Successful Classroom Deployment of a Social Document Annotation System

Sacha Zyto, David R. Karger[1]
MIT CSAIL
Cambridge, MA 02139, USA
{sacha, karger}@mit.edu

Mark S. Ackerman[1]
Univ of Michigan, Ann Arbor
Ann Arbor, MI 48109, USA
ackerm@umich.edu

Sanjoy Mahajan[1]
Olin College of Engineering
Needham, MA 02492, USA
sanjoy@olin.edu

ABSTRACT

NB is an in-place collaborative document annotation website targeting students reading lecture notes and draft textbooks. Serving as a discussion forum in the document margins, NB lets users ask and answer questions about their reading material *as they are reading*. We describe the NB system and its evaluation in a real class environment, where students used it to submit their reading assignments, ask questions and get or provide feedback. We show that this tool has been successfully incorporated into numerous classes at several institutions. To understand how and why, we focus on a particularly successful class deployment where the instructor adapted his teaching style to take students' comment into account. We analyze the annotation practices that were observed—including the way geographic locality was exploited in ways unavailable in traditional forums—and discuss general design implications for online annotation tools in academia.

Author Keywords

hypertext, annotation, collaboration, forum, e-learning

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces. - Graphical user interfaces.

General Terms

Design, Experimentation, Human Factors

INTRODUCTION

Early hypertext research offered the promise of annotating texts for educational purposes with the detailed discussion necessary to understand complex material. The Web amplified that promise. But it has not been fulfilled.

There is at present no collaborative annotation tool in widespread use in education. Past work revealed significant barriers to their adoption. For example, Brush's [3] study of an online annotation system reported that because students printed and read documents and comments offline, faculty had to force discussion by requiring replies to comments. It

has been unclear whether the annotation systems were too limited, the technical ecology around them was too rudimentary, or the educational system was not adequately prepared. Perhaps in consequence, research on the topic has lain relatively fallow for the past decade.

In this paper, we offer evidence that the time may be ripe for a renewal of research and development on collaborative annotation systems. We report on NB, an annotation forum that has been successfully deployed and used in 55 classes at 10 universities. Students use NB to hold threaded discussions in the margins of online class material.

Our contribution is twofold. First, we provide evidence that the socio-technical environment of the classroom has evolved to the point where the barriers that were encountered by earlier annotation tools have lowered enough to be overcome by motivated teachers and students. While these changed circumstances do not yet hold in *all* circumstances, we will argue that they are common enough to be worth designing for.

Our second contribution is to assess specific features of NB that we believe contributed to its being adopted and valued by its users. Our design of NB's "situated discussions," contrasting with the traditional "linked hypertext" model, was motivated by the following design hypotheses:

- That the ability to comment in the margins, without leaving the document, would enable students to comment "in the flow" while reading, reducing the deterrent loss of context involved in commenting elsewhere;

- That the in-place display of comments in the margins would draw students' attention to relevant comments while reading, and encourage them to respond;

- That the physical location of comments with their subject matter would provide a valuable organizational structure distinct from the chronological organization typical of discussion forums, helping students aggregate related threads and consider them together;

Taken together, we believed these characteristics would drive a virtuous cycle, encouraging more students to participate more heavily, thus providing more helpful material for other students, yielding additional incentive to participate.
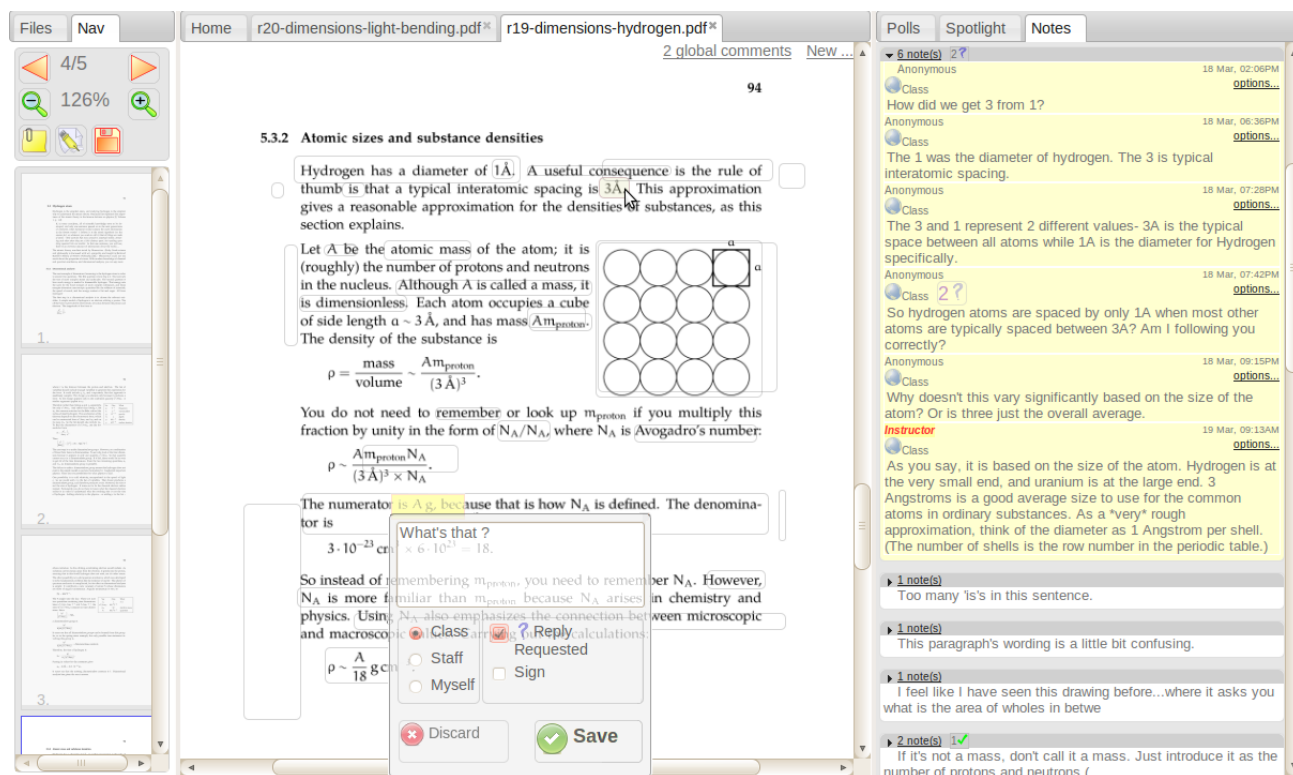
Figure 1. NB document view. *Left*: Thumbnails, *Center*: Document (with pop-up annotation editor on the bottom), *Right*: Annotations

