Creating Context: Design-based Research in Creating and Understanding CSCL

Christopher Plekss Hoadley

SRI International and Stanford University
Christopher.Hoadley@sri.com

ABSTRACT

One of the biggest challenges in helping students learn via CSCL is embedding their work in appropriate social contexts and helping create a culture of inquiry and collaboration. This article describes how design-based research allowed the deliberate evolution of a set of tools and practices to help students collaborate effectively. The SpeakEasy, one of the earliest Web-based discussion boards, was evolved from prior discussion tools, adapted to an Internet-based science learning environment, and evolved to work with both online and offline classroom projects and practices. Research conducted as part of the evolution shows how social cues can be used to help students develop an integrated understanding of science. Implications for the design of socio-technical systems are discussed.

Keywords

Design, research methodologies, threaded discussion, science education

INTRODUCTION

This article describes a general approach to combining research and design in the creation of collaborative learning in an online tool. Rather than providing simple prescriptions on how to use the Internet for learning as determined by simple either/or comparisons, the "hard research" that is so lauded by the media, this article advocates a different approach to research and development in CSCL environments that is based on combining design and research. Research is important, and the defining characteristic of research is an empirical stance, a willingness to "listen to the data" and to look for patterns that hold true across time and space. However, as is true with most educational research, the simple studies and simple answers ("Which is best, A or B?") can be misleading. The tricky part of doing educational research on CSCL is in the details—interventions may take on widely varying forms depending on the teacher (if any), the learning context, and even the particular geographic location. In technology research in particular, many researchers ask questions that bely the role of context. "Is tool A better than tool B?" is a foolish question if one doesn't ever examine what is *done* with tools A and B. It's as if one tried to answer the question, "Are books better than pencil and paper in classrooms?" by running a carefully controlled study in which half the classrooms used each without regard to purpose. I advocate an alternative approach. In the past, work combining software design and research in education has been described as design studies, action research, or design experiments (Brown, 1992; Collins, 1992; diSessa, 1991), which I group under the label *design-based research methods*. In this article, I describe one program of research on a particular CSCL tool in which comparison studies were used to answer important questions, but every experiment was highly embedded in a web of efforts that blended creative technology and curriculum generation, proactive implementation, and iteration. This cycle of activities ensured that the comparisons examined by the research team made sense, and that the interventions we tested represent the best possible examples of their kind we could provide. By describing this example program of research, I hope to provide evidence that design-based research methods can and should play an important role in CSCL. In particular, intertwining design and research is especially important for establishing *collaborative contexts,* or activities and cultural structures that support collaboration leading to learning. Unfortunately, in much CSCL research essential components of collaborative contexts are implicitly codesigned by the developers of the technologies or the researchers, but they are not adequately recognized or reported in the research, reducing the applicability of the findings.

Design as a context of research

When we discuss *design*, we imply certain ideas about the character of the activities we engage in. First and foremost, design is purposeful and creative. In the story below, our purpose was fundamental to our approach: we were seeking ways to ensure that young students (in our case, 12- to 14-year-olds) were able to learn science, not only to develop theories about CSCL or learning in general. We were troubled by the deficits that seemed rampant, including disconnected knowledge that students might parrot but didn't understand and certainly couldn't apply to their own lives (diSessa, 1988; Linn, Songer, & Eylon, 1996). This actually set us apart from pure technologists in that our major goal was not to find application of technology, but to enhance learning.

A second defining feature of design is that design is open-ended. This is usually thought of as what makes design challenging (as compared to, for instance, "problem solving," Newell & Simon, 1972). However, open-endedness proves to be an advantage in educational technology research because it means our designs are well suited to the types of open-ended questions our research addresses, such as, “How can we best use technology to support reasoning in thermodynamics?” (as compared to, “Are computers better than filmstrips?”).

Good design is iterative. The process of creating something to address a goal is repeated many times as the designed artifact or process is tested, observed, and refined. The iterative nature of design is often missing in research, but is vital in testing our interventions. By repeatedly creating, implementing, enacting, and improving our interventions, one begins to understand intuitively and empirically what works and what doesn’t, and also which features of the design are essential and which are irrelevant to the goals. In typical design, especially typical software design, this type of refinement is an informal way of doing research—“user testing” can encompass experimentation that would pass muster with the most stringent research methodologists, but usually it is far more informal. The sage researcher uses mixed methodologies combining informal and formal methods according to costs and benefits (Neilsen, 1994). In the case of the research team I joined, we used this refinement cycle as an opportunity to listen to our data and to conduct studies that were robust because they were meaningful and were grounded in the extensive contextual knowledge that came from participating in the design process that created the intervention in the first place. As with scientific research in general, we used studies to test hypotheses and to ground us as we constructed falsifiable models and theories from our data.

