Spotlights: Designs for Directing Learners' Attention in a Large-Scale Social Annotation Platform

JUMANA ALMAHMOUD, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, USA

FARNAZ JAHANBAKHSH, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, USA

MARC FACCIOTTI, Biomedical Engineering, University of California, Davis, USA

MICHELE IGO, Biomedical Engineering, University of California, Davis, USA

KAMALI SRIPATHI, Genome Center, University of California, Davis, USA

KOBI GAL, Ben-Gurion University of the Negev, Israel

DAVID KARGER, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, USA

A new approach to online discussion, which situates student discussions in the margins of the course content, can enhance student engagement with course materials. However, in high-enrollment classes, the large number of comments can overwhelm and intimidate students. Some become frustrated by the volume of potential online interactions and by a perceived lack of immediate relevance to their studies. Likewise, instructors are disappointed when outstanding discussions, that they deem valuable for all to see, get lost in the clutter.

To address these challenges, we propose visual *spotlighting* mechanisms for increasing the saliency of selected comments. We piloted and deployed multiple designs in two high-enrollment biology courses at a large public university in the United States. Interviews, surveys, and a controlled experiment show that spotlighting relevant comments in heavily annotated texts positively affects students' engagement, measured in terms of their attention to comments, and their reported sense of validation and pride. Students also reported their preferences for certain spotlighting designs.

540

CCS Concepts: • **Human-centered computing** → Collaborative and social computing systems and tools.

Additional Key Words and Phrases: spotlights, annotations, comments, ux, social annotation

Authors' addresses:, jumanam@mit.edu; Jumana Almahmoud, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA, 02139; Farnaz Jahanbakhsh, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA, farnazj@mit.edu; Marc Facciotti, Biomedical Engineering, University of California, Davis, Davis, USA, mtfacciotti@ucdavis.edu; Michele Igo, Biomedical Engineering, University of California, Davis, Davis, USA, mmigo@ucdavis.edu; Kamali Sripathi, Genome Center, University of California, Davis, Davis, USA, ksripathi@ucdavis.edu; Kobi Gal, Ben-Gurion University of the Negev, Beer Sheva, Israel, kobig@bgu.ac.il; David Karger, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA, karger@mit.edu.

1 INTRODUCTION

Social annotation tools, where readers discuss a document by highlighting sub-segments and commenting in its margins, have become popular in online learning settings. In high-enrollment classrooms, extensive comments (in the hundreds to thousands per document) and discussions may contain a treasure trove of engaging and informative ideas. But the sheer scale of the discussion may overwhelm participants. In such cases, how can the reader find annotations with which they would like most to engage? And how can instructors identify and steer students to high-value comments?

In this work, we propose the idea of *spotlighting*: using visual cues to increase the salience of specific comments in order to draw student attention to them. We explore a design space of possible spotlighting techniques, and use a mixed methods approach to measure the impact of different spotlighting techniques on students' attention. Our focus is the (online) classroom, where faculty may want to draw students' attention to particularly informative or stimulating comments on course readings. However, our work has implications for scaling social annotation in many contexts.

We carried out experiments on a biology class at a large public university with hundreds of students making thousands of comments on course readings. Combining controlled experiments with interviews and surveys, we compare the impact of various spotlighting techniques on student reading and commenting activities. We show that spotlighting can effectively draw students' attention to where faculty want it and determine which spotlighting techniques have the strongest effect on attention. In particular, we explore the paradigm of *innotation:* leveraging the mutability of digital documents to embed annotations *in the content itself.* This offers a step towards empowering students and faculty to *collaboratively edit* course material to enhance its value to students. From interviews, we gain further understanding of what signal students get from spotlighting, as well as the impact of spotlighting on student motivation.

2 BACKGROUND AND MOTIVATION

Beginning January 2021, we conducted weekly meetings with stakeholders for discovery, need-finding, and critique [37]. Our stakeholders included three instructors and two teaching assistants (TAs) who used the social annotation tool NB (see section 2.2) in their high-enrollment biology class. Issues associated with managing and navigating the many comments dominated the interviews. The general feeling was that the volume of comments, while potentially rich with opportunity for engagement, were overwhelming to both students and faculty. Most connected this feeling to high numbers of lackluster comments and modest levels of peer discussion. Compounding these issues was the sense that when students attempted to engage with one another, that fruitful discussion and explanation could get "lost" in the sea of other, less attractive comments. One instructor expressed a desire to direct students' attention to the most instructive questions or discussions in the forum. Instructors also believed that directing students to read more of their peer's comments could contribute to knowledge construction and a more collaborative learning experience [47]. Giving instructors the ability to "highlight" instructionally valuable threads, thereby selectively elevating some comments/discussions, could be one way to achieve these aims.

2.1 General practices

For design insight, we explored and synthesized current annotation practices. Our goal was to learn organic ways users emphasize important parts of their notes on dense material. We examined videos on YouTube that teach material through annotation [53], collecting screenshots, comments,

and replies. Some videos focused on annotating a particular field (such as Biology or Literature), while others explained material through annotating and reading live with the viewers.
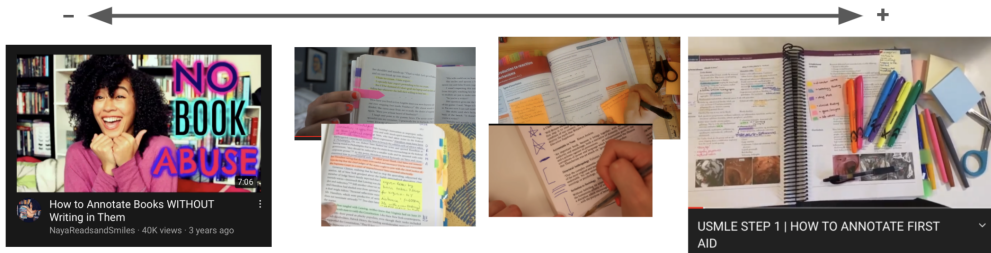

Fig. 1. A spectrum of practices in how users annotate

A thematic analysis [9] showed that users varied in their preference of the intensity of the annotation on text. Figure 1 shows a spectrum from learners who almost separate the annotation layer from the reading material to learners who use multiple highlighting techniques on the text. One challenge that readers have to work with is the restricting affordances of the material they are annotating [44]. Here are some approaches adopted by users to overcome small spaces when annotating text:

- Dedicate a space for notes in the book, or a separate paper away from the main reading
- Overlay the reading with sticky notes [36]
- Write on the margins and footnotes echoing findings by Marshall [44]
- Write between the lines
- Overlay a tape on words and lines and replacing them with others (e.g., word translation)
- Use symbols or colors with legends to explain the meaning [63]
- Combine two techniques or more [70]

We also looked at the affordances provided through other, digital, tools. For example, PDF reading tools provide sticky notes to overlay text. Other tools for text editing allow for switching modes or screens to read comments and notes. We also looked at free form annotation tools [16, 61] that allow the user to scribble anywhere on the document. In addition to that, previous researchers looked into ways to increase the affordances when reading electronic material by allowing users to shift text to annotate [3, 65, 75] using pen and touch techniques—this relates to the *innotation* technique we discuss below.

All the methods we've described show *all* annotations at all times. Using any of these techniques for all annotations in a high-enrollment course would quickly render the document unreadable, burying the text under the annotations. We reasoned, however, that we could use some of these approaches on a more selective basis to *spotlight* key annotations, as suggested by our stakeholders. The design section (Section 4.1) presents the different spotlighting designs that we developed in our design sessions with our stakeholders.

2.2 NB

We conducted our experiments by modifying the social annotation tool NB [78]. In the last decade, hundreds of instructors have incorporated NB into their classes. Our instructor-collaborators had
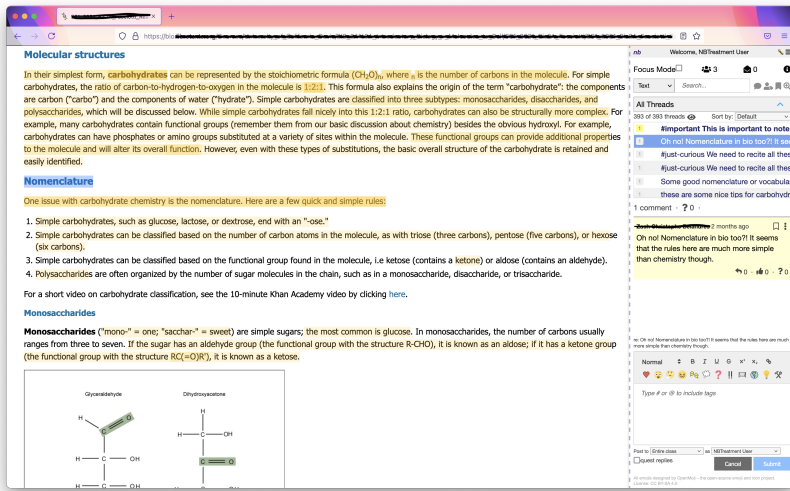
Fig. 2. NB: The NB interface occupies the entire right sidebar of the browser window. It contains the comment list, comment viewer, commenting box, and the controls for other features such as the filters and a search bar. The density of yellow highlighting in the text is proportional to the number of comments anchored on the highlighted text.

already used NB for six years in their own classes with over 10,000 students, and thus were familiar with the platform [1].

NB (figure 2) provides a sidebar on pages. A user can click-and-drag to select document text, then type a social annotation in the comment box in the sidebar. Students can post comments publicly, to the instructor only, or private to themselves. They can make public comments anonymous to their peers but not the instructor. Annotated text is *highlighted* in yellow. A list of first "header" lines of existing comments shows in the sidebar in document order. Clicking a highlight or a subject line expands the indicated comment in the sidebar along with all its replies. Each header line also shows a set of icons on the left side. For example, the yellow icon holds a number counting replies to a comment and the orange icon with the letter 'i' indicates a reply from an instructor. Above, NB provides filters on hashtags, authors, and reply requests. Users have the option to hide the highlights using the eye icon.

3 RELATED WORK

3.1 Discussion in the Classroom

Students' participation in online discussion could be driven by the desire to learn and/or to satisfy the requirements of a graded assignment [59]. An important aspect of students' participation in the classroom is their level of engagement. [24, 39]. Research suggests that collaborative discussion can help students develop greater understanding of course materials [6, 14, 17, 18, 42]. Even the simple act of reading other student comments appears beneficial [54]. As a result, some educators have incorporated online forums to facilitate online discussions [5, 12, 29, 46, 52, 62] and sought to find ways to encourage student participation. For example, Ghosh and Klein [27] studied how the balance between instructor and student comments influenced peer-to-peer communication. Others

---

[1]https://github.com/haystack/nb

have explored how instructor endorsements of comments or conversations influenced student participation [46, 52, 71].

3.2 Social Annotation

In addition to traditional forums that provide ways to discuss readings in separate environments, more recent work has explored *social annotation (SA)* systems that anchor discussions on selected (typically highlighted) parts of a document [19, 26, 34, 56, 78]. Anchoring discussions on a document makes those discussions salient at the right time—when a student is reading that content—and provides in-document context that can simplify the discussion. Marshall [44] categorizes SA into implicit and explicit annotations. She defines *implicit* annotation as simply marking up the text with highlights or drawing underlines. *Explicit* annotation adds new meaning and connections to the work by enhancing the annotation with additional text. SA tools can support both explicit and implicit annotations. Users of SA tools can annotate online resources, share knowledge, and socialize.

SA has gained popularity in education and has been shown to be useful for learning [8, 25, 26, 31, 32, 35, 78].Social annotations has also been found to encourage peer-to-peer learning [25, 34]. Critical thinking, meta-cognitive abilities, and reading comprehension are all aided by SA-based learning exercises [32, 56].

Early open source tools for social annotations include WebAnn, Hylighter and NB [11, 32, 78], which were designed for SA in educational settings. A more recent open-source tool, Hypothesis [23, 28], is designed for general SA on the web but has recently shifted focus to the learning domain. Even more recently, closed-source commercial platforms such as Perusall [49, 60] and Wattpad [1, 64] have emerged. These SA tools create anchored learning environments in which students' discussions are positioned immediately above or beside online resources [26].

Almost all SA tools use a *sidebar* to list the anchored comments. Some, such as NB and Hylighter, default to showing the sidebar at all times. Others show the sidebar on demand when users click on a highlight (Hypothesis, Perusall). While these tools all show the reading and the sidebar simultaneously, Wattpad darkens the reading when a comment is clicked, making it fade into the background and bringing users' focus to the sidebar alone. Some tools, such as Hypothesis, show each comment in its entirety; others such as NB and Perusall show only the first "header" line of each comment in a list, expanding to the full comment when the header line is clicked. Using header lines allows more comments to be packed into the same space. Showing full comments allows for reading comments without clicking but requires users to scroll more often, making it more difficult to skim (figure 3).

SA tools provide users with a variety of cues to indicate the presence and meaning of anchored annotations. For example, some SA tools use color intensities from overlapping highlights as a measure of the number of comments within paragraphs [2]. Tools such as Hylighter and Perusall use different colors to indicate the meaning of the highlights. Such color change is one of the spotlighting patterns we study below. In Perusall, yellow highlights correspond to threads started by students and blue highlights correspond to threads started by an instructor. In Hylighter, yellow denotes a reader's highlights and blue denotes other students' highlights (overlaps are green).

3.3 Scaling Social Annotation

A general challenge with online discussion, which is particularly relevant to SA in high-enrollment classes , is managing scale [68]. When hundreds of students annotate the same resource, the number of comments that students can interact with may become overwhelming and lead to disengagement.

