Interaction Analysis of a Biology Chat

Gerry Stahl

Abstract. This is an analysis of data from initial attempts to combine (a) technology from the Virtual Math Teams (VMT) Project, (b) helping agents, (c) collaborative small groups, and (d) accountable-talk prompting in order to scaffold biology student online chats about videotaped results of a biology experiment. Analysis of the response structure of the chat log of a student group reveals characteristics of their interactions in terms of building collaborative knowledge. In particular, the mediation by the VMT technology, helping agents and accountable-talk training is analyzed to determine their influences in promoting productive learning-oriented interaction. A design-based-research analytic perspective provides suggestions for redesign of the socio-technical approach based on the findings from the interaction analysis. Redesign in response to the analysis results in clear improvement, as seen in analysis of the response structure of a chat log from a second test cycle.

Analyzing Response Structure

This chapter takes a specific analytic approach, developed within the Virtual Math Teams (VMT) Project (Stahl, 2009). The VMT research team adapted video-based interaction analysis of face-to-face discourse (Jordan & Henderson, 1995) to analyze synchronous text chat by students in their mid-teens as they interact in the online VMT environment, discussing issues raised in school mathematics. We found that, from a structural viewpoint, the most important aspect of discourse is its temporal sequentiality; the field of Conversation Analysis has analyzed this extensively, beginning with (Sacks, 1962/1995) and summarized more recently by (Schegloff, 2007). We adapted such sequential analysis to student chat discourse in the VMT environment at the foundational level of "adjacency pairs" of mutually responsive postings (Stahl, 2006c)—which we take as the unit of interaction—and at the "longer sequence" level (Stahl, 2011)—which we feel is the key level of description for knowledge building in computer-supported collaborative learning (CSCL).

In this chapter, I apply the method of analyzing text-chat response structure that we developed in the VMT Project to chat among students discussing a biology experiment conducted in an early version of the environment formerly known as ConcertChat (now VMT). The text chat was integrated with class discussion, a worksheet and videos. In addition, the software was extended with a software agent, which interacted with the students as a chat participant. I ignore most of the larger context of the experiment (see chapters by Dyke, et al. and Howley, et al., this volume) and focus on what is visible in the chat log. I look at a representative case from each of the first two cycles of experimentation.

In undertaking this paper, I decided to do my own methodological experiment within the biology educational experiment. I wanted to see if sequential analysis could be used effectively as a quick-and-dirty method of evaluation within a design-based-research cycle. Design-based research is a wide-spread approach within educational research for designing technological and pedagogical interventions through iterative cycles of design, prototyping, user trial, analysis and re-design. In the biology experiment, an intervention had been designed for biology classrooms; software agents had been prototyped within a version of the VMT collaboration environment; the intervention was tried in middle-school classrooms; and it was now time to analyze the results. While some experimenters may have been hoping that analysis would show the benefits of agent support or accountable-talk training, my aim was to discover what most needed re-design in the next cycle.

Although design-based research is a much used and discussed approach to educational research, there is no established method for conducting the analysis phase of the iterative cycles. Researchers both friendly to and opposed to Conversation Analysis (CA) have argued that CA sequential analysis is inappropriate in design-based research. Adherents of CA argue that CA cannot be applied to design efforts because it is interested in seeing what emerges of interest from an unguided analysis of the participants' discourse—which is unlikely to be relevant to a designer's goal-oriented concerns. On the other hand, researchers from other approaches, such as quantitative coding of discourse, insist that qualitative CA takes too long and is too costly to fit into the workflow and focused research questions of re-design cycles. My experiment was to see if I could conduct a quick sequential analysis that would cheaply and effectively point the way for re-design. That was the practical goal of my methodological experiment.

Theoretically, I was interested in understanding what "really" occurred in the interaction between students and agent. I wanted to "bracket out" the assumptions of the people who set up the biology experiment as well as assumptions about what went on in the heads of the students or the programs of the agent, based on reports from outside the discourse data. As a researcher of group cognition (Stahl, 2006a), I am interested in the effect of the intervention on the group processes, the interaction visible in the chat log. I wanted to see how much I could learn about the group process by viewing the structure resulting from sequential analysis. I wondered what I could fathom of the group knowledge building from micro-analysis of the discourse details, i.e., from how the participants articulated their responses to each other. The goal of accountable-talk training and support is presumably to change certain aspects of the talk by the students, and this is what I wanted to observe directly—not indirectly from statistical verification of hypotheses based on testing responses of individual students outside of the group-interaction context.

Obviously, the behavior of the students will be affected by countless factors, many of which could be studied in theory with various methods and data-collection efforts: the personalities and backgrounds of the students, the programming of the agents, the funding of the schools, the history of American education, prior testing results and future test schedules, etc. But I wanted to see how far I could get in making grounded re-design recommendations by just looking with some care at a small sample of interaction data.

Furthermore, I was only concerned about the group unit of analysis, that is the interactions among group members, not the status of any one individual member. Fortunately, because the group interaction for a period of time during the experiment was mediated by the VMT system, all group interaction among the students and the agent passed through the chat tool and was captured in the chat log exactly as it appeared to the participants. This gave me a complete and reliable log of the group interaction without all the complications and interpretive issues of videotaping and transcribing. As described below, I modified the chat log representation and then constructed a representation of the sequential interaction (Figure 1). Simply looking at this representation allowed me to make some tentative conclusions about the nature of the interaction and to point these conclusions out to others. The conjectures based on this representation guided a careful look at the details of how the specific chat postings involved were designed by their posters, the groups of students.

The problematic aspects of interaction revealed in my quick response analysis of a student chat in the original intervention were taken into account in redesigning the intervention in a second cycle of design-based research a year later. I conducted a similar quick response analysis of a student chat in cycle 2 and was able to see a significant improvement in the behavior of the agent as well as in the discourse of the student group.

Method

1. Following the first classroom intervention, I was supplied with the logs of 16 chats, in spreadsheet format. The chats each lasted about a half an hour and contained the chat postings of three students and an agent. The 16 chats were divided among three conditions: in one condition the agent prompted students (indirectly) to ask each other to make specific accountable-talk moves; in a second condition the agent prompted students (directly) to make specific accountable-talk moves; in the final condition the agent did not make any accountable-talk prompts, but only guided the students through the steps of the assignment (as was also done in the first two conditions).

2. I read through each of the 16 chat logs that I was given and I wrote down a couple sentences of my initial reaction to the quality of the interaction. It struck me that similar patterns of interaction were arising in the 16 logs, and so I decided to analyze one chat in detail to get at key common patterns. I selected log C01 as representative and promising for illustrating the common patterns. This case was from the first condition, in which the agent gave indirect prompts. Clearly, other analyses with different research questions and approaches would want to contrast the different conditions (e.g., Howley, et al., this volume), but from my focus on response structure it seemed particularly useful to look closely at one typical example.

