Learning from Other Perspectives:
Design and Analysis of an In-Place Annotation System

by

Sacha Zyto

Submitted to the Department of Mechanical Engineering
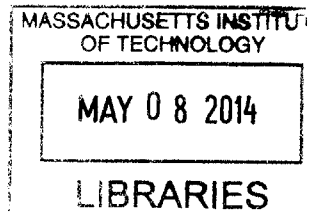in partial fulfillment of the requirements for the degree of
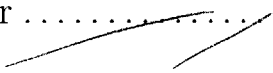
Doctor of Philosophy
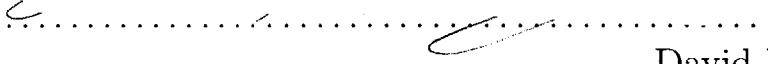
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2014

Author . . . .
Department of Mechanical Engineering
Oct 8, 2013

Certified by . . . .
David R. Karger
Professor
Thesis Supervisor

Accepted by . . . .
David E. Hardt
Chairman, Department Committee on Graduate Theses

Learning from Other Perspectives:

Design and Analysis of an In-Place Annotation System

by

Sacha Zyto

Abstract

I designed and studied NB, an in-place collaborative document annotation system 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*. Questions, replies and comments from students and faculty members are displayed in place and provide new perspectives on the content. NB also provides comment browsing interfaces that help the staff cope with reading assignments in large classes.

I describe the NB system and its evaluation in real class environments, where students used it to submit their reading assignments, ask questions and get or provide feedback. I show that this tool has been successfully incorporated into numerous courses worldwide, and that students prefer to use NB to read their notes, rather than printing out copies that are missing these annotations. The data I collected indicates that NB encourages students to comment on the class material, even students who are not verbally active in class.

To understand how and why, I focused on a particularly successful class deployment where the instructor adapted his teaching style to take students' comments into account. I analyzed the annotation practices that were observed – including the way spatial locality was exploited in ways unavailable in traditional forums. I then surveyed 30 faculty members from classes where NB was substantially used and set up an A/B experiment in an edX course, where only half of the students had access to NB. Contrary to previous literature results, in-class participation, in-place annotations and forum annotations do not necessarily compete with each other. From those observations, I derive general design implications for online annotation tools in academia.

Thesis Supervisor: David R. Karger
Title: Professor

Acknowledgments

Thank God, I was privileged to enjoy every day of my long journey though graduate school. This was possible thanks to a few wonderful people who were always available for me. Hence, I would like to express them my profound gratitude for making this journey an unforgettable experience.

Thanks to my wife Ora for always being by my side, understanding and supportive, and pushing me to move forward. Thanks to my advisor David for his wonderful guidance. Thanks to my kids Betsalel, Yehuda, Naomi and Malka for giving me so much *nachas*. Thanks to my parents and the Gronstein family for their constant support.

Thanks to my thesis committee: Sanjoy Mahajan, for his continued support, help and personal involvment, without which this research wouldn't have been the same. Dave Gossard for encouraging me to look further into what started as a casual conversation back in 2006.[1] Rob Miller for teaching me so much about HCI, and his continuous excellent advice. Maria Yang for her excellent suggestions, and support and for managing to convince me that it was time to graduate.

Thanks to my lab members and collaborators: Mark Ackerman for teaching me how to observe real users and precious help and advice for our CHI 2012 paper. m.c. schraefel, Michael Bernstein, Katrina Panovich, Max Van Kleck, David Huynh, Adam Marcus, Erik Bakke, Ted Benson, Peter Lai, Vineet Sinha and Harr Chen for their help and for showing me how amazing the field of HCI can be. Thanks to the my friends from the UID lab, including Juho Kim and Zoya Bylinskii. Thanks to our UROPs, interns and contributors: Colin Zwiebel, Eyas Alsharaiha, Stephanie Greene, Carolyn Chang, Danny Sanchez and Jen Liu.

Thanks to everyone I met over these wonderful years: Jean-Jacques Slotine, Leslie Regan, Franklin Reynolds, Zoe Antoniou, Dimitris Kalofonos and Jake Solomon. Thanks to all my collaborators at Harvard, CopyrightX, OEIT, and faculty in all universities who have shared so much of their time to help explain their results.

---

[1]Can we develop a program to get feedback from students while having them help each other?

Contents

List of Figures

11

List of Tables

Chapter 1

Introduction: The heavy toll of communication inefficiencies in class

If *the function of education is to teach one to think intensively and to think critically [...],*[1] then the responsibility to make education fulfill its function belongs to the entire community of education professionals, every actor according to their competences and their role. This thesis suggests a possible role for human-computer interaction (HCI) practitioners in education, in order to make education more effective. More specifically, based on both my personal experience of successively being a student, teaching assistant (TAs), and lecturer, and on what I found by interviewing other students, TAs and lecturers, I analyze the limits of existing communication channels commonly found in academia, and provide a tool that can be used for engaging students and faculty in fruitful discussions related to the material. This chapter discusses the limits of traditional communication channels, in order to come up with requirements that a communication channel should meet in order to facilitate effective discussion and learning.

---

[1] Martin Luther King Jr.

1.1 Traditional communication tools in class and their affordances

This section discusses the numerous ways for class participants to communicate, and their affordances relevant to effective learning and teaching.

1.1.1 In-person communication

In the traditional academic context, class participants can communicate in person in a number of ways. In every case, because the communication happens directly from person to person, there is no delay in receiving the information.

Lecture During a lecture, the instructor traditionally conveys information to the students, and sometimes prompts the students to ask or answer questions. If a question comes at the right time during lecture, then setting the context is easy: it's about what has just been covered. However, there are many factors preventing genuine conversation in a lecture setting: limited lecture time, high number of participants, fear of performing in front of other participants, a phenomenon which, according to [45], gets amplified with increasing class size and diversity. Besides, the norm is that every conversation should involve the instructor: student-to-student conversations tend to disturb the instructor and the other students, and make the participants lose focus quickly [36, 24, 25].

Lab In contrast, during labs, social norms typically encourage or expect student-to-student conversations to happen, at least among the group of students who are sitting at the same lab station. This works so well that some lecturers have adapted their lecture class to make it similar to a lab context: This is one of the key ideas behind the flipped classroom or peer-instruction methods [41]: Students are supposed to have read the material by the time they come to lecture, so that they can deepen their understanding by focusing on activities during lecture.

Recitation Recitations are generally a good opportunity to engage in real discussions. They often don't have as many participants as lectures. On the negative side, they typically happen only once a week. Hence, if students want to use recitations as an opportunity to get answers to their questions, they need to have them written down somewhere by the time they come to recitation. Setting the context of a question can be challenging for the student, who has to explain the detailed context of her question, and for the instructor, who might feel pressured to provide an answer on the spot.

Office hours Office hours are set times during the week when students can ask questions to the instructor or TA. There are variations. For instance, for teaching Discrete Signals Processing, Oppenheim has adopted a system of *open hours*, where everyone is welcome to attend. Similar to recitations, they suffer from the problem that they happen typically only once a week, so they don't allow for extended discussions at arbitrary times.

Study groups Student commonly form study groups to work on homework assignments or review for the quiz. They are a good occasion to ask questions and challenge each others' understanding. On the negative side, they often need to be scheduled. Instructors do not attend them so on the positive side, meetings can be more informal and student can think better than when they feel they're being observed by the instructor. On the negative side, they miss a source of authoritative answers and instructors miss the opportunity to see their students' misunderstandings.

Other opportunities There are other in-person communication channels available, which we could qualify as opportunistic encounters. Those would be conversations in a hallway or meeting at the cafeteria, for example. We haven't yet gotten enough data from class participants to report on their role in effective communication. However, because of their unexpected nature and the fact that participants might not have access to the resources they need to delve into the question at hand, they certainly can't be relied upon as a consistent source of effective communication.

1.1.2 Paper-based communication

Written communication is present in many different forms: Homework assignments, quizzes and term papers or project reports are occasions for instructors to gauge the understanding and level of engagement of students at regular intervals, typically several times a month. Genuine conversation is typically not possible on that particular medium since the student is always writing the first message and the instructor is offered one reply, ranging from a simple check mark to a whole paragraph. The timing of participation and reply is typically too slow to benefit the student. It also requires the student to have the mental discipline to go though a graded homework assignments and carefully read the instructor comments. On the instructor side, writing comments for each student is quite time-consuming. The workload can be split by having teaching assistants grade the assignments, at the cost for the instructor of not having a first-person look at the students' work.

Course evaluations are a way for students to give feedback to their instructors. The major problem with those is that they happen typically once, at the end of the term. Hence, the only students who benefit from them are the ones in the following course. Informal frequent feedback (for example, asking students to write down the one thing they understood the least in lecture at the end of each lecture) can be an effective way for the instructor to realize that a point wasn't clear in lecture in re-explain it later.

1.1.3 Digital communication

Unlike the communication tools that we reviewed so far, most digital communication tools have an significant advantage in that they are always available. Email is a fast and versatile communication channel. Email lists and forums offer the occasion for students to help one another, without direct intervention from the instructor. Furthermore, one advantage compared to study groups is that the instructor can follow along, thereby getting a precise understanding of what students are struggling with. Unlike direct conversation, setting the precise context of a question can be

particularly challenging since text is often the only affordance.

This is in contrast with direct communication where participants can take advantage of multi-modal communication to explain the problem (i.e. explaining the issue while pointing at the material for example). One well-known problem with forums however, is that users who have a intrinsic incentive to visit them are the ones who have questions, not the ones who have answers.

In addition to the cost of setting the context, forums and email list suffer from another severe issue. They require the user to visit a specific website. As soon as the user switches her attention from the material, the new application (forum site) starts competing for the center stage position in her mind. To make things worse, using the browser to visit a forum often turns out to be an opportunity for cyberloafing, which according to Ugrin and Pearson [52] accounts to between 60% and 80% of the time spent at the computer.

Finally, chat and video-conference tools (e.g., Skype, Google hangout or adobe connect) aren't typically part of traditional communication tools, although they have gained popularity in MOOCs. In the class we surveyed (cf section 4.3.2), video-conference modalities were used as a substitute for recitations.

1.1.4 Summary

We just reviewed the main channels class participants can use to communicate, and discussed the trade-offs they offer. Based on the characteristics we enumerated for each channel in the list above, we generated seven factors that can impact effective communication:

1. Can participants express themselves as much and as often as they wish (e.g. at least without a hard limit imposed by the channel itself)?

2. What are the costs of setting the context (e.g. explaining to which parts of the material the question is referring)?

3. At what times is communication possible (e.g. only once a week, vs. any time of the day)?

4. How fast can participants expect their message to be noticed by other parties?

5. How quickly can participants expect to get a reply?

6. Hard barriers: who can participate in theory (e.g. other students, the instructor)? Soft barriers: who does participate in practice?

7. How easy is it to recall the exchanged information when needed?

The last factor mentioned doesn't have at effect on effective communication *at the moment* per se. Yet, it has a large effect on effective learning, since the concept that had been understood during the conversation will typically be required much later (e.g. at the exam, or when reviewing for it). This factor affects almost all the communication channels we reviewed since they are separate channels by definition. Therefore, they're typically not fused together to provide a unified entry point into subsequent searches.[2] This means that the information is an extra step away from being available when needed. Students who aren't aware of that resource might simply not review it. Students who are still aware of it might also not use it, simply because of the additional search burden.

1.2   Effects on students

In the previous section, we surveyed various communication technologies available in a traditional academic context. Inspecting that list reveals that each channel suffers from at least one drawback. This means that some adverse factors cause questions not to be asked, which leads to seriously negative effects on students, both on a personal level (e.g. failing to develop self-confidence or efficient learning techniques) and on a professional level (lack of knowledge and skills). We now examine those effects in more detail.

---

[2]Technically, lecture wouldn't suffer from this issue, since conversations happening during lecture typically typically refer to the subject being lectures at the moment, so that they can appear on the lecture notes themselves, at the right place

1.2.1 Lack of active learning

The first unfortunate effect of asking questions is a lack of active learning. As explained in section 3.1.2 and in [28], passive learning yields only minimal and shallow understanding. In contrast, learning at a constructive or active level by asking questions is a requirement to mastering the concepts at a level where they can transfer to new domains.

1.2.2 Lack of critical reading skills

Passively reading the material doesn't develop critical reading skills. An example often quoted by Sanjoy Mahajan is that "Math is no spectator sport": In contrast to reading a novel, reading technical material requires critical skills. In a novel, it's acceptable or expected for the reader not to understand everything as she reads, so that the plot keeps on unfolding. In contrast, technical literature follows a logical order, where a concept explained first is supposed to be read and understood before moving on to the following one, which can build on the concepts explained before.

1.2.3 The same mistakes are repeated over and over

Another negative consequence of passivity in learning is that students will stumble upon the same mistake multiple times, either because they were not able to ask what was the correct answer the first time, or because the solution didn't make an impact on them. At the end of the term, those misunderstandings will have accumulated and the course subject isn't as well-understood as it could have been, which will have negative consequences for other classes where the subject was a prerequisite, and later in professional life.

1.2.4 Students lose or fail to develop their self-confidence

As witnessed in uncountable occurrences and given the popularity of talks on the *impostor syndrome* on campuses [31], students commonly doubt their own capacities

21

and achievements. In this section, we examine three different instances of missed opportunities to build one's self confidence that could be achieved by asking questions.

1. They don't explain to their peers: A question not asked means that students missed an occasion to answer the question, therefore missed on an opportunity to prove themselves that they understand the material enough to provide a satisfactory answer to the question.

2. They feel that they are the only ones being "stupid": A surprisingly popular feature of annotation forums is that discussions provide reassurance to readers that other forum members are struggling with the same problem as they have, or as they often put it, that they're not "stupider" than the others. Thus not asking a question deemed trivial prevents students with the same question from realizing that they were not the only ones struggling with this so-called trivial question.

3. They miss a chance to participate when they don't want to "perform": Besides the fear of being inadequate mentioned earlier, another popular fear is performing in public, which is why some students never ask question in class. Participating on a forum helps students realize that they can provide a useful contribution, which will be seen by everyone, and provide a bridge towards being able to participate in person.

1.3 Effects on faculty

Among the communication tools mentioned in section 1.1, labs and office hours are the only ones where re-creating the context is simple and where the instructor can witness a real exchange among participants. Hence, out of all the tools mentioned in section 1.1, there are only two where the instructor has access to the real conversation. This difficulty raises the the following issues.

1.3.1 Inability to obtain timely feedback on material

Faculty members are typically extremely familiar with the material they're teaching. This commonly leads to not being able to assess where students are struggling by going too fast on points that aren't obvious to the students, or, at the opposite extreme, forcing themselves to explain a concept in excessive details, causing the students to be bored during lecture.

