Given that our instructional manipulation did not improve achievement or increase gestural activity during problem solving, we further examined the possible effect of our instructional conditions on the quality of diagrams produced by participants. Recall that the gestures produced by the interviewers during instruction always depicted a right-hand direction and right-side-up orientation toward the molecular diagrams. Specifically, we were interested in knowing whether participants were more likely to adopt an instructor’s perspective when learning from gestures. First, we examined the directionality of drawn structures as above. Interestingly, we observed a trend in the dataset that did not meet statistical significance at alpha .05 a greater number of “right-side approach” diagrams were observed when instructions were accompanied by gesture, \( t(1, 43) = 2.34, \ p = .10 \) (see Figure 4). Although gestured instructions produced no observable increase in gesture use or achievement, it did appear to affect the students’ perspective when drawing molecular diagrams even in a study with low N, as here.

![Figure 4](image_url)

Figure 4. Participants’ mean number of right-side depictions by instructions and diagram availability.

Previous investigations have demonstrated that gestures maintain spatial imagery during communicative activities (Wesp, Hesse, Keutmann, & Wheaton, 2001). Our investigation suggests that this previous finding may extend to contexts in which individuals are reasoning about spatial-relationships in isolation. Moreover, our results indicate an important interaction between gesture use and external representations. That is, individuals appear to use gestures to represent spatial information when external representations of that information are unavailable. Here we observed that when such diagrams are removed from sight, students produced significantly more gestures than when the diagrams were made available. However, many students used gestures while problem solving in an idiosyncratic fashion across experiments. Some participants gestured infrequently, some gestured frequently, and still others gestured randomly during the experiment. Of note, the highest achieving students were observed to produce the largest number of gestures on nearly every task.

Despite the increased gestural activity in the diagrams removed condition, we did not observe a large effect of gesture use on achievement across conditions. Given that we observed a moderate correlation between gesture use and achievement, it nevertheless remains a possibility that increased gesture use may improve achievement. In this investigation, we observed an appreciable amount of variance in gesture production and it may be that participants did not produce a sufficient amount of gestures to support problem solving. Participants may have not produced gestures for several reasons, one being that they have pre-existing strategies for solving translation problems, as has been documented elsewhere (cf., Stieff & Raje, 2010), which they favored over the strategy presented during instruction. As it stands, these results suggest that our instructional intervention may not have been robust enough to encourage students to use gestures during problem solving, which warrant future investigations that include more extensive interventions.
References


Acknowledgement

The preparation of this paper was supported by the National Science Foundation, NSF Award #0723313 (UMD). We would like to gratefully acknowledge Lianne Schroeder for her assistance with data collection and data analysis.
Audience Effects: 
A Bidirectional Artifact Analysis of Adolescents’ Creative Writing

Alecia Marie Magnifico, University of Illinois at Urbana–Champaign, amagnif@illinois.edu

Abstract: Much of the literature that developed rhetorical and cognitive viewpoints on audience suggests that when writers write with an audience in mind, imagined readers affect the writing. This paper investigates these audience effects by examining a creative writing workshop and asking how conversations with readers lead to revisions. Drawn from a case study of an 11th grade classroom, this bidirectional artifact analysis traces writing through students’ drafts, revisions, and feedback. The resulting work shows how revision and writing development ties to interactions with readers—members of the students’ audience.

Introduction and Research Questions
Much of the literature that developed rhetorical (e.g. Ong, 1975; Berkenkotter, 1981) and cognitive (e.g. Bereiter & Scardamalia, 1987; Flower & Hayes, 1981) viewpoints on audience includes an implicit assumption: Real and imagined readers affect writers’ writing. Such studies have led to consequences in school writing curricula, including pre-writing (e.g. Pressley, 2005), writers’ workshop (e.g. Atwell, 1998), and authentic pedagogies that connect students with real-world audiences (e.g. Duke, 2010). Similar to Learning Sciences work that aims to link classroom learning with authentic practices and contexts (e.g. Brown, 1992; Shaffer, 2006), these pedagogies posit that writing for readers reinforces literacy as communicative and helps writers plan, draft, and revise.