3. In order to make the interaction flow visible, I rearranged the spreadsheet to have the postings of each participant in its own column. The newer version of VMT produces logs in this format automatically for students, teachers and analysts. We often also have columns for time elapsed since the previous posting and time when a posting was starting to be typed. These figures sometimes help to determine which previous posting a new posting is responding to. In the current log, such detailed reasoning was not generally necessary.

4. I next sketched the response structure of the chat (see Figure 1). I drew an arrow from each posting to the prior posting to which it was responding interactively, for instance to what question is an answer responding? This already gave a visual impression of some aspects of the patterns of responses. These patterns are central to the interactional dynamic of the group.

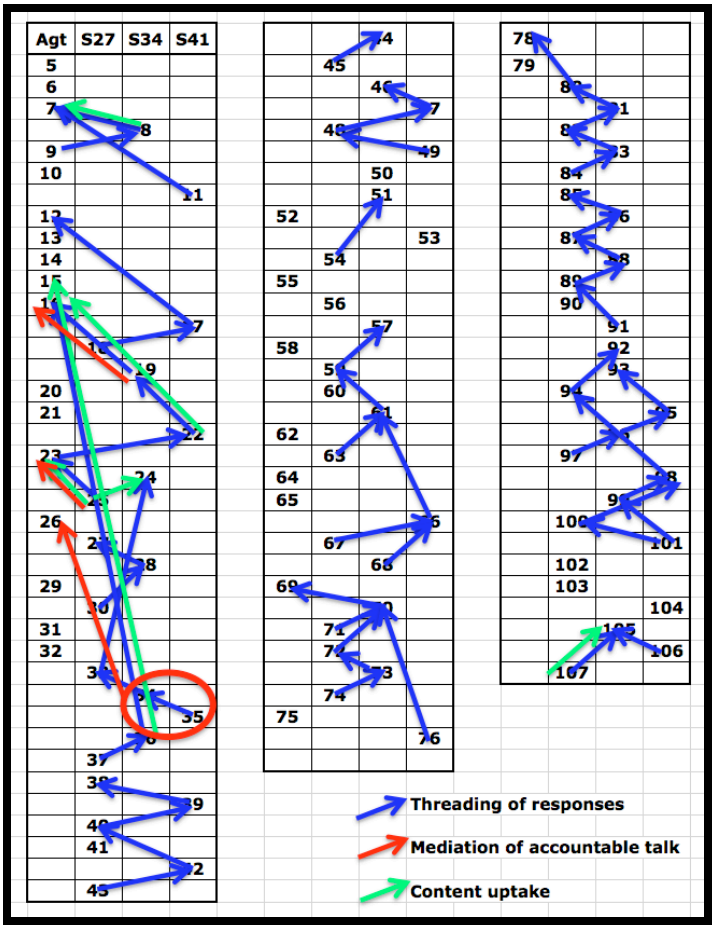

Figure 1. Sequential response structure of chat C01. Note that only interactions between actors are represented, not instances of a posting by one actor building on his, her or its own previous posting.

5. An important phase of interaction analysis is the exploration of the data, line-by-line, in a *data session* with other researchers (Jordan & Henderson, 1995). This inherently dialogical or multi-vocal approach can bring in multidisciplinary perspectives and balance one-sided views. A data session can be most effective once some initial analysis has already been undertaken by one of the researchers. After the data session, suggestions have to be synthesized and followed up with further detailed data analysis. There can be multiple cycles of group and individual analysis. The data session for this chapter's analysis included experienced online educators from the Math Forum and two analysts from other chapters (Rosé and Goggins). The session suggested a more complex representation of the response structure, it refined interpretive details, and it situated the case study in a deeper understanding of the experimental context. In particular, the data-session discussion proposed the representation of response structure of accountable talk (Resnick, O'Connor & Michaels, 2007) shown in Figure 2, which was used in refining Figure 1.

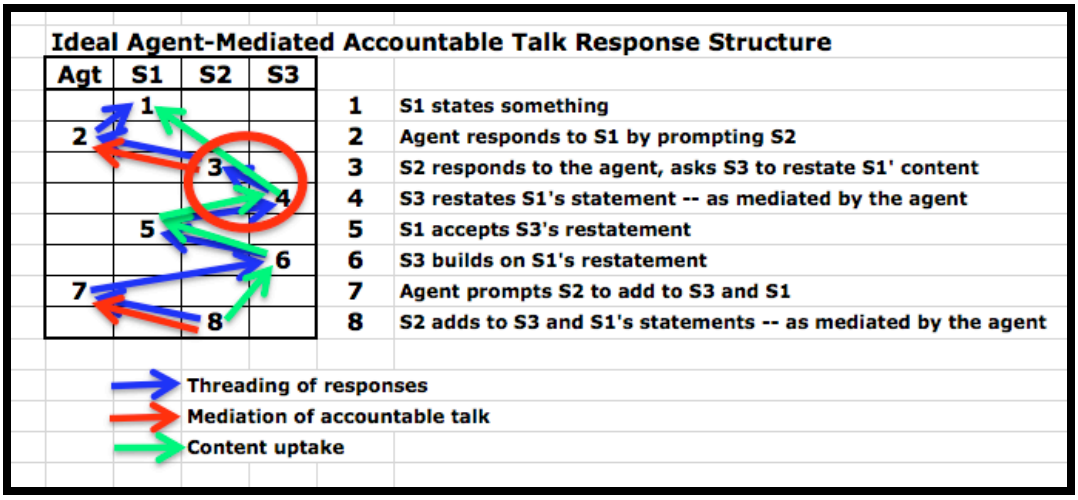

Figure 2. Sequential response structure of accountable talk.

6. Once I had a preliminary view of the response structure of the discourse in the chat, I could start to formulate tentative observations about the case study. These observations led to looking at the textual content of the postings. This showed the nature of the group interaction in more detail. The evolving analysis (see next section) also revealed the understandings and reactions of the students to their situation. This highlighted the response of the student group to its given task and to the actions of the agent, to the accountable-talk training and to the software environment.

7. As I summarized my observations (see discussion section), I felt that they generally applied to the other chats as well. By grouping the problems in relation to different design decisions in the experiment, I was able to propose several general suggestions for future re-design (see conclusions section). Other analysts, taking into account other data, additional knowledge of the constraints on the experiment, and alternative research questions will undoubtedly reach different—hopefully complementary—conclusions. I was interested in seeing what insights an interaction analysis of a single case study could provide for the long-term design-based-research effort. I wanted to do this analysis strictly on the basis of the chat data from a single case study, without being concerned about the many constraints, practicalities, and concerns that influenced the experimental design in all its complexity.