As mentioned in [57], students can use clickers to provide feedback on the lecture pace, or use a system of "muddy cards" [43]. Yet there is little they can do to provide timely feedback on how detailed course notes should be (timely here meaning fast enough so that the faculty knows what to cover in more or less detail in the upcoming lecture). As an exercise designed to get students to acquire critical reading skills, Mahajan asked students to submit a *reading memo*, i.e. an annotated paper copy of the reading, following a practice from Edwin Taylor [49]. Although it fulfilled the purpose of getting valuable feedback on the material (for instance in order to suggest changes on a book preprint), the annotated material didn't come in time in order to help Mahajan adjust the contents of his upcoming lecture based on that feedback, something that he could do using our tool (cf. section 4.2.2).

1.3.2 Inability to motivate students to come prepared to class

Another related issue is that traditionally, most of the material is covered during lectures, and during that time students are minimally involved (see section 3.1.2). Given the limited amount of lecture time, this leaves little time for extra activities where students can gain a deep understanding of the material. This makes lecture uninteresting for a large fraction of the students, who decide to skip lecture. On the other hand, if instructors had a way to know what student really need to learn before lecture happens, they could focus on those points during lecture. This would require to have students study the material before lecture, and discuss it, so that the instructor can prepare for her lecture based on those discussions, something that

Mahajan (as we mentioned in the previous section), as well as at least another faculty members were able to do using our tool (see sections 5.2.6 and 5.6).

1.3.3 Inability to detect student mistakes from an early stage

Based on personal conversation with faculty at MIT, the worst inefficiency of traditional communication tools is the lag between a student not understanding a concept and the instructor realizing it. For students express their concerns and feedback only at the last lecture, when they fill out the survey from the department. Its results are typically available a month or two later. The results are useless in helping the instructor assess the understanding of the *current* material, in order to correct misunderstandings at an early stage.

1.3.4 Duplication of work

Finally, another major inefficiency is that answers have to be given multiple times, and not by the person who benefits the most in giving it: Indeed, it is quite common to encounter several students who have the same question or misunderstanding. Traditionally, the instructor or a TA will have to address the question multiple times, say by having the same conversation or writing a similar email several times. It would be much more beneficial for students to answer their classmates' questions. By doing so, not only would they be learning actively, they would also help with the staff workload so that the staff can address the harder questions.

1.4 The case for a situated forum

Among the communication channels we reviewed in section 1.1, none has all of the seven characteristics that facilitate effective communication: All channels except lecture (and maybe labs) suffer from the problem of adding an extra step to recall the information. In addition, in-person channels mostly require communication to happen at a fixed set of times. On the other hand, digital communication channels suffer

24

from having to explain the context, are a source of distraction, and often require users to visit a site or use an particular application, which student might not even visit unless given extrinsic rewards (e.g. participation grade) [40].

Unlike other digital channels, a situated forum doesn't suffer from those drawbacks. A situated forum is an online forum where discussions about a topic appear precisely where the topic is discussed in the original document. Recalling the information is easy, since the conversation appears exactly where the topic appeared in the original material. Participation barriers are dramatically lowered: First, questions are visible by everyone, therefore making every reader a candidate to reply. Second, explaining the context is obvious, making it much easier to explain a problem. This dissertation describes the design, and evaluation of such a situated forum, as well as the lessons learned, and perspectives for improvements and future research.

As we point out in sections 4.2.2 and 5.6, introducing a new channel in the class ecosystem is a delicate operation. Its usage can be greatly affected by the role instructors give it, and it can affect other channels in return. Hence, understanding the usage of a in-place forum can't be dissociated from analyzing its usage in the context of the class ecosystem, in order to find the best practices, and make them known to the community.

1.5   Contributions

1. We have identified some of the major issues and inefficiencies that prevent instructors from getting timely feedback from the students, and that prevent students from being able to participate actively on the material.

2. We have designed a system to make discussion on a topic seamless and assessed its success according to several metrics: usage logs, students and faculty surveys. Our results underline the great importance of planning the interaction of the tool within the class ecology.

3. We have observed novel interaction techniques: in-place discussions break from

the constraints of having comments strictly organized by threads, and participants reply in the flow.

In summary, we designed and deployed a situated discussion forum, showing that online in-place collaborative lecture-note annotation systems can succeed in a classroom setting. Its success contrasts with experience using the technology of previous decades. This success is mainly reflected by a greater and broader student participation and timely feedback to students and faculty.

1.6 Thesis overview

Chapter 2 surveys some of the education and computer-supported cooperative work (CSCW) literature on online forums, with an emphasis on an earlier situated forum experiment [20]. Chapter 3 describes our design motivations and the implementation of the NB system. Chapter 4 describes our data gathering techniques and summarizes our experimental designs. Chapter 5 describes our findings and their implications. Chapter 6 discusses possibles venues for future work, before some concluding remarks in Chapter 7.

Chapter 2

Related work

Many classic web tools provide a base for collaborative work. Wikis let any registered user edit their contents, which is not desirable in a class setting since the difference should be clear between the authoritative contents (authored by the faculty) and the discussions on that content. Also, wikis lack the proper support for managing students' submissions efficiently (i.e an instructor mode).

Blogs do provide a distinction between authoritative contents and comments, but the comments are usually situated at the end of the page, thereby forcing readers to switch their attention between the content and the comments.

To the best of our knowledge, current course management systems, whether commercial such as Blackboard [1], open-source such as Sakai [2] and Moodle [3] or university-specific such as Stellar [4] do not offer the possibility to attach annotations to an arbitrary part of any document.

Supporting in-place annotations and improving online educational material have received a lot of attention from the HCI, CSCW, and e-learning communities. Presenting all the related projects that were developed in those communities would be outside of the scope of this thesis. We have focused on the ones that were most relevant to us, either because they described studies of comparable projects in the field of digital annotations or because we applied their conclusions to the design of NB. We first review the relevant work related to domain of collaborative discussions in an educational setting, and then additional references that were influential in the design

of NB.

2.1   Collaborative discussion tools in education

2.1.1   Traditional annotation tools (not in-place)

While there is relatively little current work, the past abounds with studies of collaborative discussion tools for education. It is accepted that students understand material better after discussing it [23, 29]. 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) [22]. *CSILE* (1984) and its successor *Knowledge Forum* (1995) [46] 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. 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 [51] 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.[1] (At the same time at MIT, one 91-student class using NB generated over 14,000 posts.)

---

[1]An important caveat is that Stellar is not a particularly *good* discussion system. Over the past few years, a forum tool called Piazza has begun to see widespread adoption; as part of this dissertation, we have not had the opportunity to analyze its usage, which outperforms that of Stellar.

28

Improving on this "detached" situation, CaMILE [34] 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 toward 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.

2.1.2 In-place annotation tools

The *WebAnn* project [20] 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 (2001), some factors limited the benefits of the tool. Mainly, students 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* [53] 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.

2.2 Other influential work

The rest of this chapter focuses on related work which doesn't fit per-se in the context of collaborative tools in education but whose design and analysis were influential in the design of NB.

2.2.1 Enriching traditional contents

XLibris The idea of enriching traditional digital contents without losing the context had already been explored 15 years ago: XLibris [47] is a tablet-based program developed for annotating electronic documents as intuitively as if they were on paper. Striving to reproduce the affordances of paper-based annotations in order to support *active reading* explains the key design choices that the Xlibris designers adopted, namely supporting free-form annotations (pencil strokes) and a minimalistic user interface. In addition, notes can be shown in the *Reader's Notebook*, which is a collage view of all the user's annotations, along with a snippet of the original document, in order to be able to understand the annotations in their context. However Xlibris doesn't store annotations in a centralized repository and therefore wasn't designed to be a collaborative tool, but rather to be an active reading machine and to manage a collection of documents and personal notes.

Anchored Conversations by Churchill et al. [30] enhances Microsoft Office documents with a persistent chat feature. Users can anchor chat frames in various places of their documents thereby allowing short comments to be much more effective, because they are presented in their context. This feature, in turn facilitates remote collaboration.

Classroom Presenter by Anderson et al. [18] lets faculty and students enhance Powerpoint slides with digital-ink annotations taken during the lecture, using a tablet PC. It can be used in the classroom to present slides in a more interactive fashion or to initiate participatory exercises: Students draw their answer on their Tablet PC.

Those answers are visible to the faculty, who can in turn present them to the whole class, a typical design studio practice.

An important result of this work was a typology of the purposes for which the faculty use digital ink: attentional marks, diagrams, and text. The surprising fact was the importance of digital ink used as an attentional mark, and that those marks (more precisely, their timing) were an important help to the student. The second observation was that instructors tended to use system features in a very parsimonious fashion, preferring to use the commands that required the fewest steps, albeit not being the optimal ones for their task. This last observation advocates for a minimalist design.

SparTag.us, by Hong and al. [37] lets users tag any region of a web page. Compared to online bookmarking and tagging systems available at the time such as del.icio.us, Diigo, and Google Notebooks, SparTag.us strives to be as non-disruptive as possible, by letting users tag a web page literally as they read it.

Multivalent Annotations The systems described above were intended to enrich the content of a traditional document. This is in contrast to Phelps and Wilensky's multivalent annotations [44] system, where annotations can be used to alter the original document structure.

Stet [5] is a software used to to collect in-place comments on arbitrary webpages. It was used to collect comments on the revision of GNU General Public License [6], which is, according to [7], the most widely used free-software license.

2.2.2 Stimulate audience participation

eClass An important aspect of improving student participation during lecture consists in helping students to follow the lecture contents rather than being busy with scribbling lectures notes. This was one of the specific goals of *eClass* project [26].

Previously known as Classroom2000, eClass has been jointly developed at the Georgia Institute of Technology and at University College, London. eClass captures

31

the audio and video contents from the classroom in addition to the slides, and makes them available from a browser. Since students spend a lot of time copying down course material (an average undergraduate student totals about 24 hours per week in lectures, labs and recitations) and also a lot of time studying it, there has been much research devoted to improving teaching methods and contents available to students. Among the many projects that were developed in this field, eClass became a reference partially because of the size of its deployment (over 100 classes taught using it during the first 3 years after its introduction).

The first main positive conclusion from the eClass survey [26] was that students notes became more concise, indicating that students felt less compelled to copy everything the instructor wrote on the board in order to understand it later (the scribe effect). By analyzing the notes students took, the authors were able to infer that students were following along, since the notes they took were much more succinct and specifically focused on the things they thought were worth writing down.

Another positive conclusion was that eClass didn't significantly incite students to skip lectures. It did not lure them into thinking that it replaced direct interaction with the teacher.

On the negative side, it required the classroom to have some kind of specialized equipment, such an electronic whiteboard and a recording device (camcorder), which raised portability and cost issues. Besides, it required a log-in procedure and minor setup, which lecturers generally prefer to avoid because it takes time at the beginning of the class.

One important aspect that was the eClass study revealed was the pattern in how students access multimedia class content: they flipped through the lectures (i.e. skipping back and forth through the multimedia contents to find short key points that are not clearly understood). This suggested that indiscriminate video recording was not desirable, and that the indexing mechanisms should be improved. Finally, another reported issue was privacy: First, knowing that lectures were being recorded didn't encourage students to participate in class, since they felt that their participation in class was now part of a recorded performance. Besides, the instructor would

often need to manually delete a sequence (slip of tongue, comment that would be inappropriate to keep on the record), which defeated part of the automation process.

Multitasking. Finally, it should be noted that tools designed to be used during lecture inherently compete with the instructor for the students' attention. This can have various adverse effects. For instance, in [42], presenters reported feeling uncomfortable with a large fraction of the audience using their laptop during a conference, and the speaker not knowing whether they were using the laptop for a related discussion on a backstage IRC channel, or for doing something completely unrelated. Another effect is the impact of multitasking on memorization: In [36], Hembrooke and Gay evaluated how letting students use a laptop during lecture would impact their memorization of the lecture content. The purpose of their study was to provide general cognitive psychology results, which would be independent of what the students were using the laptop for. The study yielded three main results:

1. Students involved in multitasking during the lecture (i.e. using their laptop) did significantly poorer on memory-based quizzes immediately following the lecture. This could be understood by a famous result from cognitive psychology, known as Broadbent's theory of selective attention [24, 25]. It states that ignoring information happens from an early sensory stage. Hence, the information that was ignored will not make its way to the long-term memory.

2. Whether the activity they were engaged in was related to the class content or not didn't seem to affect their performance: the main factor was the duration of the browsing sessions they were engaged in.

3. On the other hand, asking students to do an active task that forced them to manipulate the concepts on which they were going to be quizzed (sorting words among categories etc...) in semantically relevant ways did improve their performance compared to students who received no instructions.

Chapter 3

The NB system: motivation and description

As we mentioned in the introduction, traditional communication channels in academia suffer from many inefficiencies. Being a TA and lecturer, we first-hand witnessed those inefficiencies, in addition to hearing about them from feedback we received from other TAs and instructors. Our very first motivation to develop NB stemmed from our frustration with traditional communication tools. In this chapter, we first present how creating an improved communication channel can improve active learning. Then we dive into the description of the NB system, underlining the usability decisions which motivated our design decisions.

3.1   Why should we try to improve communication channels?

As we just mentioned, the inefficiencies of traditional communication channels result in a lack of communication, which, in turn, makes the learning process more passive. In contrast, we strove to design a system that would benefit from the advantages of in-person discussion, where the context is obvious, and of online forums, which are always available, and study how users benefit from it. As in the rest of education

research, the most tangible benefit would be improved learning outcomes. However, as we detail in the upcoming section, assessing the impact of a particular technology on learning outcomes is not always possible.

3.1.1 The challenges of measuring effects on learning outcomes

Given that our research has been developed in an academic context, there has been a tension between running experiments to prove or disprove the efficiency of a certain type of interaction, and making sure all students have access to the optimal learning resources. For example, most of the time, it would have been unrealistic to require that one arbitrary half of a class would use NB and the other one to use a traditional forum tool, just for the sake of assessing NB benefits, at least in the courses which were part of an official curriculum aiming to deliver a degree for which students paid tuition. On the other hand, it was possible to approach such conditions when the class was already set up like that. For example, some faculty members decided to require comments on specific weeks, which was an excellent occasion to examine how requiring comments affected the comments. In another course, only half of the class had access to NB, while both halves had access to a traditional forum. This was the occasion to compare comments made on NB with comments made on a traditional forum, and how introducing NB affected participation on the traditional forum.

Another difficulty was to decide which metric we should use to assess any benefit from NB. Using absolute grade is too noisy. For instance it doesn't account for the different levels of familiarity with the material. This explains why one of the metrics that we tried to use (cf section 4.3.3) was the CSEM[1] gain, a normalized measure of the student's improvement during the class. The CSEM gain is defined as the relative improvement between the pre-test and the post-test. As discussed in section 4.3.3, the CSEM gain is the metric against which we found the largest correlation with some NB-related activities (namely the proportion of high-level explanations).