In this work, we give evidence supporting of all of our hypotheses. We report substantial usage of Nb in many classes. To understand *how* and *why* the tool was used, we examine one "best case" use of Nb in which 91 students in a 1-semester class produced over 14000 annotations. Given that most of those comments had *substantive content* [8] and that the professor and students alike praised the system, this appears to be a successful classroom deployment of an annotation system. Since only limited successes have been previously reported in HCI, hypertext, or education literature, we assess the factors that led to this successful use and their implications for innovative educational uses and future textbooks.

MOTIVATION AND RELATED WORK

While there is relatively little current work, the past abounds with studies of collaborative discussion tools for education. Space limits us to projects we found most influential. It is accepted that students understand material better after discussing it [5, 6]. This suggests that discussion forums can be useful in an academic setting. Their use in this context can be traced back to the *Plato* system (1960) [4]. *CSILE* (1984) and its successor *Knowledge Forum* (1995) [10] explore mechanisms for encourage students to achieve *knowledge building* and *understanding at the group level*.

These tools all support discussion of class reading materials, but the discussions occur in a separate environment. As we will argue below, this is a drawback: a reader might not be aware that a topic she is considering has been discussed, so might miss the opportunity to contribute to or benefit from the discussion. Actually navigating to the discussion causes loss of context, making it harder to follow the discussion or return to the material. A study of forum use in a class in 2002 [13] found that discussion threads tended to branch and lose coherence, with many leaves of the discussion rarely read, and observed that "the typical nonlinear branching structure of online discussion may be insufficient for the realization of truly conversational modes of learning." This was 10 years ago, and one might believe that the current generation takes better to discussion forums. But an examination of MIT's classroom discussion system, Stellar, showed that the 50 classes with the *most posts* in the Spring 2010 semester produced a total of 3275 posts—an average of 65.5 per class—and a maximum of 415.[2] (At the same time at MIT, one 91-student class using NB generated over 14,000 posts.)

Improving on this "detached" situation, CaMILE [8] offered anchor-based discussions: its HTML documents can embed hyperlinks from *discussion anchors* - places where *the authors* thought a discussion could be appropriate. Although this does not offer *readers* the flexibility to discuss arbitrary points, it is a significant step towards overcoming the limitations of traditional online forums by trying to situate them nearer the context of the document being discussed. However, reading those annotations still requires navigating to a different context.

---

[2]An important caveat is that Stellar is not a particularly *good* discussion system. Recently, a forum tool called Piazzza has begun to see widespread adoption; we have not yet had the opportunity to analyze its usage, which clearly outperforms that of Stellar.

The *WebAnn* project [3] let students discuss any part of a document. More significantly, it recorded annotations *in-place* in the document margins, allowing readers to see the document and the discussions on the same page. Setting the context this way meant that comments could omit lengthy explanations since they would be visible at the same time as that material. The expected consequence was that a wider audience would read and participate easily in the discussion. However, at the time of the WebAnn study (Spring 2001), some factors limited the benefits of the tool. Mainly, students unanimously printed the lecture material, and worked on the printout. They then returned to the online site only to record the annotations they had "planned out" on their printed copies. This introduced large lags between comments and replies that inhibited organic discussion, and meant that many comments arrived too late to benefit other students while they were reading.

As people have become more comfortable online, some of the obstacles impacting tools such as WebAnn may have shrunk. With this in mind, we deployed NB to assess the present-day (and future) appeal of a collaborative annotation system, and have produced evidence that in-margin discussions can now be an effective part of teaching. Deployed at roughly the same time, Van der Pol's *Annotation System* [14] is another web-based annotation framework that has been successfully used in an academic context, and was used to quantify how both tool affordances and peer-feedback can facilitate students' online learning conversations.

SYSTEM DESCRIPTION

NB is a web-based tool where users can read and annotate PDF documents using standard web browsers. After logging in, a student typically selects a document and starts reading. As shown on Figure 1, the document is augmented by annotations that the students and faculty have written, which appear as expandable discussions on the right-hand-side panel. Hovering someplace in the document highlights the annotations covering that place, whereas clicking somewhere on the document scrolls to the corresponding annotations. Annotations in NB are either anchored to a particular location in the document or are general comments on the document.[3] To add an annotation somewhere in the document, users click and drag to select a region they want to comment on. This region is highlighted and an in-place annotation editor pops up (bottom of Figure 1).

Users can choose whether their comment should be visible to everyone in the class (the default), or to the teaching staff only, or to themselves only. The can also choose whether the comment is anonymous (the default) or signed. Once a comment has been saved, its author can delete it or edit it as long as there hasn't been a reply. He can also change its properties (visibility, anonymity). Users can tag each other's comments with the following tags: *Favorite, Hidden, I agree, I disagree*, and *Reply requested*.