Analysis of the Chat-Response Structure

Figure 2 shows a representation of the response structure of an ideal accountable-talk interaction, as hypothesized by the experimenters. The blue arrows indicate that the agent responds to the students (line 2 and 7) and that the students in turn respond to the agent (lines 3 and 8). There is also a sequence in which the students respond to each other (lines 3, 4, 5, 6). This produces a tight group interaction including the agent and the students. The green arrows indicate that subsequent postings often involve uptake of content from previous postings (e.g., lines 4, 5, 6, 8 by the students). The role of the agent does not involve content, but mediates the student uptake of content by means of accountable-talk prompts (lines 2 and 7, pointed to by the red arrows). Let us see the extent to which the data of actual interaction among students and the agent includes similar patterns of response.

Figure 1 indicates three instances of mediation of accountable talk (red arrows): (i) the response at line 19 to line 16, (ii) the response at line 25 to line 23, and (iii) the response at lines 34 and 35 to line 26. Let us consider each of these in turn.

(i) The agent requests in line 16: "`Please discuss what you predict will happen in these two conditions`." Student S034 complies after a lengthy two-and-a-half minutes of silence by asking the group, "`what do you think'ds going to happen?`" At this point, the agent interjects some information about a third condition and asks the students to move on to discussing that. The timing of this seems questionable if the goal is to encourage extended knowledge-building interaction among the students. Student S041 then ignores the agent's latest contribution and responds ironically to student S034's request for a prediction: "`the world is going to end in 2012`."

(ii) The agent quickly picks up on S041's prediction by introducing the indirect prompting for accountable talk in line 23: "`S027, now would be a good time to ask S034 to build on what S041 is saying`." This all confuses S034, who states, "`im so confused!`" But S027 dutifully instructs S034 to explain S041's remark by building on it and explaining it to S027: "`034, would you like to build on to what 041 is saying? and me too!`" The first part of this follows the script prompted by the agent, but S027 adds his sympathetic addendum, aligning with S034 by agreeing that he is also confused about what is being asked of them.

(iii) The final mediation is similar to the first. In line 26, the agent requests: "`When you are in agreement, write down your predictions and explanations for Conditions A, B and C on your worksheet`." A minute later, after S027 complains again of not knowing what to do, S034 says, "`someone predict something`." Student S041 responds again to student 034: "`THE WORLD IS GOING TO END IN 2012!`"

As the green arrows indicate, almost all uptake of content is associated with these three mediated interactions. Line 8 merely introduces the student, repeating the word "`name`": S034 responds to the agent's "`I didn't get your names yet`" with "`my name is [S034]`." Line 107 responds to line 105's birthday greeting with "`is it ur birthday?`" These are not knowledge-building moves, but are social interactions, not directly relevant to accountable talk about curricular content.

There is some evidence that the agent is responding to student postings. The agent's line 7 succeeds in getting S034 to give his or her name and the agent then responds to that by assigning a role to S034. At line 23, the agent responds to a posting by S041 by asking S027 to ask S034 to build on what S041 said. This is an instance of the indirect mediation. While the timing is appropriate to ask S027 and S034 to discuss a posting by S041, the agent clearly fails to understand the significance of the posting. The agent assumes that S041 has made a prediction about the biology experiment, and not a sarcastic joke. This could have sent the group off on a distracting tangent, but in fact only confused the students about the agent's behavior and the meaning of the agent's requests.

If we look at the blue arrows in Figure 1, we see that the only times that the agent responded to the students were in lines 9 and 23. In line 9, the agent started to assign roles that were ignored by the students. In line 23, the agent requested an accountable-talk script to build on a joke.

A look at the high-level visual structure for Figure 1 indicates that the agent dominated the discussion in the early part, but then was ignored for most of the remainder of the chat. Toward the end, there was a significant pattern of interaction among the students, who seemed to be engaged as a group. A closer look at the content of the individual students' postings suggests that S034 is trying hard to accomplish the class task. S027 seems generally lost. S041 is not interested in the biology and is more oriented to clowning around. There is

no apparent correlation of their individual behaviors to the roles assigned to them by the agent.

The period from posting 5 through 18 lasted about four minutes. This period is totally dominated by the agent, which posted over 260 words while the three students responded with a total of 9 words, mostly just stating their names. The agent did not acknowledge their responses or appear to respond to them, except as noted above. Although delivering instructions to the students through the agent may have been motivated by an attempt to establish dialog between the agent and the students, it positioned the agent as an authoritative source of knowledge and commands, while positioning the group of students as a set of largely passive listeners, thus discouraging student discursive agency.

Of course, it made no sense for the agent to ask the students to "build on" to the sarcastic answer in line 22. This response by S041 shows that he/she already did not take the agent seriously. By not interacting with the students in a way that makes sense to them, the agent fails to establish itself as a serious participant in the group discourse. Caught in the middle between human interaction with the other students and obeying the authoritative orders of the agent, S027 follows the agent's command, but adds his protest against the agent's leadership in line 25.

S027 and the other students then stop orienting to the agent and the agent is ignored for the next 10 minutes until it again provides an unhelpful indirect prompt for accountable talk at line 69. Instead of responding to the agent prompt, S027 asks who is 34 and says "`ooh. hi`" when S034 responds. The students go on to work together to fill in the worksheet. One student provides the answers and the others try to figure out how to copy those answers into their own worksheets.

The agent continues to give commands, but they are generally ignored. When in line 69 the agent prompts once more for accountable talk, the students agree that the agent is being an insufferable nuisance. They evaluate the whole supported chat experience by agreeing that "`this would be so much easier just in a group`," meaning just sitting together without any computer or agent support and filling in their worksheets. Their only subsequent response to the agent is to celebrate when it leaves.

Discussion: Issues Observed

In the initial experiment, students were placed in small groups of three students and an agent in a chat room. This is a setting that calls for intense text-based interaction. The patterns in Figure 1 are already visually suggestive. The agent does not significantly respond to (i.e., interact with) students. The student responses to the agent are problematic. After trying to be responsive, the students give up and start to engage in their own discussion. The later periods of student interaction show considerable back-and-forth responses as they elicit responses, provide responses, and then acknowledge the responses to each other in various ways. Student responses are tightly situated in the on-going discourse, whereas the agent speaks like an academic textbook, with no sense of contextualization and little apparent attempt at interaction.

The educational experiment is an attempt to support collaborative learning with (a) the VMT software environment, (b) software helping agents (c) a social small-group setting, and (d) accountable-talk prompts. It is a CSCL intervention that aims to scaffold collaborative learning with these forms of computer support and communication structuring.