Design narratives and their importance

One of the fundamental ideas in the scientific paradigm is replicability; a scientist’s report of an experiment should include enough information to permit others to repeat the experiment. Unfortunately, this is often impossible in CSCL research, for two reasons: First, because our interventions are culturally embodied, the complexity of human nature may prevent us from adequately and completely describing our research context. This problem is well explored in the field of ethnography, and researchers turn to richer and richer descriptions (so-called “thick descriptions”) of a research setting to communicate factors that may be relevant. A second, related idea is that educational research is often naturalistic and may be quasi-experimental, correlational, or descriptive. That is to say, often as researchers, we do not have the ability to control every variable, every iota of human experience in and around a classroom or learning environment, much less the out-of-school experiences students and teachers bring to their classroom lives. Because we cannot precisely engineer cultural context, we may not be able to exactly replicate an experiment. For these reasons, we may not be able to replicate others’ findings since we may not be able to recreate exactly the conditions that they encountered. What this all means is not that empirical research in CSCL is hopeless, but rather that there is a high art to identifying which factors are most relevant to this particular situation and to communicating results in a manner that appropriately contextualizes them. While findings are not universal in the tradition of physical science research, they are often helpful to others in similar (but distinct) contexts. Rather than inscribing laws in some book of truth, the goal is to conduct research which leads to locally grounded theories and findings, and through application by experienced practitioners in other contexts, to uncover just how localized or generalizable research findings are.

In the context of design-based research, we must endeavor to meet the challenge of replicability by adequately describing our research. Not only is the researcher obligated to fully describe the tools he or she may have built, but also relate as fully as possible the context in which the tools are being studied, the activities and practices offered to the users, and, most importantly, the evolution of the context over time in response to the tools. Consider how infrequently educational technology research (even some CSCL research) carries this type of description; the usual study presents a technology fully formed as if it had risen from the oceans like Venus herself; describes, at best, little of how the technology was introduced into the research setting; and may not even describe how the technology was used before judging its “effectiveness” in learning by means of some (possibly unrelated) post-test.

Contrast the typical research paper with the notion of a design narrative. Narrative is a structure for conveying a series of related events, a plot. Narrative may omit details, but important agents, events, causes, and results are relayed. A design narrative describes the history and evolution of a design over time. It may not be as complete as, for instance, videotapes of the entire design process and all uses of the designed artifacts, but it does communicate compactly and effectively how a design came into being. By relating the design’s changes over time, a design narrative can help make explicit some of the implicit knowledge the designer or designer-researcher used to understand and implement the intervention. Would that all interventional research included this kind of rich description of the “treatment” so that one might infer whether the results were applicable elsewhere.

Narrative is only one way of making sense of design-based research. In a number of cases, controlled studies helped inform the design decisions the research team and I made in implementing the interventions described here. Where appropriate, I allude to the experimentation or other data used to make our decisions. However, the goal of this article is not to provide a methodologically rigorous presentation of the myriad studies that informed our design, but rather to give the general shape of the design process and to describe what we learned in the large; by necessity, in covering more than eight years of work, I resort to a more sweeping and less detailed description.

Below, I make use of the design narrative form to describe the evolution of some of the collaborative technologies we researched and highlight the complementary roles of design and research. By reflecting on the evolution of the designs and research over time, one can see the strengths of this complementarity. The outcomes of our endeavors included locally applicable design principles (local sciences in diSessa's terminology) that help point the way to important overarching findings that isolate relevant factors in technologies' use (a more global science). We performed our methodological duty in trying to test some of our most important hypotheses with some of the strictest methodological techniques in place: controlled comparisons with random assignment, even double-blind coding of outcomes. For the details of these controlled comparisons, I refer readers to the cited papers. To really convey what happened, though, requires a story.

Designing for collaboration

This article is about some designs of technologies and activities that fostered collaborative aspects of learning, predominantly in the Knowledge Integration Environment (KIE) research project (Bell, Davis, & Linn, 1995; Hoadley & Bell, 1996) which developed software for Internet-based middle school science education. Collaboration research adds design complexity, is particularly sensitive to variations in context, and any intervention reverberates through the setting changing both the individuals and the social context. Time is required to see how the intervention settles into a more stable state as both individuals' practice and the group practices adapt to the new tools and possibly reach equilibrium. Here, I give a design narrative of work that provided rich contexts for studying how technology could scaffold student learning and knowledge integration in science. I will try to point out how technology, activity, and local culture interrelated in our studies and how our design stance helped our research, and vice versa. The central message is that by engaging in design on both a technical and a social level, one can arrive at valuable insights in how to foster computer-supported collaborative learning. This central point has been argued by others at a theoretical level (Koschmann, 1996); here, I argue it from the point of view of our research on Internet project-based learning tools.

THE SPEAKEASY DISCUSSION TOOL: A DESIGN NARRATIVE

The story of the SpeakEasy discussion tool takes place over a span of approximately eight years. SpeakEasy was one of the first two Web–based threaded discussion tools (along with HyperNews) that are so familiar to Internet users today, predating the introduction of the first Netscape browser. SpeakEasy has several unique features that have proven useful in fostering learning in science classes. In our last study, SpeakEasy discussion doubled the prevalence of correct conceptions in the student population and significantly improved partially correct conceptions. (Hoadley, 1999; Hoadley & Linn, 2000) To some extent, the point of this narrative is to describe how powerful technology can be in improving how students talk to and learn from each other. A second message is how beautifully subtle the relationships between tools and collaboration can be.

The story begins in 1992, before widespread adoption of the World Wide Web. Initially three people (Sherry Hsi, Christina Schwarz, and I) contemplated an interesting question: Could multimedia technology solve a problem educational researchers had; namely, that collaborative analysis of videotape was cumbersome and required same-time–same-place meeting in front of a videotape player? Hsi had recently seen some interesting uses of multimedia for messaging while interning at the Apple Multimedia Laboratory, and we each believed that we could help support asynchronous video analysis through similar technology.

