that they may be of poor quality. If we consider that the effectiveness of highlights relies upon the von Restorff effect [18], which predicts that an item that stands out from its background is more likely to be remembered than items that do not, we have to acknowledge that poor quality highlights may interfere with readers' ability to make sense of, and learn from, highlighted text [14]. Readers may simply jump from one highlighted passage to the next, without assessing the quality or relevance of the highlights, as was found in [3].

Silvers and Kreiner [17] studied the effects of relevant and irrelevant highlighting on text comprehension in a two-phase study with three conditions: no highlighting (control), relevant highlighting, and irrelevant highlighting. They found that the participants performed similarly in the relevant and control conditions, but the irrelevant highlighting had a significant negative effect on comprehension. The same negative impact of irrelevant highlights was observed even after warning participants that the highlights may be irrelevant. We were intrigued by these results and decided to test them in a digital context. We also wondered if individual differences between readers may have influenced the results.

The way readers interact with information is affected by their cognitive styles [1, 12]. One of the most studied cognitive styles is Field Dependence-Independence (FDI) [19] [16, 20]. Messick [12] describes the difference as follows: "The field independent person tends to articulate figures as discrete from their backgrounds and to easily differentiate objects from embedding context, while the field dependent person tends to experience events globally in an undifferentiated fashion" (p. 5). FDs use external cues to guide their

information processing, while FIs use internal ones. As a result, FDs are more likely to use the existing structure of a field, whereas FIs create their own [21].

While FDI was originally a measure of perceptual ability (to structure or restructure visual fields), it has been found to affect other tasks, such as problem solving [20], cognitive restructuring ability [6], and attention to relevant cues [2]. In the context of reading, FDs are more likely to use the pre-existing structure of a text while FIs are better at focusing their attention on relevant information and ignoring distractions [7]. It is especially difficult for FDs to focus on the most important information when they are presented with distracting cues, such as irrelevant highlights [7]. When a field is well-structured, however, FDs can perform as well as FIs [21], suggesting that relevant highlighting could be used as an aid for FDs.

Given the prevalence of passive highlights in current social reading systems, we decided to test the assumptions that seem to underlie their design: that highlights of any kind improve the reading experience. Our research questions were as follows:

- Does the existence of relevant highlights improve text comprehension?
- Does the existence of irrelevant highlights reduce text comprehension?
- Are the effects of highlighting on comprehension influenced by cognitive style, specifically FDI?

3 METHODS

To answer these questions, we conducted a within-subjects experiment with 29 participants. Following the approach used by Silvers and Kreiner [17], participants were asked to read texts in three conditions: those with irrelevant highlights, those with relevant highlights, and a control with no highlights. We updated and improved the methods used in [17] by using lengthier texts presented in digital format with no more than 15% of the text highlighted, as suggested in [9]. We designed our own comprehension tests based on a recognized model of text comprehension [8]. Details of the methods are presented in the following sections.

Texts were three general interest articles from *Scientific American* of approximately 3,000 words each and presented in plain HTML format with charts and images removed. In the relevant highlighting condition passages that contained concepts or facts central to the overall meaning, or gist of the article were emphasized. Relevant highlights were created where at least two of the three researchers highlighted the same passage. Irrelevant highlights emphasized passages that at least two of the three researchers identified as informative, but peripheral to the main themes of the text. We then carefully trimmed the highlights down to include no more than 15% of the text in accordance with the findings of [9, 10].

The highlighting condition was the manipulated variable, with each participant reading one text in each condition. The Group Embedded Figures Test (GEFT) [22] was used to measure Field Dependence-Independence. Following Demick [4], participants who scored at, or below, the median GEFT score were classified as Field Dependent (FD) ($M=10.33$, $SD=2.66$, $N=15$) and those who scored above as Field Independent (FI) ($M=10.33$, $SD=2.66$, $N=14$). The primary dependent variable was comprehension, measured

through tests administered after each text was read. Tests questions were designed in accordance with the Construction-Integration model [8] to assess microstructural and macrostructural comprehension, and were of three types: multiple choice matching questions; Sentence Verification Technique (SVT) questions [15]; and an open-ended summary question.