This paper, drawn from a case study of an 11th grade English classroom, analyzes students’ experiences with one of these pedagogies: writers’ workshop. I employ bidirectional artifact analysis (Halverson & Magnifico, under revision) to trace students’ writings through individual drafts, revisions, and feedback from peers and teacher. This work shows how writing development (or lack thereof) was tied to interactions with readers—members of the students’ audience. I take up two research questions: (1) In the context of a creative writing community, how do readers impact participants’ development of their written pieces? (2) What changes and revisions do participants make when they write for and converse with their readers?

Theory
Ideas about audience have shifted as new literacies have pushed writers into a world where they can write and read, as Lunsford & Ede put it, “among the audience” (2009) (cf. Duke, 2010). Such studies suggest a role for the audience that is both cognitive—in that writing for an audience forces writers to consider readers—and social—in that sharing work opens up communication (Brandt, 1992; Magnifico, 2010). While cognitive-process traditions show that writers consider questions about readers as they work (e.g. Flower & Hayes, 1981), audience members play a literal distributed and dialogic role in collaborative writing (e.g. Bakhtin, 1976/1994). Conversation, social knowledge construction, and perspective negotiation may result from these interactions, (Nystrand, Gamoran, Kachur, & Prendergast, 1997), helping students see writing as a communicative practice.

Methods
This bidirectional artifact analysis (Halverson & Magnifico, under revision) marries elements of narrative analysis (e.g. Labov & Waletzky, 1967/1997), discourse analysis of feedback and conversation (e.g. Wood & Kroger, 2000), and artifact analysis of students’ written texts. This combination is useful because it attends to narrative’s key paradox: “Good” stories are acceptable to readers, so they must be credible. At the same time, they must be reportable; unusual enough to reach beyond known tales. When they responded to each other’s pieces, the young writers in this study noted literary elements, and, as readers, explained the effects of these choices. While the reportability/credibility paradox was not invoked directly, student critics discussed whether their writings had achieved this balance. Was it a good story or poem? How did the piece achieve “goodness”?

Further, workshop writing relies on writers’ and readers’ critique: a comparison of drafts to discuss how revisions contribute to (or detract from) creative works. This analysis follows the logic of the critique process, examining the interplay of students’ drafts and conversations in a bidirectional way. Instead of examining successive representations, bidirectional analysis reaches backwards and forwards in time, working to understand how new representations are built as revisions of prior iterations (Enyedy, 2005).

Participants, Data, and Setting
This case follows nine students (six female, three male) and their teacher, Mr. Caswell (all names are pseudonyms), through a creative writing workshop in their International Baccalaureate English classroom. Their school, a K-12 college preparatory school, is located near a medium-sized Midwestern city. Observations totaled

© ISLS
15 hours over ten weeks, and interactions were recorded and transcribed. Students worked through three “cycles” of creative writing, drafting alone and then critiquing their work in small groups. In Cycle #1 and #2, students wrote a prose piece and a small poetry collection, revising one of these works in Cycle #3. For all cycles, students also submitted analytical writing detailing how they used literary techniques in their writing. My own role in this space shuttled between observation and participant observation. I observed and recorded field notes when the students were writing quietly, but consulted with students who needed advice as well.

Data Analysis

The data for this analysis are drawn from the students’ drafts and small-group workshop critiques. A timeline provides the central metaphor for this analysis, and thus, I created timeline representations for each piece of writing, chronologically arranging all of the collected artifacts: students’ drafts, written feedback, and transcripts of their workshop conversations. From these timelines, I worked backwards and forwards to code how the writings changed over time (Halverson & Magnifico, under revision; Enyedy, 2005): I marked revisions, traced each change (where possible) to a feedback suggestion, and examined uptake in future drafts.

The drafts and critiques pictured in Figure 1 represents Noah and Kira’s rich, complex critique over time. The goal of this messy representation was to trace how writers revised in response to feedback.

The graphical representation (Figure 2) shows the logic of this method. Solid arrows represent critiques that refer back to the text of prior drafts (e.g. “Your word choice in the first line is confusing”), while dashed arrows represent suggestions that point ahead to potential revisions (e.g. “In the next draft, try using more description”). In an “effective” critique, both of these elements are present in the feedback.