Implementation Details

---

[3]We have found that general comments are rarely used, and do not discuss them further.

At the time of the study, the server-side of NB was based on python, a PDF library and a postgresql database. Since then, NB has been re-implemented using the Django framework in order to improve portability and maintainability. NB uses a RESTful data API to exchange data between the client and server. This allows third parties to use the NB framework and implement their own UI. The NB server is open to use by any interested faculty at http://nb.mit.edu/ .

Deployment

To date, NB has has been used by in 49 classes by 32 distinct faculty at 10 institutions including MIT, Harvard, California State, U. Edinburgh, KTH Sweden, Olin College, and Rochester Institute of Technology. The majority of classes are in the physical sciences but a few are in social sciences and humanities. Of the 32 faculty, 8 were using the tool for the first time this semester. Of those who started earlier, 9 faculty (28%) made use of the tool in multiple semesters (for a total of 18 re-uses), indicating that they have continued to adopt it *after* a semester's experience of its usage. This seems a coarse indication that they believe that the tool is helping them meet their teaching goals. Informal positive feedback from many of the faculty has supported this indication.

The tool saw substantial student use in many classes. Table 1 shows that total number of comments submitted in the top 15 classes. 13 of these classes received more comments than the maximum (415) captured in *any* usage of Stellar, MIT's forum tool. The top five each collected more comments than the top 50 classes using Stellar *combined* (3275).

Class	comments	per user
Approximation in Science & Eng.	14258	151
UI Design and Implementation (*)	10420	83
Math Methods for Business (*)	4436	61
Mathematics for CS (*)	3562	23
Mathematics for CS (*)	3270	34
UI Design and Implementation (*)	2703	61
Signals and Systems	1996	39
Electricity and Magnetism	1254	17
Mathematics for CS (*)	1045	26
Pseudorandomness	880	40
Dynamics	789	21
Adv. Quant. Research Methodology	570	9
Math Methods for Business (*)	530	12
Concepts in Multicore Computing	336	21
Moral Problems and the Good Life	233	8

Table 1. Usage of NB in other classes. Starred classes are re-uses by a faculty member who had already used NB.

USAGE ANALYSIS

Given that NB is seeing some adoption, we wished to investigate *how* and *why* NB is being adopted and used in the classroom. Due to space limitations, we focus the remainder of this article on the single most successful use of NB, in *Approximations in Engineering*, Spring Semester 2010 at MIT. The teacher was Sanjoy Mahajan, our fourth author. The reader might worry that we are skewing the data, but we believe this choice is justified for three reasons:

- Our objective here is to demonstrate, not that NB *always* works but that NB *can* work in a real-world setting, which shows the research direction worth pursuing.

- Mahajan was made an author *after* his usage of NB; he had no special incentive to make the NB succeed, aside from an interest in teaching well.

- Our data from many of the other high-usage classes is qualitatively similar, as we aim to report in an extended version of this article.

Approximations in Engineering had 91 undergraduate students. The thrice-weekly class lectures came from a pre-print version of Mahajan's textbook. He assigned sections of the book, usually about 5 pages long, for each lecture. The previous four times he had taught the course, Mahajan required students to submit a paper-based "reading memo"—annotations taken on the sides of the lecture pages—at the beginning of each class. This method was popularized by Edwin Taylor [12]. Mahajan required students to make a "reasonable effort", defined in the syllabus as follows: "For reasonable effort on a reading memo one comment is not enough unless it is unusually thoughtful, while ten is too many".

NB replaced the previous paper-based annotation system. Mahajan left the reading memo model and instructions unchanged but modified the deadline: instead of requiring that annotations be delivered in class, he made the online annotations due 12 hours before class, intending to peruse them prior to lecturing (we discuss the consequences of this change in the The Instructor Perspective section). There were no Teaching Assistants (TAs) for this class.

Method
Our analysis is based on log data, user questionnaires, and a small focus-group interview. The log data included user actions down to the action level and were kept in a standard log file. All annotations that users produced were stored, with the users' consent. User questionnaires were administered at the end of the semester to both students and faculty.

In total, we obtained over 1.4 million user actions and, as mentioned, 14258 annotations from this class. These actions include page seen, comment created, time spent with NB both active and being "idle", and so on. We also obtained, to be discussed later, questionnaires from students and interviews with the instructor. The questionnaires consisted of Likert scale ratings concerning satisfaction and how NB might have helped or hindered understanding. In addition, they included open-ended comments about each question where they could explain their ratings.

Analysis of the log data followed standard quantitative procedures. As well, some of this data was analyzed by coding it for specific characteristics, such as being a substantive comment, on randomly selected samples of the data. The details of these codings and the samples are discussed in the Usage Analysis section below. The coding was done by the first author. The second and third authors reviewed the coding schemes and also the results.

Analysis of the open-ended comments on the user assessments was done by carefully reading the comments for themes and patterns, as is standard practice with qualitative

data [9]. These themes were discussed by all the authors, then re-read to examine the agreed-upon themes in more detail.

FINDINGS
In this section, we assess the usage of NB by examining the corpus of annotations and its creation process. We present evidence that substantial amounts of *collaborative learning* [8] occurred within NB. The annotations were primarily *substantive content* [8] regarding the course. Discussion threads were extensive. Students became active participants in questioning and interpreting the course material, with a large majority of questions by students answered by other students. Students interleaved annotation with reading, benefiting from the opportunity to see content and respond to content while in the midst of reading, instead of navigating to a different discussion site. Exploiting the geographic situatedness of annotations, students posted comments that addressed several distinct but co-located threads simultaneously.

Collaborative Learning
Assessing CaMILE [8], Guzdial and Turns identified 3 criteria that were deemed necessary to promote collaborative learning: broad participation, sustained discussion, and focus on class topic. We observed all three of these criteria. We cover each in turn.