Discussion platforms have employed various techniques to help manage scale issues in large online discussions. *Pinning* is the practice of fixing critical posts at the top of a forum's list of posts,

(a) Hypothesis

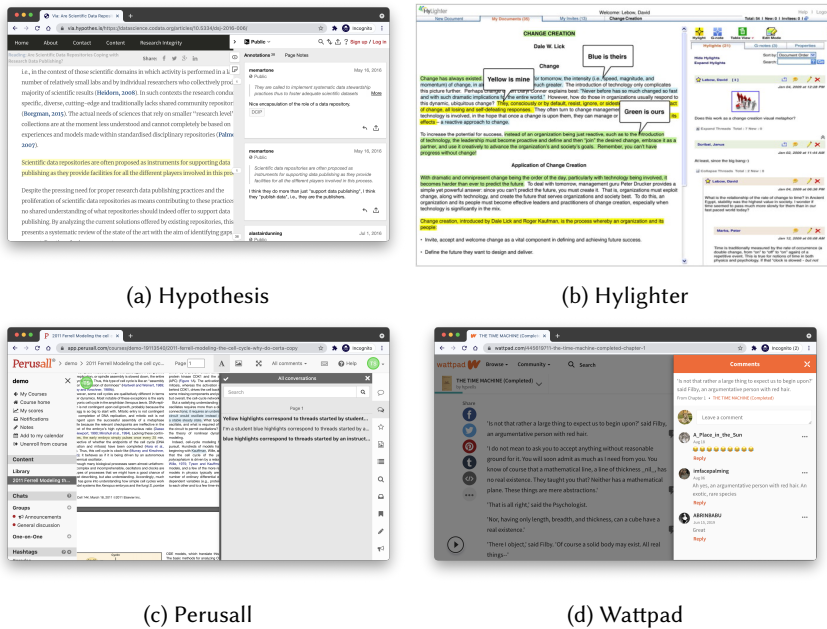(b) Hylighter

(c) Perusall

(d) Wattpad

Fig. 3. SA Tools

breaking the typically chronological order of posts. Various *social moderation* techniques have also been developed, such as reranking content by popularity (Reddit) or more complex factors (Slashdot). Instead of reranking, some forums (such as Slashdot) can also *filter* based on a score, hiding posts below a given threshold, thereby reducing visual clutter.

In an educational context, another natural alternative for managing issues of scale is to *section* students into smaller groups [50] and to show students comments only from their own section. This strategy reduces the volume of comments visible to each participant to a level that can be roughly controlled by choosing the section size. However, siloing students into discrete groups means that many valuable comments might only be seen by a small subset of the student population and that connections between students, which might have spurred productive discussions, don't happen. Each of these outcomes deprive students of potential learning experiences.

Social annotation can itself be seen as a scaling mechanism, since it anchors each comment to a specific part of the text, allowing the reader to focus only on comments germane to the text they are reading at a given time. But a sufficiently large number of comments will overwhelm this coping mechanism. To (further) scale social annotation, the traditional approaches of pinning or reranking are unappealing, since the goal is to situate each discussion in its assigned place in the document.

In this work, we explore a different design solution to help students deal with large numbers of annotations by increasing the salience of specific comments. In contrast to filtering, spotlighting allows us to draw student attention to specific comments *without* necessarily hiding any others.

3.4 Visual Design

More broadly, various techniques have been developed to increase the *salience*—visual prominence—given to particular display elements. Previous studies looked at how user engagement measures such as focus are affected by the saliency of relevant information [22, 48]. Techniques to increase

salience [20] include highlights, colors, and blinking [66]. Other techniques to draw users' attention to an element on the screen is by minimizing the visibility of the content surrounding that element [38, 76]. We studied prior work on the design of online advertisements to explore how different design features contributed to better comprehension and higher click-through rate [43]. Key variables include advertisement size, color, content, and animation. For example, [41, 67] found that increasing the size of the ad design increased the number of clicks. They also found that animation resulted in a quicker response by users. Another study [51] found that while banner ad size significantly increases the total number of clicks, it does not affect the click-through rates. More specifically, [55] found that the blue static ads with a size 300 × 250 generated more clicks than larger red ads or animated designs. Other work looked into how moderate customization of ads to match the native format of the hosting page is effective in grabbing more attention [33]. Finally, while ad designs are supposed to grab attention, [4, 43] discovered that overuse of salient cues leads to users developing banner blindness.

4 SPOTLIGHTS

In this work we refer to our proposed designs for directing attention in social annotation systems as *spotlighting*. We distinguish spotlighting from *highlighting*, the standard practice of placing a visual indicator (such as a surrounding box or a colored background) on the annotated part of the document.

4.1 Design Options

We begin our design exploration from the physical annotation and online SA tools summarized in the background and related work sections. We mixed brainstorming and prototyping with discussions with our stakeholders. We shared different spotlighting ideas with our stakeholders. Options included popup and overlay windows over the text, scribbles on the margins, and movable blocks. Some types show more text from the comments than the others, and some require clicking to view the comments while others don't. We quickly ruled out designs that *obscured* the content, such as overlaid sticky notes, as they are too disruptive to the core reading experience. What remained were the following high-level categories:

emphasizing highlights In systems that use highlighting to indicate the presence of a comment, we can change the highlighting in ways that distinguish it—for example, by changing its color.

sidebar marks Since many SA tools list comment "headers" in a sidebar, special state indicators can be added to some headers. In NB unread comments already get bold headers, questions seeking a response get a question-mark icon, and replies by the instructor get an "i" icon.

marginalia A comment can be spotlit by expanding it out of the sidebar into the margins of the document.

***innotation* (<u>**in**line</u> <u>**an**notation</u>)** If the document is mutable (as are HTML documents), we can shift content to make space to place a comment *in the document itself*. Innotation can be provided in a *block* (placing the comment in a rectangular region like a figure, near the annotated content) or *inline*—injecting the comment into the text immediately following the annotated content, like a parenthetical comment (we use a different font to distinguish it from primary text).

Most of these methods can be combined: a single comment can be simultaneously spotlit (or not) by a highlight emphasis, sidebar marks, and marginalia or innotation.
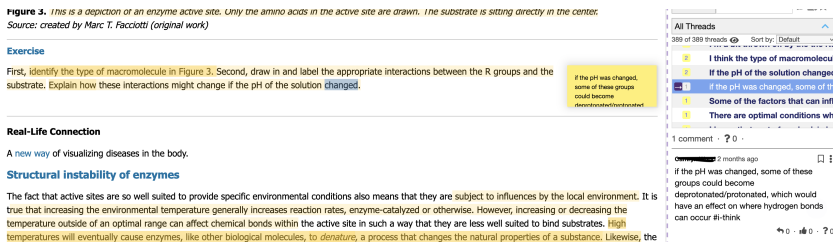
Fig. 4. Shorter Yellow sticky note [blocks] due to the position of the highlight in a narrow space. Only some of the text is shown

4.2 Design Choices

We aimed to compare multiple choices from the design space, but were subject to bounded resources for implementation and concerned about the power of our statistical tests if there were too many conditions. At the highest level, it seemed there were three truly different approaches to spotlighting: emphasizing the highlight, marking the header line in the sidebar, or placing the comment itself in view (marginalia or innotation).

Of these three approaches, marking header lines in the sidebar is essentially the same as the widely used approach [10, 77] of marking special comments in any traditional discussion forum—it leverages none of the special nature of social annotation. Spotlighting within the text itself is particular to social annotation systems and has thus been explored less.

We therefore chose not to focus on sidebar marks in our study, although we will have some observations regarding it from our study results. In fact, NB *already* uses a variety of "spotlighting" marks to the left of sidebar headlines—for example to indicate instructor comments and reply requests. Thus, we felt it would be inconsistent *not* to place spotlight indicators on these header lines. We therefore introduced spotlight icons for the sidebar to accompany the other mechanisms: ⬚ for blue inline innotations, ⬚ for block annotations, ⬚ for green highlights and ⬚ for marginalia on PDF documents. An icon is shown in figure 4 to the left side of the spotlit comment. These icons are always present for spotlit comments, not varied as part of our experiments on the effects of spotlighting.

Of the remaining annotation-specific approaches, marginalia and innotation seemed more similar to each other than either did to emphasizing highlights: both place the comment *content* somewhere near the highlight itself. Thus, in a pilot study (see Appendix C) on PDF documents (that cannot reflow), we compared highlight emphasis to marginalia (both with header line markers). Our main study involved a class that used HTML documents that can be reflowed [75], so we compared highlight emphasis to innotation. In particular, we chose three distinct spotlighting methods:

Yellow block notes (BLOCK). In this design, a yellow-background box is generated adjacent to the highlighted text and with the comment text inside it (figure 4. The yellow background recalls sticky notes, but we reflow the text instead of obscuring it as sticky notes do. To limit the perturbation of the page, the box is made no taller than the paragraph being annotated, with overly long comments truncated to fit (clicking the box expands the full comment to be read in the sidebar).

Inline blue script (IN). This design pattern was inspired by learners who annotate by writing between the lines. We insert the comment right after the annotated text (figure 6), reflowing the following text to make room. We designed this pattern to look like a handwritten note, making it different from the original content. Users prefer to recognize annotations as distinguished excerpts [44, 58]. As with block spotlights, to bound visual disruption we

We note that for bacterial photophosphorylation pathways, for each electron donated from a reaction center [remember only one electron is actually donated to the reaction center (or chlorophyl molecule)], the resulting output from that electron transport chain is either the formation of NADPH (requires two electrons) or ATP can be made but NOT not both *#lost I thought the movement of electron through the ETC couples of the movement of protons across t...* . The path the electrons take in the ETC can have one or two outcomes. This puts limits on the versatility of the bacterial anoxygenic photosynthetic systems. But what would happen if a process evolved that used both systems? More precicely, a cyclic and noncyclic photosynthetic pathway which could form both ATP and NADPH from a single input of electrons? A second limitation is that these bacterial systems require compounds such as reduced sulfur to act as electron donors to reduce the oxidized reaction centers, but they are not necessarily widely found compounds. What would happen if a chlorophyll $_{ox}$ molecule would have a reduction potential higher (more positive) than that of the molecular O_2/H_2O reaction? Answer: a planetary game changer.

Fig. 5. Inline spotlight at the end of a sentence within a paragraph. Hovering on the spotlight shows more text in place

Fig. 6. Inline spotlight with passive and hovered state

Since redox chemistry is so central to the topic we begin with a quick review of the table of reduction potential - sometimes called the "redox tower" or "electron tower". You may hear your instructors use these terms interchangeably. As we discussed in previous modules, all kinds of compounds can take part in biological redox reactions. Making sense of all of this information and ranking potential redox pairs can be confusing. We have developed a tool to rate redox half reactions based on their reduction potentials or E_0' values. Whether a particular compound can act as an electron donor (reductant) or electron acceptor (oxidant) depends on what other compound it is interacting with. The redox tower ranks a variety of common compounds (their half reactions) from most negative E_0', compounds that readily get rid of electrons, to the most positive E_0', compounds most likely to accept electrons. The tower organizes these half reactions based on the ability of electrons to accept electrons. In addition, in many redox towers, each half reaction is written by convention with the oxidized form on the left followed by the reduced form to its right. The two forms may be either separated by a slash, for example, we write the half reaction for the reduction of NAD^+ to NADH: $NAD^+/NADH + 2e^-$, or by separate columns. We show an electron tower below.

Fig. 7. Emphasized Green Highlight (EM)

define a maximum length and truncate longer comments. Students can hover to show more of the text and click to show the whole comment in the sidebar. This type of spotlight is the closest to content it annotates. The color choice was influenced by past work [55].

Emphasized Green Highlight (EM). Highlighting in green is a contrast to the yellow highlights available on NB. With this pattern, the annotated content is spotlit as opposed to spotlighting the comment itself (figure 7).

For tractability and statistical power, we chose to explore these methods in isolation, not in combination.

Our designs can create as many as *three* signifiers of a single spotlit annotation in a document: the highlight, markers on the header line in the sidebar, and the innotation or marginal comment. When the user hovers over any one of these, we highlight the others to show the connection, a technique known as brushing [45]. Clicking on any one of the traces expands the comment and reply thread in the NB sidebar.

5 THE STUDY

In this section, we describe a controlled experiment carried out to *quantify* the impact of spotlighting on students' experience while reading with NB. We designed a between-subjects experiment and multiple within-subject conditions [40, 57]. We then followed up our experiment with interviews and surveys. Some of the questions that we wanted to explore through our experiments were:

(1) Does spotlighting comments increase engagement with these comments, and how much?
(2) Which type of spotlighting yields more engagement?
(3) Are there groups of students that are more affected by the spotlights?
(4) Does the quality of the spotlit comments affect how students interact with them?

There were 26 readings assigned during the semester, generally four per week. For each reading, the instructor identified a small group of "high quality" comments that they would like to spotlight. We then spotlit these comments for *some but not all* students at random. Because the readings were HTML documents that could be reflowed, we used the greenlight and block- and inline-

innotation spotlighting methods. We then used the logs to measure how often different students clicked comments when they were and were not spotlit, and applied statistical tests to determine whether spotlighting led to a significant difference in click rate. As a control for instructor selection bias, we also applied the same sometimes-spotlight approach to *randomly* selected comments. We elaborate on our methodology, and on how we controlled for various other factors, below. Before deploying this study, we have deployed a pilot study that we describe in Appendix C.