(a) The first problem is that the lesson design does not succeed in fostering collaboration. The students are each given their own worksheet to fill out and then they are each tested

individually. There is no meaningful group task or group goal to be accomplished collaboratively. The questions to be addressed by the students are not open-ended issues to encourage group inquiry and discussion, but questions with instructor-defined correct answers that the students can solve individually. Consequently, there is little evidence of real knowledge building taking place collaboratively. The most that occurs is that a student who knows the correct answer will give it to students who do not know it. Rather than this taking place as accountable talk, it naturally takes place in the form of students copying each other's answers to fill in their individual forms, without caring much about understanding the science—i.e., a common school process understood by all as cheating rather than collaborating or learning. The VMT environment was designed for shared tasks, with a shared whiteboard provided as a shared external memory that can be even more important for communication and joint work than the text chat (Çakır, Zemel & Stahl, 2009). Rather than this, the experiment uses the whiteboard to display once more a static cartoon of accountable talk, which appears to have been completely ignored by the students. The whiteboard could have contained the worksheet, to be filled out collaboratively by the team. That group artifact could then have been evaluated for the grading, rather than threatening the students with individual quizzes (causing expressions of test phobia). The shared whiteboard (or additional tabs with web browsers or other whiteboards) could also have been used to present data of the biology experiment, rather than having the students have to start up other applications (causing further confusion).

(b) The second problem involves the design of the agent interventions. First of all, the agent was in effect non-interactive. The agent may have been carefully programmed to intervene in an interactive way, but it does not come off that way in a sequential analysis of the chat—which is more important than the intentions of the programmer. To the students, the agent's timing did not appear to be effectively coordinated with the student discourse or responses. Inevitably, the agent postings introduced confusion for the students rather than clear structure. They were incredibly verbose—within the chat medium, which is known for its conciseness of expression. It might have made more sense to explain the process in class before breaking into online chat groups. Helping agents should probably not be used to automate teacher-centric instructors, but should get out of the way of student interaction until the students express a need for help. When an agent does intervene, it has to know what is going on well enough to judge what kind of response might be helpful. The agent behavior programmed here was an extreme example of "over scripting" and the opposite of the recommended "SWISH approach" (Dillenbourg, 2002; Dillenbourg & Jermann, 2006).

(c) A third problem involves social identity. Teenage students are mainly learning social skills, despite teacher efforts to have them learn curricular content. So when they are put together to interact in small groups it is essential to them that they know as much as possible about each other. In the VMT Project, we tried to put together students with no prior knowledge of each other so that we researchers could know everything the students knew about each other, so that we could interpret their interaction logs on a par with their understanding of the group interaction. In this biology case study, the students knew each other very well and had well practiced relationships. By assigning the chat participants anonymous identifiers, the experiment interfered with their exercise of these important and motivating social relationships (see chapter by Cress & Kimmerle, this volume). The students spent much time and attention in overcoming this circumstance (e.g., chat lines 17/18 and 27/28/30), positioning them in opposition to the conditions imposed upon their daily routines by this experimental intervention.

(d) Finally, accountable talk needs to take place at a sophisticated level of discourse. Like all effective discourse, it must be highly situated in the on-going discussion. That is the skill

of a teacher who has mastered accountable talk moves, to know just when and how to prompt. A complicated prompt cannot just appear out of the blue and hope to be helpful in building shared understanding. This poses a major technical challenge for software agents at many levels; it may require many cycles of design-based research to evolve an effective interaction behavior for helping agents that can effectively prompt for accountable talk by students.

Suggestions for Redesign

The biology experiment is cutting-edge research. The components that it brings together each require groundbreaking advances in the knowledge of their domain. It is not a matter of simply applying well-understood techniques.

(a) It took years of research by a large international, interdisciplinary team to develop the integration of pedagogy, problem, and technology for the Virtual Math Teams Project in the domain of collaborative online discourse of school mathematics—and there is still much investigation to be done there. Similar explorations will be needed for the domain of online discourse of school biology. A primary issue in guiding student inquiry in small online groups is how to avoid intruding in the important processes of small-group collaboration among the students; the case study just analyzed shows that there is a long way to go in achieving this with the approach tried. Our past research emphasizes how important yet difficult guidance or scaffolding of collaborative knowledge building is to achieve. In the VMT Project, we often had an adult facilitator in the chat room with the group of students. We trained the facilitators to avoid intervening too much in the interaction, mainly answering questions and helping with technology issues. A study of this showed the subtlety of supporting student group agency rather than interfering with it (Charles & Shumar, 2009).

(b) Involving software agents as participants in open-ended collaboration is quite different from the approaches that have been so successful in automated tutors of individual students being trained in well-defined algebra procedures within tightly constrained interfaces. In collaboration with Carolyn Rosè's research group, we started to explore the interaction of software agents with students in online discussions in the VMT environment with experiments in a mathematics classroom (Stahl et al., 2010). Here we discovered how invasive agents tend to be. Even with "wizard of Oz" experiments in which human researchers played the role of software agents, the presence of the "agents" radically transformed the online interaction. The students oriented their discussion to the agents instead of to each other and to the math problems. Much more experimentation seems necessary to design less invasive agent behaviors, even in theory. In addition, it may be necessary to study successful examples of accountable-talk prompts or interventions by skilled teachers, using the micro-analytic techniques of Conversation Analysis before trying to design software algorithms to replicate such expert behavior. In particular, we need to know how to effectively time interventions and how to adapt the linguistic structure of interventions to the on-going discourse.

(c) Designing effective CSCL interventions and introducing new technologies to scaffold interaction is a complex undertaking. It requires many cycles of iteration. The data analyzed here functions as an initial, pilot iteration. It was probably premature to run multiple conditions and to expect to see effects in subsequent testing of individual students. If anything, the VMT environment, the software agents, and the accountable-talk prompts seem to have each done more to interfere with any possibility of collaborative discussion of biology than to promote it.

(d) The theory of accountable talk has intuitive appeal to scientifically well-trained, mature, rational adults, whose thinking is heavily influenced by explicit textual expression.

However, theories relevant to CSCL stress the social, situated, and linguistic nature of cognition (Stahl, 2012). To introduce accountable-talk moves into the highly situated, socially interactive text-chat interaction of school children will involve much more than providing canned prompts of the form used in the case study. It will require understanding the situated, sequential, social, interactional character of student chat, developing agents that can follow these subtle processes through real-time analysis of cryptic, ironic, juvenile postings and can formulate agent postings that engage in the co-construction of shared understanding. It is even possible that actually accomplishing that would exceed the theoretical possibilities of artificial intelligence to engage in intersubjectivity with humans. But before we can reasonably speculate on that, it seems important to understand the nature of effective knowledge-building discourse and productive accountable-talk prompting; again, micro analysis of prototypical examples of such interaction need to be carried out.

The point now is to take the lessons learned back to the drawing board for extensive redesign: (a) First, integrate more aspects of the biology experiment into the collaboration-support software environment by allowing the group to see the diffusion experiment results in a shared view and to embed its inquiry reasoning and its group conclusions in the VMT shared whiteboard. This can make better use of the collaboration tools of the software as a collaborative medium. (b) Second, develop the agents to follow the student discourse and to just intervene when needed. This involves real-time natural language processing of the student postings, which is a complex, subtle, and situated skill, which may exceed the current state of the art. (c) Third, encourage collaboration among friends by letting the students know each other's identities and having them work for a group product, rather than filling in individual worksheets and taking individual tests. This would transform the exercise from one focused on individual learning to collaborative knowledge building. (d) Fourth, figure out how accountable-talk prompts can be contextualized as part of natural verbal interaction. This will involve development of this approach beyond the current conceptualization of the technique.