[1]Conceptual Survey of Electricity and Magnetism

However, the CSEM gain is a metric which is more typical in a physics course, and isn't used pervasively. In order to be able to use the results from all classes who used NB, we found that the best approach was to rely on the existing literature results, namely that active learning promotes effective learning [23, 29, 39, 28]. Among the frameworks, we will be using ICAP hypothesis [28], for its simplicity and how general it has proven to be across fields. The same results, which underline the importance of active learning, could have been explained by using other popular frameworks, such as Bloom's revised taxonomy [39].

3.1.2 How active learning promotes effective learning: The ICAP Hypothesis

The ICAP hypothesis states that there are four levels of student engagement: *Passive, Active, Constructive*, and *Interactive*. Each level yields better results in terms of understanding and concept retention than the previous one, with the *ICAP* acronym symbolizing those levels from highest to lowest. Chi summarizes the four levels as follows [8, 28]:

1. Passive: *[...]when students are oriented toward or receiving instruction (this is what can be considered as paying attention)*, and gives the following examples: listening to a lecture without taking notes, watching a video or observing a demonstration, studying a worked example or reading silently. The expected learning outcome for students participating at this level is described as *minimal or shallow understanding*. In NB, this would be simply reading the contents.

2. Active: *[...]when students are doing something with their hands (or bodies) with the materials*, and gives the following examples: Copying the solution from the board, underlining the important sentences, manipulating or measuring test tubes, pointing, rehearsing or repeating definitions. The expected learning outcome for students participating at this level is described as *shallow understanding*. In NB, this would be reading the contents and highlighting the important

37

points as *private annotations*,[2] often but not always providing a couple of words to "tag" what was highlighted. This practice has been observed extensively in NB. For example, when NB was used in the edX class *CopyrightX* in Spring 2012, we observed that over 30% of the comments made in NB were private (1119 comments out of 3662). A much larger proportion of those private comment tended to be simple hightlighing mark of just a couple of words (372 comments out of 414 i.e. 90% were in found in those private comments).

3. Constructive: *[...] when they generate some information beyond what was presented in the learning materials*, and gives the following examples: drawing a concept map or a diagram, self-explaining or elaborating text sentences or solution lines in an example, posing questions, providing justifications, forming hypotheses, comparing and contrasting. The expected learning outcome for students participating at this level is described as *deeper understanding that might transfer*.[3] In NB, this would mean asking a question, or replying to one, but not in the context of a discussion (participants participating multiple times in turn).

4. Interactive *[...] restricted for now to refer to two or more students engaging with each other through dialog*, and gives the following examples: explaining jointly with a peer, building on each other's contributions in a wiki way, arguing with a peer (requesting and providing justification), reciprocally teaching a peer and responding to a peer's questions, discussing a joint product (concept map) with a peer. The expected learning outcome for students participating at this level is described as *understanding that might create novel ideas*. In NB, this would be substantive discussions.

Now that we reviewed how active learning improves effective learning, we present the NB system, with a particular emphasis on the features targeting to make users' reading be more active.

---

[2]a private comment is a comment that can only be seen by its author

[3]i.e. that might be used to solve a different problem than the one that was seen in the original content

3.2 Main interaction modes and design decisions on NB

NB is a web-based tool where users can read and annotate PDF or HTML documents using standard web browsers: there is no need to download an application or even a browser plug-in in order to read and annotate the material. After logging in,[4] a student typically selects a document and starts reading. As shown in Figure 3-
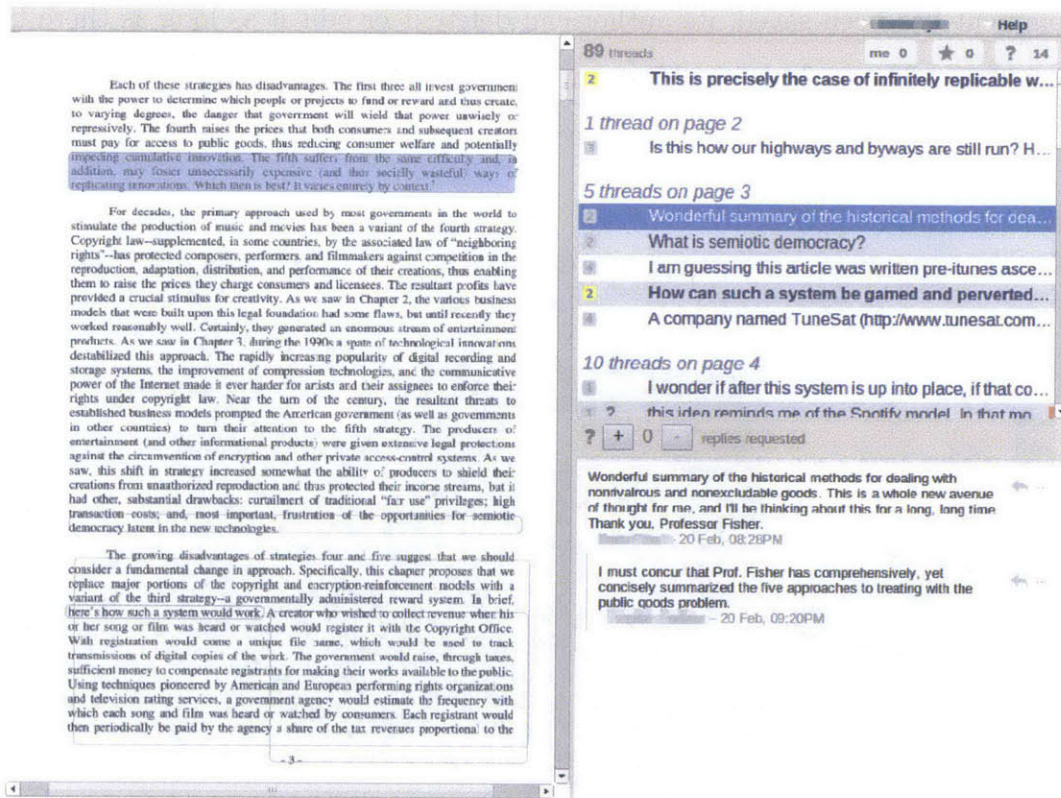
Figure 3-1: NB document view
Left: Content, *Top-right*: Thread list. *Bottom-right* Current thread

1, the document is augmented by annotations that the students and faculty have written, which appear as threaded discussions on the right-hand-side panel. Hovering somewhere in the document highlights the annotations covering that spot, whereas clicking there scrolls to the corresponding annotations. Annotations in NB are either

---

[4]The browser maintains the user's credentials in a cookie for a extended period of time, so that on a personal computer, the logging procedure doesn't need to be repeated over and over, as we saw that the startup procedure should be minimal.

anchored to a particular location in the document or are general comments on the document.[5] To add an annotation somewhere in the document, users click and drag to select a region on which they want to comment on. This region is highlighted and an in-place annotation editor pops up (bottom-right of Figure 3-2).

Users can choose whether their comment should be visible to everyone in the class (the default), to the teaching staff only, or to themselves only (a private annotation). 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. The author can also change its properties (visibility, anonymity). Users can tag each others comments with the following tags: *Favorite, Hidden, I agree, I disagree*, and *Reply requested.*
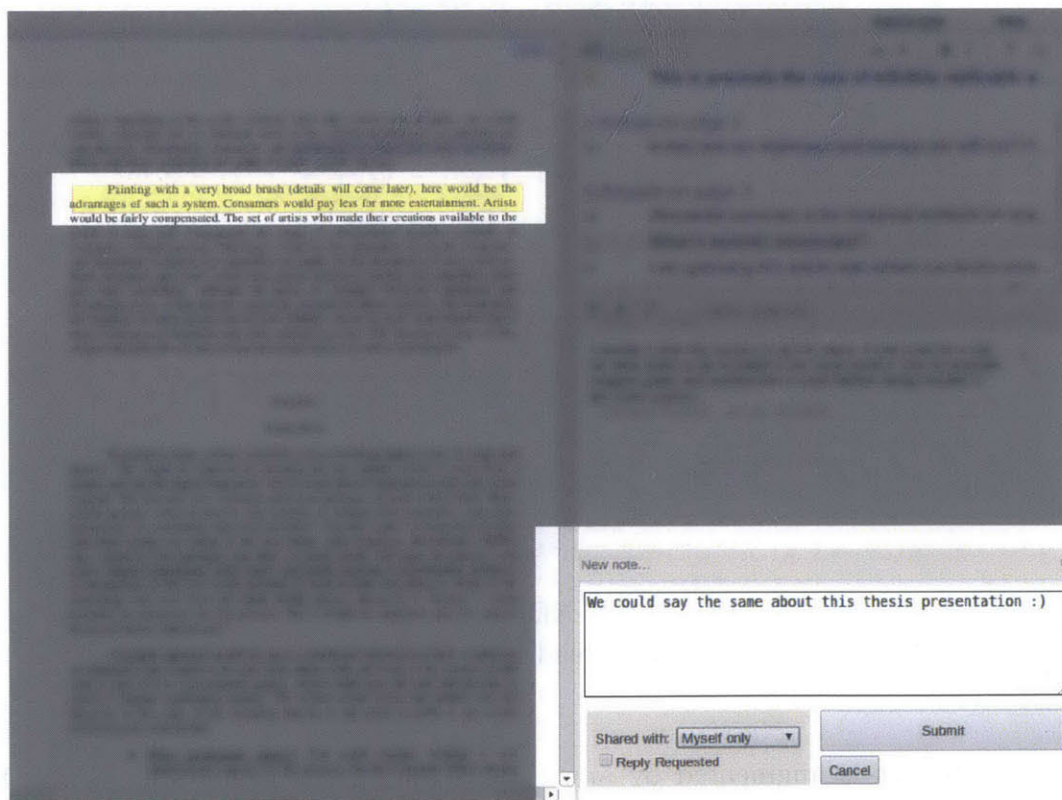
Figure 3-2: Creating an annotation
The user selects a region to comment on and an editor automatically pops up. For clarity, the parts of the screen that aren't relevant to authoring a new note have been grayed out.

---

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

The above design immediately comes with two important benefits: unlike the traditional digital communication tools mentioned in 1.1.3, NB neither suffers from the cost of having to set the context nor from the risk of cyberloafing. The context is clear: it's just what has been highlighted. This means that there's no need to explain the question nearly as much as what would have been necessary on a forum.

3.2.1 Reading discussions and commenting in the flow

In particular, users are able to see the discussions about what they're currently reading. This means finding answers on the spot to questions they may wonder about, or being able to answer on the spot questions that others may have. It it also trivial for users to participate to a long discuss as they're reading, therefore automatically increasing their learning engagement to the *active* level, the highest level in the ICAP hypothesis (cf. 3.1.2).

Moreover, even if a user completely ignores what is being said in the comments, the superposition of text selections provide a map of "trouble spots" which indicate the passages that generate a lot of discussion.

3.2.2 Support for non-sequential uses

As we mentioned above, NB primarily helps its users read and participate to discussions in the flow. At the same time, our design includes some support for non-sequential use. For instance, the homepage contains a list of the questions that have the largest amount of pending requested replies, as shown in Figure 3-3. Users can participate to those discussions without having to scan through the pages for pending questions. Once a reply is offered, the student who asked the question and everyone who marked the question as reply requested receive a notification email, and are prompted to give some feedback on the reply.

Also, NB supports collage views. Collage views are meant to present a sequence of comments, along with a snippet of the document where each comment was made. A simple use of collage views is to be able to glance at all the unread comments on

a whole course, something that can't be done with simple filters when viewing single documents (Figure 3-4). Another use is to select comments on given assignments by a given student, a feature provided to the faculty members as the *spreadsheet view* (Figure 3-5). The spreadsheet view allows instructors to glance at what a student wrote on a assignment in order to be able to assign a grade for that student and that assignment in a few seconds. Extensions of this concept are described in section 6.6.
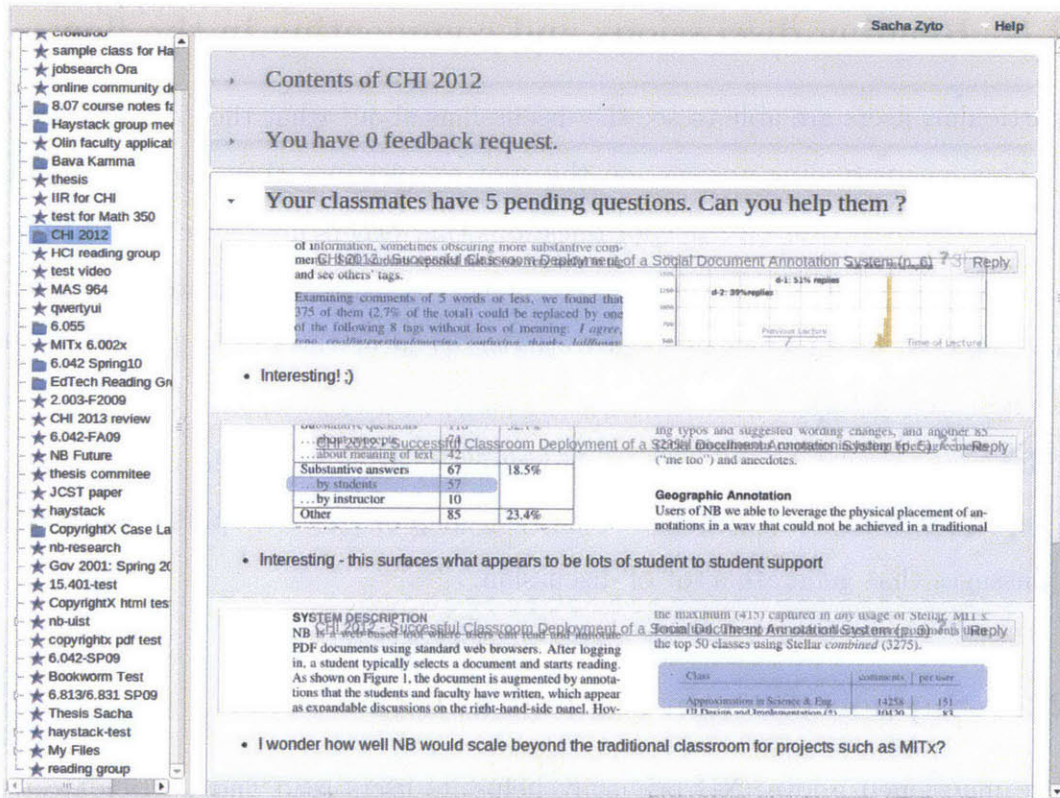
Figure 3-3: Replies requested pane
The *replies requested* pane allows users to see the most shared pending questions from their NB desktop, so they can reply to them in place

3.3 Implementation

3.3.1 Avoiding upfront costs

Based on Van Kleek and Bernstein's results on *information scraps* [21, 38], we designed NB in order to eliminate any upfront costs to being able to read the material