5.1 Methods

NB was used in an undergraduate biology class at a public university in the United States. The study took place during the Summer term (from June 21, 2021 to July 30, 2021). This course is the standard introductory biology course from the academic year, but taught at double speed. Student commenting on NB is mandatory and graded. A minimum of 3 comments or replies by each student on each reading was required. Students had about four readings each week totalling 26 readings. Enrollment was 205 students. In the following sections we describe the experimental design and the interviews and surveys we conducted.

5.1.1 Experiment. For each reading, we monitored the accumulation of comments. When the total number reached a threshold of 50 (of about 600 expected in total), we notified the instructor that it was time to select comments for spotlighting. The threshold guaranteed that there would be enough comments in place to let the instructor select some they considered to be of high instructional value. At the same time, the threshold was low enough to ensure that there were still plenty of students—about 180—who did their reading *after* the faculty selected spotlights, so that students would encounter those spotlights during their reading. The instructor had the freedom to pick any comment they thought was worth spotlighting from the 50+ comments in a reading, and aimed to select at least 6.

Spotlights have multiple characteristics that could influence user behavior. We identified them as the independent variables in the experiment: type of spotlight, content of the spotlit comment, and familiarity with spotlights (whether students have been exposed to them before). To measure these effects, we first randomly chose a small *control group* that was shown no spotlights the entire term. After 6 lectures with no intervention to set a baseline, the remaining students were, in each lecture, randomly assigned to be treated by one type of spotlighting or to *no spotlighting at all in that lecture*. Each student got *one* type of spotlighting each lecture, to avoid an unpleasantly inconsistent user experience, but the type was randomized by lecture. Thus, each student randomly shifted through all the different spotlighting types over time. Note that the no-spotlight students differed from the control group because they had seen spotlights in other lectures—we wanted to understand whether this would affect their reading when no comments were spotlit.

Once we had assigned students to a spotlighting type for a lecture, we examined the effect of that spotlighting. For each student, we randomly selected some of the instructor-selected comments to be spotlit. Other intructor-selected comments were presented normally, unspotlit. Randomness meant that each instructor-selected comment was seen spotlit by some students and unspotlit by others. We fixed that decision, as well as the spotlighting type, for each student and comment so it would not change from reading session to reading session. We then compared click rate on each comment when it was in the spotlit versus unspotlit conditions. Table 1 shows the experiment design with the different conditions for students.

We speculated that some students might be influenced by the content quality (potentially biased by instructor selection) of spotlit comments. To address this issue, half the comments spotlit to each student were randomly chosen from the instructor's selections, and the remaining half were randomly selected from the other comments. This allowed us to compare the effect of spotlighting

Table 1. Experiment Design and Conditions

Was student ever exposed to spotlighting?	For a given lecture, was student exposed to spotlighting?	Spotlighting type	For a given comment, comment was spotlit to the student?	Number of (student, comment) pairs in the click dataset
No (Control)	-	-	-	112342
Yes(Treatment)	No	-	No	100656
	Yes	Inline	Yes	3238
			No	97391
		Block	Yes	3281
			No	99580
		Emphasis	Yes	3319
			No	98586

on instructor-selected comments versus typical comments. For each student we randomly chose to spotlight 6 or 10 comments in total for each reading.

Note that our experimental "subjects" here are not really the students in isolation but rather the "potential interactions" between a student and a particular comment. It is these student-comment *pairs* that we subject to different conditions to see how they affect whether a potential student-comment interaction becomes actual.

Various confounding factors exist on NB that could potentially affect students' attention. These include cues such as sidebar icons that indicate a reply from an instructor, a question mark icon to indicate a question that needs to be answered, and an icon with the number of replies. We describe how we handled these confounding factors in the results section. In the Discussion and Appendix we argue further that our experimental design allowed us to control for these confounding factors.

5.1.2 Interviews. After the class concluded, we interviewed six undergraduate students from the course. Each interview lasted 45 minutes and followed a semi-structured format. The main questions about their reading habits on NB and reactions to the spotlights were the same for all interviews. The first set of questions asked students about how and when they read and comment on NB. The second set asked them about the workload and nature of the reading tasks on NB. The final set was about the spotlights and their perception of this new feature.

5.1.3 Surveys. Based on the interview results, we designed a Google form survey for the entire class during the last week of the academic term and right after the deadline for the last reading assignment. The survey included questions about their commenting habits their perceptions of the reading annotation assignment, and the impact of spotlights, and requests for suggestions and feedback. The control group did not receive any questions about the spotlights (i.e., third and fourth sections).

5.2 Experiment Results

5.2.1 Analysis. Our experiment yields two datasets for two types of interactions that students had with the comments: click and reply outcomes. Each datapoint in the dataset consisted of a student and comment pair, whether the student interacted with the comment (e.g., whether the student clicked on the comment in the click outcome dataset), and the characteristics of the comment and the reading on which the comment had been posted at the time of capturing the interaction (e.g., the number of replies that the comment already had when the student replied to it, in the reply

outcome dataset). If a comment was posted after a student last interacted with the reading, we did not include that comment in the data for that student because the student was never exposed to that comment. These datasets contain only the first-time interactions of a student with a comment, for instance, the first time a student clicked on a comment and not their subsequent clicks on it. The click outcome and the reply outcome datasets both contained 518393 datapoints. There were 32276 positive click outcomes (the student clicked on the comment). Thus, the overall click rate on comments was 6.23%. The number of positive outcomes for the reply dataset was 2454, of which 1528 were first replies to a comment. For each reading, some students arrived too early to see spotlights, but over the semester all students were frequently exposed to them—details are in Appendix B.2.

To understand the effect of spotlighting on interaction outcomes, we first partitioned the data by the type of spotlighting that a student had received in a particular lecture. We first compared each spotlight condition with the control group that had never experienced spotlighting.

We built four partitions: **Block (Yellow block notes), Inline (Inline blue script), Emphasis (Emphasized green highlights), and None**, with None containing the data from those students that had not received any spotlit comments in a particular lecture but that nevertheless had seen spotlit comments in previous lectures. To analyze the effect of our treatments, we added the data from the students in the control group, who had never experienced spotlighting, to each of the partitions. Comparing the treatment and the control group in any of these new partitions would then inform us whether being exposed to any spotlit comments in a reading (being part of the treatment group) impacted a student's propensity to click on any comment within the reading, irrespective of whether the comment is spotlit.

In order to keep our model clean, we selected comments for *hypothetical* spotlighting in the **None** and **Control** conditions in the same way as the others. However, these hypothetically spotlit comments were not *actually* spotlit in these conditions. This made it straightforward for us to measure effects, by comparing the relative effect on clicks of sptlighting in the different conditions *including* **None** and **Control**

With this design, the *spotlit* variable in our dataset denotes whether a comment was *chosen* to be spotlit, but may or may not have been, depending on whether it belonged to a treatment group (Block, Inline, or Emphasis) or the controls (None and true Control).

Next, we fit the following model to the data in each of the four partitions:

$$\text{interaction outcome}_{sc} = \beta_1 1(spotlit)_{sc} + \left(\beta_2 1(spotlit)_{sc} \times 1(treatment)_s\right) + \delta_s + \gamma_c + \epsilon_{sc} \quad (1)$$

where the interaction $outcome_{sc}$ is whether the comment c was clicked by student s in the click dataset (or replied to, in the reply dataset). The indicator variable $1(spotlit)_{sc}$ denotes whether the comment c was marked as spotlit to student s, regardless of whether it was displayed as a spotlit comment. The model also includes the interaction between *spotlit* and *treatment*, with the indicator $1(treatment)_s$ denoting whether the student s belongs to the treatment or the control group. δ_s and γ_c are student and comment fixed effects respectively. The student fixed effect is correlated with $1(treatment)$ because, per our experiment design, when a student is assigned to one of the treatment or the control groups, they remain in that group forever and therefore a change in one of the student or treatment variables is correlated with a change in the other. Therefore, to avoid collinearity, we did not include the main effect of *treatment* as an independent variable in the model [2]. The coefficients of interest are β_1 and β_2. ϵ_{sc} is the error term.

---

[2]In fact, treatment is a comment-invariant characteristic and any explanatory variable that is constant across comments for all students gets swept away by the fixed effects transformation [74]

Table 2. Linear regressions predicting click outcome, with standard errors clustered on student and comment. In each of the partitions that include a treatment group (Block, Inline, and Emphasis) and the control group, the interaction between treatment and spotlighting is significant, indicating the effect of being assigned to the treatment group for the spotlit comments is to increase the click ratio.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	0.001	0.006	0.228	0.82
	spotlit × treatment	0.051	0.015	3.536	**<0.001**
$N_{datapoints} = 215,203$, $N_{comments} = 4,804$, $N_{students} = 164$				Adj. $R^2 = 0.11$	$RMSE = 0.23$
Inline vs Control	spotlit	0.001	0.006	0.090	0.93
	spotlit × treatment	0.073	0.015	4.926	**<0.001**
$N_{datapoints} = 212,971$, $N_{comments} = 4,804$, $N_{students} = 165$				Adj. $R^2 = 0.11$	$RMSE = 0.23$
Emphasis vs Control	spotlit	-0.002	0.006	-0.312	0.76
	spotlit × treatment	0.057	0.015	3.954	**<0.001**
$N_{datapoints} = 214,247$, $N_{comments} = 4,804$, $N_{students} = 166$				Adj. $R^2 = 0.12$	$RMSE = 0.23$
None vs Control	spotlit	0.003	0.006	0.422	0.67
	spotlit × treatment	0.007	0.012	0.607	0.54
$N_{datapoints} = 212,998$, $N_{comments} = 4,804$, $N_{students} = 163$				Adj. $R^2 = 0.12$	$RMSE = 0.23$

We used the function "feols" from the R package "fixest" to define the model. Although the lefthand side of the equation is a binary outcome, when reporting the main results, we use the function "feols", which is used for OLS regression, for ease of interpretation. In the Appendix, we show that we obtain consistent results when we fit a generalized linear model (logistic regression) to the data.

To account for correlated errors in click choices by one participant and for one comment, we clustered standard errors on student and comment [3]. Although it is possible that clicks vary systematically across readings (i.e., for some reason, the comments in a reading attract more clicks than the ones in another reading), we do not include reading fixed effects in the model because we already include comment fixed effects and there is perfect linear dependence between the two. In other words, comment fixed effects control for all characteristics of a comment c that do not change across students, including the reading to which the comment belongs.

5.2.2 Effects on Student Clicks. Table 2 shows the results of the regressions for each of the 4 samples in the click outcome dataset. In each of the treatment groups where students were shown spotlit comments, we see that **the interaction between spotlighting and treatment has a significant positive effect**. The coefficients for the interactions show that the effect of being assigned to the treatment group for spotlit comments (i.e., for spotlit comments to be displayed in any of the spotlighting fashions) was to increase the click ratio. Figures 8, 9, and 10 demonstrate this interaction term for each of the regressions on the partitions containing the treatment groups. Overall, the click rate on spotlit comments (averaging over the different types) increases from its baseline of 9.7% to about 15.4%.

The regression performed on the partition containing the control and the None groups suggests that having been previously exposed to spotlights (the None group) does not have a statistically significant effect on what comments a student chooses to click in a reading with no spotlights.

---

[3]Failure to control for within-cluster error correlation can lead to misleadingly small standard errors, and therefore misleadingly narrow confidence intervals, large t-statistics, and low p-values [15].
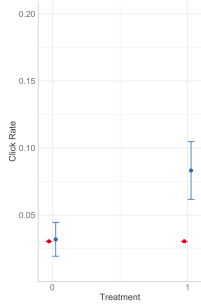
Fig. 8. Interaction plot of treatment and spotlighting in the partition of the data that includes Block type spotlighting and the Control.
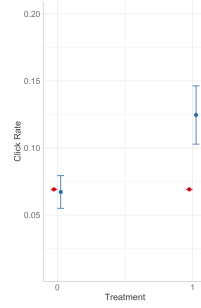
Fig. 9. Interaction plot of treatment and spotlighting in the partition of the data that includes Emphasis type spotlighting and the Control.
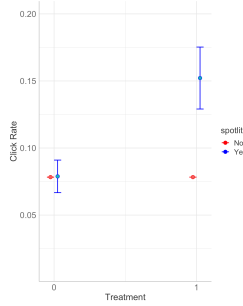
Fig. 10. Interaction plot of treatment and spotlighting in the partition of the data that includes Inline type spotlighting and the Control.

We determined that **all spotlighting treatments affect click likelihood**. To understand how they fare relative to each other, we built a model with click outcome as the dependent variable, and dummy variables for each of the 3 spotlighting conditions. Similar to the model described above, we also included comment and student fixed effects to account for unobserved heterogeneity between the different groups. We fit this model to those datapoints in the treatment conditions (Block, Inline, Emphasis, and None) where the comment was spotlit to the student—or simply chosen as spotlit but not displayed as such in the None condition. To account for non-independence of clicks by a user or on a comment, we clustered standard errors on student and comment. The model was as follows:

$$\text{click outcome}_{sc} = \beta_1 1(\text{Block treatment})_{sc} + \beta_2 1(\text{Inline treatment})_{sc} + \beta_3 1(\text{Emphasis treatment})_{sc} + \delta_s + \gamma_c + \epsilon_{sc}$$
(2)

We found that **the Inline (Inline blue script) treatment increases the click likelihood the most** ($\beta_2 = 0.07, SE = 0.01, p < 0.001$), while the **impact of the Block and the Emphasis treatments are similar to each other** ($\beta_1 = 0.05, SE = 0.01, p = 0.002$ for Block; $\beta_3 = 0.05, SE = 0.01, p = 0.009$ for Em).