Results

The results I present focus on the case of one student, Noah, who was chosen because his work with different partners provides representative examples of both predominant critique patterns in the classroom. Noah wrote five poems over the course of Cycles #2 and #3 and discussed two of these poems extensively in class. He focused on Pitter Patter, a poem that describes a rainstorm, through two conversations with Kira, and After the Battle, a poem that recasts gym class dodgeball as an epic battle, through two conversations with Nasha.

Noah and Nasha: Abstract Critique

Most of the students had little experience with creative writing workshops, and often commented on each other’s work in a general, abstract way. Noah paired up with Nasha during two class periods to discuss After the Battle. These meetings featured vague criticisms, arguments about words and style, and few direct references to the text. When Nasha challenged his word choice, for example, Noah often argued to justify his ideas:
Nasha: I noticed that you... um. You use big words that don’t really fit in the context.
Noah: Goddd. I like this one. There’s a continuity to this one.
Nasha: Your rhyming stinks.
Noah: I don’t try to.
Nasha: If you’re going to make this a lyrical poem, and I’m assuming that’s what it is, is that what is this?
Noah: It doesn’t have to be, it doesn’t have to rhyme...
Nasha: Exactly, so don’t make it rhyme. Disregard rhyming entirely, because it’s so awful.

(Classroom transcript 4/24/2009.)

Nasha’s trouble—Noah’s use of “big words”—surfaces in many critique conversations about Noah’s poetry and invokes poetic genre; for instance, this criticism cites the potential of “mak[ing] this a lyrical poem,” which Nasha “assumes” is After the Battle’s genre. The critique challenges Noah to conform more closely to traditional poetic language, but Nasha remains abstract, providing neither context nor textual references. She never explains which “big words... don’t really fit in the context,” or which rhymes are “awful.” Noah argues, again at the level of genre, that the words have “continuity” and that lyric poems “[don’t] have to rhyme.” Lacking clear referents in the text, the critique becomes defensive. These characteristics represented much of Noah and Nasha’s conversation over two class meetings, as demonstrated by the descriptive statistics in Table 1.

Table 1: Descriptive Statistics: Noah and Nasha’s Critique Conversations

<table>
<thead>
<tr>
<th>Category</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time available</td>
<td>1.00.33 (two 30-min conversations)</td>
</tr>
<tr>
<td>Written suggestions</td>
<td>2</td>
</tr>
<tr>
<td>Oral suggestions with clear textual antecedents</td>
<td>6</td>
</tr>
<tr>
<td>Oral suggestions that refer generally to the poem</td>
<td>4</td>
</tr>
<tr>
<td>Total revisions with clear antecedents in suggestions</td>
<td>3</td>
</tr>
</tbody>
</table>

Over the course of two meetings (1:00:33 total time), Nasha makes two written suggestions—word substitutions. She provides six oral suggestions with clear referents in After the Battle, and four criticisms that generally refer to the poem. As a result, Noah makes three revisions, two of which are Nasha’s suggested word substitutions. Given the high-level comments and arguments about genre that characterized much of their conversation, it is possible (although not clear from the data) that Nasha led Noah to consider poetic language more carefully. What is clear is that these abstract discussions led to little textual change over time: Despite Nasha’s concerns, Noah revises little.

Noah and Kira: Concrete Critique

Some small-group partnerships offered specific context and suggestions for revision. For the duration of Cycle #2, Noah and Kira worked together twice, closely reading each other’s work. While they did challenge each other’s interpretations, they often resolved these arguments with textual analysis. Preceding this segment, Kira wrote comments on Noah’s poem, showing she did not understand the setting; driving during a downpour:

Kira: I wrote a lot of stuff, did you have any questions?
Noah: Yeah. Um... Yeah... I guess. I’m in a car.
Kira: Yeah... like, I get that, I found out that you were in a car...
Noah: Down here.
Kira: Here. I was like, oh! Wait! You’re IN the car, okay. So maybe like up here, to like create that perspective right from the top, say like ‘of my rubber tires,’ or like, do something to say that you ARE part of this, you’re in this car, not just... because when I read this, I just imagined this, like, a street and a car, like a car going by on a street, like I didn’t imagine you were in the car...
Noah: That works too...
Kira: [Then] the car came back, and I got it! ...Um, so yeah, so I wrote that here, perspective, create your perspective from the beginning and that makes the poem easier to sorta picture.
(Classroom transcript, 4/1/2009).