42
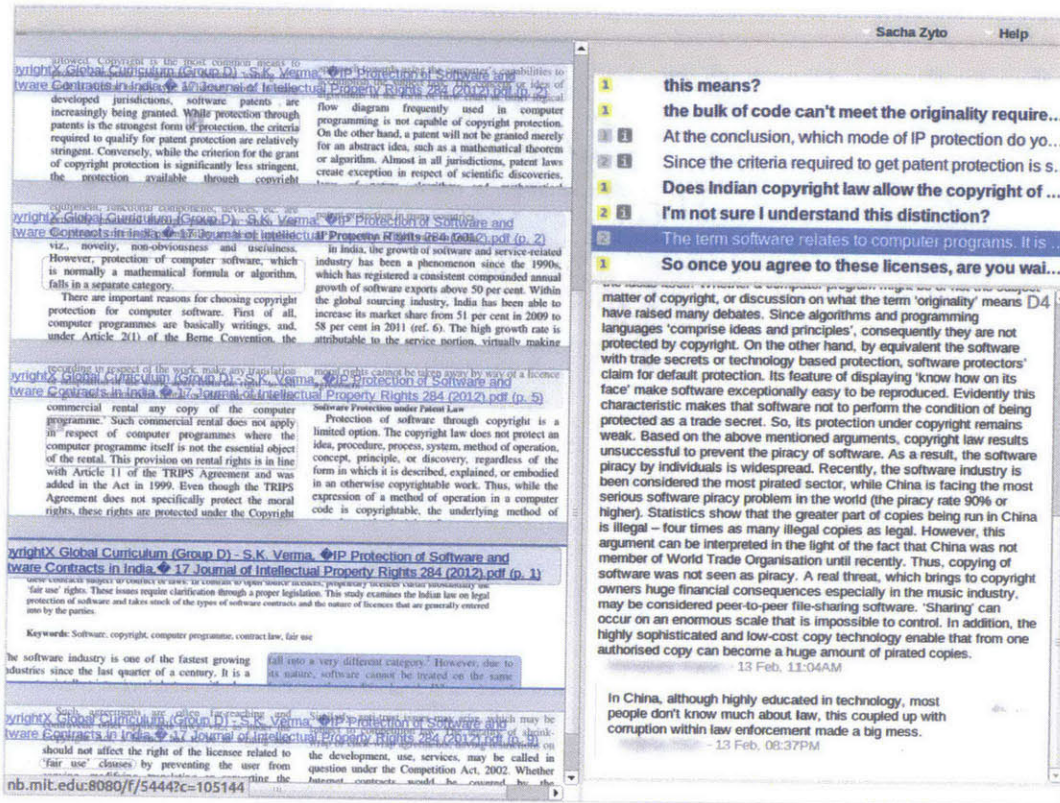
Figure 3-4: Collage View
NB can display arbitrary collections of notes along with their context, for instance a view of all the comments a particular user hasn't read for a class

and participate to discussions. In the traditional flow, students receive an invitation at the beginning of the course by email. They can use the link provided in the invitation to log in, access the content and immediately participate. Also, the number of steps required to annotate is minimal: With a drag motion, the user selects a region to annotate, which triggers an editor to pop-up, so that she can directly type her annotation, and click on the "save" button. NB uses browser cookies [9], so that next time the user visits the NB website, she will be automatically logged in.

43
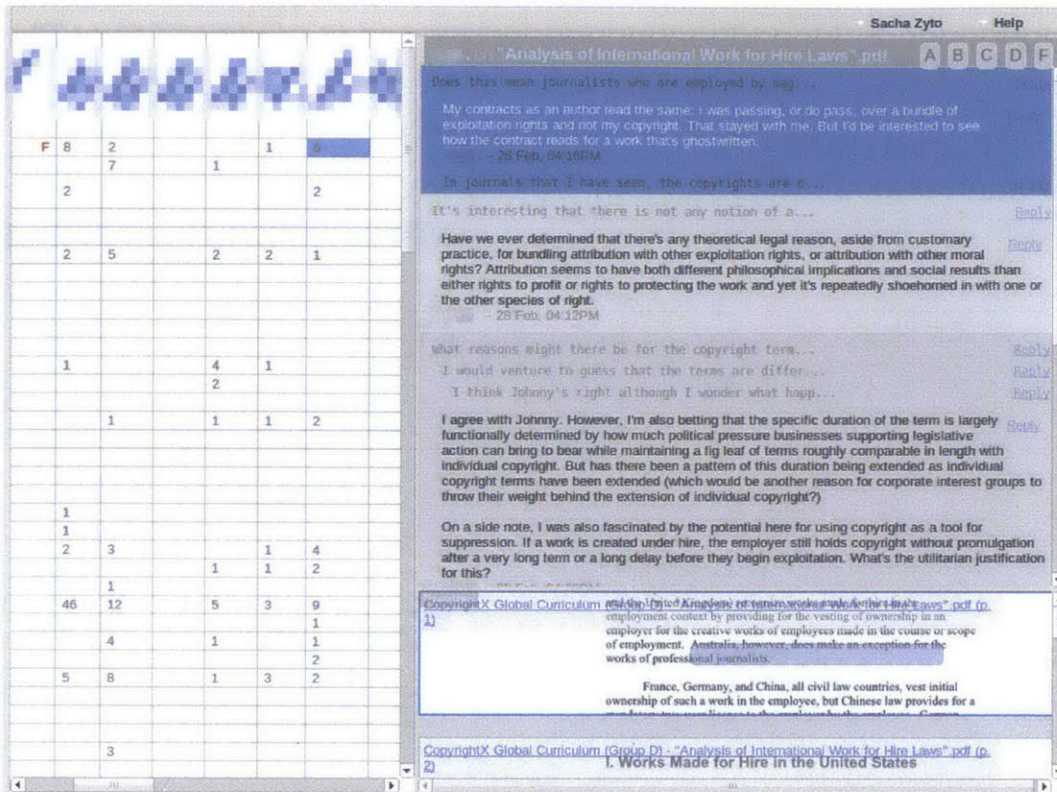
Figure 3-5: Spreadsheet View
On the left pane, in rows the students, in columns the assignments. Numbers displayed are number of comments. On the right pane, instructors can glance at all the comments by a given student on the given assignment in their original context, and assign a participation grade.

3.3.2 Internal and external consistency

Another important element in order to allow users to use NB as seamlessly as possible was the use of external and internal consistency.[6] Looking at the leftmost half of the screen in Figure 3-1, one notices the consistency with a regular PDF viewer. Meanwhile, the rightmost half is consistent with a typical email client. Internal consistency is used by reusing the same annotation behavior for different types of documents (PDF, HTML, and video).

---

[6]In HCI, external consistency is a term used to refer to several applications sharing the same paradigm. It is a powerful learnability mechanism: Once users are familiar with the paradigm by using it in one application, they can spot it on another application and immediately know how to use it. Examples include Cut-and-paste or scroll bars. Internal consistency refers to a single paradigm being reused across a given application. Examples include the use of a same style or common symbols for elements which share common properties.

44

3.3.3 Front-end architecture

The front-end of NB is a typical website. It used the techniques used by modern sites: HTML5, javascript, ajax, and CSS. One important design decision was not to require any plugin, therefore making NB available on most browsers and platforms.

3.3.4 Back-end architecture

At the time of its first deployment, the server side of NB was based on Python [10], a PDF library and a Postgresql database [11]. Since then, NB has been re-implemented using the Django framework [12] in order to improve portability and maintainability. NB uses a standard HTTP API to exchange data between the client and server. This allows third parties to use the NB framework and implement their own user interface. For instance, during the Fall 2010 term, a faculty member took advantage of that feature and generated PDF files consisting of the annotated version of his course notes. Those PDF files could then be used as an alternate UI for both reading *and* authoring: Users could post a reply to an existing comment by clicking on that comment, as it would open a browser window at that exact place on the NB website.

3.3.5 Availability

The NB service is open to use by interested faculty at `http://nb.mit.edu`. The source code for both the server and the web client is available at `https://github.com/nbproject/nbproject`. It is released under the MIT license, with the exception of the annotated PDF generator, which is under the GNU Affero GPL.

Chapter 4

Usage assessment

Given that NB was seeing some adoption, we wished to investigate *how* and *why* NB was being adopted and used in the classroom. After describing the techniques that were available to conduct our assessment, we focus on some particular instances where the tool was used successfully.

4.1 Available techniques and their strengths

The measurement techniques described below enabled us to gain insight into various dimensions of NB usage. The most interesting results were often seen when some techniques supplemented each other. Typically, qualitative observations, such as a visual inspection of some of the comments, or quotes from the students can suggest hypothesis to test later quantitatively.

4.1.1 Annotation statistics

Randomly selected samples of the data were analyzed by coding for specific characteristics, such as being a substantive comment. The details of these codings and the samples are discussed in section 4 below. The coding was done by the first author of [58], while the second and third authors reviewed the coding schemes and the results.

4.1.2 Log events

In total, for the class we reported on [58], we obtained over 1.4 million user actions and 14,258 annotations. These actions include page seen, comment created, time spent with NB both active and being "idle", and so on. Analysis of the log data followed standard quantitative procedures.

4.1.3 Student surveys

User questionnaires were administered at the end of the semester to both students and faculty, using web-based surveys. 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 for each question where they could explain their ratings.

4.1.4 Focus group

Focus groups are the occasion to explore more open-ended questions with the users. The main advantages are its informal settings, namely a casual conversation between the interviewer and the user. Their limitation is that the participants can influence each other to the point that a participant may agree to something that she didn't really experience in real life. In addition, the physical presence of the interviewer, who is often related to the project, often biases participant to try to "say nice things" about the system.

4.1.5 Faculty survey

Similarly to student surveys, we conducted several faculty surveys. Initially, our purpose was to know what faculty members liked and disliked about NB, and ask what feature they wished to see in NB in subsequent terms. After our experiment described in [58] and in section 4.2, we designed a faculty survey (cf. Appendix A) in order to confirm whether the conclusions we had drawn in [58] could be generalized to

faculty members in a statistically significant fashion. Besides surveying faculty with online survey, we were privileged to be able to meet with faculty members who had used NB and ask them about their experience using it.

4.1.6   Labeling and content analysis

Labeling and content analysis aim to look at the actual contents of the comments, and categorize it into general themes. The approach we used was suggested by Mark Ackerman, and consisted in first, reading the comments once, and try to understand both the main themes as well as the unexpected things, and refine our ontology as we went along. In cases where there were just too many comments, we only examined two samples: a learning sample in order to develop our ontology, and a test sample in order to assess how well our ontology would generalize and to perform inter-rater reliability.

4.2   First experimental design: Usage observation in "best-use" class

We first focus on the single most successful use of NB at the time we published [58]: *The Art of Approximation in Science and Engineering* taught by Sanjoy Mahajan in Spring 2010 at MIT. In this first study, our objective wasn't to demonstrate that NB *always* works but rather that it *can* be successfully used in a real-world setting, which shows the research direction worth pursuing. The Art of Approximation in Science Engineering had 91 undergraduate students. The thrice-weekly class lectures came from a draft 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 on the printed lecture pages—at the beginning of each class, following a method developed by Edwin Taylor [49]. 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 section 5.6). There were no Teaching Assistants (TAs) for this class.

4.2.1 NB being valued by students

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. At the end of the term, students were asked to fill out an optional web-based poll. We wanted to know more about their annotation 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 interactivity 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 a 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 in 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.*

51

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, and a way to make students realize they're not "impostors" [31]:

- *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.*

4.2.2 A very positive 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 the class ecology to to is a key success factor Guzdial and Turns [34] 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 section 4, 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 (33 readings in total), 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 the instructor is explaining the material, as discussed in [49]. This point was stressed in the reading memos instructions, and created an atmosphere where students valued the chance to make comments on material written by the faculty.

The WebAnn study [20] 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 (see Figure 4-1). In the sample lecture we analyzed, we found three requests to use simpler examples, two requests to review/explain a concept during class (Mahajan replied that he would try), and four notes mentioning something that had been seen during lecture. 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.

Figure 4-1: Impracticality of paper-based reading memos
The sheer amount of page-flipping makes it almost impossible for the instructor to use the comments in a timely fashion

4.3 Follow-up experimental designs

Is data from a single class convincing? Clearly, we benefited from a very talented and motivated faculty user of our system. One might fairly ask whether "other" faculty could expect to see the same benefits. Table 5.1 demonstrates that many other faculty at several other institutions were able to achieve significant adoption, some exceeding the best case studied in [58], even though few of them had previously made use of reading memo requirements.

Of course, some preconditions apply to successful usage of NB. As one reviewer of [58] 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 on-line 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.

The class we report on in [58] used an early prototype of NB. Many users complained about UI issues: slowness with pages that had lots of notes, bugs, and limited browser support (our newest versions have addressed these issues). Successful usage occurred *despite* these deterrents. In particular, students seemed satisfied with two main aspects: the ease with which one could make context-specific annotations (just drag on a region of the text and start typing), and the presence of those annotations on the side of the material.

We now focus the rest of our analysis in this chapter on other classes, first trying to report whether adoption in these other classes was determined by the same factors as the one analyzed in [58], or entirely different ones, then whether we could find a correlation between usage of NB and learning outcomes, and finally we report on the use of NB in a MOOC environment.

4.3.1 Generalization by follow-up with faculty members

Based on the preliminary conclusions from *The Art of Approximation in Science and Engineering*, it seemed that NB was a useful enhancement to their teaching practice. Still, there were a number of pending questions, most notably how those results could be generalized. In order to find out, we decided to run a faculty survey when the number of faculty members using NB reached a number for which we could draw statistically significant results, which occurred in Spring 2013. The high-level purposes of the faculty survey were as follows:

- determine if faculty share the concerns we had envisioned when we developed NB (cf. Introduction: The heavy toll of communication inefficiencies in class) and when we collected feedback (as we wrote [58]).

- determine if there were other concerns.

- clarify which success factors in *The Art of Approximation in Science and Engineering* were particularly important.

Consequently, we adopted a hybrid approach between directed questions and open questions. The directed questions were aimed to follow up from our study in [58] and to get a sense of how well NB fulfilled the purpose for which it had been introduced. The open questions aimed at understanding whether the reasons we had found for developing NB so far were shared by the faculty at large, if there were other objectives we may have not be aware of, and if there was any unexpected behavior. The entire faculty survey is in Appendix A. 47 faculty were invited to fulfill the survey, all from classes which had been produced at least 100 comments over the Spring 2013 term. N=100 was chosen in order to filter out instances of NB that had just been used as a test. Our of those faculty, 29 participated to the survey. All participating faculty gave their consent for their comments to be quoted.