Table 3. Linear regressions predicting reply outcome, with standard errors clustered on student and comment. Conditioned on a comment being clicked on, its assignment to a treatment condition has no statistically significant effect on the probability that it is replied to.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	-0.009	0.012	-0.746	0.46
	spotlit × treatment	0.009	0.014	0.641	0.52
$N_{datapoints} = 13,191, N_{comments} = 3,333, N_{students} = 157$				Adj. $R^2 = 0.02$	$RMSE = 0.14$
Inline vs Control	spotlit	-0.005	0.013	-0.367	0.71
	spotlit × treatment	0.011	0.015	0.714	0.48
$N_{datapoints} = 13,114, N_{comments} = 3,361, N_{students} = 153$				Adj. $R^2 = 0.03$	$RMSE = 0.14$
Emphasis vs Control	spotlit	-0.010	0.013	-0.812	0.42
	spotlit × treatment	0.008	0.016	0.529	0.60
$N_{datapoints} = 13,916, N_{comments} = 3,342, N_{students} = 155$				Adj. $R^2 = 0.03$	$RMSE = 0.14$
None vs Control	spotlit	0.001	0.001	-0.879	0.38
	spotlit × treatment	0.002	0.002	0.990	0.32
$N_{datapoints} = 212,428, N_{comments} = 4,804, N_{students} = 163$				Adj. $R^2 = 0.00$	$RMSE = 0.04$

5.2.3 Spotlight Effect on Student Replies. We wanted to understand the effect of spotlighting on student replies. To remove the possible confounding effect that previous replies can have on a future reply (the presence of previous replies on a comment could conceivably encourage or discourage future replies), we filtered the reply outcome dataset to only include the first replies to a comment. We also limited the dataset to those student and comment pairs where the student had clicked on the comment, to understand the probability of reply to a comment conditioned on it having been clicked on. We used the same model as outlined in the previous section with reply outcome as the dependent variable and fit the model to each of the 4 partitions (The combination of the 3 treatment conditions as well as the None condition with the control group) separately.

Table 3 shows the results of the regressions for each of the 4 partitions. We observe that conditioned on a comment having been clicked, the treatments have *no statistically significant effect* on the reply outcome. However, because we saw in the previous section that the treatments have a positive effect on clicks, they result in an increase in the reply likelihood as well, although this effect is driven entirely by an increase in the click rate[4].

5.2.4 Spotlighting affects different reading styles differently. We performed exploratory analyses to understand what types of students are most influenced by our treatments. From the logs, we noticed that some students read the discussions mostly by clicking through the comments on the sidebar where the salience of spotlighting is low. On the other hand, other students mostly read the discussions by clicking through the highlights on the page where spotlit comments are visibly differentiated. We therefore, built 2 partitions, one with students who performed 60% or more of their clicks on the sidebar, and one with students who performed 60% or more of their clicks on the highlights from 6 readings from the pre-experiment week before introducing the spotlights. This means the partitions were determined *before* any changes in behavior brought on by spotlighting.

We then performed the same analyses as above for measuring the effects of spotlights on each of the two partitions separately. The results show that the **spotlighting treatments significantly**

---

[4]Bayes' theorem: $P(\text{reply}) = \frac{P(\text{reply}|\text{click}) \times P(\text{click})}{P(\text{click}|\text{reply})} = \frac{P(\text{reply}|\text{click}) \times P(\text{click})}{1}$

affect the students who click through the highlights. These effects, however, are **not significant for those students who mostly click through the comments on the sidebar**. This makes sense, as our spotlighting designs increase the salience of spotlit comments *in the document*, not in the sidebar. The full analysis is included in the Appendix Section A.

This analysis provides a post-hoc justification for our decision (Section 4.2) to focus on varying designs for in-document spotlighting by including sidebar spotlighting in all our designs, even though we could have disabled it to explore more designs. Since in-document spotlighting affects *readers of the document*, while header-line spotlighting affects students *skimming the headlines*, these two design elements can be studied independent of one another.

5.2.5   Quality of Comments. We wanted to understand whether the quality of comments influenced the impact of spotlighting. The results reveal that the effect of spotlighting on low-quality comments is not much different from on high quality ones: in both cases, spotlighting caused a substantial increase in click rate. The full analysis is included in the Appendix Section B.3

5.3   Interview Results

We analyzed interviews using thematic analysis [9]. The initial analysis process of the interview data started by identifying themes through multiple iterations and peer debriefing [21] with our collaborators from the large public university using an inductive approach [7]. Our peers in the debriefing sessions were two instructors and two teaching assistants who taught the class before and were familiar with NB, making them knowledgeable peers to ensure the reliability of our findings. The following are the main themes we identified, which helped design the survey we distributed to the class to elicit more data.

5.3.1   How students perceived spotlights. In this study we did not tell the students any particular reason comments were spotlit. We asked students about what they thought the spotlights meant; for the most part students either thought the spotlights meant important comments or instructor-endorsed comments. For example, a student mentioned: "It [spotlighting] is definitely useful, since it points out the important part that students should focus on." Another student reported: "I assumed the instructor selected the earlier ones in the week." Another student appreciated the quality of the spotlit comments, despite the fact that 50% of those comments were randomly selected: "the [spotlit] comment had more thoughts and they were better comments." Another student also added: "spotlights give comments credibility, I know they're picked by the instructor." A student elaborated: "green highlight is helpful, I want the instructor to tell us if it's an important paragraph." This last comment demonstrates a drawback of highlight=emphasis spotlighting (as opposed to innotation): it may suggest to the student that it is the *highlighted text*, rather than the comment on it, that is important.

5.3.2   Preferred method of spotlighting. Some students appreciated specific types of spotlighting patterns. For the yellow sticky notes, a student expressed: "I like the yellow postit, I like it shows text in longer format." Another student also mentioned: "I think that it's better than highlighting the actual passage as it doesn't seem to be that distracting." One student mentioned that she did not appreciate the number of the highlights on the reading, and she preferred having comments in the blue inline script instead: " I like this [blue inline script], I don't like like the highlights they're overwhelming." Another student suggested combining two types of spotlighting depending on the density of highlights in the paragraphs as they both tend to be helpful to her: "I prefer the yellow post-its. But when there are a lot of comments I would like to have the post-it combined with the green highlights."

5.3.3 Number of spotlights. Students in the interview appreciated the small number of comments being spotlit in the experiment. One student was specific: "I prefer total of 5 spotlit comments to feel they're important and to not lose their locations in the text."

5.3.4 Seeing one's comment as a spotlight in the reading. We asked students how they feel about their comments being spotlit. All students in our interviews appreciated the idea: "If I see my comment spotlit it'll make me proud specially with online learning, it feels nice that the instructor recognized me!" Other students also expressed how validating it is to have her comments spotlit.

5.3.5 Using Spotlights to deal with volume. Students reported reading spotlit comments helped them manage the load: for instance, one student stated, "I make sure to read those flagged comments, post-it or bolded, on the page. I then skim the sidebar quickly, the subjects lines," referring to the preview of comments provided in the sidebar of NB. Another student reported that "when there is a lot number of comments or I'm in a rush, I read the comments put on the side" (margins and blocks). The same student elaborated: "I think that it's [spotlights] better than highlighting the actual passage as it doesn't seem to be that distracting." Another student mentioned looking at the spotlights for the conversation that happens under them and that she would like to see an indicated on spotlights for the number of replies "I always read those [spotlights], they have replies that help me, but it doesn't show the replies."

5.3.6 Replying to Spotlit Comments. We asked students if they reply to these comments. 5 out of the 6 students reported that they tend to shy away from replying because they feared showing a lack of knowledge on such a highly visible comment. While spotlights were effective in directing students' attention to comments and clicking on them as we have shown in our experiment section it seems that other social factors affect replying.

5.4 Survey Results

Based on the interview responses, we designed and circulated a survey to the class to gather more information. We received a total of 38 responses to the survey (treatment=29, control= 9). According to our collaborators, a 15% response rate is expected given the timing of distribution (end of a Summer term). We gained interesting insights from these responses that helped us explain some of our findings from the experiment.

5.4.1 Perception of Load. We asked students about their perception of the number of comments they are presented with when they do the reading. On a Likert scale, 11 students indicated 5, which is "far too many to read," 15 reported 4, 6 reported 3 and only 4 students indicated 2, which is close to "very few."

22 students reported that they read the comments through clicking on the highlights on the text as opposed to reading the comments through browsing the list on the sidebar. 7 students reported that they read the whole article then go back to interesting areas and click on the highlighted text to read comments.

5.4.2 Spotlights. All students in our treatment pool (n=29) encountered spotlit comments in their readings. For each spotlighting pattern we asked students the following:

- How noticeable is this spotlighting pattern? (see table 4)
- How distracting is this spotlighting pattern? (see table 4)
- How likely are you going to read a comment that was spotlit using (this) method? (see table 4)
- How do you like this spotlighting pattern? (see table 4)

Table 4. Each Spotlighting pattern on Likert scale (N= 29 Students)

Question	Treatment	1	2	3	4	5
How noticeable	Green highlight	1	4	4	4	16
	Yellow block	2	0	4	6	17
	Inline script	1	2	0	9	16
How distracting	Green highlight	10	7	8	4	0
	Yellow block	4	10	10	2	2
	Inline script	6	4	3	7	9
Likelihood to read	Green highlight	1	1	7	10	10
	Yellow block	1	1	8	5	13
	Inline script	2	4	6	6	11
How do you like	Green highlight	0	1	15	7	6
	Yellow block	0	0	13	5	10
	Inline script	5	5	12	4	3

Then we asked students an open ended question about anything they would like to tell us about each spotlighting method. We list their answers below.

Green highlights.
In the open ended question for this spotlighting pattern, some students expressed how they liked this method: "The color contrast demonstrates the importance of the comments and pulls my attention to it." Another student also appreciated the contrast: "I think it is a great contrast against the yellow highlights that have already been made." A student elaborated, "It makes the highlighted portions very noticeable and more likely for me to read." One of the students mentioned how they focus on them more "since it pops out more." Other students gave suggestions to make the pattern less distracting for their experience or more customized. For example, one student suggested "Additional colors" for the highlights, which echos some of the suggestions we received from our interviews.

In-line blue script.
Some students expressed how "subtle" this spotlighting method was: "it was subtle text but still crucial." Another student suggested: "I think the spotlighting method is not super noticeable. I think having it blue but a bit larger font will help emphasize that particular section." Another suggested a better place for this pattern based on her experience: "It would be better to have the blue inline type comments at the end of a paragraph, so as not to distract from the reading itself."

Yellow blocks.
Most of students' reactions were positive: "I liked the post it note annotations best because they were least distracting when trying to focus on the actual text. They were easy to read as an aside." Another said "The post it notes were a bit confusing at first, but ultimately was a big help." Another student mentioned that they liked the online reading experience with the block innotations: "I really like the post-it addition. Although the reading is online, it helps me understand that section of reading further."

5.4.3 Comparing spotlighting patterns. We asked students about the most noticeable and most distracting patterns, 51.7% reported green highlighted text as the most noticeable and 64.3% reported the blue inline script as the most distracting. We then asked students about their preferred spotlighting method (students could choose more than one pattern). Figure 11 shows students responses. We followed up by asking participants why they picked their preferred method. The following is an aggregated list of their responses:

- For yellow post-it notes, one student expressed that this pattern serves as a hint from the instructor "The yellow post it boxes seem more crucial because they kind of seem like a teacher is hinting at something more important." Another student liked how they fit into the reading "I prefer the post it boxes because it is the clearest presentation- easy to read and captures the reader's attention the best." A student expressed that it plays nicely with all of the highlights on the reading by being not too distracting "It's the least distracting from the bunch." Another student appreciated the neat experience provided with yellow sticky notes: "post-it boxes: separate the comments with the reading content, which make the reading more neat and comments are clearly expressed."
- For green highlights, a student mentioned that they like how noticeable they are: "I prefer the green highlighting because I think it is more noticeable." Another compared how it is easier to read than other methods: "I think the highlight is the easiest to read and having the text be blue or post it notes on the screen can be distracting when going through the article." Other students also added: "Subtle, but still distinct enough from the other, standard highlights to indicate that it is a spotlight pattern." Another student suggested a combination of the green highlights and the yellow sticky notes: "I like the green highlight as it pops out of the reading and I like the post it as it separates the important parts of paragraphs."
- Finally, for the inline blue script, students liked how it fits nicely with the reading "The blue letter is better than highlighting in just another color." Another student added, "they set apart the text/comment from the normal yellow highlights." One student mentioned that they like how it is in the right location close to the reading compared to the other methods: "The post it is on the side of the page which makes me lazy to read. The blue inline script is directly beside the text so it's more noticeable and easier to read." One student picked all spotlighting patterns as his favorite but elaborated with his comment: "I think in text methods are great [blue inline script]! However, I think for students who like to read the entire reading on their own first may be distracted. The post-it or the green highlighted can help them identify that this is additional and they can read the additional notes after. It can help them identify comments easier."

5.4.4 Number of spotlit comments. Almost all students (28 out of 29 students) indicated that they are likely to read spotlit comments. When we asked how many comments they think should get spotlit to make the reading more helpful but not distracting, most students reported numbers from 3 to 10, which was within the number of spotlights presented to students in the experiment. One student reported: "I found that the amount of comments that were spotlit in the class seemed adequate. I don't know exactly how many that is, but the amount that I saw seemed appropriate; not too many or too few."