Methodologically, this stage of research calls for observations of pilot studies in order to guide design in the various aspects of the project. A single case study, looking in detail at the interactions, can provide insight into what *group-cognitive processes* (Stahl, 2006a) take place empirically—in ways that quantitative comparisons of different conditions generally cannot. This can provide important correctives to what designers assumed would take place based on their best preconceptions. Statistical controlled comparisons and quantitative measures of changes in individual test results at this initial stage would likely produce results that would at best be confusing, but more likely be misleading when interpreted on the basis of researcher preconceptions of what transpires in student interaction. This response analysis from cycle one has tried to provide a detailed case study that analyzes the actual interaction (among humans and agents) to reveal processes that are fundamental to human interaction under such conditions and are therefore likely to take place in other cases. It has tried to show how interaction analysis focused on the response structure of interaction can provide insight into group-cognitive processes and can indicate how experimental interventions do or do not support the group interaction. It contributed to guiding the redesign of this design-based research effort at this early stage of educational design.

Cycle two of design-based research

Due to the practicalities of conducting an experiment in public schools and due to the level of re-design called for by the lessons of the analysis of the first cycle of user testing, it took a year before the next cycle's user testing could be conducted. In this section, I take a similar approach to seeing what a quick sequential analysis can yield with the data from the second cycle.

1. As described in Dyke, et al. (this volume), the new intervention had students working in four conditions. I decided that the revoicing condition would be the most interesting. I wanted to see the effect of the agent prompting students to revoice their chat postings.

2. I read through each of the 5 chat logs in the revoicing condition and I wrote down a couple sentences of my initial reaction to the quality of the interaction. I selected log F01 as the one that seemed to have the richest student interactions. I wanted to see how the agent postings—particularly revoicing prompts—affected the accountable talk of the students.

3. I rearranged the spreadsheet to have the postings of each participant in its own column.

4. I next sketched the sequential response structure of the chat (see Figure 3).

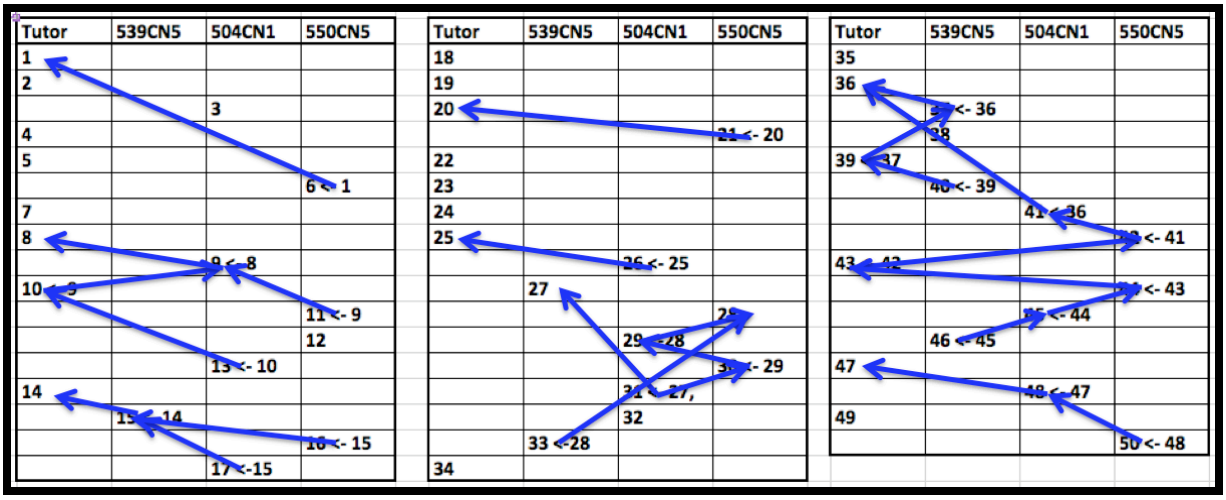

Figure 3. Sequential response structure of chat F01.

5. A visual scan of the response structure shows that the tutor (first column) is still very dominant in the discourse. Of 50 postings, now only 10 are by the tutor agent, but most of them are lengthy, whereas many of the student postings are only a word ("yes", "ok", or the student's name). Primarily, most of the student postings are in response—either directly or indirectly to the tutor. However, there are now several brief interactions among the students and even a couple of quite involved interactions (posts 27-33 and 41-50).

6. If we look at the content of the posts, we see that the whole discussion remains closely on-topic, following the agenda of the tutor. The tutor takes a strong instructionist teacher role. The students seem to accept this and respond to it much as they might to a classroom teacher. Although this was not the case in all of the chats, the one analyzed here seems quite successful in terms of student responses to the agent.

7. The student-to-student interaction (stimulated repeatedly by the tutor) progressed well. All the students participated (at least when prompted by the tutor), they discussed each other's proposals and they all agreed to a group answer after each of the extended interactions. This may have been encouraged by the formulation of the task, which was presented as a group task, to come up with an explanation that everyone agreed with.

8. The focus on accountable talk was reduced to the idea of revoicing—at least in terms of the tutor programming in this chat. The tutor only posted two explicit revoicing moves: postings 39 and 43. In both of these, the tutor proposed an alternative (and more scientifically formal) way of describing a biological process and the student simply said, "yes" to the

proposed revoicing. So the agent's move did not significantly expand the accountable discourse of the students. However, for whatever reason, the students in this group did seem to act in a generally accountable way by including and respecting each other and by describing biological phenomena.

9. Although some of the other groups expressed the kind of confusion about what was going on generally, about the role of the tutor and about the intelligibility of the tutor's postings that was rampant in the first year, the group in chat F01 did not. They accepted the tutor and responded to its postings as reasonable instructional statements. The timing of the tutor postings was also much improved. Student discussions were not often cut off by the tutor trying to follow a schedule. The tutor even seemed to react to student postings in ways the students could accept.

10. In conclusion, one cycle of re-design was adequate for eliminating the worst problems of agent intrusiveness, at least in the case of this one group, which I selected as most promising based on a skim of the logs. The ultimate goal of the theory of accountable talk is to have groups of students being accountable for their own discourse. It may be that at the level of ninth grade biology most students still need strong instructionist guidance and modeling before they can effectively adopt accountable talk practices in student-centered scientific discourse.

My quick analysis of a sample from the second cycle suggests that the major technical problems were adequately identified by my quick interaction analysis of the first cycle log and that they have been substantially addressed by the extensive re-design effort that it called for. The ground has now been laid for subsequent cycles exploring the complex issues of scaffolding group cognition among young students of science.