Purposes for introducing NB We used questions 1-a [1] and 4-c [2] to produce the following typology for the reasons faculty used NB in their class (presented in decreasing order of occurence):

Improve in-class experience: Encourage in-class discussion by having students come prepared, and tailor the contents of lecture based on the comments

- *To encourage students to read the papers before class, and to get them thinking about it before the class discussion.* [User 2]

- *Primarily, I wanted them to interact with the reading [...].*[User 3]

- *Generating group discussion on readings - in and out of class.*[User 13]

- *I wanted to spend class time discussing and working on concepts, not presenting equations that they see for the first time. So I wanted them to read beforehand [...].* [User 14]

- *as i use peer instructions in my lectures i thought that NB would increase the motivation for reading in advance* [User 15]

- *I used NB to encourage students to read and ask questions so that I could better prepare for class. I also used NB in larger classrooms to help me connect with students-I tried to respond to at least 10% of the comments/questions and I tried to respond to different students each time.*[User 19]

- *Discuss online paper that were later to be discussed in class.*[User 20]

- *I wanted (1) to make sure students read fundamental texts before class; (2) students have an opportunity of sharing thoughts on difficult texts that would be further discussed in class.*

- *To prepare a more effective class environment.* [User 25]

---

[1]Why did you decide to use NB?

[2]For which main purpose(s) did you introduce NB in your class?

- *I wanted to use a tool that encouraged students to continue the discussions we had as a large group outside of our meetings and to see what things they understood/broadly agreed on and where there were misunderstandings/disagreements that would be a good use of shared time.* [User 27]

- *Create discussion before class* [User 29]

- *[...] for me as an instructor to have a sense for student understanding of the reading before in-class discussions*

Allow students to benefit from other students questions and address scaling problems

- *To allow students to benefit from others' questions about the lectures.*[User 2]

- *spare staff from having many queries to answer.* [User 6]

- *To enable students to ask questions directly in the PDF and to enable me to answer them such that everyone can see it.* [User 21]

- *I had previously asked students to simply email me with queries. Having their queries visible to all, and in context seemed like a good idea.* [User 28]

- *To facilitate communication among students that were trained in different fields within the class* [User 30]

Get students comments on a draft textbook or lecture notes

- *Secondarily, I wanted feedback on the draft textbook itself.* [User 3]

- *to get more feedback for the textbook I am writing.* [User 4]

- *(3) Get feedback from students- these are draft notes for a book under development* [User 8]

- *I was working on a draft book, and wanted to use NB to distribute the book to students and get their comments.* [User 11]

- *Wanted student feedback to improve lecture notes.* [User 22]

This was also a theme previously encountered with Manajan's class (The Art of Approximation in Science and Engineering), as the reading were chapters of his textbook. However, we had underestimated how popular this reponse would be.

Get timely feedback that students are reading the material, and where they're finding trouble spots

- *[...]and discuss, with me and with each other, what they were and were not understanding [...]* [User 3]

- *[...] because of the interaction through NB, I could assume that the assigned material was assimilated* [User 4]

- *So I would know that my students were reading[...]* [User 7]

- *Also to "force" students to keep a certain learning pace and not wait all the way until the exam approaches.* [User 9]

- *To get information about what parts of the reading were unclear, and what the students' questions were.* [User 11]

- *I wanted a way to push students to actually read the assignments and know whether or not they did [...]* [User 18]

- *See the King and Sen article [13]. Basically: quick feedback, peer teaching, and interaction.* [User 26]

The theme of getting timely feedback has been recurring: in early discussions with faculty members at MIT that led to the concept of NB, in the faculty interview with Sanjoy Mahajan after he used NB in *The Art of Approximation in Science and Engineering* and in this survey. Getting timely feedback appears to be a win-win outcome: it helps the students not to get behind, and helps faculty improve their course, which was especially useful when the course material consisted of a preprint. Since NB

currently logs pages that were read (see section 4.1.2), this suggests providing the instructor with a percentage of the class who read the material on time. Although results from our study in [58] advocate for giving more freedom to the students and not enforcing two separate deadlines, we nevertheless observed that in another class providing information about who read by the time of the class motivated the instructor to enforce a stricter policy about reading / commenting deadline in the following term: students were allowed one week to ask questions or post replies, and an extra week just to post replies.

Encourage collaboration *in the flow*

- *To make an existing writing response more collaborative.* [User 5]

- *Collaborative reading* [User 8]   .

- *collaborative skills* [User 10]

- *To get students to collaborate more outside of class and to have them ask and answer each others questions as they studied the lecture notes. The annotation directly on the notes is much better than other chat rooms/forums for this interaction. NB primarily had the lecture notes used in class. Students could then post questions directly in the notes for others (or myself to answer). If a student emailed me a question, I directed them to post it in the forums, so that others could try to answer, or at least benefit from seeing our discussion about the concept.* [User 12]

Promote improved/effective learning and critical reading skills

- *Improve learning by facilitating peer activity while being able to guide the entire class process by (mild) NB interference and class discussion* [User 9]

- *article analysis* [User 10]

- *I want students to learn how to read and learn from technical materials such as textbooks. To do so, they need to practice asking questions of the material.*

60

NB makes that possible, and enjoyable, because students get answers to their questions. [User 14]

- *Helping the students to understand how to read text books in technical fields. [...] helping students understand difficult and / or subtle concepts* [User 15]

- *To force students' attention on the detail of literary text and their critical commentary about it.* [User 16]

- *[...]wanted to give them a guided entry to the primary literature.* [User 17]

- *[...] as well as helping them interpret the readings.* [User 18]

- *I wanted to improve both the likelihood of students completing the assigned readings and the level of absorption of the readings that students do complete. [...]* [User 24][3]

Results Analysis. As we can see in the previous section, one of the major faculty goals is to *Promote improved/effective learning and critical reading skills.*[4] User 24 carried out his own research initiative to determine whether annotating collaboratively would help alleviate the problem of students being more passive when doing the assignments, and reports that[5] "NB succeeded at providing an additional level of interactivity, which in turn made students more engaged with the reading."

This can be understood in light of the ICAP hypothesis, presented in section 3.1.2. As we can see from the faculty quotes, it seems that in the light of the ICAP framework, NB lowers the barriers for students to reach higher levels of participation (I,C and A, rather than just P).

Note that unlike WebAnn, where in-forum discussion and in-class discussion competed with each other, encouraging discussion became a "goal" of NB, and was rated

---

[3]This user provided a reference to a blog page describing how he used NB to carry out an experiment about how students acquire critical reading skills

[4]This was one of the main purposes which motivated Mahajan to use NB in *The Art of Approximation in Science and Engineering,* so that students can understand that "Math is no spectator sport."

[5]https://sparc.colorado.edu/moonhawk-kim-collaborative-reading-of-academic-articles/

as a success (6 out of 7 to question 4.d). As described by User 27, and the ones in 4.3.1, NB fulfills 3 important roles:

1. preparing for good dialog during class, by having students come prepared.

2. adjusting the contents of the lecture based on forum discussions.

3. providing a platform for follow-up questions after the class: Lecture time is limited both in space and time. NB lifts those traditional limits.

4.3.2 Deployment in an multi-section MOOC environment

In Spring 2012, NB was used for the first time in a MOOC course called CopyrightX, taught by Prof. Fisher at Harvard. The approximately 500 initially enrolled students for the classes were split into 4 curricula (A,B,C,D). Curricula A and B were only able to use the edX forum whereas curricula C and D had access to both the edX forum and NB. Within each curriculum, the approximately 125 students were dispatched into sections of typically 25 students, each TF (teaching fellow) being responsible for a section. Given that only half of the class had access to NB, it provided, at least in theory, almost ideal conditions to study NB's influence.

	#Authors (edX)	#Authors (NB)	#Threads (edX)	#Threads (NB)	#Comments (edX)	#Comments (NB)
A	104	-	426	-	1049	-
B	88	-	250	-	1198	-
C	82	66	290	1332 (573 public)	1047	1800 (1033 public)
D	91	90	343	970 (629 public)	1357	1860 (1496 public)

Table 4.1: Usage summary across the four CopyrightX curricula

Table 4.1 shows the main annotation statistics. From those statistics, the main observations was that the curricula that used NB produced more comments even on the edX forum, therefore suggesting the two forum modalities helped each other: using NB added targeted discussions without reducing the edX forum discussions.

We also used annotation statistics to determine how the annotation practices evolved over the term. In order to compensate for the dropout rate, which can be quite important on MOOCs, we looked at the weekly number of comments per active user. An active user is defined as a user who signed on to attend the mandatory weekly recitation on the class teleconferencing tool (adobe connect). We also removed the 2 first and last weeks to eliminate external factors due to learning of the tool or assignments being changed close to the exam week, as well as week 8, when classes got canceled as a consequence of the Boston marathon bombings. The results are shown on Figure 4-2. As we can see, the curricula which had both the edX and NB (C and D) consistently submitted more comments than the curricula that only had the edX forum (A and B).
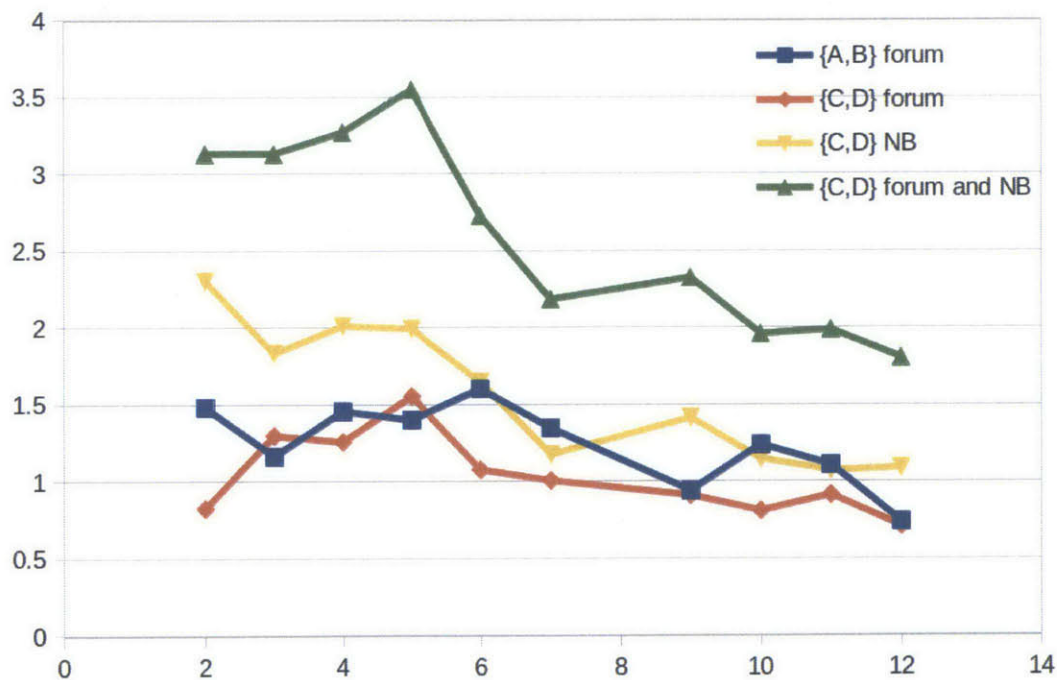
Figure 4-2: Evolution of weekly number of comments per active CopyrightX user {A,B} *forum*: participation from curricula A and B on the edX forum (same as their total participation since had no access to NB, {C,D} *forum*: participation from curricula C and D on the edX forum, {C,D} *NB*: participation from curricula C and D on NB, {C,D} *forum and NB*: total participation from curricula C and D

Qualitative analysis We performed a qualitative analysis of a random sample of the threads, by labeling 100 randomly selected threads from the edX forum and 100 randomly selected threads from NB. We used the following label ontology inspired from [34]: learning objectives, homework (grading, strategy), the collaboration tool itself, infrastructure (e.g., homework, class pace, lecture quality), and off-topic (anything else).

Category	edX Forum	NB
Learning objectives	67	53
personal notes	0	44
review questions	8	3
infrastructure	12	0
off-topic	13	0

Table 4.2: Qualitative analysis of 100 randomly selected threads from CopyrightX 2012

As shown on table 4.2, substantial amounts of collaborative learning occurred within both platforms (edX forum and NB), and that users took advantage of the specificity of each platform to carry out tasks that each platform would facilitate: in-context discussion using NB, and holistic discussions using the forum. At the same time, in both platforms, the annotations were primarily substantive content regarding the course (i.e. learning objectives). Discussion threads were extensive and students became active participants in questioning and interpreting the course material. Besides peer-to-peer discussion, users made substantial use for NB in order to enter personal notes.

Differences across sections Unlike *The Art of Approximation in Science and Engineering* where all the students were taught by the same instructor, CopyrightX was split into 20 sections. Each section was supervised by a different TF and had separate requirements, determined by the TF, which fell into three categories:

1. Sections where participation wasn't mandatory

2. Sections where participation was mandatory but without an explicit quantity of required comments per week

3. Sections where a minimum number of comments was imposed in order to get full participation credit.

We observed important participation variations across sections: Both in NB and the edX forum, there was over a factor 10 in the number of comments between the most and the least prolific sections. The requirements enumerated above are the first order factor in terms of participation. On the other hand, looking at the effect of those requirements on the quality of the discussions would be the subject for future analysis.

4.3.3 Correlation with learning outcomes

The results presented in this section are based on a ongoing collaboration with Professor Eric Mazur at Harvard. The major part of the data analysis has been carried out by Kelly Miller from Professor Mazur's team and was presented at AAPT2013 [17]. NB had been used several times in a flipped classroom setting, where students come to lecture having already read the course material, so they can focus on group activities during lecture in order to develop concrete understanding of the material. As mentioned in section 3.1.1, those classes used the CSEM gain to evaluate the progress in understanding of the material made by each student.

Using NB in such a setting provided an favorable experimental setup to explore whether some NB usage dimensions correlate with the CSEM gain.[6] In order to do so, quantitative and qualitative analyses were performed. The material consisted of six textbook chapters, on which the 28 enrolled students authored 900 comments organized in 400 threads. The qualitative coding was performed by three coders, with a inter-rater reliability of 70%.

The coding scheme consisted of three main variables:

---

[6]Correlations with other performance measures were performed as well: namely with the exam grade and the in-class concept-tests performance. However, none of those measurements provided correlations as strong as the one observed with the CSEM gain, as shown on table 4.3

1. The annotation type, which was one of the following: comment, open question, closed question, or explanation.

2. The annotation quality, in term of the course cognitive domain (namely, physics), adapted from the TIMSS 2011 assessment frameworks: no physics (0), knowing the cognitive domain (1), applying the cognitive domain (2), reasoning about the cognitive domain (3).

3. The extent of the argumentation, using a scheme adapter from [54]: claim (1), claim+warrant (2), claim+warrant+backing (3), claim+warrant+backing+qualifier(4), see Figure 4-3 for examples.

for example, shows the pattern of field lines created by a pair of oppositely charged particles (that is, a dipole). Sixteen field lines emanate from the positively charged particle on the left and sixteen field lines terminate on the negatively charged particle on the right. Notice the correspondence between this pattern and the corresponding vector field diagram in Figure 24.3a and the pattern created by the fibers in Figure 24.3b.