Alternatively, some students did not suggest a number; instead, they provided a rule-based approach to which spotlights should be shown. For example, one student suggested only showing important and recurring ones: "Only the most crucial concepts and ones that multiple people seem
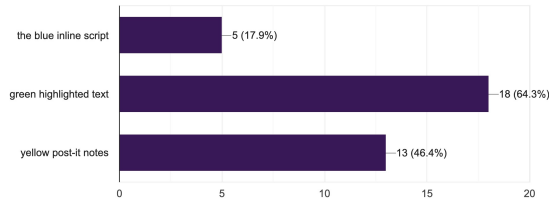
Fig. 11. Students answers to the question: Which spotlighting method(s) do you prefer?

to question." Another student suggested, "I think comments that the instructor replies to since those tend to be more important/informative." Another student suggested "maybe you should do it based on replies or upvotes or something." Finally, a student suggested a number dependent on the reading length: "I think it should be limited to 1 or 2 per section of the reading."

5.4.5   How spotlighting students' comments made them feel. All the open-ended responses to this question were positive. One student mentioned how validating it is to see their comment being spotlit: "it would probably feel validating—that what I was contributing was important." Another student also added, "it makes me feel that I'm on the right track to grasp the material. A good confidence boost." One student described that it would make them feel "honored" to see their comments being spotlit. Another student expressed: "I don't know if any comments were spotlighted. However, if any were I would feel proud and happy that my comments were in some way helpful."

While for the most part students responded positively to their comments being selected, some felt differently. A student mentioned that seeing their comments being spotlight put them "on the spot": "it would make me worry if I "miswrote" something when I was thinking, which would make someone misread the comment." The same student mentioned that even replying to spotlit comments is intimidating: "replying to spotlit comments is intimidating as my replies might be seen more." This suggests we may want to add a "don't spotlight me" option for students in the class.

5.4.6   Usefulness of spotlights to learning as reported by participants. As a final question to the spotlights section in the survey, we asked students if they think it is useful to spotlight comments and why. 28 out of the 29 responses reported spotlighting as useful. We list examples of where spotlighting was useful, as reported by the students:

- Studying: A student reported: "It's useful for a student who wants to understand what they are reading and allow them to broaden their mindset on that specific topic." Another student mentioned: "I think it is very useful to spotlight comments to help students further understand material and bring their attention to comments other students had. These comments may be something they are thinking of as well and it can be very beneficial." Another student emphasized: "I think it helps highlight the important comments that will help students more often and are more relevant to future content." Students also mentioned the spotlights were useful for comprehension and identifying important parts in the reading to focus on.
- Discoverability of relevant comments: Some students reported that spotlights helped in browsing comments by reducing redundant comments and managing the large volume of

comments in the reading: "I think it's very useful. It makes those comments more easily accessible, more likely to get replies, and could emphasize important points in the reading." Another student elaborated on how this helped them focus: "it can allow us to focus on useful/important comments instead of browsing through all fifty or so comments- which is very time consuming." Another student mentioned that it helps to spotlight redundant or recurring questions: "Yes, I think if there is a reoccurring question or concept that keeps getting brought up it is more beneficial and efficient to highlight one of those comments as opposed to have multiple different threads of the same concept." Another student reported that they appreciate increasing the salience of some comments to get more reads by students: "It is important to highlight important comment in order for other students to read them."

- Sparking discussion: Spotlights were not just useful for reading, but also for discussion: "It highlights strong comments that would be a useful read and potentially sparks conversation."
- Enhance the reading material for personal use: A student added that they would like to select the comments they wanted spotlit "It's based on someone else's (or the algorithm's) opinion on why the comment is important. I would like to read it myself and verify that it's useful to me. This can be frustrating when you understand much of the material but continuously run into repetitive and basic strings of information." This student appears to be imagining spotlighting to be like a personal bookmarking functionality. NB offers personal bookmarking, and it is an appealing idea to spotlight each students bookmarks to them. However, this can only be done *after* a student reads the comment, so would not help direct them to important ones initially.

5.4.7 Spotlights encourage students to read comments. We asked the participants (control and treatment) about their suggestions on how to encourage students to read each other's comments. Interestingly, students in the control group suggested spotlighting comments without ever having seen it (as they have reported in the survey). For example, one student suggested that the instructor star or mark relevant threads: "Have the instructor 'star' or mark certain threads of conversation that are especially relevant to the class so that students are more likely to read those threads." Another suggested spotlighting students' questions: "I would suggest that making the questions that students are asking highlighted in different colors, so people can clearly see them." A different student in the control group suggested getting students to read closer to the text. That student elaborated: "For example, after a section of text, there could be several questions that works as checkpoints to make sure the students are understanding the content."

6 DISCUSSION

This work has introduced and begun to explore a new functionality in social annotation platforms. *Spotlighting* increases the salience of particular annotations in order to draw a reader's attention to them. According to related work, such as the case study by Gao and colleagues [25], learners found that a large number of comments was a challenge, distracting to navigate on a social annotation tool. Spotlighting offers an approach to dealing with scale problems in social annotation systems, by surfacing important comments from a sea of less-relevant ones. Compared to filtering (itself relatively rare in annotation tools), which *hides* less important content, spotlighting represents a better application of the Focus+Context visualization paradigm [45] since it preserves that "context" of other annotations while directing "focus" to the spotlit ones.

In this work, we focused on very basic questions of spotlighting: what is the design space for spotlights, and what is the raw impact of spotlighting.

From the design perspective, we observed how social annotation systems offer a much richer set of spotlighting affordances than traditional discussion platforms: in addition to spotlighting

comments in a list (the sidebar), there are many ways to spotlight comments *in the document being discussed* by mutating the highlight, or by placing comments beside (marginalia), over (sticky notes—though we did not test this approach) or even inside (innotations) the document content. In contrast to sidebar spotlighting, placing spotlights in the document increases their salience for users *at the moment they are reading the annotated content* which is often exactly when you want those annotations to be noticed.

To study the raw impact of spotlighting, we designed controlled experiments that provided evidence that spotlighting significantly increases the chances that readers will notice and read a comment. We showed how the impacts of several different spotlighting designs compared, and gave evidence that the impact is stronger for readers who arrive at comments through the document than for students who explore comment through the sidebar list. Our instructors confirmed that spotlighting is a capability they desire in the classroom social annotation platform, as it helps them direct students attention where they think it ought go. Students also appreciated spotlighting; they liked having their attention directed by instructors, and at least some found it validating and motivating to have their comments spotlit.

Although we studied spotlighting in the context of SA for education, nothing about our techniques depends on this specific setting. Our designs would transfer directly to any other setting in which (any number of) users need to cope with a large number of annotations. This includes commenting for collaborative editing in tools such as MSWord or Google Docs. These tools currently use similar affordances to the SA tools—comments in the margins or, for scalability, with header lines in separately scrollable sidebar. It might also include citizen science or citizen journalism platforms which collect annotations from large numbers of participants—such as the MPs expenses project [5]. In any of these situations, there naturally arise needs to prioritize some annotations above others for attention. This may be based on global considerations of general importance, or may spotlight different annotations to different users based on relevance—such as mentions or replies to that user.

6.1 The User Experience

Existing social annotation systems must address the question of how to visually signal the association between a highlighted text and its comment in the margins or sidebar [30, 38, 69, 73]. One common approach is to maintain vertical alignment between these text and its annotation as one signal of connection. Brushing (increasing salience of the comment when the highlight is hovered, and vice versa) and linked browsing (scrolling to the comment when the highlight is clicked, and vice versa) are other common association mechanisms. Innotations offer a more direct indication of association, as they are embedded in the document directly beside their associated content. The strong performance of innotation in our experiments suggests that this technique, which has not been widely studied or deployed in SA systems, deserves additional attention.

Our study results yielded some findings that are consistent with prior work in the field of visual design and advertisement. For instance, the results showed that blue inline script received the highest clicks. We speculate that the blue script received the highest number of clicks due to the effective message design [41, 51]; its blue color [55], and the partial comment that it shows in place. More importantly, the spotlight design is also close to the native text of the document, which is an effective strategy to gain users' attention and trust in foreign UI patterns that are not part of the page [33].

On NB (and other SA tools mentioned previously) users could view comments either by clicking on the highlights on the document or by browsing through the list in the sidebar. Our analysis shows that **students who tend to interact with the document (highlights) directly are more**

---

[5]https://onlinejournalismblog.com/2009/06/19/the-guardian-build-a-platform-to-crowdsource-mps-expenses-data/

likely to be affected by the spotlights. The majority of the students in our experiment tended to interact with annotations through the document and thus be more strongly affected by the spotlights.

In addition, our analysis showed that because spotlights have a positive effect on clicks, they result in an increase in the reply likelihood as well. However, from our interviews with students, we sensed that many other factors play into the task to reply. One factor could be the nature of the task; students are required to post three comments on multiple readings every week. Some students felt that it was overwhelming and that they would rather write good comments about things they know rather than reply to make sure they score a good grade. Another factor could be confidence, as some participants in our interview reported that they feel shy or not confident enough to provide answers to other students which prevents them from replying.

6.2 Spotlighting Considerations

One of the first things to consider when spotlighting is the *type and design of the annotated document*. For PDF documents that have a fixed presentation (see Appendix), we focused on spotlighting by changing highlight colors and placing comments beside the document. In contrast, the reflowing possible with HTML documents permitted to explore *innotations* that embed new (comment) material directly in the document itself.

Drawing attention to something often means drawing it away from other things. It is important to consider how strongly we want to steer students away from unspotlit comments. We opted for static color changes and content embeddings and avoided using animation [41, 67] for the spotlights. Our goal was to bring comments to students' attention to increase engagement, not to excessively distract them [48].

Our designs only focus on spotlighting a few comments and how that would affect the reading experience. Our results show that spotlighting only a few comments, can work in getting students' attention to read and interact with those comments. While we have not tested the limits of spotlighting, we tested spotlighting 50 comments in one reading in our pilot study (see appendix C). That did not positively affect the click rate, and students did not appreciate the experience. Further work needs to be conducted focusing on testing the bounds on the number of spotlights. Almost all students (37 out of 38 students) liked the number of spotlit comments in our experiment (6 to 10 comments). However, this number could change depending on many factors, including the length of the reading and the density of the document or discussion (annotations).

We did not combine different spotlighting types in our study, as we wanted to understand the effect of each type without confounding our designs. However, we believe that **combining two types of spotlights** creates even more opportunities. For example, combining the green highlight approach with innotation would presumably provide an even stronger spotlight than either alone. Combining methods could also solve some problems that individual methods have. For example, we might want to use the yellow blocks approach, which stands out most strongly, when spotlights are sparse, switching to inline annotations as they become denser and finally to green highlights when they are exceedingly dense. Alternatively, one could use the inline text at the end of titles and paragraphs and inline blocks for self-contained comments. It is worth recalling our results though, that some spotlighting methods have stronger effect than others. While inline spotlights had the highest effect on reading, students from the interviews reported that the green highlights were the most noticeable, an important point to consider when mixing different types.

6.3 Spotlighting effect on the authors of the comments

Spotlighting had a mostly positive effect on the authors of the spotlit comments. Students reported that it is validating, motivating, and gives them a sense of pride for contributing to the content.

Students felt a sense of agency towards the reading material as their comments are integrated into the text as opposed to being treated as a marginal comment. This opens a new design space for SA in which readers suggest edits in the form of inline comments integrated within the document. It is also important to allow users to opt-out from getting their annotations spotlit. As we saw from our results, some authors might feel pressured seeing their comments spotlit and integrated into the document. In a way, some users prefer keeping a low profile in the discussion, and the design should respect that.

6.4 Limitations and Future work

Some spotlighting patterns such as marginalia, yellow blocks and inline blue script provide a longer preview of the comment. For that reason, we think that students might have read those comments without clicking. We were not able to log such passive reading interactions in our experiment. We suggest eye-tracking as a way to address this limitation. However, this limitation only makes our results *more conservative* as some innotation spotlights may have been read (increasing the rate for spotlit comments) without being able to detect it—meaning our estimate of the increased reading rate for spotlit comments may be an *under*estimate.

Some spotlighting methods might not work for users with vision problems. For example, colorblind users might not be able to alternate-colored highlights. Other methods such as innotation might serve better. One way to address this challenge is to let users choose the spotlighting method to accommodate their needs.

Due to the huge number of highlights, some students tend to turn off highlights and end up not seeing the spotlights. In the future, we are planning to test the effect of keeping spotlit comments regardless of whether highlights are disabled in the document.

We plan to explore different drivers for spotlighting besides instructor choice. For example, spotlighting recent comments or comments that contain replies to the reader.

7 CONCLUSION

Through our collaboration with instructors, we learned that they seek ways to bring students' attention to relevant or interesting comments on course readings. This paper investigated the design space and influence of spotlighting specific comments to direct students' attention to them.

We conducted experiments on a biology class at a large public university with hundreds of students making thousands of comments on course readings. We compared the impact of various spotlighting techniques on student reading and commenting activities through controlled experiments, interviews and surveys. We demonstrated how instructors might use spotlighting to successfully bring students' attention to where they want it. We also discussed which spotlighting tactics had the most powerful attention-getting effects. We found that inline blue script increased the click likelihood the most. From our qualitative data, we learned that students appreciated spotlighting. We also saw that spotlighting comments had a positive impact on students' motivation.