Broad Participation
The 91 students created over 14000 annotations during the semester (averaging 153), while the instructor created 310. The average number of annotations authored per student per assignment was 3.67. This quantity *increased* over the course of the semester: a linear regression of this quantity over time shows it increasing from 2.73 to 4.2 per assignment, an increase of 1.57 ($p < 10^{-5}$). Although annotating was required, we take this increase over time as a sign of voluntary participation beyond the minimum requirement, suggesting that students found the tool useful.

The instructor also posted problem sets, on which *no annotations were required*. Nonetheless, 217 annotations were made on this material, in another demonstration of voluntary usage.

Sustained Discussion
Of the 14258 annotations, 3426 (21.4%) were *isolated threads*—single annotations with no reply, while the remaining 10832 (78.6%) were part of *discussions*—threads containing an initial annotation and at least one reply. For assignments, there were on average 13.9 discussions per page and 3.48 annotations per discussion. As shown in Figure 2, the thread length distribution exhibits a smooth decay, with over 400 discussion of length 5 or more, i.e. 1.4 lengthy discussions per page of material on average.

Focus on class topic
We read and categorized all 413 comments in 187 discussions for a typical 5-page reading assignment (a lecture on dimensional analysis, given in the middle of the term). We used
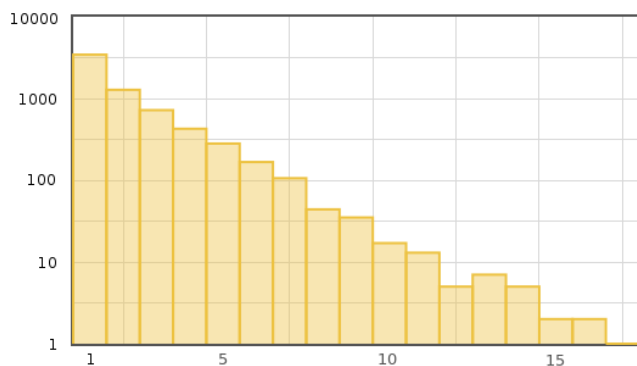
Figure 2. Distribution of the number of comments per discussion

Type	Number	Percentage
Substantive commentary	95	26%
Substantive questions	116	32.1%
…about concepts	74	
…about meaning of text	42	
Substantive answers	67	18.5%
…by students	57	
…by instructor	10	
Other	85	23.4%

Table 2. Breakdown of 363 class-learning comments (87% of the total).

Guzdial and Turns' [8] coding scheme of 6 categories in order to label the type of comments.[4] We found that annotations related to the objectives of the course (*class-learning* in Guzdial and Turns' coding scheme) represented an overwhelming majority of the comments —363 comments (88.1%) found in 164 discussions (87.7%). To gain further understanding, we subcategorized these 363 class-learning comments. Table 2 summarizes their breakdown.

Question Answering
A primary use of the tool was to ask *substantive questions about the material*, i.e. the result of a genuine thought process, stemming from an active and critical reading of the notes: 116 comments (32.1%) in 89 discussions (55.1%). These 116 were classified as 74 (20%) requests for help to understand a concept and 42 (12%) requests for clarification about the wording in the material.

A notable result is that these occurrences included a high rate of *substantive student-to-student teaching*: 57 replies (15% of comments, 85% of total replies) in 43 (48%) discussions aimed at providing a conclusive answer were posted by students. This was greatly appreciated by the instructor (see the The Instructor Perspective section).

Besides the student-to-student teaching, the instructor provided answers in 10 discussions (11%), and 2 questions were answered by their own author, leaving only 19 discussions (21%) without a conclusive answer. Of these, 9 were vague expressions of confusion, 2 were asked as "staff-only," 3 were

---

[4]namely: learning objectives, the technical tools (e.g programming environment), homework (grading, strategy), the collaboration tool itself, infrastructure (e.g., class pace, lecture quality), and off-topic (anything else).

Total questions	116
Resolved by student in same thread	59 (50.8%)
Resolved by student in different thread	14 (12%)
Resolved by faculty	10 (8.6%)
Not resolved	11 (28%)

Table 3. Breakdown of questions asked and their resolution.

asked after the assignment deadline, and 5 simply went unanswered. This is summarized in table 3.

In four discussions, we observed another important *study group* phenomenon: Students trying to propose several hypotheses and look for support from their peers, often ending their sentence with a call for confirmation ("right ?").

Besides the 183 substantive questions and answers (50%), we found 95 comments to the author/instructor (26%) regarding typos and suggested wording changes, and another 85 (23%) miscellaneous comments including brief agreements ("me too") and anecdotes.

Geographic Annotation
Users of NB we able to leverage the physical placement of annotations in a way that could not be achieved in a traditional forum. Of the 116 *substantive questions* voiced in the remaining 46 discussions, we found out that 13 of them (14%) were answered by a student, but on a *nearby* thread on the page. Each page in our sample had at least two threads that referred to another thread located nearby. In that sense, NB enabled a new behavior compared to regular (i.e. non-situated) forums: Participants can use the spatial proximity of threads to implicitly address questions that were posed in the surrounding threads. In the most impressive instance, a student replied to 6 surrounding questions by providing a single detailed explanation of *why the motion of the electron around the proton in the hydrogen atom can't be described by classical physics*. Although this was explained in the textbook, the explanation generated lots of confusion among the students (indicated by a multitude of annotations). Those very annotations prompted that student to re-explain the whole reasoning in his own terms.