Our design goal was straightforward: allow discussion of videotape among researchers who weren't in a single location at the same time. Like many design problems, this one capitalized on the potential of technology to make possible what had previously been impossible. We designed our initial prototype in HyperCard and dubbed it the Multimedia Forum Kiosk, or MFK. We examined prior interfaces such as Internet newsgroups (at that time, primarily an academic communication medium) and email mailing lists. We adopted an unofficial motto of "better than Net news" because we hoped to create a more reflective, less impulsive dialogue and, to the extent possible, avoid needless "flaming" (reactive, inflammatory comments that were more confrontational than the participant would contribute in a face-to-face discussion). Another important example we considered was Scardamalia and Bereiter's tool, CSILE (Scardamalia & Bereiter, 1992; Scardamalia & Bereiter, 1994; Scardamalia, Bereiter, McLean, Swallow, & Woodruff, 1989). We appreciated the ways in which CSILE encouraged reflective discourse, but we wanted to incorporate a more general discursive model than CSILE's (which was primarily science-focused), to foster a sense of community or awareness of others in the dialogue (CSILE didn't directly support social awareness), and to integrate video into discussions (Hoadley & Hsi, 1993).

Our tool had many common features (including a top-level organization by topic and threaded discussion) and several features which made it unique (Hoadley, Hsi, & Berman, 1995). First, it provided two collaboration spaces, one, the *opinion area*, allowed one comment per person on the topic which could be revised over time, while the second, the *discussion area*, allowed threaded discussion but did not allow revision of prior comments, only response. Secondly, the tool made use of *semantic labels*, or labels from a fixed set of choices (we borrowed this idea from Scardamalia and Bereiter, but while their categories were specific to scientific discourse, ours were aligned with a more general model of small group discussion: see Bales, 1969). Third, we made extensive use of social cues throughout the interface based on a theory of social representations. All comments were represented by face icons and all topics were introduced by a topic author. This tool underwent at least three major redesigns, with at least two incarnations as the Multimedia Forum Kiosk, and at least two incarnations as the Web-based tool, SpeakEasy. In this design narrative, I do not describe every design change (or even all of the major ones) but rather choose some to illustrate how our stance of design experimentation and design-based research led to new insights about generating collaboration for science learning.

Usability vs. conditions of use

Naively, we assumed that usability would be the primary indicator of success in our design. After creating the initial prototype, we tested the tool with subjects from an education research department using think-aloud analyses, time-usage analyses, and interviews. Our initial analysis did in fact demonstrate that the tool was usable—our test subjects were given no instruction and still managed to uncover and use every feature of the system, from reading and navigating comments to contributing their own comments in both the opinion area (nonthreaded) and discussion area (threaded). In one case, the think-aloud protocol provided direct evidence that our semantic types prompted reflective thinking and prevented a "flame." Interestingly, one of the first lessons we learned in this study (though we hardly noted it at the time) was the importance of seeding a discussion—adding not only a topic for discussion, but some sample viewpoints that would help initiate the discussion. By usability metrics, our system was a success already; people quickly figured out what it was for and how to use it, even people who hadn't used Internet newsgroups or, indeed, any online discussion tools other than email (Hsi, Hoadley, & Schwarz, 1992).

Our first field trial was less encouraging. The system was set up in a department lounge, and we undertook the time-consuming task of providing everyone in the department (around 30 to 50 people) with accounts, which involved laboriously taking and digitizing their photos. Since every member of the department had to pose for a photo, each person had at least some contact with the developers where we could explain to them the purpose of the system, more than we had done for our lab subjects. The topics we included were of general interest to members of the department (including both informal topics and comments on research data video segments). The accounts didn't even require learning a password, just choosing one's name from a pull-down menu. The location was frequented by most members of the department, as it contained mailboxes and was the location for well-attended, weekly community teas. We therefore hypothesized that people would participate in the online discussion, since the system addressed a known and professed need, and it didn't appear to provide any barriers to use.

Even so, the participation was underwhelming—each of four topics had approximately a dozen comments after an entire semester. We asked our friends and colleagues why they didn't participate, and heard answers like, "I don't have time," or, more tellingly, "I don't expect to do that sort of thing in the lounge." In fact, many people were discussing research in the lounge face-to-face, but the online system didn't really fit into the activity structure of the place. When people did use the system, they were often the only person in the lounge at the time (e.g., someone working late who had stopped in for a break or a cup of coffee). Once someone began using the tool, the arrival of additional people might spawn conversation over this artifact. This was an instructive first lesson on inserting our tool into existing settings and practices.

Designing functional activities and implementing conditions of use

Like many research and design projects, ours was subject to external constraints. What started as a small, unfunded project for researcher communication was repurposed. We received the first grant to study the system through a coalition of engineering schools, united in improving their undergraduate curricula through technology. This first grant was the result of what turned into an ongoing process of shopping our technology around, doing demos and presentations for anyone who would listen while trying to get expert design feedback from colleagues in HCI, education, and technology (Hoadley, Hsi, & Linn, 1993).