The number of field lines that emanate from a positive charged object is arbitrary; we could have chosen some number other than 16 for Fig. 24.2. However, in a given field line diagram, the number of field lines is always proportional to the magnitude of the charge carried by the object. If, for example, sixteen field lines emanate from an object carrying a charge $+q$, then thirty-two lines emanate from an object carrying a charge $+2q$ and eight lines terminate on an object carrying a charge $-q/2$.

24.2 Field line density   3

Figure 24.2 Electric field line diagram for an electric dipole.

Student 1:
"Are any two oppositely charged particles considered a dipole when brought near each other?"

level 1 closed question
(knowing cognitive domain)

Student 2:
"I think any two particles that are oppositely charged to form a pair are considered to be a dipole"
level 1 explanation
(claim without warrants, backing or qualifiers)

Student 3:
"An electric dipole is a system for which a dipole moment exists. This is a measure of the separation of positive and negative electrical charges in a system of charges, that is, a measure of the charge system's overall polarity. So the important thing is to define your system! Two oppositely charged particles within the same molecule may be considered a dipole, and 2 oppositely charged particles that don't belong to the same molecule can also be considered a dipole as long as you include them in your system seeing how they have a dipole moment. "
level 4 explanation
(claim with warrants, backing and qualifiers)

Figure 4-3: Illustration of different levels of argumentation

The two main observed results in this study were as follows:

1. There is a strong positive correlation ($R^2 = 0.57, p < 0.005$) between the CSEM gain and the proportion of high level explanations (level 3 in annotation quality).

2. There is a very strong positive correlation ($R^2 = 0.21, p < 10^{-5}$) between the annotation quality and position in the thread even when controlling for which student is writing the annotation.

The next step in this ongoing research is to determine whether the two observed strong correlations are also causation. For the first item, we have designed the follow-up experiment for a subsequent semester.[7] The experiment consists in splitting the class into sections and selectively seeding some sections with comments containing a higher than average proportion of high-level explanations, relying on social norm [40] for users to provide an higher than average proportion of high level explanations, and examine if that section's participants score a higher CSEM gain. For the second point, our follow-up experiment would consist in seeding different sections with an incomplete thread (namely the whole thread except the last explanation), and see if the next offered explanation is of higher quality than average in a statistically significant fashion.

NB metric	CSEM gain	Exam performance	ConceptTest performance
number of annotations (average: 6.5 comments per user per chapter)	-	-	-
amount of time spent actively reading	-	-	-
comment quality	0.46 *	0.39 *	0.36 *
proportions of explanations	-	0.43 *	0.47 *
proportions of high-level explanations	0.57 **	-	-

Table 4.3: Correlation between NB metrics and class performance metrics (-) means no stat. significant correlation, (*) stat. significant correlation with $p < 0.05$, (**) stat. significant correlation with $p < 0.005$.

---

[7] As of the time of this writing, we have begun an experiment with the Fall 2013 AP50, where we have been using different seeding conditions in different sections

Chapter 5

Findings and Discussion

5.1 General usage statistics

For a research project, NB benefited from quite a large deployment, both in number of classes and students, and in number of terms it was used. We first present a summary of the usage statistics, trying to differentiate between ephemeral versus substantial use of the tool, and discuss the evolution of NB usage. One limitation of the analysis is that NB changed a lot in five years. Yet, we believe the analysis is relevant, since the main functionality of allowing discussion within the context of a document has remained the same.

5.1.1 Ephemeral versus substantial use

In Fall 2011, when we wrote [58], NB had been used in 49 classes by 32 distinct faculty members at 10 institutions including MIT, Harvard, California State, University of Edinburgh, KTH Sweden, Olin College, and Rochester Institute of Technology. The majority of classes were in the physical sciences but a few were in social sciences and humanities. Of the 32 faculty members, eight were using the tool for the first time that semester. Of those who started earlier, nine faculty (28%) made use of the tool in multiple semesters (for a total of 18 re-uses), indicating that they continued to adopt it *after* a semester's experience of its usage. This seems a rough 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 5.1 shows the 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 5.1: Usage of NB in the top 15 classes, in number of comments as of Fall 2011 Starred classes are re-uses by a faculty member who had already used NB.

5.1.2 Evolution of usage and annotation trends

Since NB launch in 2009, we observed a rapid increase in classes using NB, as shown in table 5.2. We define a *substantial usage* of the tool as a course that produced 100 or more comments. As of October 2013, NB has been used substantially in 202 courses across 36 institutions. This represents over 169,000 comments in 112,000 threads, authored by 8690 users.

Year	classes with substantial usage		all classes	
	number of classes	number of comments	number of classes	number of comments
2009	3	14933	16	15014
2010	11	24380	37	24706
2011	18	14307	79	15194
2012	57	31609	170	34137
2013	132	116729	481	121805

Table 5.2: NB usage evolution over five years
Classes with substantial usage are the ones with 100 comments or more

5.2 Annotation outcomes

In the rest of this chapter, we present evidence that substantial amounts of collaborative learning [34] occurred within NB, along with some phenomenon that are unique to a situated forum: First, 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. Second, exploiting the geographic situatedness of annotations, students posted comments that addressed several distinct but co-located threads simultaneously.

5.2.1 Presence of factors promoting collaborative learning

Assessing CaMILE [34], Guzdial and Turns identified three criteria necessary to promote collaborative learning: broad participation, sustained discussion, and focus on the class topic. For our analysis, we focused on *The Art of Approximation in Science and Engineering* course described in section 4.2, and observed behavior satisfying all three of these criteria. We cover each in turn.

Broad Participation The 91 students created over 14,000 annotations during the semester (averaging 153), while the instructor created 310. The average number of annotations authored per student per assignment was 3.67.

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

71

of voluntary usage.

5.2.2 Sustained Discussion

Of the 14,258 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 5-1 ,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.
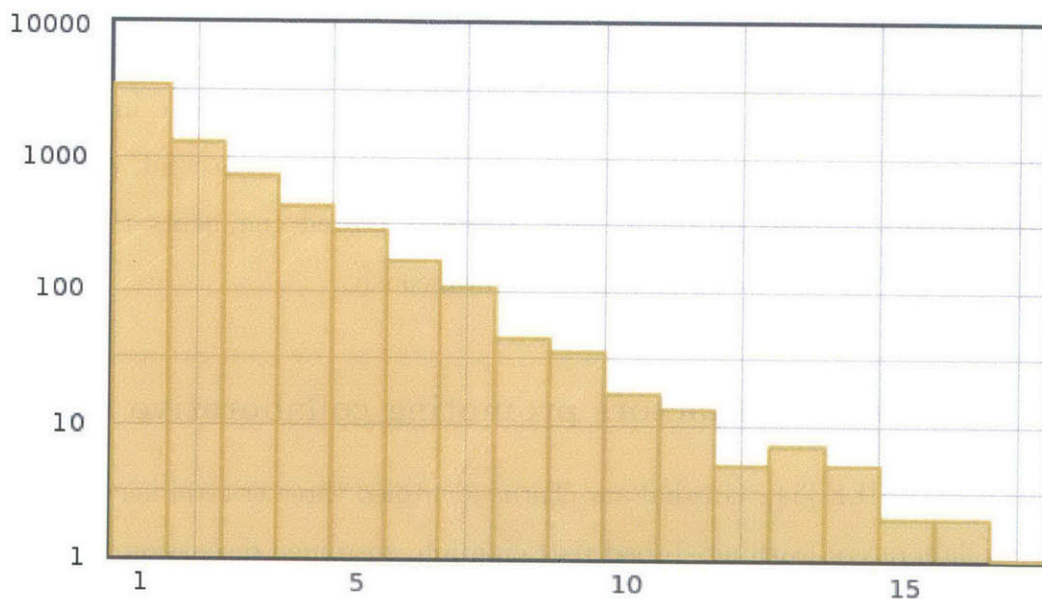

Figure 5-1: Distribution of the number of comments per discussion

5.2.3 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 Guzdial and Turns' [34] coding scheme of six categories in order to

label the type of comments.[1] We found that annotations related to the objectives of the course (*class learning topic* in Guzdial and Turns' coding scheme) represented an overwhelming majority of the comments —363 comments (88.1% of the total comments in that assignment) found in 164 discussions (87.7% of the total discussions in that assignment). To gain further understanding, we sub-categorized these 363 class-learning comments. Table 5.3 summarizes their breakdown.

5.2.4   Question resolution and student-to-student teaching

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 4.2.2 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 asked after the assignment deadline, and 5 simply went unanswered. This is summarized in table 5.4.

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

---

[1]namely: class learning topic, 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).

Type	Number	Percentage	Example
Substantive commentary	95	26%	My earlier statement, about stiffness not solely being dependent on the scaling of bond energies from hydrogen, makes me question the other conclusions like speed of sound and energy from sugar. Its just hard for me to believe that things are that simple. [...]
Substantive questions ...about concepts	116 74	32.1%	 I didn't follow exactly why since q is relevant that it produces a force, that the charge appears only in the combined equation given
...about meaning of text	42		Do you mean to say "uses the hydrogen atom..."?
Substantive answers ...by students	67 57	18.5%	 I think just means that since hydrogen atoms are prevalent in many complex substances, it's useful to know the energy of a hydrogen bond.
...by instructor	10		Being polar doesn't affect the intermolecular spacing much in water. The molecules are still as closely packed as possible (basically, until their electron clouds touch and repel each other). But in ice, being polar is responsible for the open structures that water molecules form [...]
Other	85	23.4%	I'd use: "All from the basic understanding of hydrogen!" add: "approach"

Table 5.3: Breakdown of 363 class-learning comments

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 5.4: Breakdown of questions asked and their resolution.

another 85 (23%) miscellaneous comments including brief agreements ("me too") and anecdotes.

5.2.5 Increase of weekly annotation volume over the term

We mentioned that the average number of comments per student per week 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.

5.2.6 Timely feedback to students and instructor

As we mentioned in the sections on student feedback (4.2.1) and faculty feedback (4.2.2 and 4.3.1), the ability to get timely feedback was noted and appreciated. In particular, two early faculty adopters of NB, Sanjoy Mahajan and Katrin Wehrheim made a similar description of one of the major teaching changes that NB made available, namely being able to look at what the main questions were the night before lecture and organize the lecture based on those questions. The feedback we collected from the Spring 2013 faculty survey confirmed these observations, since the theme of getting timely feedback was mentioned 7 times among the 29 faculty replies.

5.3 Observed use of specific benefits of in-place forums

So far, the observations made in this chapter could have been observed in a traditional forum. In this section, we detail our observations specifically related to the fact that the content and the discussions are located at the same place.

5.3.1 Use of context

Given the select-a-region-and-annotate design principle we adopted for NB, it is no surprise that over 95% of the observed comments take full advantage of the fact that they appear in context, in the sense that they could not be understood as-is without the accompanying selection. As we mentioned in the sections on student feedback (4.2.1) and faculty feedback (4.2.2 and 4.3.1), this in-context usage was appreciated by everyone. However, a few questions remain about the 5% of non-contextual comments: They are virtually always encountered at the very beginning of documents, where authors typically select the very first line or title to create a holistic comment related to some of the contents in the document which can't be pinpointed to one precise location. A more careful review of those comments revealed three typical use cases:

1. Comments that are truly holistic, for instance a personal reflection about the whole reading.

2. A workaround for not being able to highlight several places in the document.

3. Directions intended to be seen by the readers before they start reading the document: Indications on how to do the homework for instance, or use the tool.

An extension to the current annotation framework that would address the two first use cases is discussed in section 6.6. As for the third use case, we think the current situation is an unexpected yet appropriate use of the tool, since it fulfills the intended goal of attaching a comment that readers will see first.

5.3.2 Use of geographical proximity of threads

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 six 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 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 six 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 six relevant comments would receive the reply? And how could posters on the other five 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 at least partially by referring to neighboring threads in the discussions.

The question remains of whether the posters in the other five threads noticed the answer, and if they noticed it in time. For example, they didn't get an email notification saying that someone participated to a neighboring thread. At the time, they would be able to see that there were comments surrounding the region of the content

where they had a question. Since the version that was used in this class, we adopted a simple heuristic to signal new comments, similar to the affordance used in many email clients: A thread with new comments appears in boldface. Therefore, the next time people visit the lecture notes, they can notice if there are new comments neighboring the one on which they asked a question. We believe that a possible research direction would consist in letting users connect similar threads, so that participants can be notified when a reply arrives in a neighboring thread.

Finally, 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 content that needed clarification in the upcoming lecture.

5.3.3 Students abandon paper and annotate early

Although the number of assignments in our class differed from the WebAnn experiment [20], we found that the number of annotations per author per assignment were very similar: a bit more than 4.[2] 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: reasonable effort included posting only questions.

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. [20], 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

---

[2]In a study of how peer-feedback can increase the relevance of online discussions, van der Pol [53] (chapter 4) also reported 4.7 annotations per student per (weekly) assignment.

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.

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. Students who printed too early might never see some comments at all. To address the problem, Brush and al.[20] 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 at the same time, and get clarification on confusing points in the lecture notes in a timely fashion (cf. the NB being valued by students section). NB yielded a much greater proportion of replies than WebAnn, without imposing WebAnn's differential deadlines or specific requirement to reply.
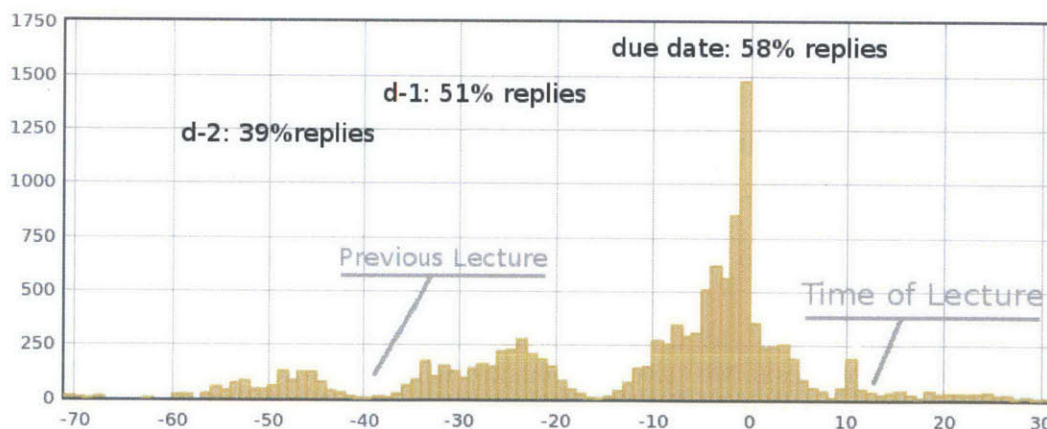

Figure 5-2: Distribution of intervals (in hours) between the comments creation time and the corresponding assignment deadline

The ongoing nature of the interaction is confirmed by Figure 5-2, which presents the number of comments posted as a function of the time (in hours) between a comment's creation 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%), 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 5-2 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. With NB, there is also 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.