Kira notes the poem’s unclear setting and premise, and in response, suggests “creating [the] perspective” “from the beginning.” She points to clear consequences, invoking herself as a reader who had to “[find] out that you
were in [the] car,” and who thus misunderstood the context. Kira pinpoints her lack of clarity, noting that *Pitter Patter* does not mention the speaker until late in the poem. It shows the downpour to readers, but not the speaker’s perspective on the scene. To remedy this lack of clarity, Kira suggests revising the first stanza “to like create that perspective right from the top... do something to say that you ARE part of this.” She notes that because she didn’t understand this perspective, she failed to understand the poem’s scope—Noah’s experience, not an external description of a scene. While most of Kira’s suggestions change few words, they add sensory details, precise language, and Noah’s personal perspective to *Pitter Patter*. In contrast to his collaboration with Nasha (Table 1), his critique conversations with Kira (Table 2) contain explicit suggestions and justifications for revision. As demonstrated in the quotation above, Kira aligns her critiques with Noah’s text, pointing out textual elements and the consequences of his choices.

### Table 2: Descriptive Statistics: Noah and Kira’s Critique Conversations

<table>
<thead>
<tr>
<th>Category</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time available</td>
<td>1:18:42 (two 38-min conversations)</td>
</tr>
<tr>
<td>Written suggestions</td>
<td>6</td>
</tr>
<tr>
<td>Oral suggestions with clear textual antecedents</td>
<td>9</td>
</tr>
<tr>
<td>Oral suggestions that refer generally to the poem</td>
<td>0</td>
</tr>
<tr>
<td>Total revisions with clear antecedents in suggestions</td>
<td>12</td>
</tr>
</tbody>
</table>

Over the course of their two critique conversations (1:18:42 total time), Kira makes six explicit written suggestions that include word substitutions, image clarifications, and perspective suggestions; nine oral suggestions that have clear antecedents in *Pitter Patter*, and no general criticisms. As a result of these exchanges, Noah makes 12 revisions, including new or expanded stanzas, several clarified images, and a strengthened first-person perspective.

In contrast to his collaboration with Nasha, Noah’s conversations with Kira led to many more revisions. Tracing Noah’s work shows that while most students made relevant comments (ones that were potentially useful to learning about concepts like genre), not all of these suggestions led to real revision in future drafts. Rather, Noah’s responses to critiques present a clear pattern. He revises when a reader shows that his language does not convey his ideas—for example, Kira’s notes about *Pitter Patter’s* perspective—and suggests useful changes. In this way, audience response seems critical to Noah’s creative and representational processes.

### Limitations

Any analysis is necessarily limited by available data. Because bidirectional artifact analysis uses artifacts and transcripts as data sources, it cannot capture the full range of Noah’s revisions (some of which are not rooted in the feedback), or individual cognitions. While I cannot claim that these critiques fully account for the development of Noah’s poems, tracing the artifacts of and conversations around this work point to a link between feedback structures and revision patterns, as well as pedagogical strategies for creative writing.

### Discussion and Significance

Much early cognitive and rhetorical work on audience claims that attention to audience is one hallmark of expert writers (e.g. Berkenkotter, 1981), who ask hypothetical questions about their audience as a planning strategy (e.g. Flower & Hayes, 1981). Extending this claim to literacy learning suggests that learning to write for an audience contributes to writing skill. This paper examines the effects of writing for and interacting with readers and represents an attempt to learn more about how audience carries this weight. In this classroom, a workshop design allows for social and dialogic audience feedback (Bakhtin, 1976/1994), and makes the contributions of students’ readers more clearly evident. This bidirectional analysis draws concrete links between students’ feedback and revisions, and thus, broadly maps the affordances of a workshop design in a classroom space: Discussions with real readers proxy for the reactions of outside readers (Lunsford & Ede, 2009) and help learners to see their writing as communicative and meaningful. Classroom writing can become a design experiment that teaches authentic practice (Brown, 1992).