Initially, we took our tool into engineering classrooms on several college campuses, both graduate and undergraduate. At this time, we also started installing the tool elsewhere: a self-paced study center for undergraduates, a museum, the lobby of a college building. Partially through discussions with users, partially through comments students left in the system, and partially by comparing participation in the different settings, we realized that there were important preconditions for use (Hoadley, Hsi, & Linn, 1994; S. Hsi & C. M. Hoadley, 1994). The public installations turned out to be too idiosyncratic for us to understand what made some people use them and other people not, but the classroom experiences started giving us some consistent messages. First, we realized that students' use of the tool was directly related to their ability to access the kiosk running the software (remember, this was prior to widespread use of even the Mosaic browser), the degree to which the topics were perceived as relevant and interesting, and the degree to which the tool was integrated with their course. (Hoadley et al., 1994; S. Hsi & C. Hoadley, 1994) These findings seem obvious in hindsight, but addressing them is easier said than done, and involved significant exploration in our contexts. For instance, we thought of classrooms and public spaces as easy to access, but they were not because of the social discomfort caused by working on the kiosk in these spaces. Instead, laboratories provided a much more approachable venue, since students were used to being collocated with other students working on independent activities. Likewise, the perceived relevance and interest of the topics we posed came out differently than we expected. Topics that were highly controversial, or better yet, topics with diametrically opposed seed comments, were engaging. Generally, extreme viewpoints provoked reaction. Topics that we thought would be interesting to students (like discussing the strengths or weaknesses of the course) were too vague and provoked little interest.

Regarding integration with the course, we saw different instantiations of integration that supported the tool via the course and vice versa. In some cases, students felt they were better able to solve homework problems if they read and participated in the online discussion because the topics closely paralleled the technical content in class, and in other cases students participated because the instructor summarized comments in class and reacted to them, indicating a strong interest on the part of the professor. In many cases, anonymity played a big role in the participation, as students had few if any ways to communicate anonymously with their instructors besides our system. In some other cases, the asynchronous nature of the communication medium proved important; for instance, students with limited proficiency in English were able to participate in the discourse by taking extra time to read comments and prepare responses in English. The integration with the course also took some interesting twists. While some instructors actually provided participation grades for contributing comments to the system, we had nearly equivalent participation when an instructor read, summarized, and responded to student comments in class (this was a large course with nearly 100 students, and other opportunities to influence instruction were rare). The kiss of death, though, was superficial integration with the course—even if students were introduced to the system in class, if the instructor never mentioned the system again and didn't give grades on it, most students would opt not to participate. The few who did participate in these circumstances, interestingly, were often women or minorities. Without the in-class discussions and one-on-one interactions the kiosk provoked, the kiosk itself would have been a different intervention. Identifying the nature and scope of the intervention when the cultural changes provoked by our tools and activities were co-constructed simultaneously with use of the tools and activities made traditional before-and-after testing less meaningful. This coevolution of phenomena proved to pose a methodological challenge that would crop up repeatedly, one that is probably intrinsic to the problem of studying collaboration (Barab, Hay, & Yamagata-Lynch, 2001; Hoadley, 1999; Roth, 2001).

The overall lesson we learned, one which supports the strong design stance in our research, was that implementation and adoption required a lot of social design: designing activity structures that made sense in the local context, and implementing those designs either through our own participation in the community or (as was especially the case with sites we worked with at a distance) through communicating with leaders like faculty, teaching assistants, lab managers, and students about how to create their own successful activity structures in which the tool's use made sense. Had we simply scattered the tool to the four winds and tested outcomes, we might never have realized what conditions of use needed to be met, nor would we have been able to proliferate those conditions as a theme and variations in a wide variety of contexts. When testing new tools, as we were, any sort of research on effectiveness would have been meaningless without giving the tools a chance to succeed by helping establish best practices of use. This point bears repeating. Certainly, though one may study the outcomes of technologies in all the naturally occurring variations of use that might arise in the field, these studies may not answer the question we really want to know, which is: What will happen if the tool really takes root? Like the hypothetical study comparing blackboards to notebooks, we might get a lot of data but it doesn't address meaningful questions about how best to educate or support learning.

Evolving with the background (technology and culture)

Later in the development of the system, we began experimenting with our discussion tools in the Computer as Learning Partner middle-school science classroom (with 12- to 13-year-old students (Linn & Hsi, 2000). Initially, this experimentation began with the Multimedia Forum Kiosk technology and science-oriented topics (Hoadley, 1999; Hsi, 1997). There were important interactions between our tool and the culture of the classroom, interactions that evolved as tools influenced use and use influenced culture. Some elements of the local culture already supported use. For instance, students in this classroom (which had a 2:1 ratio of students to computers) were familiar with computers, and each student had some prior experience working on a computer. Likewise, the teacher had previously started a tradition of coming in to work on labs or computer work during lunch and immediately before and after school; the system benefited from these practices. Other aspects of the culture evolved in ways that we would not have predicted. For instance, the fact that the system was based on a sole kiosk (we actually had two computers in a single kiosk, but each student had an account on only one of the two machines) led to some interesting cultural outcomes. Initially, the single kiosk enhanced interest and face-to-face collaboration—students would gather around the kiosk and read over each others' shoulders as comments were made. The relative rarity of the kiosk machines made them more attractive, and soon "kiosk groupies" would frequently visit the machine as a social group outside of class time. Unfortunately, the emergence of these groupies began to erode access to the discussion for other students; the stronger the social bond between the groupies became, the harder it was for those not in the clique to access the machine. The teacher, who was aware of the problem, began to try different ways to ensure access, including a signup sheet for time on the kiosk and strategic shooing when clumps of people began to form around the machine. The teacher did not dissuade all groups from clustering around the machines, but rather based his actions on who else was in the room and whether they were likely to be encouraged or dissuaded by the current group near the kiosk (Hsi, 1997). This type of very nuanced design activity was only possible because the teacher was aware of activity around the machine (in part with the help of the researchers) and had a number of techniques to try to encourage equitable access. It is likely that in other circumstances different social issues would have arisen and required different interventions to allow all students to participate in the online discussion. Eventually, we moved to the Web-based SpeakEasy system which eliminated the problem of a single point of access, but raised other issues about which students had access to the Internet at home or other out-of-school locations.