Achieving such a holistic response in a traditional discussion forum would be very challenging. For a student to realize there were 6 distinct threads addressing the same question, she would have to keep a large number of discussions in working memory, or else rely on someone explicitly organizing discussions by (possibly non-obvious) topic. It's also unclear where the answer would go—which of the 6 relevant comments would receive the reply? And how could posters on the other 5 threads realize that their question had been answered, again without being able to remember large chunks of the discussion forum content or relying on someone else's topical organization? The spatial layout of the notes provides an implicit topical organization not available in traditional forums, and students clearly exploited it.

The geographic layout of the annotations also revealed particularly problematic parts of the text. Heavily annotated regions provided "heat maps" showing where lots of confusion

was present. Mahajan and other instructors reported exploiting this visualization to identify lecture content that needed clarification.

Tagging

As was observed in the context of the usage of *digital ink* [1], comments were often used to *tag* a section in the text with labels such as "confusing", "like", "dislike", and "easy". Those comments used lots of screen real estate to convey small bits of information, sometimes obscuring more substantive comments. Still, students reported that it was very useful to tag and see others' tags.

Examining comments of 5 words or less, we found that 375 of them (2.7% of the total) could be replaced by one of the following 8 tags without loss of meaning: *I agree, typo, cool/interesting/amazing, confusing, thanks, lol/funny, me too, what's this ?*. A tagging interface could have presented this information in less cluttered and more informative form, e.g. by color coding.

Continuous Ongoing Discussion

Although the number of assignments in our class differed from the WebAnn experiment [3], we found that the number of annotations per author per assignment were very similar: a bit more than 4.[5] However, these annotations classify differently than in WebAnn: the larger number of replies per author per assignment (2.53 vs 1.58 in WebAnn) indicates that students who used NB engaged in more conversations with one another. This difference is even more notable given that the WebAnn experiment *required* each student to enter at least one reply per assignment, whereas the class using NB had no such requirement.

One possible explanation for this difference might be the difference in online versus offline usage of the two tools. NB users rarely printed the lecture notes—our end of class poll estimated only 16.9% (N=26 and SE=5.16) ever did so. In contrast, WebAnn users printed lecture notes systematically. Common practice (cf. [3], p. 4) was to print and annotate a paper copy of the notes, and at some later convenient time "transfer" the annotations online. There are plausible rationalizations for this offline usage. WebAnn users lacked ubiquitous access to the Internet and the WebAnn software (which involved a special browser plug-in). The user experience with 2001-vintage Web applications was poor, and students had less experience working online.

Regardless of the reason, WebAnn's offline usage created a large lag between the time an annotation was first recorded (on paper) and when it could be read and a reply generated. And students who printed too early might never see some comments at all. To address the problem, Brush et al. [3] found it necessary to enforce two separate deadlines: Tuesdays at noon for submitting initial comments, and Wednesdays before class for (required) replies.

In contrast, the fact that many NB users were reading online (so getting up-to-date views of annotations) drove ongoing discussion and rapid responses. Students using NB particularly appreciated the fact that they could read, comment, and reply all at the same time, and get clarification on confusing points in the lecture notes in a timely fashion (cf. the "Student feedback" section). NB yielded a much greater proportion of replies than WebAnn, without imposing WebAnn's differential deadlines or specific requirement to reply.
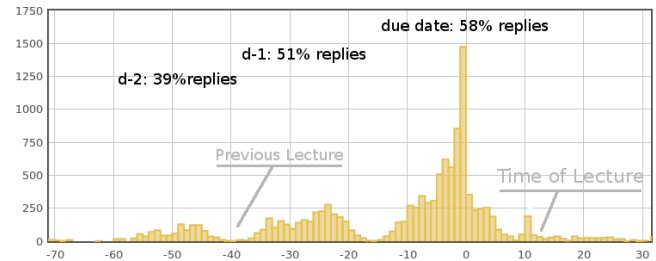

Figure 3. Distribution of intervals (in hours) between the comments creation time and the corresponding assignment deadline

The ongoing nature of the interaction is confirmed by Figure 3, which presents the number of comments posted as a function of the time (in hours) between a comment creation time and the deadline for the corresponding assignment (10PM on the day before lecture). We can observe 3 main clusters, corresponding to annotations authored by students who began working on their assignments respectively 2 days before (1047 annotations i.e. 7.7%)[6], 1 day before (2682, i.e. 19.7%) and on the due date (7344, i.e. 53.7%). The remaining comments (2599, i.e. 19%) were authored mostly later, either as part of extensions, or when a old discussion was revived, typically before an exam.

In summary, Figure 3 shows that NB participants didn't experience the problem of *discussion seeding* that WebAnn did - i.e. assignments done right before the deadline, which produce rushed single comments rather than helpful discussions. Clearly, there is a peak of activity in the few hours before the deadline, but since many comments have been entered already, there are many opportunities for discussion. In fact, even annotations entered by "early-bird" students 2 days before the deadline were spread out enough to enable discussions on that very same day: 39% of comments entered on that day were replies.

Annotating in the Flow

A strong motivation for our design of NB was the hypothesis that discussion can be improved if it is situated in the context of the document. Letting readers comment without leaving the reading environment meets the goals of keeping the user "in the flow" of their work, rather than interrupting it [2]. It also means that readers can encounter and respond to comments and question as they read, instead of having to go hunting for relevant comments.

---

[5]In a study of how peer-feedback can increase the relevance of online discussions, van der Pol [14](chapter 4) also reported 4.7 annotations per student per (weekly) assignment.

[6]We find it remarkable that so many students began working on their assignment on day $d - 2$, which corresponded to the evening before the previous lecture

Given this hypothesis, we tried to measure whether such "in-flow" annotation happened. More specifically, we looked whether *opportunistic replying* occurred, namely writing a reply to a comment while reading the lecture notes. We took two approaches.