A convenience sample of 29 undergraduates (14 males, 15 females) was recruited via listservs and posted advertisements. Sessions lasting 1.5 hours were run with groups of 10 participants randomly seated in a large computer lab. Computers were pre-set to display the articles. Instructions, pre- and post-session questionnaires, and comprehension quizzes were provided in print format.

Participants completed a consent form, a demographics and reading habits questionnaire, and the GEFT (on paper). Then they were told their task was to read three articles and complete a comprehension quiz for each. The scenario was to imagine the articles had been assigned for an upcoming class discussion for which they had limited time to prepare. Participants were given five minutes to read each article. This was done to encourage efficient reading strategies, such as the use of highlights, and to simulate typical online reading behaviors [13].

The three articles were counterbalanced to reduce ordering effects but all participants experienced the conditions in the same order: relevant, control, irrelevant. This was to limit a negative carry-over effect that we observed in the pilot test when the irrelevant highlights were assigned first and subsequent highlights were ignored. After reading an article, participants were given seven minutes to complete the corresponding comprehension quiz. They were not allowed to look at the test until they had finished reading the article and were not able to refer to the article once they started the test. After completing the reading tasks, participants completed a post-session questionnaire asking about their experiences and their thoughts on the usefulness of the highlights. They received a \$20 honorarium for participating.

The 3 researchers scored the open-ended summary responses independently using a rubric, compared scores, and reached a consensus score for each to ensure inter-scorer reliability. Data were analyzed using ANOVA and an alpha level of 0.05 was used for all statistical tests.

4 RESULTS

Comprehension scores for each condition were compared to measure the effects of highlighting on comprehension. Comprehension was measured using multiple choice, open-ended summary, and SVT questions. An overall measure, which averaged the scores from the three measures, was also used. The highest possible comprehension score for each measure is 100. Differences were found between measures of comprehension used in the study. Most previous studies have only used multiple choice as a measure of comprehension; these results found differences between the measures of comprehension, which may provide an argument for pairing other measures of comprehension with multiple choice.

Table 1 shows the mean overall comprehension scores for the three conditions, and are reported for all participants, FIs, and FDs. Across all three groups, the highest means were achieved in the control condition and lowest in the irrelevant.

Group	n	Condition					
		Relevant		Irrelevant		Control	
		M	SD	M	SD	M	SD
all	29	59.13	17.02	60.99	14.09	53.09	12.76
FD	15	62.70	15.43	54.30	13.58	64.06	11.94
FI	14	55.31	18.36	51.80	12.19	7.71	15.86

Table 1: Descriptive statistics of the overall measure.

The mean comprehension scores for all participants on the multiple choice and summary questions did not vary significantly across the three conditions; however, there was a significant effect of condition on the comprehension scores in the SVT question ($F(2,56)=4.360$, $p=.017$). *Post hoc* tests, using the Bonferroni correction, indicated that SVT scores in the irrelevant condition ($M=53.09$) were significantly lower than in the control ($M=60.99$) ($p=.006$), but the scores in the relevant condition ($M=59.13$) did not differ from the other conditions. When we combined the comprehension scores into an overall measure, there was a borderline significant effect of condition ($F(2,56)=2.998$, $p=.058$) showing the same trend as for the SVT scores.

The mean comprehension scores of the FDs were relatively consistent across conditions. One-way within-groups ANOVAS showed no evidence of significant differences between comprehension scores across conditions for any of the measures.

For FIs, some differences did arise by condition. One-way within-groups ANOVAs show no evidence of significant differences by condition for the multiple choice and summary questions, but there was a significant difference for the SVT question ($F(2,26)=3.405$, $p=.049$). A *post hoc* pairwise t-test, using the Bonferroni correction, indicates that scores in the irrelevant condition were significantly lower than the control ($p=.023$). Comprehension scores in the relevant condition were also lower than the control ($p=.053$).

In a post-session questionnaire participants were asked to describe what they thought of the highlights and whether some were more useful than others. Reactions to the highlights were mixed. Of the 29 participants, eight indicated that the highlights were very helpful, 13 claimed the quality of the highlights varied, while 8 found the highlights unhelpful.