Broadly, Mr. Caswell’s classroom shows that a collaborative workshop, even one housed entirely within a school space apart from the “real world,” can create a “real” audience. The most successful groups came to understand that writing is a tool to share meanings, images and ideas. Before Kira pointed out Noah’s failure to clarify his perspective, he was likely thinking of his poem as an assignment, not a representation of his experience to share with readers. This analysis shows that Noah was more likely to revise around these perspectival critiques, perhaps because they changed his understanding of the poem itself. These insights about
the communicative nature of writing are important and rare in schools, where most traces of “writer” and “reader” have been subsumed in “right” and “wrong” answers to teachers’ questions (Nystrand, et. al., 1997).

At the same time, the open workshop format became a constraint for many students who did not ask questions, negotiate meanings, or justify reasonable paths for revision self-sufficiently. Examining these conversations in a bidirectional way highlights productive critique moves and shows that some students (Kira) constructed these on their own—and consequently inspired significant revision. Despite similar classroom resources, Nasha’s feedback was never as concrete or generative as Kira’s. This analysis thus suggest that (1) mentors should model revision techniques, and (2) mentors should model critique techniques for students to use (e.g. mirroring outside readers). Examining how critique stretches between and across drafts and students suggests how pedagogy can encourage a productive, revision-focused writing environment.

**Audience as a Social and Cognitive Tool**

These data show that these young, learning writers’ individual cognitions only took them so far. Whereas cognitive process theorists (Flower & Hayes, 1981) found that the abstract sense of audience—“to whom am I writing?”—helped experts to begin their work, social contact with real readers aided these students. There is little evidence that these young writers thought about audience before they began to write, but many used the designed practice of getting feedback from readers to identify problem areas, refine ideas, and revise. In short, workshop writing teaches young writers that writing is more than what a writer thinks that she is saying, but what other readers and listeners perceive as well. Developing this understanding motivates an orientation of questioning meaning, revising to feedback, and refining ideas through conversation. Real-world, disciplinary knowledge is built in this way as well—but these techniques are unavailable in schools when communicative, social paradigms for learning are absent.

**References**


The use of text and process mining techniques to study the impact of feedback on students’ writing processes

Rafael A. Calvo(1), Anindito Aditomo(1,2), Vilaythong Southavilay(1), Kalina Yacef(1),
University of Sydney, Australia,
Email: rafael.calvo@sydney.edu.au, aadi4954@uni.sydney.edu.au, vstoto@gmail.com, kalina.yacef@sydney.edu.au

(1) The University of Sydney, Australia,
(2) The University of Surabaya, Indonesia

Abstract: Understanding the impact of feedback in complex learning activities, such as writing, is challenging. We contribute a combination of writing environments and data and process mining tools that can provide new ways of measuring this impact. We use the tools in a field experiment in an engineering course (N=45). Responses (timing, amount and types of text changes) were examined using log data and process mining techniques. Two experimental conditions were used: reflective followed by directive feedback (A) and vice-versa (B). We found that both forms of feedback were read multiple times. Students required longer times to respond to reflective, compared to directive, feedback. The type of feedback, however, made little difference to the types of revisions that students performed. Overall, our findings point to the difficulty of encouraging students to reconsider and revise what they have already written.

Introduction

Writing is considered to be a critical form of learning activity at all educational levels. Writing is also a particularly complex activity, and it is generally believed that both novice and experienced writers benefit from feedback provided by others in order to improve their writing. This paper presents a new approach to study the impact of feedback and findings from a field trial in engineering education. This paper examines the impact of different types of writing feedback (directive vs. reflective) on students’ writing process, including the types of revisions students make to their document.

Feedback can be defined as information provided to a person about his/her performance in a task. In educational contexts, the provision of feedback is intended to increase not only a student’s performance, but also the likelihood of learning from the task. Intuitively, feedback should almost always improve learning and performance. Research has shown that the relationship between feedback, performance and learning is nothing but simple (Hattie & Timperley, 2007; Shute, 2008). A meta-analysis found that while feedback improved performance on average, there was a large variation in the effect sizes, and in a third of the studies feedback had a negative impact (Kluger & DeNisi, 1996). To explore possible moderators, Kluger and DeNisi formulated a theory called Feedback Intervention Theory (FIT). FIT draws from control theory and cybernetics to state that feedback interventions cause a person to compare the feedback with a standard or goal. Perceived discrepancies between the feedback and the standard will motivate the person to reduce the discrepancy.