Our first approach considered the distribution of annotation times over a "reading session", i.e. over a single time span that users would spend when doing their reading online in order to prepare for lecture. We used log data to identify the beginnings and ends of sessions. We focused attention on sessions of length between ten minutes and one hour, assuming that shorter sessions may have reflected quick look-ups of specific bits of information, and longer sessions may have included substantial multitasking or idle time or logging errors. We looked at the 6544 annotations that were made during those typical reading sessions. We scaled the times of those annotations as a fraction of the total time spent reading and plotted the distribution. Overall, this distribution is flat, showing that annotations were being authored throughout the course of typical reading sessions. We did the same for the subset consisting of 3676 replies, and found that it too was flat, suggesting that readers were replying to comments in the midst of reading. Figure 4 shows this distribution for replies (the distribution for annotations is similar).
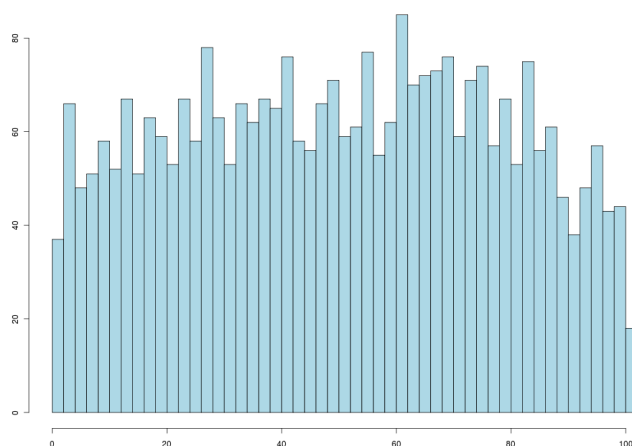

Figure 4. Distribution of relative creation times for replies over the course of a reading session.

Our second approach considered reading activity on single pages, and determined whether the (relative) time a reply was authored was linearly related to the position of the thread on that page, which would suggest that replies were written as the reader traversed from beginning to end of the page.

Again, we normalized the time of writing as a fraction of the total time spent reading each page (we logged entries and exits to each page), and correlated that normalized time to the position of the annotation on the page (all readings in the class were single-column, so reading ran linearly from top to bottom). We filtered out pages where students spent less than 10 seconds or more than an hour, and data points where the normalized time wasn't in the $[0, 1]$ range (due to measurement errors such as clock differences between client and server).

This resulted in analyzing a set of 3826 replies, for which we found a linear regression slope of 0.47 ($p < 10^{-15}$), and a adjusted $R^2 = 0.1125$. This implies that a statistically very significant portion of the user's placement of replies can be "explained" by the user placing them at the position indicated by a linear read through the text.

USER ASSESSMENT

The general utility of NB was also demonstrated in student and faculty feedback. Students reported that using NB helped them learn. They felt the level of class discussion to be quite high and valuable to them in understanding. Anchoring the discussion in the material motivated students to return to the material, which they argued benefited their learning. The instructor reported that NB helped him to teach better and also observed that it let students be involved in a genuine discussion while trying to understand the material.

Student feedback

At the end of the term, students were asked to fill in an optional web-based poll. We wanted to know more about their annotating practices (for example, whether they print the material or annotate while reading it online) and how NB had helped or hindered their understanding of the material. Of 91 students, 37 (40%) responded. However, not all students completed the survey, so we report varying N's below.

Overall, students valued NB. They were asked how they felt that NB had impacted their learning during the term, on a 5-point scale (1: very positively to 5: very negatively). The response was positive with a mean of 1.72 (N=37).

We also analyzed the comments that accompanied the ratings. We found three themes:

Significant Discussion and Learning
First, students appreciated seeing others' efforts, including the answers to their own questions by other students but also questions asked by peers. Some students felt that they were engaged in a helpful discussion about the material:

- *Never had this level of in-depth discussion before. . .*
- *It was cool to see what [sic] other people's comments on the material.*
- *I really enjoyed the collaborative learning. The comments that were made really helped my understanding of some of the material.*

Students liked being able to get questions answered in timely fashion:

- *I was able to share ideas and have my questions answered by classmates*
- *Open questions to a whole class are incredibly useful. Everyone has their area of expertise and this is access to everyone's combined intelligence*
- *Due to the considerable number of people in the class and the requirement to make annotations, responses are prompt and predominantly helpful*

This led to a general sense that NB allowed much more inter-activity in the reading:

- *The volume of discussion and feedback was much greater than in any other class.*

The student-to-student teaching as well as automatic email notifications when an reply was posted seemed to make the feedback time acceptable: On a scale ranging from 1 (strongly agree) to 7 (strongly disagree), students reported an average of 3.04 (N=27), i.e. "Somewhat agree" to the statement "When I ask a question using NB, I usually get a timely reply".

Situated Annotations
Although the comments above show that students appreciated the in-depth discussions, these could equally have taken place in a traditional forum (though they often do not). However, other comments showed how students specifically valued the situating of the discussion on the text:

- *The commenting system on NB is really useful because it allows us to challenge the text and each other and to see feedback from others taking the class.*

- *Being able to read the comments of others allows me to review the text more than once based on these comments*

The first quote, referring to "challenging the text," shows how the primary material was kept central to the discussion, unlike in a separate discussion forum. The second emphasizes the role of comments that are present *while reviewing the text*.

Indeed, students felt that NB provided additional motivation to do the readings and interact with them:

- *[NB] forced me to read the text and interact with it.*

- *It forced me to read the "textbook" which I don't usually do. It forced the professor to break it down into chunks, making material more concise and less repetitive/tedious*

Understanding where problems are
Earlier we discussed the "heat map" effect of seeing where comments cluster densely. Students were asked to rate whether NB helped them understand where their classmates had a problem on a 7-point Likert scale (1:strongly agree, 7: strongly disagree). The class had agreement that this was often true (mean=2.03, $N = 28$).