There was an even distribution of FDs and FIs among these response groups, suggesting that there was no effect of cognitive style on attitudes towards the highlighting conditions. Positive responses tended to emphasize that highlights helped with focus and overall understanding. P28 reflected that “Reading the highlights helped to get the gist of what [the] article was talking about but didn’t help with the little details”. P15 added “I tended to focus on the highlights.”

P21 stated, “I thought the highlights were distributed between helpful and useless”. P14 added, “I found the highlights to be very helpful, especially in the first article [the relevant condition], the third article highlights [the irrelevant condition] made me skim the surrounding information”. P5 stated, “While skimming, I felt like I had to read them [the highlights], which was obnoxious when they weren’t helpful”.

Among those who had a negative response to both highlighting conditions, there was a split between those that claimed to have ignored all the highlights and those that looked at them, but did not find them useful. Several participants used words such as “annoying” and “distracting” to describe the highlights. Two participants said that they do not value highlighting. P10 said “I didn’t really notice the highlights, because I was focused on reading. I believe highlights are most helpful, when done by the reader.” P23 added, “I didn’t find the highlighting useful. If I wasn’t the one to highlight, the highlight just gets in the way”. Four participants said they ignored the highlights. P9 said “did not pay much attention to the highlights, felt they were distracting as they pulled my focus away from the article when I was reading.” P4 “noticed them but didn’t really analyze them.”

5 DISCUSSION

How do relevant and irrelevant highlights affect comprehension? The results provide no evidence that relevant highlights improve comprehension for any of the groups, and some evidence that irrelevant highlights negatively affect comprehension. Mean comprehension scores were lowest in the irrelevant highlighting condition across all groups in most measures. Statistical testing provided some evidence to support this trend, indicating that comprehension scores were significantly lower in one measure for the irrelevant condition than in the control condition for all participants and FI groups. For FIs, there is even limited evidence that relevant highlights may be more detrimental to comprehension than no highlights at all.

These results validate the earlier work of Silvers and Kreiner [17], showing no benefit of relevant highlights and some negative effect of irrelevant highlights. Furthermore, this study shows that the effects of highlighting observed by Silvers and Kreiner extend to longer texts in digital reading environments. Together, these results suggest passive highlighting provides little benefit in terms of comprehension for these participants. For most individuals, reading texts without highlights results in higher comprehension than when presented with any kind of highlights.

A unique contribution of this work was its exploration of the effects of cognitive style. In examining the effects of relevant and irrelevant highlights within the FD and FI groups, our results showed that the comprehension of FDs was not significantly affected by the different conditions, while FIs seemed to be negatively affected by both irrelevant and relevant highlights. Both FDs and FIs performed best without highlighting. There was some variation in comprehension between groups and measures; however, mean comprehension was generally highest in the control and lowest in the irrelevant condition. Differences between relevant-control and irrelevant-control pairings were insignificant for FDs. While the relevant-control comparison was also insignificant for FIs, the irrelevant-control pairing was significant in the SVT measure. This finding suggests that readers’ cognitive styles may be an important factor in the understanding of texts with passive highlighting. Contrary to what we would expect from the literature, FIs, rather than FDs, were affected by passive highlights. Given that FDs rely on external structuring to process information, we were surprised that the highlighting conditions had no effect on FDs’ comprehension. We expected a positive

effect of relevant structuring in the form of highlights, and negative effects of the control, which did not structure the text, and the irrelevant, in which the text was poorly structured. Similarly, we assumed that FIs would be unaffected by highlighting, because these individuals use internal rather than external processes to make sense of information. However, the comprehension scores for FDs were similar across conditions. The comprehension scores for FIs indicated that the irrelevant highlights were an impediment. There was even limited evidence that the relevant highlights interfered with comprehension.

One explanation for these two surprising results is that FDs and FIs engaged at different levels with the highlights. FDs' comprehension was not influenced by the highlights. FDs are known to accept the presentation of a text as given, so these individuals may have expended little or no effort assessing the relevance of the highlights. It was also found that highlights negatively affected FIs' understanding of the texts. These individuals may have spent great effort assessing the quality or relevance of the highlights, distracting them from their task in this study. Given that these results are contrary to our expectations, we must question how FDs and FIs interact with passive annotations. Future research could provide answers by measuring participants' visual attention through eye tracking.