ACKNOWLEDGMENTS

REFERENCES

[1] 2015. Wattpad - Where stories live.   https://www.wattpad.com/
[2] 2020. NB.   https://nb2.mit.edu/

[3] Maneesh Agrawala and Michael Shilman. 2005. DIZI: A Digital Ink Zooming Interface for Document Annotation. In *Human-Computer Interaction - INTERACT 2005 (Lecture Notes in Computer Science)*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer, Berlin, Heidelberg, 69–79. https://doi.org/10.1007/11555261_9

[4] Michael J. Albers. 2007. Information salience and interpreting information. In *Proceedings of the 25th annual ACM international conference on Design of communication (SIGDOC '07)*. Association for Computing Machinery, New York, NY, USA, 80–86. https://doi.org/10.1145/1297144.1297163

[5] Martin A. Andresen. 2009. Asynchronous Discussion Forums: Success Factors, Outcomes, Assessments, and Limitations. *Educational Technology & Society* 12, 1 (2009), 249–257. Publisher: International Forum of Educational Technology & Society.

[6] Charles C. Bonwell and James A. Eison. 1991. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183 ($17. https://eric.ed.gov/?id=ED336049 ISSN: 0884-0040.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Place: United Kingdom Publisher: Hodder Arnold.

[8] Monica Brown and Benjamin Croft. 2020. Social Annotation and an Inclusive Praxis for Open Pedagogy in the College Classroom. *Journal of Interactive Media in Education* 2020, 1 (May 2020), 8. https://doi.org/10.5334/jime.561 Number: 1 Publisher: Ubiquity Press.

[9] Emeline Brulé. 2020. Thematic analysis in HCI. https://medium.com/usabilitygeek/thematic-analysis-in-hci-57edae583ca9

[10] A. J. Bernheim Brush, David Bargeron, Jonathan Grudin, Alan Borning, and Anoop Gupta. 2002. Supporting interaction outside of class: anchored discussions vs. discussion boards. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL '02)*. International Society of the Learning Sciences, Boulder, Colorado, 425–434.

[11] A. J. Bernheim Brush, David Bargeron, Anoop Gupta, and J. J. Cadiz. 2001. Robust annotation positioning in digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*. ACM Press, Seattle, Washington, United States, 285–292. https://doi.org/10.1145/365024.365117

[12] Elizabeth J. Burge, Daniel Laroque, and Cathy Boak. 2000. Baring Professional Souls: Reflections on Web Life. *Journal of Distance Education* 15, 1 (2000), 81–98.

[13] Loraine Busetto, Wolfgang Wick, and Christoph Gumbinger. 2020. How to use and assess qualitative research methods. *Neurological Research and Practice* 2, 1 (May 2020), 14. https://doi.org/10.1186/s42466-020-00059-z

[14] Ricardo Caceffo, Guilherme Gama, and Rodolfo Azevedo. 2018. Exploring Active Learning Approaches to Computer Science Classes. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 922–927. https://doi.org/10.1145/3159450.3159585

[15] A Colin Cameron and Douglas L Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of human resources* 50, 2 (2015), 317–372.

[16] M.A. Chatti, T. Sodhi, M. Specht, R. Klamma, and R. Klemke. 2006. u-Annotate: An Application for User-Driven Freeform Digital Ink Annotation of E-Learning Content. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. 1039–1043. https://doi.org/10.1109/ICALT.2006.1652624 ISSN: 2161-377X.

[17] Michelene T. H. Chi. 2009. Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science* 1, 1 (Jan. 2009), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

[18] Arthur W. Chickering and Zelda F. Gamson. 1987. Seven Principles for Good Practice in Undergraduate Education. *AAHE Bulletin* (March 1987). https://eric.ed.gov/?id=ED282491

[19] Elizabeth F. Churchill, Jonathan Trevor, Sara Bly, Les Nelson, and Davor Cubranic. 2000. Anchored conversations: chatting in the context of a document. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*. ACM Press, The Hague, The Netherlands, 454–461. https://doi.org/10.1145/332040.332475

[20] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. 2009. A review of overview+detail, zooming, and focus+context interfaces. *Comput. Surveys* 41, 1 (Jan. 2009), 2:1–2:31. https://doi.org/10.1145/1456650.1456652

[21] John R. Cutcliffe and Hugh P. McKenna. 1999. Establishing the credibility of qualitative research findings: the plot thickens. *Journal of Advanced Nursing* 30, 2 (1999), 374–380. https://doi.org/10.1046/j.1365-2648.1999.01090.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2648.1999.01090.x.

[22] Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: Conception, Theory and Measurement. *Comput. Surveys* 51, 5 (Nov. 2018), 99:1–99:39. https://doi.org/10.1145/3234149

[23] dwhly. 2018. Home. https://web.hypothes.is/

[24] Jennifer A Fredricks, Phyllis C Blumenfeld, and Alison H Paris. 2004. School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research* 74, 1 (March 2004), 59–109. https://doi.org/10.3102/00346543074001059 Publisher: American Educational Research Association.

[25] Fei Gao. 2013. A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education* 17 (2013), 76. https://www.academia.edu/4502063/A_case_study_of_using_a_social_annotation_tool_to_support_collaboratively_learning

[26] Fei Gao, Tianyi Zhang, and Teresa Franklin. 2013. Designing asynchronous online discussion environments: Recent progress and possible future directions. *British Journal of Educational Technology* 44, 3 (2013), 469–483. https://doi.org/10.1111/j.1467-8535.2012.01330.x _eprint: https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2012.01330.x.

[27] Arpita Ghosh and Jon Kleinberg. 2013. Incentivizing participation in online forums for education. In *Proceedings of the fourteenth ACM conference on Electronic commerce (EC '13)*. Association for Computing Machinery, New York, NY, USA, 525–542. https://doi.org/10.1145/2482540.2482587

[28] Carlos C. Goller, Micah Vandegrift, Will Cross, and Davida S. Smyth. [n.d.]. Sharing Notes Is Encouraged: Annotating and Cocreating with Hypothes.is and Google Docs†. *Journal of Microbiology & Biology Education* 22, 1 ([n. d.]), ev22i1.2135. https://doi.org/10.1128/jmbe.v22i1.2135 Publisher: American Society for Microbiology.

[29] Bette Gray. 2004. Informal Learning in an Online Community of Practice. *Journal of Distance Education* 19, 1 (2004), 20–35. https://eric.ed.gov/?id=EJ807836 Publisher: Canadian Network for Innovation in Education (CNIE).

[30] Raphael Hoffmann, Patrick Baudisch, and Daniel S. Weld. 2008. Evaluating visual cues for window switching on large screens. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, Florence, Italy, 929. https://doi.org/10.1145/1357054.1357199

[31] Wu-Yuin Hwang, Chin-Yu Wang, and Mike Sharples. 2007. A study of multimedia annotation of Web-based materials. *Computers & Education* 48, 4 (May 2007), 680–699. https://www.learntechlib.org/p/67080/ Publisher: Elsevier Ltd.

[32] Tristan E. Johnson, Thomas N. Archibald, and Gershon Tenenbaum. 2010. Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior* 26, 6 (Nov. 2010), 1496–1507. https://doi.org/10.1016/j.chb.2010.05.014

[33] A-Reum Jung and Jun Heo. 2021. Does cluttered social media environment hurt advertising effectiveness? The moderation of ad types and personalization. *Journal of Research in Interactive Marketing* ahead-of-print, ahead-of-print (Jan. 2021). https://doi.org/10.1108/JRIM-11-2020-0238

[34] Jeremiah H. Kalir. 2020. Social annotation enabling collaboration for open learning. *Distance Education* 41, 2 (April 2020), 245–260. https://doi.org/10.1080/01587919.2020.1757413 Publisher: Routledge _eprint: https://doi.org/10.1080/01587919.2020.1757413.

[35] Jeremiah Holden Kalir, Esteban Morales, Alice Fleerackers, and Juan Pablo Alperin. 2020. "When I saw my peers annotating": Student perceptions of social annotation for learning in multiple courses. *Information and Learning Sciences* ahead-of-print, ahead-of-print (April 2020). https://doi.org/10.1108/ILS-12-2019-0128

[36] Remi H. Kalir and Antero Garcia. 2021. *Annotation*. MIT Press, Cambridge, MA, USA.

[37] Kate Kaplan. 2019. 5 UX Workshops Cheat Sheet. https://www.nngroup.com/articles/5-ux-workshops/

[38] Azam Khan, Justin Matejka, George Fitzmaurice, and Gordon Kurtenbach. 2005. Spotlight: directing users' attention on large displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 791–798. https://doi.org/10.1145/1054972.1055082

[39] Kerri-Lee Krause and Hamish Coates. 2008. Students' engagement in first-year university. *Assessment & Evaluation in Higher Education* 33, 5 (Oct. 2008), 493–505. https://doi.org/10.1080/02602930701698892 Publisher: Routledge _eprint: https://doi.org/10.1080/02602930701698892.

[40] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Wiley Publishing.

[41] Hairong Li and Janice L. Bukovac. 1999. Cognitive Impact of Banner Ad Characteristics: An Experimental Study. *Journalism & Mass Communication Quarterly* 76, 2 (June 1999), 341–353. https://doi.org/10.1177/107769909907600211 Publisher: SAGE Publications Inc.

[42] Celia C. Lo. 2010. Student Learning and Student Satisfaction in an Interactive Classroom. *The Journal of General Education* 59, 4 (2010), 238–263. https://doi.org/10.5325/jgeneeduc.59.4.0238 Publisher: Penn State University Press.

[43] Ana Margarida Barreto. 2013. Do users look at banner ads on Facebook? *Journal of Research in Interactive Marketing* 7, 2 (Jan. 2013), 119–139. https://doi.org/10.1108/JRIM-Mar-2012-0013 Publisher: Emerald Group Publishing Limited.

[44] C. Marshall. 1997. Annotation: from paper books to the digital library. In *DL '97*. https://doi.org/10.1145/263690.263806

[45] Allen R. Martin and Matthew O. Ward. 1995. High Dimensional Brushing for Interactive Exploration of Multivariate Data. In *Proceedings of the 6th conference on Visualization '95 (VIS '95)*. IEEE Computer Society, USA, 271.

[46] Margaret Mazzolini and Sarah Maddison. 2003. Sage, guide or ghost? The effect of instructor intervention on student participation in online discussion forums. *Computers & Education* 40, 3 (April 2003), 237–253. https://doi.org/10.1016/S0360-1315(02)00129-X

[47] Iain McAlpine. 2000. Collaborative learning online. *Distance Education* 21, 1 (Jan. 2000), 66–80. https://doi.org/10.1080/0158791000210105 Publisher: Routledge _eprint: https://doi.org/10.1080/0158791000210105.

[48] Lori McCay-Peet, Mounia Lalmas, and Vidhya Navalpakkam. 2012. On saliency, affect and focused attention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 541–550. https://doi.org/10.1145/2207676.2207751

[49] Kelly Miller, Brian Lukoff, Gary King, and Eric Mazur. 2018. Use of a Social Annotation Platform for Pre-Class Reading Assignments in a Flipped Introductory Physics Class. *Frontiers in Education* 3 (2018), 8. https://doi.org/10.3389/feduc.2018.00008

[50] Kelly Miller, Sacha Zyto, David Karger, and Eric Mazur. 2014. Improving online class forums by seeding discussions and managing section size. In *Proceedings of the first ACM conference on Learning @ scale conference (L@S '14)*. Association for Computing Machinery, New York, NY, USA, 173–174. https://doi.org/10.1145/2556325.2567866

[51] Aidin Namin, Mitchell L. Hamilton, and Andrew J. Rohm. 2020. Impact of message design on banner advertising involvement and effectiveness: An empirical investigation. *Journal of Marketing Communications* 26, 2 (Feb. 2020), 115–129. https://doi.org/10.1080/13527266.2017.1393767 Publisher: Routledge _eprint: https://doi.org/10.1080/13527266.2017.1393767.

[52] Dip Nandi, Margaret Hamilton, James Harland, and Geoff Warburton. 2011. How active are students in online discussion forums?. In *Proceedings of the Thirteenth Australasian Computing Education Conference - Volume 114 (ACE '11)*. Australian Computer Society, Inc., AUS, 125–134.

[53] Jakob Nielsen. 2012. The Most Important Usability Activity. https://www.nngroup.com/articles/the-most-important-usability-activity/

[54] P. Nokelainen, J. Kurhila, M. Miettinen, P. Floreen, and H. Tirri. 2003. Evaluating the role of a shared document-based annotation tool in learner-centered collaborative learning. In *Proceedings 3rd IEEE International Conference on Advanced Technologies*. 200–203. https://doi.org/10.1109/ICALT.2003.1215056

[55] Michael North and Michael Ficorilli. 2017. Click me: an examination of the impact size, color, and design has on banner advertisements generating clicks. *Journal of Financial Services Marketing* 22, 3 (Sept. 2017), 99–108. https://doi.org/10.1057/s41264-017-0028-3

[56] Elena Novak, Rim Razzouk, and Tristan E. Johnson. 2012. The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education* 15, 1 (Jan. 2012), 39–49. https://doi.org/10.1016/j.iheduc.2011.09.002

[57] Gary W. Oehlert. 2010. *A First Course in Design and Analysis of Experiments*. http://conservancy.umn.edu/handle/11299/168002 Accepted: 2014-12-01T17:00:35Z.