Issues for further multivocal analysis

a. Design-based research for designing technology

Too often, research reports are written to give the impression that a well-defined hypothesis was tested and that everything went according to plan, resulting in the reported findings. The widespread popularity of design-based research in educational technology design is a testament to the fact that research in real classrooms rarely simply follows a preconceived experimental plan. Rather, understanding about how to design effective educational technology emerges gradually from iterative attempts to refine prototypes in response to unanticipated issues that only become apparent in messy trials. The initial attempt to promote accountable talk in a biology classroom through the use of conversational agents ran into myriad circumstances that modified the ideal experimental plan. Dyke, et al. (this volume) listed some of these. Cress & Kimmerle (this volume) argued that the experimental situation, as actually implemented, did not support the social aspects of interaction that are so important to the students. The preceding sequential interaction analysis of one group's chat log from cycle one indicated that the agents were not very "conversational" in the resultant situation. Howley, et al. (this volume) further investigated the social, linguistic and sequential structure of the chat interactions, both to see how the agents and students positioned each other as knowledge-building partners and to track the temporal unfolding of the chats. These analyses begin to inform the design of the software agents and of the educational intervention generally, suggesting approaches to be tried in cycle two and in subsequent iterations. Other types of analysis can no doubt offer additional suggestions for redesigning features of this multi-dimensional intervention.

b. Scripting of the software agents and situated interaction

Just as the experiment as a whole is situated amid the complex constraints on conducting experiments in typical public school classrooms, so the postings of the agent and students are situated in the unpredictable and subtle constraints of the social and linguistic interaction that unfolds in the chat room. In particular, each posting must make sense as following previous postings. Furthermore, when someone has difficulties making sense of the sequence of postings in this context then there is a need for "repair" processes. The sensitivity of a posting to preceding chat posts motivated my decision to look at the adjacency-pair structure, as a key indicator of the extent to which posts—particularly those of the agent—were meaningfully related to preceding and subsequent posts by students. My analysis revealed that agent posts in cycle one were not adequately situated in this sense. Furthermore, the agent showed no ability (or even inclination) to repair problems of meaning making when they arose.

In a chapter I wrote for a book on scripting (Stahl, 2006b), I cautioned that scripts should be conceptualized as situated resources rather than implementable plans for action. For instance, rather than scripting the agent to instruct the students to watch the video at precisely 8 minutes 15 seconds after the start of the chat, the agent should try to find an appropriate moment roughly 8 or 9 minutes into the chat for doing this, depending on what the students are doing at that point. I cited Suchman's (1987, p. 181) recommendation that computer support compensate for its limitations by: (1) extending its access to the actions and circumstances of the user; (2) clarifying for the user the limits of the computer's access to the users' rich interactional resources; and (3) providing a wider array of alternative resources, particularly to help the users respond to unforeseen breakdowns. Suchman was talking about the design of help systems for large copying machines. Compared to that, the conversational agents have the significant advantage of having access to all actions in the chat room—they have the same access that the students have to each other's actions. However, the agents have been programmed to project an anthropomorphic personality, pretending that they have meaning-making and language-understanding capabilities far in excess of what they can actually do. Suchman warned explicitly against doing this because it inevitably confuses the relationships and leads to misunderstandings and frustrations. As Cress & Kimmerly emphasized, a classroom is a highly social setting for the students, and introducing a new social partner with no social skills may not be an effective approach. Finally, the agent is designed to perform multiple roles, scripting the macro-level phases of work as well as the micro-level accountable-talk moves. When the students reject the agent, they are left to their own resources.

c. Sequential interaction analysis of small groups

While the design-based-research approach is often recommended for educational technology, this approach does not generally specify a method for analyzing the results of trials. In the past, I have suggested adapting Conversation Analysis to provide insight into how teachers and students are actually making use of a prototype, rather than quickly counting surface features of interactions or coding utterances based on the designer's or researcher's conceptualization of the intervention. Although we have found data sessions based on VMT sessions to provide quite useful design feedback in a matter of hours, many researchers claim that qualitative analysis is too time consuming to give timely feedback. That is why I tried in this paper to see how much insight into central problems of an intervention could be gleaned from a quick adjacency-pair analysis of one typical chat session.

For the data from cycle one, I skimmed through the chats and got a sense of the problematic nature of the sessions, much like the feelings that the authors of the related chapters expressed. I selected a chat session that seemed to have relatively clear examples of

the problems. Specifically, I selected a session in the "indirect" condition, which was the condition of greatest interest for the experiment. I then sketched an initial version of Figure 1. Based on the visual appearance of the figure and the content of the connected adjacency pairs of posts, I drafted an initial version of this chapter, arguing for the need for changes to the agents and to the intervention in subsequent iterations. During a data session with some of the other chapter authors, refining Figure 1 and our understanding of what took place interactionally in the chat, we agreed on directions for further analysis and experimentation. In this way, the sequential interaction analysis with the graph of adjacency pairs provided a quick sense of where major issues lay, which needed to be addressed in re-design. Thus, it played a role similar to so-called "discount methods" in human-computer interaction, where designers need fast feedback at low cost.

d. Accountable talk and off-task student practices

Throughout the history of CSCL, researchers have conducted educational interventions with expectations that the students would engage in knowledge building, inquiry, transactivity, collaborative learning, warranted argumentation and other lofty conceptions of scientific intellectual discourse. These expectations were operationalized so that research assistants could reliably interpret student utterances as falling into different coding categories. Inevitably, few utterances could be coded in the highest categories; a large percentage fell outside the scheme, and they were called "off topic."

To conclude this paper, I would like to raise the ethnomethodological question: what are the students doing when they are off topic? If they do not *do* being-a-student by engaging in recognizably accountable talk, how do they do it? Is it due to some personal characteristics of these students that they engage in "cheating" rather than in following the instructions of the agent? Perhaps if we break free of the conceptualizations imposed by the experiment's world-view, we can understand the off-topic behaviors in a positive light. As Cress & Kimmerle (this volume) suggest, the teenage students are engaged in social activity with one another. Their social relations support their discussions of curricular topics and their talk in the classroom feeds into their social relations. Any arrangements that interfere with their social relations—such as hiding everyone's identities—will interfere with the possibility of any kind of interaction and will generate attempts to repair the problem. In addition to the social practices involved in relating with their peers, the students are involved in established classroom practices, oriented largely around earning good grades. While the researchers were looking for accountable talk in the details of interaction, the students were oriented toward completing the individual worksheets and taking tests. Thus, many of the early interactions graphed in Figure 1 involved finding out the identities of the chat room occupants and many of the later interactions involved filling out the worksheets. While the exchange of information needed for the worksheets did not fit the profile of accountable talk, it did contribute to a rather efficient sharing of worksheet answers. Although it did not meet the ideals of collaborative knowledge building or group cognition, the three students did effectively divide up the tasks and share the results, in the classic form of cooperation that most people tend to use. If the task had been to fill out a group worksheet, neither the students nor the researchers would have viewed this procedure as cheating. Classrooms are not blank slates waiting for a researcher's conceptualizations; they are richly structured with established practices that can only be discovered by analysis of interventions in the wild.