Another aspect of our intervention coevolving with culture happened later, as the culture of technology changed outside the school. When we switched to the SpeakEasy tool from the MFK software (mid-semester), our students brought their prior practices easily to the networked version of the tool; and student participation rates escalated slightly but insignificantly. We found no differences in student comment length or quality. This switch occurred around the beginning of the KIE project, near the introduction of Netscape 2.0. At that time, we began to introduce Internet technology to the class by choosing a few students from each class period to be technology guides. At first, we saw an average of one or two students per class period (out of approximately 25 to 30 students in each class period) who had any experience with the Internet at all. Almost none of these students had experience with the World Wide Web. In an after-school session lasting about an hour, we gave the guides an introduction to the Web that included instruction on what hyperlinks looked like, how to click on them, and how to use the "Back" and "Forward" buttons to retrace their prior steps. The rest of the students got an abbreviated version of this tutorial and were encouraged to seek help from peer guides.

When students began to use the online discussion tool, they often perceived it to be a completely different social setting, with different expectations, than their familiar face-to-face counterparts such as in-class time or on the playground. Hsi documented how this worked to our advantage, as students expressed amazement not only that their peers could discuss science topics with them, but also that their peers had different ideas than they did about scientific phenomena. This eye-opening experience was described by many students in clinical interviews, and many students contrasted the rules of the new space with those in other social spaces, explicitly denying that they would ever have the same conversations with the same people (their peers at the school) face-to-face (in class or out). The ability of the teacher to "stake out" this new social territory as being for intellectual, student-centered, science-oriented discussion was a powerful point of leverage on the students' social interaction (Hoadley, 1999; Hsi, 1997).

Over time, this advantage dissipated due to changes in the cultural surround. Within three years of this initial run, the Internet went from being unknown to being ubiquitous. Not only did a majority of students come to class with knowledge of hyperlinks and browsers, they had favorite search engines, Web sites, and deeply held beliefs about what types of activity one would perform on the Internet. Our initial training needs decreased (no need to explain what blue, underlined text stood for) and student access from home and from the popular nearby library skyrocketed. However, students came to class with strong expectations about what online discussion was like. Increasingly, students would mention AOL chat rooms, email, and other online discussions in their interviews about the SpeakEasy, and it became more and more difficult to ensure that students held to the norms we tried to set in SpeakEasy. The teacher spontaneously began to differentiate the tool when introducing it to the class, by describing how special it was, how experimental, and so on, and by explicitly contrasting it with AOL chat. Maintaining the sense of our online discussions as new social territory required deliberate effort.

Likewise, we were aided by invoking cultural norms specific to the classroom environment. Students might not have had a good idea of what scientific explanation, argument, and questions looked like before coming to this course, but this was a genre the teacher could invoke as the students learned these concepts during the semester. This prospect in particular suggests how delicately intertwined the nature of the cultural practices and the nature of the tool itself are, and how locally (and temporally) specific they are. While one might think the 1990s are an exception to the rule due to the rapid growth of the Internet, in fact the technological and techno-cultural surround are always changing. Fads, new technology developments, and local culture will always mean our interventions are aimed at a moving target of existing culture.

Shaping collaboration through feature improvement

Given the plethora of external influences changing students' practices, do we as designers of technology have any leverage on the situation, any ways we can influence learning through the technologies? The answer is a qualified yes. In our work, we saw, again and again, how small changes in technology could have large and pervasive impact on behavior and practice. One of the most dramatic examples of this in our work occurred when the middle-school classroom got new computers; the classroom upgraded from Macintosh LC II computers to new, faster Power Macintosh clones. This change occurred mid-semester, so students had already begun working with our technology environment. Every detail of the user interface was the same, from the KIE software down to the operating system environment; only the speed of the computers differed. Overnight, student writing in their online assignments almost doubled, compared to their own work earlier in the semester and to prior semesters of student work. This experience serves as reinforcement of the idea that technology use *will* change over time, even if the tools we are studying don't themselves change. Likewise, it proves that the most powerful changes may come from the least expected places. Often, it is not what the computer makes possible, but what it makes easy, that proves to have the greatest impact. Because the rest of the research team and I had intimate contact with the environment under study, we could make mid-course corrections and help the students adapt to the technologies we provided and improve the affordances of our tools.