In general, these results challenge the assumption that passive, or "social" highlights, benefit readers by helping them make sense of, and, learning from, texts. For some readers, passive highlights, whether relevant or irrelevant, may be a form of distraction and annoyance. This does not preclude benefits of highlights on different reading processes and outcomes, such as reading efficiency, navigation, re-finding information, and engagement, which may be the focus of our future research.

6 CONCLUSION

This study furthers our understanding of the effects of highlighting on reading comprehension through three contributions. No previous work on relevant and irrelevant highlights explored the effects of 1) readers with different cognitive styles, 2) the length of texts, or 3) presentation in digital reading environments. This study validates previous work [17], by showing that relevant highlights do not positively affect comprehension, while irrelevant highlights may have a negative effect on comprehension. The results of the study are limited to passive highlights. Future work could use techniques such as eye tracking to collect more data on reading processes and consider measuring different reading outcomes, such as efficiency and engagement. It would also be useful to compare the effects of passive highlights with active highlighting. Work built upon the findings of this study will be significant in the development

of digital reading environments that may better support readers' understanding and engagement with texts.

REFERENCES

- [1] Lynna J Ausburn and Floyd B Ausburn. 1978. Cognitive styles: some information and implications for instructional design. *Educational Communication and Technology* 26, 4 (1978), 337–354.
- [2] Elizabeth Berger and Leo Goldberger. 1979. Field dependence and short-term memory. *Perceptual and Motor Skills* 49 (1979), 87–96.
- [3] 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, Berlin, Germany, 589–598.
- [4] Jack Demick. 2014. *Group Embedded Figures Test: manual*. Technical Report. Mind Garden, Inc., Menlo Park, CA.
- [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] Donald R Goodenough. 1976. The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin* 83, 4 (1976), 675–694.
- [7] J Kent-Davis and Kathryn F Cochran. 1989. An information processing view of field dependence-independence. *Early Child Development and Care* 51, 1 (1989), 31–47.
- [8] Walter Kintsch. 1998. *Comprehension: a paradigm for cognition*. Cambridge University Press, Cambridge, United Kingdom.
- [9] Robert F Lorch, Jr. 1989. Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review* 1, 3 (1989), 209–234.
- [10] Robert F Lorch, Jr, Elizabeth Puzles Lorch, and Madeline A Klusewitz. 1995. Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology* 20, 1 (1995), 51–64.
- [11] Catherine C Marshall. 1997. Annotation: from paper books to the digital library. In *Proceedings of the Second ACM International Conference on Digital Libraries*. ACM, New York, NY, 131–140.
- [12] Samuel Messick. 1976. *Individuality in learning*. Jossey-Bass Publishers, San Francisco, CA.
- [13] Jakob Nielsen. 1997. How users read on the web. (Oct. 1997). <https://www.nngroup.com/articles/how-users-read-on-the-web/>
- [14] Sherrie L Nist and Mark C Hogrebe. 1987. The role of underlining and annotating in remembering textual information. *Literacy Research and Instruction* 27, 1 (1987), 12–25.
- [15] James M Royer, Barbara A Greene, and Gale M Sinatra. 1987. The Sentence Verification Technique: a practical procedure for testing comprehension. *Journal of Reading* (1987), 414–422.
- [16] Stephanie Shipman and Virginia C Shipman. 1985. Cognitive styles: some conceptual, methodological, and applied issues. *Review of Research in Education* (1985), 229–291.
- [17] 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.
- [18] Hedwig von Restorff. 1933. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung* 18, 1 (1933), 299–342.
- [19] 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.
- [20] Herman A Witkin and Donald R Goodenough. 1981. Cognitive styles: essence and origins field dependence and field independence. *Psychological issues* 51 (1981), 1–141.
- [21] Herman A Witkin, Donald R Goodenough, and Philip K Oltman. 1979. Psychological differentiation: current status. *Journal of Personality and Social Psychology* 37, 7 (1979), 1127–1145.
- [22] Herman A Witkin, Philip K Oltman, Evelyn Raskin, and Stephen A Karp. 1971. *A Manual for the embedded figures tests*. Consulting Psychologists Press, Palo Alto, CA.