5.3.4 Evidences of participation 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 [19]. It also means that readers can encounter and respond to comments and question as they read, instead of hunting for relevant comments.

Given this hypothesis, we tried to measure whether 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 5-3 shows this distribution for replies (the distribution for annotations is similar).

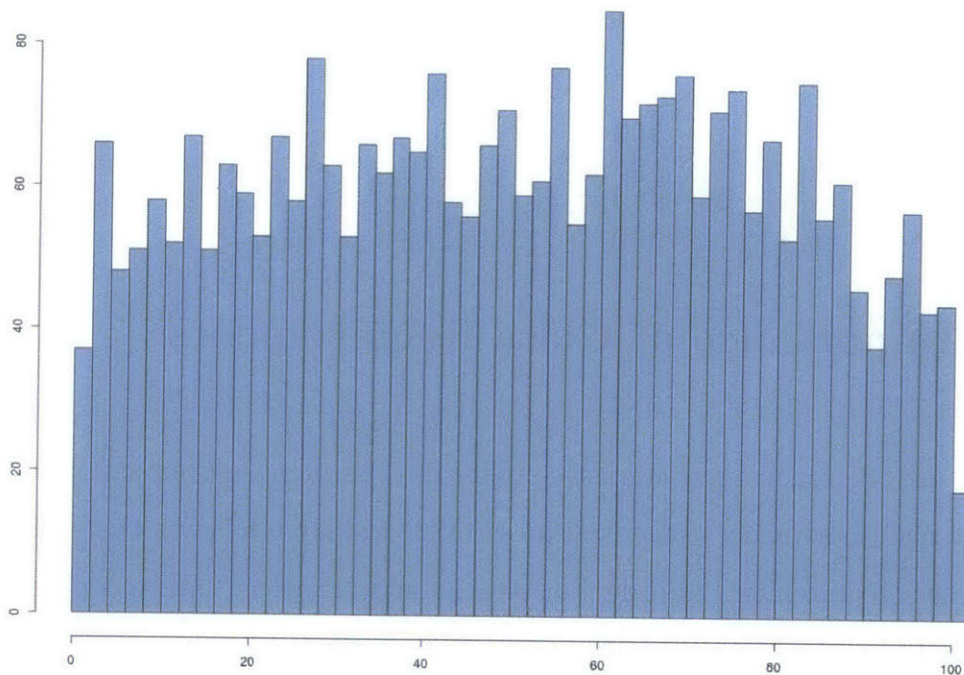Figure 5-3: Distribution of relative creation times for replies over the course of a reading session.

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.

5.4 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, in order to seed the network effect and to compensate for the natural trend for student to skip what's not required given their limited resources (e.g. available time to study). Besides, he believes that a key aspect is to have usage requirements and guidelines that *aren't too precise* because it leaves students some agency. This means demanding a reasonable and steady effort instead 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.

5.5 Re-using comments: 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. Graham, Knuth and Patashnik's *Concrete Mathematics* [33], a traditional textbook, contains 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. Similarly, Edwin Taylor's *Spacetime Physics* [50] has student questions and then answers, based on earlier reading memos. 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.

5.6 When do different class channels collaborate rather than compete?

The introduction of NB in the classroom caused a new communication channel to exist in the class, which could disrupt existing communication channel. As we mentioned in section 4.2.2, Brush and al. reported that WebAnn [20] negatively affected in-class participation, since most of the question had been asked before lecture. In certain

classes, we observed no effect between NB and an existing modality. For instance, as we mentioned in section 4.3.2 on NB usage in CopyrightX Spring 2012, NB had no significant effect on the edX forum participation. Finally, there were times where NB helped the other modality. We already mentioned in section 5.2.6 the fact that faculty members were able to plan their lecture using the comments on the reading material for that lecture. Another positive interaction is that student came better prepared to class, and were therefore able to ask better-planned questions. One indicator was the faculty replies to question (4-h) from the survey *Did NB change the way you taught? If so, how?*:

- *It puts more onus on the students. They don't just come to class and listen and go home. They know something when they come to class and can participate and ask GOOD questions in class. We went faster some days. Other days we were able to have deeper and more thought provoking discussions where in the past we couldn't even have a discussion..* [User #7]

- *Yes, class sessions changed from regular teaching to discuss NB posts. Of course it had become possible much better to focus on important and difficult points.*[User #9]

- *I have been changing the way I teach over the past few years and NB has certainly helped me make my classroom more student-centered.*[User #19]

- *Not really, but it did enrich the discussions*[User #20]

- *Not exactly, but it's true that it establishes beforehand a discussion that is normally held in class. So classes must bring something different, complementary to this previous reading experience.* [User #23]

- *For my graduate seminar, I could address the smaller questions online and devote the class time to discussing bigger, more extended issues.*[User #24]

- *I think it elevates the level of discussion and the level at which I can introduce material b/c I can make a stronger assumption about their baseline level of*

knowledge. That is, I can assume they've done the reading.[User #25]

It it hence possible to introduce a new channel which doesn't compete with the other ones, on the opposite. We'd like to suggest that a key element to that success was the instructors' high expectations about the potential for the tool to improve in the in-class experience, since *Improving the in-class experience* was the theme that came most often (12 occurrences) in faculty questions (1-a) *Why did you decide to use NB?* and (4-c) *For which main purpose(s) did you introduce NB in your class?*.

Chapter 6

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 interesting open issues for future work.

6.1   Adoption versus Learning Outcomes

The Holy Grail of an educational tool is improved learning outcomes. Assessing learning outcomes is always difficult. In [58], 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.

We mentioned in section 4.3.3 our attempt to correlate NB usage to learning outcomes. So far, we have observed strong correlations and designed follow-up experiments for Fall 2013 or Spring 2014 in order to determine whether the proportion

of high level explanations authored by students affects their CSEM score.

Finally, it's always tempting to believe students who claim that NB improved their learning, but they might be biased. However, instructors (cf section 4.3.1) themselves reported the extent to which NB had a positive impact on learning: To the question *How do you feel NB impacted learning in your class this semester?* (from 1: very negatively to 7: very positively), the average was 5.55 ($\sigma = 1.05$).

6.2   Context mismatches as an opportunity

We have previously shown in section 5.3.4 that NB helped users annotate while *in the flow*, by lowering the entry barrier to enter a comment. Namely, NB takes advantage of the context so that initiating a comment requires only a single user action (selecting a region). Remarkably enough, people sometimes took advantage of context so much that the resulting comments didn't make sense. How to interpret for instance the comment "same here", when it appears as the first comment in a thread? It illustrates a typical example of context mismatch: Readers have only the geographic context in order to understand the comment. On the other hand, while writing, the author had both geographic and temporal contexts in mind and she meant *same observation as what I wrote in my previous comment*. Rather than a shortcoming, this reveals a great opportunity. Indeed, it testifies that the tool became transparent enough for users to express themselves without thinking how (albeit forgetting that their comment wouldn't make any sense to others and to themselves later). This opens UI research opportunities related to (1) how to make those comments understandable after the fact, and (2) how to extend the notion of *context* to include additional attributes, such as the existing annotations the author of a given comment had already written at that time, or her familiarity with the material.

6.3 Optimal section size and number of comments

Two common complaints have been made by NB users. One the one hand, in classes where NB wasn't used much, students have complained that it was hard to "break the ice" and start commenting on a set of empty (or almost empty) pages, a problem encountered in virtual communities which haven't yet reached a *critical mass* [40]. On the other hand, we also observed the opposite: students complaining that "everything has already being said" on documents that contained too many comments (each page of *The Art of Approximation in Science and Engineering* contained about 100 comments per page, which was perceived as too many. In fact, the problem of having too many comments to read is a key problem in determining how to scale NB to MOOCs, where hundred of thousands of students could be annotating the same document.

As we discussed in sections 4.3.3, one approach to handle the empty-room effect is to seed the documents with comments from previous terms. However, it suffers from inconveniences. As we explained above, it raises the question of how to choose which comments and how to make sure that they appear at the right place if the underlying text has changed, requiring a robust annotation positioning algorithm, such as the ones described in [32, 27]. In addition, seeding doesn't solve the problem of having too many comments per page once many conversations start kicking in.

To solve both problems at once, we propose to use an idea based on control theory: continually adjust the size of sections, based on the density of comments in each section.

1. Initialization: Randomly split the class into N sections (for example, to ensure that each section has a fixed number of participants).

2. At each time step:

   - If a section has too many comments, decrease its size, by splitting it.

   - If a section has too few comments, increase its size, by combining it with another section.

Similarly to a traditional controlled system, we can control on the integral quantity (total number of comments) or the derivative quantity (number of comments per time unit), or a combination of both, depending on the system characteristics (response time, stability) that we wish to achieve.

In its most crude implementation, we can imagine a simple cron job [14] that periodically readjusts the section sizes. One important detail, however is to make sure that users don't lose the comments that they have seen. Hence, since we keep track of what comments have been seen, the rule to determine which comments a given user should be able see is simply *all the comments in her current section, plus the comments she's already seen*), an approach that we have already successfully tested in Fall 2013 in an ongoing seeding experiment. We can imagine further improvements such as adjusting the section size so that every student feels as comfortable as possible, by asking for personal preferences for the number of comments on a page.

6.4 Better support for annotation's specific affordances

NB users discovered and took advantage of certain capabilities of annotations that are not present in traditional forums. We could provide better support for those capabilities. Above, we discussed how geographic annotation was leveraged to answer the same question repeated in several 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 the presentation of tags through less cluttered and more informative interfaces such as color coding.

6.5 Using annotations to improve the course material

An analogy with bug-tracking software could be exploited in order to help faculty by providing *document versioning* features. Our interviews with faculty members revealed a real problem in knowledge sharing in academia: when a instructor is assigned a new class, he or she often has to start from scratch. NB could be used a knowledge repository for instructors to assess what was and wasn't clearly understood in previous terms. Faculty willing to improve the course material could mark comments as solved when they upload new, improved, versions of the files. Finally, a few comments could be kept as-is in order to seed discussion for the following term. Several faculty members didn't wait for those features and spontaneously used NB in order to review their manuscript with collaborators in other institutions. This would suggest switching from the current design where annotations are anchored to rectangular areas on top of an "immutable" document into a framework similar to Phelps and Wilensky's multivalent annotations [44] that can be used to alter the document structure.

With PDF documents, one of the major challenges would be to find the part of the original document corresponding to the selection. A particular case, yet based on commonly uploaded documents on NB would be documents generated from TeX, for which a utility such as SyncTex [15] would provide the mapping back to the source file. With HTML documents, this problem disappears, at least in theory, since there is no intermediate format if writing HTML directly, for example using a WYSIWYG editor.[1]

---

[1]The problem remains when authors use a authoring language different from HTML, such as markdown.

6.6 Annotations as an aid to authoring: trails of thoughts

This section explores more advanced concepts for NB can help the learning and authoring processes.

6.6.1 The process of becoming an expert

What feels so special about a well-written paper or a good lecture? Perhaps it's the feeling to be barely understanding and having someone to guide us, and without whom we would not understand at all. In education, this regime corresponds to the Zone of Proximal Development, or ZPD [55]. We don't have to think about putting together the concepts we're being presented with. Instead, they just fit together as they come. Behind the scene, the sequence used to present the information was carefully planned to match our expectations, and answer our questions as we go along. The two roles in this guided tour scenario are the *consumer*, assumed to be unfamiliar with the subject, and looking for guidance, and the *producer*, who, already had developed an expertise in the subject.

How did producers become producers? They worked to acquire a powerful mental model about the subject. They read the material, and read it again, thought about it, asked questions, generated alternate knowledge representations (cf. ICAP hypothesis presented in section 3.1.2), let it rest for a while. They found patterns, developed their own knowledge models. In the process, they abandoned the canonical order things were presented to them, in order to create their own representations. Those *trails of thoughts* are sequences that let them associate ideas more efficiently and powerfully. In [35] Hawkins stresses the importance of sequences as the fundamental building block in his attempt to provide a framework for a theory of intelligence. As the popular joke in graduate school goes, the fastest way to learn a subject isn't to take the class, but to be a TA for that class.

Another well known mechanism to improve memorization is grouping into mean-

ingful sequences. For example it is noticeably harder try to memorize the following sequence: B,M,W,F,B,I,I,B,M,C,I,A, compared to memorizing the same sequence made by grouping those letters three by three.[2] Similarly, when shown real chessboard configurations for a short time, chess masters were able to perform much better than chess novices at recreating the board they had observed for just a few instants. However, they didn't perform better than novices when shown random configurations[3]. Another way, already mentioned in section 3.1.2, is related to the ICAP hypothesis, namely making the learning experience more active.

We already hypothesized that the benefits of NB could be connected to the fact that it promotes more active types of learning. Hence, to further improve the learning process, we should allow users to arrange what they've read or annotated into meaningful sequences. Little has been done so far in that direction: NB's document view followed the canonical page sequence of the original text, and the collage views presented annotations in an arbitrary order, e.g. sorted by date or by how many people marked "reply required." Those orderings don't lend themselves to help the learning process because they don't create any meaning between annotations. We believe that an application that facilitates the creation of meaningful sequences and trails of thoughts between comments would help the learning process.

6.6.2 A generalization of annotation and reading practices

The concept of trails of thoughts and sequences formalizes many presentation modalities that we've encountered in this thesis; for instance: collage view and document view are two simplistic sequences. Similarly holistic comments versus in-place targeted annotations are two extremes: A trail of thoughts helps develop a holistic understanding of the material by anchoring places together from different parts of some documents in a meaningful sequence and providing meaningful transitions.

In other words, a trail of thought is a graph where the nodes are in-place references and edges are the transitions that help connect the edges in order to gain a meaningful

---

[2] BMW, FBI, IBM, CIA
[3] Thanks to Rob Miller for these two examples

representation from the material. However, trails are different from a summary in the sense that they can have a flavor: rather than trying to summarize the material, their purpose is to use it in order to back up an idea being developed.

In addition to facilitating generating of new ideas rooted to existing content material, such an application could provide the teaching and machine-learning communities with a new tool to explore how people create meaning and gain expertise. We could also imagine using it as a diagnostic or help tool for people having learning disabilities, by comparing the patterns in their trails of thoughts graph with the ones in people without the disability and devise alternate teaching mechanisms. Similarly, we could use it to understand better the kind of connections that gifted learners make, and help other users to avoid rat holes, or detect and signal procrastination.