It is important to note that our design process was principled and relied on a specific, tentative model of how collaboration would foster learning. We recognized that poorly implemented collaboration could hinder learning as much as help (Linn & Burbules, 1993). Our model of productive discussion (after Hsi & Hoadley, 1997; Pea, 1992) dovetailed with the knowledge integration approach taken elsewhere in our research program. We faced two challenges: first, to ensure participation in discussion; second, to ensure the discussion was productive, meaning that it demonstrated the features hypothesized to be necessary (and possibly sufficient) for learning via discussion. Briefly, these features are: inclusiveness and participation (all members of the discussion are able to participate), the externalization of a repertoire of understandings or models of the domain (often different initial viewpoints), differentiation processes (where old models lead to new variants), linking (consideration of which models are coherent or incoherent), and selection (privileging or selecting the models that have the most explanatory power and coherence). In addition, as a component of a larger set of interventions (initially, the Computer as Learning Partner microcomputer-based laboratories (Linn & Hsi, 2000) and later the Knowledge Integration Environment suite of tools and activities) we had a responsibility to contribute to the overall goals of the project. We explicitly tried to help students develop their scientific epistemology through a coherent curriculum that included real-world experiences, laboratory experiences and inquiry, and critical examination of information resources from the Internet. Eventually, we succeeded in all these goals, although it took two dissertations to develop and implement a workable set of tools and activities, ensure that our tools were actually fostering productive discussion (Hsi, 1997), and demonstrate how this productive discussion leads to individual learning (Hoadley, 1999).

Anonymity: a highly context-dependent feature

Here, I describe the evolution of a set of technology features that helped support more equitable discussion practices among the students we worked with. Equity is an important issue, especially for middle-school science, where girls, who have higher achievement than boys in the primary grades, begin a downward trend compared to their male peers, presumably due to social factors. In particular, girls are often disadvantaged in classroom talk (AAUW Educational Foundation, 1992). Both because this is a recognized problem in participation, and because inclusiveness is an important component of our model of productive discussions, we had a deliberate goal of ensuring equitable participation by members of both genders. In our engineering work, we saw that the ability to communicate asynchronously, without needing to interrupt or take the floor to contribute, was an important force towards inclusiveness. (Asynchronous, text-based communication was also anecdotally related to the ability of non-native speakers of English to participate in the discussions in our engineering work.) We also saw that anonymity was important for participants who might not have social status but wished to express their views. This in particular conflicted with earlier theories that had driven our work: specifically, a theory that representations that included social context information and were socially engaging would promote ownership of ideas and motivate participation. It was for this reason that we had initially included face icons as part of the initial MFK system and had carried that feature through each iteration. However, we also heard that students were making use of anonymity in support of their participation, which would suggest that less social representations might be better. This became an important question for us as we investigated the role of identity in online participation and as we investigated how our system affected both genders.

The initial MFK system had a limited set of pseudonymous identities that people could use to contribute anonymously, such as Minnie Mouse. These icons were initially created to allow users to participate who had not been previously set up in the system. We also saw the possibility that they could be used to contribute anonymously and therefore made it possible to contribute using one of these pseudonymous identities even after logging in as oneself. Initially, we questioned whether consistent pseudonymity was important and several versions of the MFK were designed so that each person, when commenting anonymously, was given a separate anonymous identity, making it possible to identify which anonymous comments were made by the same or distinct individuals, even if the specific individual could not be identified. We did find in surveys that participants appreciated the ability to contribute anonymously. Some discussions were heavily anonymous (especially those discussing sensitive topics such as classroom atmosphere in the college engineering courses), while others had less anonymity. Interestingly, in one semester with the four engineering instructors, we noticed much less anonymity in the discussions of the two courses led by female professors than in the two courses led by male professors. Gender certainly seemed to be playing some role in the participation structures.

Hsi and I undertook a more careful comparison in the middle-school science classroom. Students were given free choice of anonymity, and girls contributed significantly more of the anonymous comments than boys (Hsi & Hoadley, 1997). Interviews with boys and girls revealed that the girls cited social safety (avoiding embarrassment) as the primary reason that online discussion was better than offline discussion. In what was expected to be a replication, we varied whether students were forced to attribute their comments to their real names and identities or were forced to attribute their comments. Surprisingly, we saw no significant differences between participation in the two groups, and no interactions between treatment group and gender (Hsi & Hoadley, 1997).

How could we explain these findings? In interviews with girls and boys in later semesters (with free choice of anonymity), girls often mentioned the option of anonymity as an important social feature that increased their comfort level in the discussion. Surprisingly, many of the girls who mentioned this *never made anonymous comments in any discussions.* As designers, we found this to be an exceptionally poignant example of a finding that would not have been uncovered without iterative design. We had created an interface feature that had important benefits for the collaboration without even being used! If use of the anonymity feature was independent of how the feature affected social comfort, how could we explain why some students used the anonymity feature while others did not?

It was around this time that we probed student beliefs about anonymity and attribution further. We surveyed, interviewed, and observed students to ascertain how they might view or use attribution in navigating or understanding student comments. Half of the students navigated the comments in the discussion (chose which ones to read or in which order to read them) on the basis of attribution, and students frequently stated that they liked being able to tell who had contributed a comment before and after reading the contribution. Many students explicitly said that they avoided reading anonymous comments. This contradicted the impression held by many girls that anonymity was an important safety valve to allow students to honestly and safely express ideas to their peers. It appeared that students were less likely to read anonymous comments, which defeated the inclusivity purpose of the anonymity feature, one of the central aspects of our theory of productive discussions. Students might feel empowered to contribute to the discussion if they could do so anonymously, but their ideas were not being heard by other students. Around this time, we switched from the stand-alone MFK system to the Web-based SpeakEasy.