References

Çakır, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal

CSCL medium. *International Journal of Computer-Supported Collaborative Learning. 4*(2), 115-149. Web: http://GerryStahl.net/pub/ijCSCL_4_2_1.pdf

Charles, E. S., & Shumar, W. (2009). Student and team agency in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 11, pp. 207-224). New York, NY: Springer.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen, Netherlands: Open University of the Netherlands.

Dillenbourg, P., & Jermann, P. (2006). Designing integrative scripts. In F. Fischer, H. Mandl, J. Haake & I. Kollar (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives.* (p. 275-301). Dodrecht, Netherlands: Kluwer-Springer Verlag.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences. 4*(1), 39-103. Web: http://lrs.ed.uiuc.edu/students/c-merkel/document4.HTM

Resnick, L., O'Connor, C., & Michaels, S. (2007). *Classroom discourse, mathematical rigor, and student reasoning: An accountable talk literature review*

Sacks, H. (1962/1995). *Lectures on conversation*. Oxford, UK: Blackwell.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.

Stahl, G. (2006a). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stahl, G. (2006b). Scripting group cognition: The problem of guiding situated collaboration. In F. Fischer, H. Mandl, J. Haake & I. Kollar (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives.* (pp. 327-335). Dodrecht, Netherlands: Kluwer-Springer Verlag. Web: http://GerryStahl.net/pub/scripting.pdf

Stahl, G. (2006c). Sustaining group cognition in a math chat environment. *Research and Practice in Technology Enhanced Learning (RPTEL). 1*(2), 85-113. Web: http://GerryStahl.net/pub/rptel.pdf

Stahl, G. (2009). *Studying virtual math teams*. New York, NY: Springer.

Stahl, G. (2011). *The structure of collaborative problem solving in a virtual math team.* Paper presented at the iConference 2011. Seattle, WA. Web: http://GerryStahl.net/pub/iconf2011.pdf

Stahl, G. (2012). Theories of cognition in collaborative learning. In C. Hmelo-Silver, A. O'Donnell, C. Chan & C. Chinn (Eds.), *International handbook of collaborative learning.* New York, NY: Taylor & Francis. Web: http://GerryStahl.net/pub/clhandbook.pdf

Stahl, G., Rosé, C. P., O'Hara, K., & Powell, A. B. (2010). *Supporting group math cognition in virtual math teams with software conversational agents.* Paper presented at the First North American GeoGebra Conference. Ithaca, NY. Web: http://GerryStahl.net/pub/geogebrana2010a.pdf

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.

Appendix A: Chat Log of Group C01 from Cycle One

	time	Tutor	S027	S034	S041
1	11.54.45				
2	11.55.35				
3	11.55.41				
4	11.55.51				
5	11.56.01	Hi all! Today, each of you will be both the instructor and the student. Each of you will be responsible for looking for opportunities for a specific discussion strategy to be used in the conversation.			
6	11.56.06	Lets get started by introducing ourselves. I am Alex.			
7	11.57.04	S034,S041,S027... i didn't get your names yet.			

8	11.57.16			my name is {S034}	
9	11.57.16	Hi S034 you are the Revoicer. When an explanation or idea is given which would benefit from revoicing, please encourage one of your team mates to do so, for example, by asking them to explain what was said in their own words.			
10	11.57.24	S041, you are the Challenger. When you see a statement being made (whether it is right or wrong) which would benefit from being challenged, please encourage one of your team mates to do so, for example, by asking them whether they agree or disagree with what was said, and why.			
11	11.57.25				{S041}
12	11.57.31	Ok... So the Explainer is S027. When you see a place where it would be good to get more explanation, you can encourage one of your team mates to do that, for example, by inviting that person to add on to what the other student has already said.			
13	11.57.37	You should probably all mark an X opposite your role in the book, so you won't forget.			
14	11.57.43	Its nice to meet you all. :) Please make sure you have understood what you are responsible for.			
15	11.58.00	In condition A, we placed a glucose solution inside the cell model and immerse it in distilled water. In condition B, we did the opposite.			
16	11.58.05	Please discuss what you predict will happen in these two conditions.			
17	11.59.29				who is 27
18	12.00.06		{S027}		
19	12.00.46			what do you think'ds going to happen?	
20	12.01.02	In condition C, we made a modification to condition A. We replace the glucose solution with a starch suspension. In order to detect the presence of starch, we replace the distilled water with Lugol's Iodine solution.			
21	12.01.07	You should now move on to discussing what will happen in Condition C and your explanation for this change.			
22	12.01.14				the world is going to end in 2012
23	12.01.21	S027, now would be a good time to ask S034 to build on what S041 is saying.			
24	12.02.00			im so confused!	
25	12.02.08		034, would you like to build on to what 041 is saying? and me too!		
26	12.02.32	When you are in agreement, write down your predictions and explanations for Conditions A, B and C on your worksheet.			
27	12.02.42		who is 34?		
28	12.02.53			{S034}	
29	12.03.00	You are now going to watch a video showing the cell in Conditions A, B and C.			
30	12.03.04		ooh. hi		
31	12.03.05	As you watch the video, write down your observations on your worksheet.			
32	12.03.11	Go to the Videos folder on the Desktop, and watch the video which is there.			
33	12.03.24		idk what to do.		

34	12.03.27			someone predict something	
35	12.03.56				THE WORLD IS GOING TO END IN 2012 !
36	12.04.05			oh i get it! what do you think is going to happen when you put the glucose in the distilled water?	
37	12.04.22		i think it will turn a different color.		
38	12.04.29		{S041}?		
39	12.04.45				ok dude what 27 said
40	12.05.04		in your own words		
41	12.05.10		?!?!?!		
42	12.05.26				you have to be more specific 27
43	12.05.43		ur one to talk!!!		
44	12.05.46			the answer to number one : the distilled water will move into the internal envorment of glucose solution	
45	12.06.36		kk		
46	12.06.45			and the explaination is glucose is denser	
47	12.06.57				huh?
48	12.07.05		just write what she[{S034}] said.		
49	12.07.20				for the explination
50	12.07.33			#2 the distilled water will move into the external enviroment	
51	12.08.14			watch the video	
52	12.08.38	Is everyone back?			
53	12.08.45				we have to do all the prediction , thats what NACHO said
54	12.08.48		ok, but where do i click to watch it?		
55	12.09.00	Now discuss what you observed in the video about conditions A and B. Compare what you observed to your predictions.			
56	12.09.40		nevermind i found it.		
57	12.10.01			#3 the water will move into the starch because starch absorbs it	
58	12.10.08	S027, now would be a good time to ask S041 to build on what S034 is saying.			
59	12.10.33		{S034} are these #'s all of the predictions or explanations?		
60	12.11.51		{S034}?		
61	12.12.01			yeaaaa	