[58] Kenton O'Hara and Abigail Sellen. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)*. Association for Computing Machinery, New York, NY, USA, 335–342. https://doi.org/10.1145/258549.258787

[59] Kate Orton-Johnson. 2008. The Online Student: Lurking, Chatting, Flaming and Joking. *Sociological Research Online* 12, 6 (Jan. 2008), 21–31. https://doi.org/10.5153/sro.1615 Publisher: SAGE Publications Ltd.

[60] Perusall. 2018. Perusall. https://perusall.com/

[61] Beryl Plimmer, Samuel Hsiao-Heng Chang, Meghavi Doshi, Laura Laycock, and Nilanthi Seneviratne. 2010. iAnnotate: Exploring Multi-User Ink Annotation in Web Browsers. *User Interfaces* 106 (2010), 9.

[62] Margaret Polk. 2020. Promoting a sense of learning community in asynchronous online courses. *Journal of Computing Sciences in Colleges* 36, 4 (Oct. 2020), 23–31.

[63] Carol Porter-O'Donnell. 2004. Beyond the Yellow Highlighter: Teaching Annotation Skills to Improve Reading Comprehension. *The English Journal* 93, 5 (2004), 82–89. http://www.jstor.org/stable/4128941

[64] Simone Rebora and Federico Pianzola. 2018. A New Research Programme for Reading Research: Analysing Comments in the Margins on Wattpad. *DigitCult - Scientific Journal on Digital Cultures* 3, 2 (Sept. 2018), 19–36. https://doi.org/10.4399/97888255181532 Number: 2.

[65] Hugo Romat, Emmanuel Pietriga, Nathalie Henry-Riche, Ken Hinckley, and Caroline Appert. 2019. SpaceInk: Making Space for In-Context Annotations. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New Orleans LA USA, 871–882. https://doi.org/10.1145/3332165.3347934

[66] Jinwook Seo and Ben Shneiderman. 2003. Interactively Exploring Hierarchical Clustering Results. In *The Craft of Information Visualization*. Elsevier, 334–340. https://doi.org/10.1016/B978-155860915-0/50042-1

[67] Adam Sigel. 2008. THE IMPACT OF BANNER AD STYLES ON INTERACTION AND CLICK-THROUGH RATES. https://doi.org/10.48009/2_iis_2008_337-342

[68] Karen Singer-Freeman and Linda Bastone. 2016. Pedagogical Choices Make Large Classes Feel Small. (2016), 26.

[69] Markus Steinberger, Manuela Waldner, Marc Streit, Alexander Lex, and Dieter Schmalstieg. 2011. Context-preserving visual links. *IEEE transactions on visualization and computer graphics* 17, 12 (Dec. 2011), 2249–2258. https://doi.org/10.1109/TVCG.2011.183

[70] Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, and Hanspeter Pfister. 2016. Guidelines for Effective Usage of Text Highlighting Techniques. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016),

489–498. https://doi.org/10.1109/TVCG.2015.2467759

[71] Claudia Szabo, Nickolas Falkner, Antti Knutas, and Mohsen Dorodchi. 2018. Understanding the Effects of Lecturer Intervention on Computer Science Student Behaviour. In *Proceedings of the 2017 ITiCSE Conference on Working Group Reports (ITiCSE-WGR '17)*. Association for Computing Machinery, New York, NY, USA, 105–124. https://doi.org/10.1145/3174781.3174787

[72] Cheryl Bagley Thompson and Edward A. Panacek. 2007. Research study designs: Non-experimental. *Air Medical Journal* 26, 1 (Jan. 2007), 18–22. https://doi.org/10.1016/j.amj.2006.10.003 Publisher: Elsevier.

[73] Manuela Waldner, Werner Puff, Alexander Lex, Marc Streit, and Dieter Schmalstieg. 2010. Visual links across applications. (2010), 8.

[74] Jeffrey M Wooldridge. 2015. *Introductory econometrics: A modern approach*. Cengage learning.

[75] Dongwook Yoon, Nicholas Chen, and François Guimbretière. 2013. TextTearing: opening white space for digital ink annotation. In *Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13)*. Association for Computing Machinery, New York, NY, USA, 107–112. https://doi.org/10.1145/2501988.2502036

[76] Shumin Zhai, Julie Wright, Ted Selker, and Sabra-Anne Kelin. 1997. Graphical Means of Directing User's Attention in the Visual Interface. In *Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction (INTERACT '97)*. Chapman & Hall, Ltd., GBR, 59–66.

[77] Joyce Zhu, Jeremy Warner, Mitchell Gordon, Jeffery White, Renan Zanelatto, and Philip J. Guo. 2015. Toward a domain-specific visual discussion forum for learning computer programming: An empirical study of a popular MOOC forum. (2015), 101–109. https://doi.org/10.1109/VLHCC.2015.7357204

[78] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful Classroom Deployment of a Social Document Annotation System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1883–1892. https://doi.org/10.1145/2207676.2208326 event-place: Austin, Texas, USA.

A ON WHOM SPOTLIGHTING HAS THE LARGEST IMPACT

Table 5 summarizes the results of the analyses described in 5.2.4 for each of the student groups and each of the spotlighting conditions. The results show that the spotlighting treatments affect the students who click through the highlights. These effects, however, are not significant for those students who mostly click through the comments on the sidebar.

B EXPLORING THE POSSIBLE EFFECT OF CONFOUNDING FACTORS

B.1 Visual Features Other than Spotlighting.

Although we observed that spotlighting comments affected student's clicks, it is possible that the higher click rate on treated spotlit comments (comments shown as spotlit in any of the treatment conditions) is in fact due to other visual features that happen to be more common among these comments. One such feature could be the presence of an instructor icon next to the comment signifying that the instructor has replied to it. Another is the presence of a question mark icon indicating that the comment is in fact a question. It is conceivable that the comments marked as high quality by the instructor have a higher instructor response rate or are questions that the instructor believes all students should see. Because half of spotlit comments comprise high quality comments, the question and instructor reply icon could be relatively more common among spotlit comments. Therefore, to remove the possible confounding effect of these indicators, we removed all the entries from the dataset that had these indicators and performed the same analyses as before on student clicks. The regression results are shown in Table 6

We found that consistent with the results we obtained before, **the interaction of spotlighting and treatment had a significant effect on student clicks, suggesting that spotlighting alone (with no additional icons) can attract clicks**.

B.2 Student arrival order to comment on readings

Students included in the 50 comment threshold for each reading–who commented before the instructor saw spotlights, and therefore saw none in that reading—tend to vary. Figure 12 shows a breakdown of the students who were part of the threshold. Students who commented early in

Table 5. Linear regressions predicting click outcome with standard errors clustered on student and comment, performed on two separate partitions: on the data from the students who performed 60% or more of their clicks on the highlights of the page, and the data from those students who performed 60% or more of their clicks on the sidebar. The results suggest that the treatments do not affect those students who mostly interact with the sidebar to read comments. They do however, affect those who read comments via clicking on the highlights.

Conditions	Regressor	Estimate	Std Error	t value	p value
	Students who mostly click on highlights				
Block vs Control	spotlit	-0.003	0.008	-0.358	0.72
	spotlit × treatment	0.071	0.020	3.536	**<0.001**
$N_{datapoints} = 98,672$, $N_{comments} = 4,804$, $N_{students} = 73$				Adj. $R^2 = 0.09$	$RMSE = 0.23$
Inline vs Control	spotlit	-0.003	0.007	-0.376	0.71
	spotlit × treatment	0.072	0.020	3.652	**<0.001**
$N_{datapoints} = 96,279$, $N_{comments} = 4,804$, $N_{students} = 72$				Adj. $R^2 = 0.09$	$RMSE = 0.23$
Emphasis vs Control	spotlit	-0.003	0.008	-0.404	0.69
	spotlit × treatment	0.067	0.019	3.641	**<0.001**
$N_{datapoints} = 95,061$, $N_{comments} = 4,804$, $N_{students} = 74$				Adj. $R^2 = 0.09$	$RMSE = 0.24$
None vs Control	spotlit	0.000	0.008	-0.005	1.0
	spotlit × treatment	0.022	0.014	1.598	0.11
$N_{datapoints} = 98,101$, $N_{comments} = 4,804$, $N_{students} = 74$				Adj. $R^2 = 0.09$	$RMSE = 0.24$
	Students who mostly click on the sidebar				
Block vs Control	spotlit	-0.002	0.015	-0.125	0.90
	spotlit × treatment	0.030	0.035	0.858	0.40
$N_{datapoints} = 56,979$, $N_{comments} = 4,783$, $N_{students} = 46$				Adj. $R^2 = 0.12$	$RMSE = 0.25$
Inline vs Control	spotlit	0.000	0.016	-0.009	0.99
	spotlit × treatment	0.062	0.031	1.985	0.05
$N_{datapoints} = 63,285$, $N_{comments} = 4,788$, $N_{students} = 47$				Adj. $R^2 = 0.11$	$RMSE = 0.25$
Emphasis vs Control	spotlit	-0.002	0.016	-0.136	0.89
	spotlit × treatment	0.034	0.034	0.981	0.33
$N_{datapoints} = 61,288$, $N_{comments} = 4,756$, $N_{students} = 47$				Adj. $R^2 = 0.12$	$RMSE = 0.26$
None vs Control	spotlit	-0.004	0.015	-0.263	0.79
	spotlit × treatment	0.004	0.034	0.127	0.90
$N_{datapoints} = 55,516$, $N_{comments} = 4,765$, $N_{students} = 47$				Adj. $R^2 = 0.14$	$RMSE = 0.25$

at least half of the readings account for 15 out of the 166 in the treatment group. In addition, 90% of the students in the treatment group were exposed to at least 100 spotlights (Average number of spotlights encountered by a student= 388.7, Maximum number of spotlights encountered by a student= 3108, Minimum= 0).

B.3 Quality of Comments.

We wanted to understand whether the quality of comments had an effect on the students' clicks, conditioned on whether comments were or were not spotlit. Therefore, we separated our filtered click outcome data (with the comments that had instructor reply and question icons removed) into two partitions: one in which for a pair of student and comment, the comment was spotlit to the student in any of the treatment groups, and the other, the comment student pairs where either the

Students with Early Comments
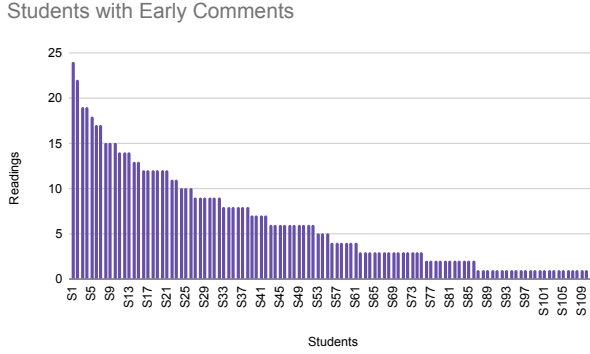
Fig. 12. Students that were part of the 50 comments threshold. The y axis is the reading, in this class there were 26 readings. The x axis is the students.

comment was not spotlit or the student belonged to the Control or the None group and therefore did not experience spotlighting.

We fit the following model to each partition separately:

$$\text{Click outcome}_{sc} = \beta 1(quality)_c + \delta_s + \epsilon_{sc} \tag{3}$$

where Click outcome$_{sc}$ is whether the comment c has been clicked by student s, $1(quality)_c$ is an indicator variable for comment quality, and δ_s is student fixed effects. Because we include quality of comments as an independent variable in the model, we do not include comment fixed effects because the two variables are collinear. ϵ_{sc} is the error term. We also clustered standard errors on comment and student.

We found that for treated spotlit comments, quality had a statistically significant effect on clicks at the alpha=0.05 threshold ($\beta = 0.020, \sigma = 0.009, t = 2.120, p = 0.04$). **For comments that were not spotlit however, quality was a very significant predictor of clicks** ($\beta = 0.040, \sigma = 0.008, t = 4.847, p < 0.001$).

The results suggests that students are able to discern comment quality before they even click on the comments, perhaps through scanning comment "headers" that are visible even before the comments are expanded.

To decouple the effect of comment quality on clicks from the effect of spotlighting, we performed the same analyses of predicting clicks by treatment and spotlighting conditions on the portion of the click dataset (with the comments that had instructor reply and question icons removed) where the comments were not assessed as high quality by the instructor. The results, shown in Table 7, reveal that consistent with the results we observed for the entire dataset, in the low quality portion of the data, treated spotlit comments are still clicked at a higher rate.

Figures 13, 14, and 15 show that although in the low quality portion of the data the treated spotlit comments get clicked at a higher rate, the untreated spotlit comments are clicked as much as the untreated unspotlit comments.