We got our big break by examining who was making anonymous comments. We found that rates of anonymity were surprisingly consistent for any given individual over time. That is to say, the percentage of comments made anonymously by a person in one discussion correlated very highly with the percentage of comments made anonymously by the same person in a later discussion. Also, the percentage of comments made by a person in a discussion correlated with rates of anonymity for other students in the same discussion. Thus, some discussions had a large amount of anonymous participation by many individuals while others did not (Hoadley, 1999).

We finally uncovered a large part of the reason for anonymous contribution through informal observation and discussion with students in the classroom. Many students (not surprisingly) would skim the comments already in the discussion before contributing their initial opinion. If the students encountered mostly (or entirely) anonymous opinions, they themselves would contribute anonymously. This happened quite frequently since we had learned to seed discussions with comments to avoid an intimidating “blank slate” discussion. To avoid presenting these views as authoritative (coming from us as researchers), we added them anonymously. This anonymity would be perpetuated as increasing numbers of anonymous opinions accumulated, further discouraging students from contributing their views under their own name. The reason that some discussions had escaped this fate was that some students preferred to contribute before reading others’ comments. These students were basing their decisions about comment attribution on their own sense of confidence rather than on the prior contributions.

Responding to this realization, we designed a simple intervention that would encourage students to participate with attribution. Resurrecting an interface design we had employed earlier, we changed the system to force students to contribute their opinion on the topic before browsing others’ opinions. We had dropped this feature when we had introduced it previously because users were reluctant to state their views without exploring the topic (especially for science topics that were new to them), but we found this reluctance could be overcome. We also emphasized in our oral introduction to the system that students should revise their opinions as often as their views changed, even during their first login session, if change was warranted. The new feature and the new instructions had three benefits: students were less likely to comment anonymously (since they were basing their decisions on their own confidence rather than peer pressure exerted by the fictitious contributors of the seed comments), students were encouraged to develop a habit of revising their opinion area comments, and we as researchers got the beneficial side-effect of having a true student pretest for the topic (which was ultimately part of the data collection technique for our individual learning measures.) Overall, student participation—reading and writing comments—remained equally high as without the new feature (actually trending toward an increase), gender balance of contributions remained high (with trends favoring girls), and anonymity (which had inhibited other students from reading the comments) dropped significantly.

In this way, through a design stance and a close involvement with the classroom, we short-circuited what might have been a long series of expensive studies that would have misled us about how anonymity could benefit the discussion. Indeed, our view on anonymity in discussion changed from believing anonymous participation was evidence of inclusiveness to believing it was a threat to inclusiveness. By designing a new technology feature and some new activities around the feature, we were able to maintain the sense of safety in the discussion by allowing the option of anonymous participation, while greatly reducing the negative impact heavy use of that option previously implied.

Consider how differently this research might have unfolded if we had instead conducted laboratory studies. Certainly, since the discussions represented sustained effort on the part of the students, we would have had to make use of a demandingly long research protocol. The investment in subject hours required to run the experiment would have probably encouraged us to carefully pilot and then fix a particular set of instructions and a particular version of the interface. The iteration we conducted on a time scale of several years would have been far less likely. There is every likelihood we would have misinterpreted the role of gender and anonymity in the interface. Even if, by some miracle, we had uncovered the inconsistencies between girls attitudes as a result of the *presence* of the anonymity option versus the effects of *use* of the anonymity option, we wouldn't have had the informal observation that led us to not only a sensible explanation, but an easy remediation. This is the power of design-based research methodologies.

In this design narrative, I have described how a particular discussion tool coevolved with various activities in a context of learning science. The moral of this story is not about the particulars of the design of an online discussion system (this is another interesting story told elsewhere, as in Hoadley & Linn, 2000; Hsi & Hoadley, 1997). Rather, it serves as an example of the crucial interrelationship between the collaborative tool and the ways in which the tool is construed and embedded in local participants' activity structures. It also shows how a detective-like attentiveness to details and causes of social phenomena by participants (in this case, by the researchers and teacher) allows for a much greater degree of robustness, as idiosyncratic barriers to productive discussion can be sniffed out and addressed through (sometimes trivially easy) intervention.

CONCLUSIONS

Through this design narrative, I have described some of the advantages of a unified approach integrating design and research. First, as discussed in the section on conditions of use, not only usability but the development of conditions of use is a prerequisite to testing the tool in context; functional activities are part of the intervention along with the tool. Next, as demonstrated in numerous examples in the narrative, designed features do have powerful impact on collaboration and learning, but this impact is often hard to predict (such as the effective feature which is never used). Iterative design in context is an exceptionally good way to uncover these unanticipated consequences. Third, since local culture is a moving target, constant redesign and course-corrections are required throughout any interventional phase in research; by documenting change over time, the research is bolstered, not confounded. Lastly, the intimacy that comes with designing and refining tools and collaborative contexts during research can lead to important insights that can guide and support the research endeavor (as with the interpretation of anonymity findings). These anecdotes help illustrate why combining design and research can be not only a reasonable reaction to the complexity of tool use in cultural context, but also a beneficial one where design and research are each strengthened by the presence of the other.