Other application for exploiting the information generated by trails of thoughts could be of use for authors: For instance, they could offer a metric of how well-written a paper is, by looking at the sequences that people use, and see how much they differ from a strictly increasing order. That could also pinpoint parts that aren't essential, or create an automatic summary by aggregating the text corresponding to the most typical sequences. However, it seems that the real gain could be realized from being able to create those trails of thoughts between ideas that are scattered across several papers. Instructors could create assignments that really force people to compare several body of works and go into details, a characteristic which was praised by faculty user #16 from our faculty survey: *This was an introductory, large-lecture Shakespeare course. In such a situation, students will frequently do anything to avoid engagement with the details of a text, including groundless speculations about historical generalization, pop psychology, and authorial biography. The notation format of NB forced them out of these comfort zones..*

6.6.3 Social trails of thoughts

How would trails of thought be affected by the social aspect of NB? Like cloning a repository on github [16], users could use someone else's trails as a basis for creating their own trails. They could also could give feedback on existing trails or make re-

quests. For instance, it could let readers interact with authors in a more participatory style: *I'd like to know more about this: Please create a "Going further" branch.*

6.6.4 UI primitives for creating trails of thoughts

"This reminds me of X"   How to help users to build trails of thoughts in practice? When learning about a new topic, learners often reason by analogy (i.e. *this reminds me of X*). Hence, we could imagine a UI paradigm that facilitates the capture of *this reminds me of...* occurrences. Often, users might not be sure of what exactly this reminds them of, so we could imagine them providing fuzzy descriptions and having other users help clarify the analogy.

Flashbacks / Flash-forwards   Successful speakers know how adjust the level of their presentation based on what the audience understands. This means adding a little reference, or a one-sentence introduction when they feel a prerequisite might be unknown. Or, when introducing a new notion, there could be flash-forward references pointing towards more advanced material that use this notion. Those could help motivate learners to understand material that seems dry and boring. For example, when learning about matrices and rotations, we could imagine flash-forward references to computer-based animation, where matrices play a fundamental role in coordinate transformations.

"As I just wrote" / "as I just saw"   As mentioned in section 6.2, we observed many occurrences of students starting a comment with "as I wrote previously." When doing so, students don't have in mind that the current way notes are being presented will make it impossible for others to know what comment they're referring to. It also suggests that one primitive to create trails could be based on displaying a short history of comments written by that author. The same paradigm can be used to reply to many comments at once, i.e. a way to quickly point at all the comments a user recently saw.

History sharing By capturing and classifying user events, the system could infer how students are using the material and the annotations. This could first be useful for the user herself: she could access the history of what trails she followed last time she was there for example. It can also be useful for the class. For instance, the system could detect common reading trends and suggest trails based on what a lot of people looked at in a similar situation, such as going back to a few sections to look at a definition.[4]

Assisting cross-references Yet another type of trail could be facilitated by assisting the user in creating cross-references. When quoting other work, authors often provide a reference as well an approximate quote. Using a full-text search index, the system could pinpoint the exact location of the quote in the related reference, so that readers can actually follow the references and get an opportunity to explore the quoted material. In cases of non-unique match, the system would suggest potential matches, and let the user specify the correct one. Finally, note that such hyperlinks could be both suggested by the the original author or by the users commenting on the document.

6.6.5 Building on top of trails

What operations could users perform with trails of thoughts? We already evoked creating, following and branching off from them. Other operations useful for manipulating concepts could be:

- Join several threads together to produce a longer trail.

- Order / reorder the threads themselves in meaningful collections.

- Insert missing transitions.

- Fan out, for example to illustrate a concept using several examples.

---

[4]Here, the word trail borrows the meaning of a dirt trail, in the sense that a trails is formed by aggregating many footsteps from other users

- Fan in, for example to gather several observations into a unified hypothesis.

- Pruning irrelevant remarks, consolidate similar ones: actions that would be often used in order to boil a long thread into a useful summary.

In summary, the main purpose for trails of thoughts would be to help turn a series of small, localized thoughts into a nice, organized reflection. It is interesting to note that this theme of research was echoed by [User 16], a faculty member who participated to our survey (cf section 4.3.1), as he replied the following to question 4-j,[5] *I will want to explore NB as a platform for peer editing in courses that will still require long-form essay writing.*

6.7   Promoting study groups

In section 5.6 we investigated when different channels can collaborate rather than compete. A legitimate question is whether offering more and more elaborate discussion capabilities on any type of document would have adverse consequences in terms of causing students to have less direct interaction with each other, for example, less work in study groups. Similar to our conclusions from section 5.6, we believe that an improved annotation software could address study group issues (for example by allowing participants to think about the material first), and actually encourage them: The system could suggest opportunistic study group matches based on the participants location, availabilities, and familiarity with the material.

---

[5] In the future, what else do you wish you could do with NB?

Chapter 7

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. Although those hypothesis were initially the product of our personal reflection, they were refined and confirmed by conversations with faculty members, and generalized by surveying students and finally faculty members who used the tool.

We posited that 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. They allow faculty to get timely feedback, and help them with the scaling problem of having a few faculty members answer questions from hundreds of students, possibly many times.

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:

1. 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.

2. 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 affordances of paper as a support for reading, described in [48].

3. 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.

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

5. 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.

6. 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. Based on our faculty survey (Spring 2013, N=29), faculty largely look forward to re-using the tool in a following term (6.21/7).

7. Previous studies argued that adoption barriers prevent online annotation. We refute this. However, instructors should be made aware that such online communal annotation tools aren't a one-size-fits-all solution. This work'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 worked our in light of our current results. However, the evidence suggests that we have reached a turning point where online social annotation systems could become a standard and valuable educational tool.

Appendix A

Faculty Survey Questions

1. Your general impressions on using NB in your class

   (a) Why did you decide to use NB? (free text)

   (b) Overall, how satisfied were you using NB in your class this semester? (1: Not at all satisfied - 7: Very satisfied)

   (c) Overall, how do you think your students felt about NB? (1: Hated it- 7: Loved it)

   (d) How do you feel NB impacted learning in your class this semester? (1: Very negatively- 7: Very positively)

   (e) What did you particularly like (if anything)? (free text)

   (f) What did you particularly dislike (if anything)? (free text)

   (g) What missing features did you need the most? (free text)

2. Structure of your class

   (a) How did you introduce NB to your students? If convenient, you may wish to copy and paste text from a class syllabus or a link to it. (free text)

   (b) Was student participation on NB mandatory? (yes or no)

   (c) What was your rationale for this decision? (free text).

3. Details of required participation (if any)

(a) What were the specific requirements? For instance was there a required number of comments per week or per chapter? Requirements on the quality of the comments? Special accommodations? Feel free to copy and paste text from your class syllabus or a link to it.

(b) How do you think students felt about the requirement?

(c) What percentage of the final grade was determined by participation on NB?

(d) Agree/Disagree: at the beginning of the semester, students felt that NB was useful (1: Strongly Disagree- 7: Strongly Agree)

(e) Agree/Disagree: at the end of the semester, students felt that NB was useful (1: Strongly Disagree- 7: Strongly Agree)

(f) Agree/Disagree: at the beginning of the semester, students went beyond the requirements (1: Strongly Disagree- 7: Strongly Agree)

(g) Agree/Disagree: by the end of the semester, students went beyond the requirements

4. Your Detailed Impressions of NB

(a) The amount of posted content was (1: Way too little - 7: Way too much)

(b) The quality of content posted by students was (1: Very low - 7: Very high)

(c) For which main purpose(s) did you introduce NB in your class? (free text)

(d) How well did NB fulfill the purposes for which you introduced it? (1: Very poorly - 7: Very well)

(e) Please detail if necessary (free text)

(f) Did you read/respond to student comments? Why or why not? (free text)

(g) Did you observe things you didn't expect? If so, please detail. How beneficial/harmful was it? (free text)

(h) Did NB change the way you taught? If so, how? (free text)

(i) Is there anything else you think we should know in order to understand the use of NB in your class? (free text)

(j) In the future, what else do you wish you could do with NB? (free text)

(k) How likely would you use NB for the next class you teach? (1: Not a chance - 7: For sure)

Appendix B

Known External Studies using NB

This section provides references to external studies that have used NB, i.e. studies for which we do not consider ourselves to be part ot the investigating staff, and that are currently known to us.

- King and Sen [13] used NB in assessing the impact of social connections and timely feedback in learning outcomes.

- Mazur and al. used NB in a peer-instruction [41] framework.

- In [56] (to appear), Wright and Newman used NB in assessing how faculty can better understand the misconceptions that students develop in a biology class, and reported on "evidence of knowledge transfer and synthesis" using NB.[1]

- In https://sparc.colorado.edu/moonhawk-kim-collaborative-reading-of-academic-articles/, Kim explains his experiment in using NB to improve students active reading skills.

---

[1]We list this study as external althought we're in the authors list because we simply helped with technical advice and informal discussions regarding NB.

Bibliography

[1] http://www.blackboard.com.

[2] http://www.sakaiproject.org.

[3] http://moodle.org.

[4] http://stellar.mit.edu.

[5] http://en.wikipedia.org/wiki/Stet_(software).

[6] http://gplv3.fsf.org/comments/gplv3-draft-4.html.

[7]

[8] http://www.nectec.or.th/icce2011/speakers/ICCE2011_michelene_20111118.pdf.

[9] http://en.wikipedia.org/wiki/HTTP_cookie.

[10] http://www.python.org.

[11] http://www.postgresql.org.

[12] https://www.djangoproject.com.

[13]

[14] http://en.wikipedia.org/wiki/Cron.

[15] http://itexmac.sourceforge.net/SyncTeX.html.

[16] https://github.com.

[17] The role of an online collaborative textbook annotation tool in a flipped introductory physics class. Technical report.

[18] Richard J. Anderson, Crystal Hoyer, Steven A. Wolfman, and Ruth Anderson. A study of digital ink in lecture presentation. In *Proceedings of CHI '04*, pages 567–574, New York, NY, USA, 2004. ACM.

[19] Benjamin B. Bederson. Interfaces for staying in the flow. *Ubiquity*, 2004:1–1, September 2004.

[20] Alice Jane Bernheim Brush, David Bargeron, Jonathan Grudin, Alan Borning, and Anoop Gupta. Supporting interaction outside of class: Anchored discussion vs. bulletin boards. In *Proceedings of CSCL 2002*, pages 425–434, 2002.

[21] Michael Bernstein, Max Van Kleek, David Karger, and Mc Schraefel. Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems*, 2008.

[22] Donald Blitzer. The wide world of computer-based education. *Advances in computers*, 15:239–283, 1976.

[23] Charles C. Bonwell and James A. Eison. Active learning, creating excitement in the classroom. *ERIC Clearinghouse on Higher Education*, 1991.

[24] D. E. Broadbent. *Perception and communication*. Pergamon Press, New York, NY, 1958.

[25] D. E. Broadbent. *Decision and stress*. Academic Press, New York, NY, 1971.

[26] J. Brotherton and G. Abowd. Lessons learned from eclass: Assessing automated capture and access in the classroom. In *ACM Transactions on Computer-Human Interaction (TOCHI)*, June 2004.

[27] AJ Brush, David Bargeron, Anoop Gupta, and Jonathan J Cadiz. Robust annotation positioning in digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 285–292. ACM, 2001.

[28] Michelene T. H. Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1):73–105, 2009.

[29] Arthur W. Chickering and Zelda F. Gamson. Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin*, 1987.

[30] Elizabeth F. Churchill, Jonathan Trevor, Sara Bly, Les Nelson, and Davor Cubranic. Anchored conversations: chatting in the context of a document. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '00, pages 454–461, New York, NY, USA, 2000. ACM.

[31] Pauline R. Clance and Suzanne A. Imes. The impostor phenomenon in high achieving women: Dynamics and therapeutic intervention.

[32] Gene Golovchinsky and Laurent Denoue. Moving markup: repositioning freeform annotations. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 21–30. ACM, 2002.

[33] Ronald Graham, Donald Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, 2nd edition, 1994.

[34] Mark Guzdial and Jennifer Turns. Effective discussion through a computer-mediated anchored forum. *The Journal of the Learning Sciences*, 9(4):437–469, 2000.

[35] J. Hawkins and S. Blakeslee. *On Intelligence*. Henry Holt and Company, 2007.

[36] H. Hembrooke and G. Gay. The lecture and the laptop: The effects of multitasking in the classroom. *Journal of Computing in Higher Education*, 15(1):46–65, 2003.

[37] Lichan Hong, Ed H. Chi, Raluca Budiu, Peter Pirolli, and Les Nelson. Spartag.us: a low cost tagging system for foraging of web content. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '08, pages 65–72, New York, NY, USA, 2008. ACM.

[38] Max Van Kleek, Michael Bernstein, Katrina Panovich, Greg Vargas, and David Karger. Note to self: Examining personal information keeping in a lightweight notetaking tool. In *In Proceeding of Conference on Human Factors in Computing Systems, ACM*, page 1477, 2009.

[39] David R Krathwohl. A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.

[40] Robert El Kraut and Paul Resnick. *Building successful online communities*. The MIT Press, 2012.

[41] E. Mazur. *Peer Instruction: A User's Manual*. Prentice Hall Series in Educational Innovation. Prentice Hall, 1997.

[42] Joseph F. Mccarthy and danah m. boyd. Digital backchannels in shared physical spaces: experiences at an academic conference. In *Human Factors in Computing Systems (CHI 2005); ACM*, pages 550–553. ACM, 2004.

[43] Frederick Mosteller. The "muddiest point in the lecture" as a feedback device. *On Teaching and Learning: The Journal of the Harvard-Danforth Center*, 3:10–21, 1989.

[44] Thomas A. Phelps and Robert Wilensky. Multivalent annotations. In *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*, pages 287–303, London, UK, 1997. Springer-Verlag.

[45] Matt Ratto, R Benjamin Shapiro, Tan Minh Truong, and William G Griswold. The activeclass project: Experiments in encouraging classroom participation. *Computer support for collaborative learning*, pages 477 486, 2003.

[46] Marlene Scardamalia. Csile / knowledge forum. In *Education and Technology: An encyclopedia*, pages 183–192, 2004.

[47] B. N. Schilit, G. Golovchinsky, and M. N. Price. Beyond paper: supporting active reading with free form digital ink annotation. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 1998. ACM.

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

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

[50] E.F. Taylor and J.A. Wheeler. *Spacetime Physics*. W. H. Freeman, 1992.

[51] M.J.W. Thomas. Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, 18(3):351–366, September 2002.

[52] Joseph C. Ugrin and J. Michael Pearson. The effects of sanctions and stigmas on cyberloafing. *Computers in Human Behavior*, 29(3):812 – 820, 2013.

[53] Jakko van der Pol. *Facilitating Online Learning Conversations*. PhD thesis, Universiteit Utrecht, 2007.

[54] Claudia Von Aufschnaiter, Sibel Erduran, Jonathan Osborne, and Shirley Simon. Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1):101–131, 2008.

[55] L. S. Vygotsky. *Mind and society: The Development of Higher Mental Processes*. Harvard University Press, Cambridge, MA, 1978.

[56] L. Kate Wright, Sacha Zyto, David R. Karger, and Dina Newman. Online reading informs classroom instruction and promotes collaborative learning.

[57] Erping Zhu. Teaching with clickers. *Center for research on learning and teaching occasional papers*, 22:1–8, 2007.

[58] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1883–1892, New York, NY, USA, 2012. ACM.