B.4 Replacing OLS Regression with Logistic Regression

In Section 5.2, we performed our analyses using OLS regression. Here, we present the results of the same analyses by estimating our models (equation 1 for both click and reply datasets) with logistic regression. To do so, we used the function "feglm" from the R package "fixest", with family function

Table 6. Linear regressions predicting click outcome on the portion of the data that does not have instructor reply or question mark icons.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	0.004	0.007	0.648	0.52
	spotlit × treatment	0.050	0.015	3.398	**<0.001**
$N_{datapoints} = 220, 105, N_{comments} = 4, 828, N_{students} = 164$				Adj. $R^2 = 0.11$	RMSE = 0.23
Inline vs Control	spotlit	0.004	0.007	0.528	0.60
	spotlit × treatment	0.071	0.015	4.761	**<0.001**
$N_{datapoints} = 217, 409, N_{comments} = 4, 828, N_{students} = 165$				Adj. $R^2 = 0.12$	RMSE = 0.23
Emphasis vs Control	spotlit	0.001	0.007	0.140	0.89
	spotlit × treatment	0.056	0.015	3.762	**<0.001**
$N_{datapoints} = 218, 942, N_{comments} = 4, 828, N_{students} = 166$				Adj. $R^2 = 0.12$	RMSE = 0.23
None vs Control	spotlit	0.006	0.007	0.921	0.36
	spotlit × treatment	0.004	0.012	0.349	0.73
$N_{datapoints} = 219, 070, N_{comments} = 4, 830, N_{students} = 163$				Adj. $R^2 = 0.12$	RMSE = 0.23

Table 7. Linear regressions predicting click outcome on the portion of the data that was not marked as high quality by the instructor.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	-0.003	0.007	-0.469	0.64
	spotlit × treatment	0.063	0.012	5.147	**<0.001**
$N_{datapoints} = 214, 099, N_{comments} = 4, 732, N_{students} = 164$				Adj. $R^2 = 0.11$	RMSE = 0.22
Inline vs Control	spotlit	-0.003	0.007	-0.395	0.69
	spotlit × treatment	0.079	0.013	6.045	**<0.001**
$N_{datapoints} = 211, 422, N_{comments} = 4, 732, N_{students} = 165$				Adj. $R^2 = 0.11$	RMSE = 0.22
Emphasis vs Control	spotlit	-0.002	0.007	-0.324	0.75
	spotlit × treatment	0.065	0.013	5.099	**<0.001**
$N_{datapoints} = 212, 924, N_{comments} = 4, 732, N_{students} = 166$				Adj. $R^2 = 0.11$	RMSE = 0.23
None vs Control	spotlit	-0.003	0.007	-0.399	0.69
	spotlit × treatment	0.022	0.009	2.335	**0.02**
$N_{datapoints} = 213, 037, N_{comments} = 4, 734, N_{students} = 163$				Adj. $R^2 = 0.12$	RMSE = 0.22

binomial and link logit. Our results are consistent with what we reported in Section 5.2. The results of the analyses for click outcome and reply outcome are shown in Tables 8 and 9 respectively.

C PILOT STUDY

In order to design our user study and experiment (Section 5), we conducted a pilot study. We deployed NB with the new designs in a biology class at the same large public university in the United States. The readings were all primary scientific literature papers in the form of PDF files. For that reason, we experimented with marginalia and emphasized green highlights as neither requires any reflow of the DOM.

The *Marginalia* pattern was designed explicitly for PDF as the content of the file can not be reflowed . We insert a widened *margin* between the document content and the sidebar, and expand spotlit comments in the margin, inside a yellow sticky note with a faded edge (figure 16). The
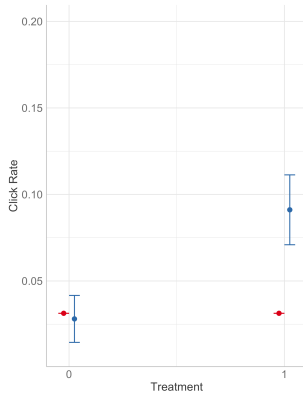
Fig. 13. Interaction plot of treatment and spotlighting in the partition of the data that includes Block type spotlighting and the Control, only including the comments that were not marked as high quality by the instructor.
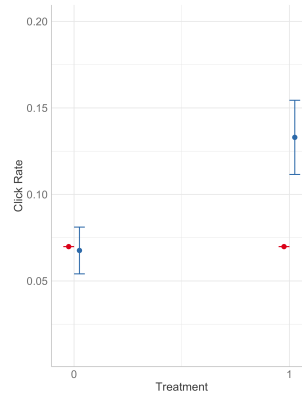
Fig. 14. Interaction plot of treatment and spotlighting in the partition of the data that includes Emphasis type spotlighting and the Control, only including the comments that were not marked as high quality by the instructor.
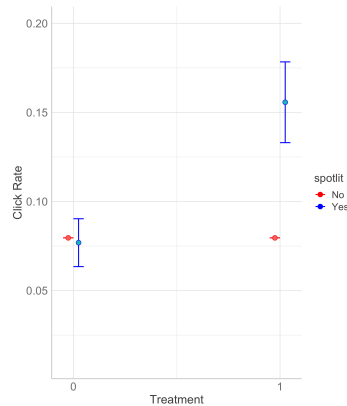
Fig. 15. Interaction plot of treatment and spotlighting in the partition of the data that includes Inline type spotlighting and the Control, only including the comments that were not marked as high quality by the instructor.

position of the note is calculated based on the location of the related highlight (annotation) in the document. Once again, we truncate long comments to keep the note from being too large in the margin. If the user clicks on the comment or the related highlight, the visibility of the marginalia

Table 8. Generalized linear regressions (logistic regressions) predicting click outcome with standard errors clustered on student and comment. In each of the partitions that include a treatment group (Block, Inline, and Emphasis) and the control group, the interaction between treatment and spotlighting is significant, indicating the effect of being assigned to the treatment group for the spotlit comments is to increase the click ratio.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	-0.006	0.088	-0.069	0.95
	spotlit × treatment	0.733	0.158	4.634	**<0.001**
$N_{datapoints} = 172,197$, $N_{comments} = 3,475$, $N_{students} = 153$					
Log-Likelihood $= -37,641.7$, Adj. Pseudo $R^2 = 0.14$					
Inline vs Control	spotlit	-0.003	0.083	-0.040	0.97
	spotlit × treatment	0.915	0.154	5.933	**<0.001**
$N_{datapoints} = 170,418$, $N_{comments} = 3,442$, $N_{students} = 157$					
Log-Likelihood $= -37,400.2$, Adj. Pseudo $R^2 = 0.15$					
Emphasis vs Control	spotlit	-0.022	0.085	-0.259	0.80
	spotlit × treatment	0.661	0.147	4.481	**<0.001**
$N_{datapoints} = 171,908$, $N_{comments} = 3,446$, $N_{students} = 155$					
Log-Likelihood $= -39,210.0$, Adj. Pseudo $R^2 = 0.14$					
None vs Control	spotlit	-0.001	0.085	-0.008	0.99
	spotlit × treatment	0.147	0.149	0.991	0.32
$N_{datapoints} = 170,528$, $N_{comments} = 3,447$, $N_{students} = 155$					
Log-Likelihood $= -37,147.7$, Adj. Pseudo $R^2 = 0.15$					

Table 9. Generalized linear regressions (logistic regressions) predicting reply outcome, with standard errors clustered on student and comment. Conditioned on a comment being clicked on, its assignment to a treatment condition has no statistically significant effect on the probability that it is replied to.

Conditions	Regressor	Estimate	Std Error	t value	p value
Block vs Control	spotlit	-0.105	0.599	-0.175	0.86
	spotlit × treatment	0.090	0.765	0.118	0.91
$N_{datapoints} = 1,772$, $N_{comments} = 288$, $N_{students} = 85$					
Log-Likelihood $= -588.5$, Adj. Pseudo $R^2 = -0.10$					
Inline vs Control	spotlit	-0.041	0.653	-0.062	0.95
	spotlit × treatment	0.347	0.721	0.481	0.63
$N_{datapoints} = 1,928$, $N_{comments} = 296$, $N_{students} = 87$					
Log-Likelihood $= -647.8$, Adj. Pseudo $R^2 = -0.12$					
Emphasis vs Control	spotlit	-0.204	0.594	-0.343	0.73
	spotlit × treatment	-0.032	0.644	-0.049	0.96
$N_{datapoints} = 2,119$, $N_{comments} = 290$, $N_{students} = 92$					
Log-Likelihood $= -668.2$, Adj. Pseudo $R^2 = -0.09$					
None vs Control	spotlit	-0.396	0.465	-0.852	0.39
	spotlit × treatment	0.429	0.547	0.783	0.43
$N_{datapoints} = 11,610$, $N_{comments} = 312$, $N_{students} = 89$					
Log-Likelihood $= -1,512.2$, Adj. Pseudo $R^2 = -0.16$					

Fig. 16. Marginalia note showing a preview of the comment with faded end. This screenshot shows a paper by Engler Et al. (2008) A one pot, one step, precision cloning method with high throughput capability. from the pilot class

will increase and it will show more text. This technique was explored only in our pilot with PDF documents (see Appendix).

We augmented the instructor's version of the NB UI with controls that permitted them to mark chosen comments for spotlighting. The instructor was free to choose the number of comments and the type of spotlight for each and had the option to mix both types of spotlights within the same reading. In the first half of the class (readings 1 to 9), the instructor picked the spotlit comments (e.g., relatable questions, linking to previous topics, and encouraging specific students). The average number of spotlights picked by the instructor was 4 spotlights for each reading. The instructor reported that he mixed the types of spotlights. He explained to us that he used marginalia to spotlight comments that he thought were critical for understanding, and emphasized green highlights for comments that illustrated commonly misunderstood material or discussion points. For the other half of the class (reading 10 to 17), 10 comments were randomly selected from comments provided by students who read the assignment early (first 2 days after the pdf was posted). These comments were spotlit (with a random spotlight type) and presented to all the students.

Students taking this class were undergrads in the Bioengineering program. The study took place during the Spring quarter (April 14, 2021 to June 10, 2021). As this was during the COVID pandemic, all teaching was online. Students' engagement on NB was mandatory; a minimum of three comments on each reading was required. New annotations or replies were all graded the same. Students had two readings each week and a total of 17 readings. There were 51 students in the class.

We examined the usage logs to understand how these students interacted with the new spotlighting feature [72].The logs included information for clicks, comments, and replies made through NB. Click data for a comment included information on the type of spotlight (if any) and *how* students accessed each comment: by clicking on the highlighted text, by clicking on the marginal note (when available), or by clicking on the header line in the sidebar. We followed our log examination with informal interviews with 8 students [13]. A goal of the interviews was to investigate the behaviours we observed from our logs. The interviews were between 45 to 60 minutes long for each student

C.1 Considerations for the User Study Design

In the following section we discuss the results from the pilot study (logs and interviews). The pilot study data set included 11,283 interactions from the 51 students over 17 readings. The total number of comments *threads* was 1696 for all readings. The average number of threads per reading was 99. The total number of spotlit comments was 104, with 50 emphasized green highlights and 54 marginalia. We synthesized our findings from the log data and the interviews and identified the following themes that were essential in designing our formal user study (Section 5).

C.1.1 Timing matters with spotlighting. Some students tended to post their comments early, and the rest did their reading and comments right before the deadline. Comments posted after the instructor looked for spotlights were of course not spotlit, so most of the spotlit comments tended to be chosen from early commenters, who were usually from section 2. Those comments tended to be posted by the same, select set of students. This positively impacted the students who authored those comments, as we show in the interview results. When we compared the time of the week in which students comment to the time in which students interact on spotlights, we found that students who tended to do their assignment midweek tended to click on spotlights more than the others, and as the deadline approached, the number of interactions with the spotlights tended to drop slightly.

C.1.2 Comments were discovered through the highlights and the sidebar. In general, students on NB access comments through the document highlights or the sidebar. Upon looking at a snapshot from one reading we notice that students click to read comments from the sidebar (grey squares) and through the highlights (yellow squares). We also notice that the spotlights that were included in the document were mostly clicked on after an interaction on the document. That is, 15 students out of the 18 students who clicked on the spotlights performed most of their previous clicks on the highlights in the document. This makes sense, since spotlights by design are closer to the reading (the highlighted text). In addition to the logs, students in our interviews reported that they tend to read the comments as they are reading the textbook by looking at highlights relevant to the passages they read. This explains why some students in the log appear to click on comments through the highlights more often than the sidebar.

C.1.3 Green highlights were clicked on more than marginalia. While both spotlight types in the pilot study were presented to students equally often, overall emphasized green highlights received 160% more clicks than marginalia. Marginalia do show an excerpt from the comment as opposed to green highlights which require students to click in order to read the comments. It is possible that students are reading marginalia and leaving them unclicked. When we asked students about their preferred type of spotlighting, we received a mixture of reactions. Some students preferred marginalia: "I would prefer margins since it would not interrupt the reading experience." Another one reported: "margins were more aesthetically pleasing." A student explained that they prefer this pattern since it provides a large preview of the comment, they get to read the comment without clicking on the highlight or the sidebar. Some students who preferred the green highlights explained: "I prefer colored highlights...click on them then I read on the sidebar. Margins are extra text and away from the article." Another student appreciated the colors of the highlights and preferred shorter spotlights on the page: "Highlight colors helped my eyes focus. I wish the margins were shorter." We kept those results in mind when deploying our next batch of spotlighting patterns in the main study.

C.1.4 Quality of spotlit comments matters. In the second half of the term, we picked random comments to spotlight, and some students picked up on this: "did not know why they were highlighted. Initially I thought it was because they were important but after reading some of the spotlight comments I stopped thinking that was the reason. I thought they were just randomly selected."

C.1.5 Spotlighting comments positively affects the authors of those comments. We asked students about how they felt about their comments being spotlit, and most students expressed that it made them feel proud and motivated to make "better" comments: "My study buddy noticed and she told me, 'I like those.' It made me happy to see my comments spotlighted. I started to make class related comments using keywords I've seen in the class." Another student reported that seeing

her comment being spotlit made her put more effort into her comments: "Yay! I tried harder with in-depth questions." A third student also added that it made them feel proud: A third student added that "if I see my comment it will make me proud".

Jumana Almahmoud, Farnaz Jahanbakhsh, Marc Facciotti, Michele Igo, Kamali Sripathi, Kobi Gal, David Karger,,