ACKNOWLEDGMENTS

Many thanks to all my collaborators on the work in this paper, especially Sherry Hsi, the KIE Research Group, and the Learning and Design Underground. I thank the funders of our work including the National Science Foundation, the Spencer Foundation, and the Evelyn Lois Corey fellowhip. Special thanks to Ondine Hoadley Plekss for editing the manuscript.

REFERENCES

AAUW Educational Foundation. (1992). *How schools shortchange girls*. Washington DC: AAUW Educational Foundation.

Bales, R. F. (1969). *Personality and interpersonal behavior*. New York: Holt, Rinehart, and Winston.

Barab, S. A., Hay, K. E., & Yamagata-Lynch, L. C. (2001). Constructing networks of action-relevant episodes: An in situ research methodology. *Journal of the Learning Sciences, 10*(1), 63-112.

Bell, P., Davis, E. A., & Linn, M. C. (1995). *The Knowledge Integration Environment: Theory and design.* Paper presented at the Computer Supported Collaborative Learning '95, Bloomington, Indiana.

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences, 2*(2), 141-178.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. New York: Springer-Verlag.

diSessa, A. (1991). Local sciences: viewing the design of human-computer systems as cognitive science. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 162-202). Cambridge, England: Cambridge University Press.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age. The Jean Piaget symposium series.* (pp. 49-70): Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.

Hoadley, C. M. (1999). *Scaffolding scientific discussion using socially relevant representations in networked multimedia.* Unpublished Ph.D. Dissertation, University of California, Berkeley, CA.

Hoadley, C. M., & Bell, P. (1996, Sept. 1996). Web for your head: the design of digital resources to enhance lifelong learning. *D-Lib Magazine*.

Hoadley, C. M., & Hsi, S. (1993). *A multimedia interface for knowledge building and collaborative learning.* Paper presented at the Adjunct proceedings of the International Computer Human Interaction Conference (InterCHI) '93, Amsterdam, The Netherlands.

Hoadley, C. M., Hsi, S., & Berman, B. P. (1995). *The Multimedia Forum Kiosk and SpeakEasy.* Paper presented at the ACM Multimedia '95, San Francisco, CA.

Hoadley, C. M., Hsi, S., & Linn, M. C. (1993). *Assessing curricular change with an electronic discourse tool.* Paper presented at the NSF Engineering Education Coalitions Evaluators Workshop, Baltimore, Maryland.

Hoadley, C. M., Hsi, S., & Linn, M. C. (1994). Innovative Assessment with the Multimedia Forum Kiosk: a Preliminary Report, *NSF Site Review, SYNTHESIS National Engineering Education Coalition*. Tuskegee AL: Tuskegee University.

Hoadley, C. M., & Linn, M. C. (2000). Teaching science through on-line, peer discussions: SpeakEasy in the Knowledge Integration Environment. *International Journal of Science Education, 22*(8), 839-858.

Hsi, S., & Hoadley, C. (1994). *The Multimedia Forum Kiosk Assessment Tool for Curricular Reform Evaluation* (Final project report): SYNTHESIS National Engineering Education Coalition.

Hsi, S., & Hoadley, C. M. (1994, April). *An interactive multimedia kiosk as a tool for collaborative discourse, reflection, and assessment.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Hsi, S., & Hoadley, C. M. (1995). Assessing curricular innovation in engineering: using the Multimedia Forum Kiosk, *Annual Meeting of the American Educational Research Association*. San Francisco, CA: American Educational Research Association.

Hsi, S., & Hoadley, C. M. (1997). Productive discussion in science: gender equity through electronic discourse. *Journal of Science Education and Technology, 10*(1).

Hsi, S., Hoadley, C. M., & Schwarz, C. (1992). *Scaffolding Constructive Communication in the Multimedia Forum Kiosk* (Unpublished course report for EMST 291B): University of California at Berkeley: Education in Math, Science, and Technology.

Hsi, S. H. (1997). *Facilitating knowledge integration in science through electronic discussion: the Multimedia Forum Kiosk.* Unpublished Ph.D. dissertation, University of California, Berkeley, CA.

Koschmann, T. D. (1996). *CSCL, theory and practice of an emerging paradigm*. Mahwah, N.J.: L. Erlbaum Associates.

Linn, M. C., & Burbules, N. C. (1993). Construction of knowledge and group learning. In K. G. Tobin (Ed.), *The practice of constructivism in science education* (pp. 91-119). Washington, DC: American Association for the Advancement of Science (AAAS) Press.

Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., Songer, N. B., & Eylon, B.-S. (1996). Shifts and convergences in science learning and instruction. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 438-490). New York: Macmillan library reference.

Neilsen, J. (1994). Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In R. G. Bias & D. J. Mayhew (Eds.), *Cost-justifying usability*. Boston: Academic Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N. J.,: Prentice-Hall.

Pea, R. (1992). Augmenting the discourse of learning with computer-based learning environments. In E. deCorte & M. Linn & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving*. New York: Springer-Verlag.

Roth, W.-M. (2001). Situating cognition. *Journal of the Learning Sciences, 10*(1), 27-61.

Scardamalia, M., & Bereiter, C. (1992). An Architecture for Collaborative Knowledge Building. In E. De Corte & M. C. Linn & H. Mandl & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (Vol. 84). Berlin: Springer-Verlag.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*(3), 265-283.

Scardamalia, M., Bereiter, C., McLean, R. S., Swallow, J., & Woodruff, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research, 6*(1), 55-68.