62	12.12.02	You should now move on to discussing what you observed in the video about condition C. Compare what you observed to your predictions. How is this different from what happened in condition A?			
63	12.12.25		#3, is that the prediction?		
64	12.12.41	Lets spend the rest of the time we have discussing your observations from all the three conditions.			
65	12.12.47	Discuss explanations of what you observed in each condition and make sure you fill in your explanations on your worksheets.			
66	12.13.00				what is explanation for #2
67	12.13.07		idk.		
68	12.13.07			its both, the first part is the prediction the second part is the explanation	
69	12.13.13	@S027, can you ask S041 to build on what S034 is saying.			
70	12.13.34			i want to shoot myself in the foot -.-	
71	12.13.50		ikr.		
72	12.14.00		this is so stupid.		
73	12.14.11			this would be so much easier just in a group	
74	12.14.19		yep yep		
75	12.14.23	@S027, can you ask S041 to build on what S034 is saying.			
76	12.14.25				you dont have the balls to do it
77	12.14.36				
78	12.14.38	Ok, I gotta go.			
79	12.14.43	It was nice talking to you all. :-)			
80	12.14.55		our tutor left us!!!		
81	12.15.10			D:	
82	12.15.23		she is a baad tutor!!!		
83	12.15.44			im pretty sure its a he	
84	12.16.09		really? haah! oh well. idc. he/she/it		
85	12.16.31		i ddnt even watch the whole video		
86	12.16.49			i watched most of it	
87	12.16.58		do you know what happened?		
88	12.17.13			i just wrote the observations	
89	12.17.42		oh. what was it after 1 hour for condition A?		
90	12.19.17		{S041}, did you get everything so far?		
91	12.19.24			condition a: 620 glucose inside; 635 glucose inside; 630 glucose inside; explanation: the membrane will hold stuff for a certain amount of time	

92	12.20.05			condition b: 540 glucose outside; 525 glucose outside; 525 glucose outside; explanation: the glucose was not being absorbed by the water	
93	12.20.10			i think .	
94	12.21.45		k, i'll fast forward and try to find the last one, when im done writing the condition b.		
95	12.21.47				you mean i know !
96	12.22.02			yea, sure i do, whatever. hahahahahha	
97	12.22.15		lol		
98	12.22.19				lol what is C
99	12.22.27			i dont even think there was one	
100	12.22.33		i sd i'd try to find it if i could.		
101	12.22.51				there is & ok
102	12.23.04		aaah there is a quiz!!!!!!!!!!!!		
103	12.23.19		im gonna fail!!!!!!!!!!!!!		
104	12.23.57				stfu !
105	12.23.59			HAPPY BIRTHDAAAY {S041} :D :D :D :D :D :D :D :D	
106	12.24.10				lol thsnks
107	12.24.13		is it ur birthday?		
108	12.24.20			[blank line]	
109	12.24.23			[blank line]	
110	12.24.23			[blank line]	
111	12.24.23			[blank line]	
112	12.24.23			[blank line]	
113	12.24.23			[blank line]	
114	12.24.24		...		
115	12.24.25				
116	12.24.35		bye bye guys.		
117	12.25.01				
118	12.25.11				
119	12.28.34				

Appendix B: Chat Log of Group F01 from Cycle Two

No.	Time	Tutor	539CN5	504CN1	550CN5
1	13.57	Hi, my name is Sage. I'll be working with you in this activity. Take a minute to introduce yourselves.			
2	14.42	I won't tell you if you're right or wrong, but I'll give feedback as you discuss things as a team.			
3	14.49			hi dajha!	

4	14.58	Let's look at a setup which is the opposite of the one we saw earlier - the iodine is now inside and the starch is outside.			
5	15.24	<< On the whiteboard, you can see a picture of this new setup.			
6	15.29				Vinny<3
7	15.38	<< Here's what happened in this setup over time.			
8	15.59	Notice how the inside and outside environments changed. Why did these changes happen?			
9	16.26			the iodine left the bag	
10	16.31	Okay, 504CN1 - why would how we use indicators be relevant?			
11	17.39				the iodine was never in the bag the iodine went into the
12	17.48				bag
13	17.51			the weight is decreasing and and the starch water changed colors	
14	17.54	Any input from 539CN5?			
15	18.59		if you loook at the diogeam it shows the big getting liter because the iodine left the bag		
16	19.21				agreed...
17	19.26			agreed...	
18	19.32	Okay, team - in a moment we'll move to another set of experiments.			
19	19.56	For these experiments, we'll use the same membrane, but with pure water on the outside and a glucose solution on the inside.			
20	20.27	Discuss the differences between glucose and starch molecules.			
21	20.46				glucose molecules weigh less
22	21.19	We'll use the tube weight as an indicator for water movement.			
23	21.23	We'll also use *glucose test strips* to detect glucose in both environments.			
24	21.37	<< On the whiteboard, you can see what happened in this setup over time.			
25	22.01	The results with glucose are different than what you saw with starch - as a team, discuss and agree upon an explanation for these results.			
26	22.50			the weight of the gluclose soltuion increased over the time period..	
27	22.59		the water has no suger in it so the strip will stay yellow. the gucose solution has duger so the strip will		

			turn green		
28	23.03				i think the glucpse molecules left the bag and went into the water in the process of that water molecules entered the bag
29	23.41			once again.....agreed with dajha ... hahaBACKSLASH	
30	24.00				always a smart thing to do :)
31	24.41			but also i agree with vicky too ;)	
32	24.50			:)	
33	25.32		and over time glucose cane out of the bag and got into the water and that is why after 1 hour and 24 hours the strip that was in the water started to change.		
34	25.38	Okay, everyone - let's move on to the last experiment.			
35	26.00	In this setup we have glucose solution on the outside, pure water on the inside.			
36	26.23	What changes do you think you'll see in this condition? Agree on a prediction for what you'll observe over time, *and* an explanation that everyone understands.			
37	26.50		the water has not suger so the strip is yellow		
38	27.36		and the glucose has suger and the strip is green		
39	27.41	So is 539CN5 saying "a glucose test strip changes color when glucose is there?"			
40	27.51		yes		
41	28.06			i think the gluclose test strip will slowly change to a bright yellow color it is on the weighing scale. also i think the bag of distilled water will weigh more because over time of sitting in the gloclose you will get some glucose molecules in the	
42	28.06				i don't think the sugar molecules will get in side the membrane i think they will be to big i think the water molecules will come out of the membrane and go into a concentrated equal enviorment

43	28.13	Would another way to say that be "molecules will move to the area of lower concentration if they can?"			
44	28.38				yes
45	29.41			agreed.....with dajha AGAIN!.	
46	29.51		i agree to		
47	30.02	Make sure each of you is clear on your team's prediction and explanation, and write it in your worksheet.			
48	30.09			WE ALL AGREE!:)	
49	30.45	All right, we're wrapping up. Thanks, team!			
50	30.53				just to restate... the sugar molecules will not go into the membrane the distellied water will come out and the test strip will turn yellow