Open-ended answers to this question also provided evidence that students found their ability to see the confusion of others to be helpful for self-assessment:

- *It's encouraging to see if I'm not the only one confused and nice when people answer my questions. I also like answering other people's questions.*

- *. . . [NB] helps me see whether the questions I have are reasonable/shared by others, or in some cases, whether I have misunderstood or glossed over an important concept.*

The Instructor Perspective

We interviewed the course instructor, Sanjoy Mahajan, to understand his motivations and practices while using NB. Mahajan reported that the impact of NB on his class was very positive. Conversely, we speculate that some of the success that NB had in his class is due to the way Mahajan modified his teaching practices to take advantage of NB.

Adapting to NB
Guzdial and Turns [8] urged exploring how the instructor's involvement impacts ". . . his or her willingness to explore further uses of information technology and to participate in educational reform". One possible reason that NB worked so well in this class could be that Mahajan adjusted his teaching style to exploit NB. As we discussed in the opening of the Usage Analysis section, Mahajan had already incorporated a "reading memo" practice into his class. He thus had a sense of how to motivate students to make annotations as well as how to take advantage of them.

Mahajan required use of NB, but his requirement were deliberately vague: students had to submit one or more comments that showed "decent effort". This was guaranteed to receive full credit, regardless of whether the author was right or wrong. Students had to provide a *steady effort* by commenting on every lecture, but were automatically allowed up to eight extensions of 1 week each. Two students interviewed in our focus group indicated that since they didn't know what "decent effort" really meant, they used their common sense in order to participate in a "decent way" (i.e. contribute an interesting participation given their other time constraints).

Mahajan also emphasized to the students that unlike problem sets, where faculty are assessing whether students get the right answer, student annotations were assessments of how well he was doing as an explainer. This created an atmosphere where students valued the chance to make comments on material written by the faculty.

Feedback

The WebAnn study [3] reported that on-line comments often competed with in-class discussions. Mahajan observed the opposite: he explained that NB was an unprecedented success for his class, because he was now able to adjust the contents of his *upcoming lecture* in order to address the confusing points mentioned on NB. Comments were due at 10pm on the day before the lecture. He would begin reading them around 11pm and adapt the material in time for his lecture starting at 11am the following day. He reported that the sheer amount of page-flipping would have made this impossible using his previous paper-based submission approach. In the sample lecture we analyzed, we found 3 requests to use simpler examples, 2 requests to review/explain a concept during class (Mahajan replied that he would try), and 4 notes mentioning something that had been seen in class. In-forum and in-class contents seemed to complement each other.

Finally, Mahajan mentioned that the "part that [he had] underestimated about NB", and which "turned out to be really important" was the extent at which students answered each other, which is why he only needed to participate in 10.4%

of discussions. This connects with our discussion above, that students found responses timely.

DISCUSSION AND FUTURE WORK

NB has provided evidence that an in-place social annotation tool can be adopted and considered of positive educational value by both faculty and students in a modern classroom. In an attempt to understand how and why this adoption takes place, we have centered our analysis on showing that NB promoted *student-to-student teaching*; and that NB's in-place nature encouraged integrating annotations during reading, making WebAnn's enforcement of separate deadlines for comments and replies no longer necessary. Here we discuss ramifications and interesting open issues.

Is Data from a Single Class Convincing?

Clearly, we benefited from an very talented and motivated faculty user of our system. One might fairly ask whether "other" faculty could expect to see any of the same benefits. While a detailed analysis of how different faculty affect outcomes must await a future paper, Table 1 demonstrates that many other faculty at several other institutions were able to achieve significant adoption, some approaching the best case studied in this paper, even though few of them had previously made use of reading memo requirements. We cannot yet report whether adoption in these other classes was determined by the same factors as the one analyzed here, or entirely different ones. At a high level, however, we can confirm that numerous faculty believed that the tool was a useful enhancement to their teaching practice.

Of course, some preconditions apply to successful usage of NB. As one reviewer noted, "Their technology is good for students in highly connected environments who all have computers and for teachers who are tech savvy and lecture using online materials rather than a textbook. As a counter example, the tweedy old-school professors at my husband's less than super-tech-savvy graduate school who all use textbooks would not be a good target for this technology." However, we believe that the necessary preconditions are already quite common and becoming more so.

Adoption versus Learning Outcomes

The Holy Grail of an educational tool is improved learning outcomes. Assessing learning outcomes is always difficult. Here, we settled for assessing adoption by faculty and, secondarily, students. Numerous faculty have voluntarily adopted the tool, and numerous students have gone beyond the requirements in using it. It is conceivable that all these faculty and students are misguided, and that NB is not in fact enhancing learning outcomes. However, we feel that so many faculty and students are likely on to something, suggesting that improved learning is happening.

The Importance of Class Requirements

Effective use of a social annotation system in the classroom isn't only about designing the right tool; motivating usage is also essential. Mahajan believes that annotations need to be a requirement. He believes that a key aspect is to have usage requirements and guidelines that *aren't too strict* (i.e. reasonable and steady effort vs. a required number of comments per week). In another class, where the requirement was set as "exactly two annotations per lecture," the students met that requirement exactly and never exceeded it. On the other hand, another class where annotations were not required at all did nonetheless see substantial usage. Clearly the question of effective motivation to annotate requires further investigation.

A Finished Product versus a Process

As some classes have begun to use NB several times, an interesting question has emerged about whether or not keep previous terms' annotations available for future student use. To the extent that these annotations clarify the material, it seems natural to preserve the "improved" lectures plus annotations for the next group's use. In practice, faculty users of NB invariably discard the old annotations. They say that the *process* of discussing the notes in the margins is considered a valuable contribution to the students' learning, which would be lost if past comments were already available for reading.

At the same time, marginal notes can provide an effective contribution to a text's narrative. Knuth's Concrete Mathematics [7], a traditional textbook, publishes in its margins a selection of the marginal comments recorded by students using a draft in the first version of the class. These comments add insight, humor, and unique student perspectives without disturbing the main narrative. We believe there would be value in tools that help instructors to *curate* annotations, selecting some to drive changes in the text, some that would be most valuable remaining as marginal notes, and some that should be removed so that future classes can rediscover them.

Better Support for Annotation's Specific Affordances

Our users discovered and exploited certain capabilities of annotation that are not present in traditional forums. We can provide better support for those capabilities. Above, we discussed how geographic annotation was leveraged to answer sometimes-multiple questions in other threads. It would be useful to capture this answering behavior in the thread structure, for example to let an author explicitly mark (multiple) threads to which they were responding. We also discussed the use of annotations as tags, and suggested there could be value in directly supporting tags presentation through less cluttered and more informative interfaces such as color coding.

CONCLUSION

Our development of NB was driven by several design hypotheses about the way an "in-place" annotation tool could outperform traditional forums as a medium for discussion of classroom materials. Situating discussions in-place allows students to annotate and question while reading, remaining in the flow instead of losing context on a different forum. It draws student attention to relevant discussion at the moment they are reading the material, instead of requiring them to consider that there might be relevant discussion and search for it (and retain the context) in a separate environment. It allows them to consider all relevant discussion threads drawn

together by physical proximity, instead of organized by posting chronology, and author answers that draw many of these threads together.

Our deployment of NB has provided evidence supporting these hypotheses. In our "best-use" class, students contributed 14,000 distinct annotations, outdoing by a factor of 4 the combined product of the 50 most active classroom discussion forums at the same university. Students and faculty gave significant positive feedback regarding the role of NB in the class. Data show that students write and read comments in tandem with reading the primary materials, and exploit the geographical coherence of annotation to draw multiple threads together into substantive discussions.

From our experience we were able to draw the following design conclusions:

- Current students do abandon paper for online reading. We hypothesize that the gain of interactivity (access to the latest comments, asking a question while reading) outweighs the irreplaceable affordances of paper as a support for reading, described in [11].

- Students interleave annotation with reading, implying that it must be kept easy to annotate while in the flow of reading. For instance, we recommend against using modes or required fields.

- Students combine response to several geographically co-located threads, implying that future tools should support marking multiple threads for simultaneous reply.

- Requiring annotations may be necessary at least at the beginning of the term, but students learn to value them and go far beyond the requirement.

- There is demand among faculty for a tool to stimulate student feedback and discussion. Feedback can happen at a timescale that allows adapting the following lecture based on the questions and comments from the previous lecture and the reading assignment.

- Student-to-student feedback is far faster than faculty feedback. Students overwhelmingly appreciate that fast response time. The design implication is that students should be able to discover questions that are currently being asked. Future system should help students differentiate between "stale" conversations and the ones that are worth reading.

- Previous studies argued that adoption barriers prevent online annotation. We refute this. Yet, instructors should be made aware that such online communal annotation tools aren't a one-size-fits-all solution. This paper's best case is an example where it worked wonderfully, but future work will need to uncover when and why it does and does not through comparative longitudinal studies.

NB offers an "existence proof" that it is possible for an online collaborative lecture-note annotation system to succeed in a classroom setting. This contrasts with experience using the technology of previous decades. Whether this is due to changes in teaching style, changes in technology, or changes in the expectations of users of that technology cannot be clearly worked out from this single case study. However, the evidence suggests that we have reached a turning point where online social annotation systems could become a standard and valuable educational tool.

REFERENCES
1. Anderson, R. J., Hoyer, C., Wolfman, S. A., and Anderson, R. A study of digital ink in lecture presentation. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (New York, NY, USA, 2004), 567–574.

2. Bederson, B. B. Interfaces for staying in the flow. *Ubiquity 2004* (September 2004), 1–1.

3. Bernheim Brush, A. J., Bargeron, D., Grudin, J., Borning, A., and Gupta, A. Supporting interaction outside of class: Anchored discussion vs. bulletin boards. In *PROC. OF CSCL 2002*, G. STAHL, Ed. (2002), 425–434.

4. Blitzer, D. The wide world of computer-based education. *Advances in computers 15* (1976), 239–283.

5. Bonwell, C. C., and Eison, J. A. Active learning, creating excitement in the classroom. *ERIC Clearinghouse on Higher Education* (1991).

6. Chickering, A. W., and Gamson, Z. F. Seven principles for good practice in undergraduate education. *American Associassion for Higher Education (AAHE) Bulletin* (1987).

7. Graham, R., Knuth, D., and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Addison-Wesley, 1994.

8. Guzdial, M., and Turns, J. Effective discussion through a computer-mediated anchored forum. *The Journal of the Learning Sciences 9*, 4 (2000), 437–469.

9. Miles, M. B., and Huberman, A. M. *Qualitative Data Analysis*, 2nd ed. Sage, Thousand Oaks, CA, 1994.

10. Scardamalia, M. Csile / knowledge forum. In *Education and Technology: An encyclopedia* (2004), 183–192.

11. Sellen, A., and Harper, R. *The Myth Of The Paperless Office*. MIT Press, 2001.

12. Taylor, E. F. Guest comment: Only the student knows. *American Journal of Physics 60*, 3 (1992), 201–202.

13. Thomas, M. Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning 18*, 3 (Sept. 2002), 351–366.

14. van der Pol, J. *Facilitating Online Learning Conversations*. PhD thesis, Universiteit Utrecht, 2007.