FIT provides a framework to predict the influence of different types of feedback on learning. For instance, feedback with criticism (or praise) towards the learner would divert attention from task relevant processes and can impede learning. Similarly, feedback that highlights one’s performance compared to others (normative feedback, such as grades) would also impede learning. In contrast, feedback that directs attention to the task should facilitate learning, especially if it contains information needed to address the problem highlighted in the feedback (Kluger & DeNisi, 1996, pp. 267-268). Feedback that includes cues about the goal or standard of the task outcome (“goal-setting interventions”) should also increase performance and learning.

Writing, the particular task that we are concerned with here, is much more complex than the typical tasks used in feedback research. Nonetheless, there are some findings consistent with FIT. For instance, FIT would predict that without distinguishing different types of feedback, the effect of writing feedback might be negligible. In line with this prediction, a meta-analysis which lumped together feedback of various kinds found that feedback did not significantly increase the effectiveness of writing interventions on learning (Bangert-Drowns, Hurley, & Wilkinson, 2004, p. 47). But when types of feedback are distinguished, FIT would predict that their impact would differ. Not many studies have examined this issue, but there is some supporting empirical evidence. One study by Nelson and Schunn (2009) examined correlations between features of peer feedback and the likelihood of the feedback being implemented. The writing task was an essay in an undergraduate, introductory course on history. This study found that task-focused feedback (such as those that included specific solutions or specific location of problems) predicted implementation, whereas feedback that focused on the writer (those with affective language such as praise and criticism) did not.
To conjecture further about the role of feedback in writing, we draw from a cognitive model of writing proposed by Bereiter and Scardamalia (1987). They proposed that writing could occur in two different modes: knowledge telling and knowledge transforming. In knowledge telling, the composition process begins with the writer picking up topic and genre cues from the task description, and writing down knowledge from memory activated by these cues. Text already produced becomes an additional source of cues to retrieve knowledge from memory. This process of memory retrieval cued by the task description and text already produced is repeated until the writer feels he/she no longer has any relevant knowledge (or until time or space constraints are met).

The knowledge transforming mode of writing is more complex. It involves the construction and continual reconstruction of a content problem space (what to say) and a rhetorical problem space (how to say it) (Bereiter & Scardamalia, 1987). The two problem spaces interact, with output from one feeding into the other. For instance, a writer in this mode would think about whether the produced text, in its current form (a rhetorical issue), conveys what they intend to say (a content issue). The writer would also think about whether others, or she herself, believe what the text is saying. This may change the way she thinks about the topic, which in turn may prompt her to find a different way to express her new view.

How does writing feedback come into play? Feedback that points to the writing task (as opposed to the writer) has the potential to prompt a writer to reconsider what she/he has written and/or how it was written. In other words, feedback could prompt processes associated with the knowledge-transforming mode of writing, as reflected in more revisions that go beyond cosmetic text changes and also in the greater time lag between feedback and first revision. However, we conjecture that certain types of (task focused) feedback may be more effective than other types at prompting knowledge transforming processes. For instance, feedback that contains specific instructions (what we call here directive feedback) may prompt a writer simply to correct the specific problems, without much consideration. In contrast, feedback that asks students to connect problems in their text with broader content or stylistic issues (what we call here reflective feedback) may prompt more substantial revisions. In this study, we examine two approaches for giving task-focused feedback: directive and reflective. Directive feedback tells the student that there is a problem in the text that needs addressing. Reflective feedback asks the student to consider whether there is a problem.

Method

Research questions and hypotheses

The main question addressed in this study is: “What are the different impacts of reflective and directive feedback on students’ writing process in terms of the timing, amount and types of text changes performed?” As mentioned above, directive feedback explicitly informs a student that their text contains a specific problem, and instructs them to address the problem. Reflective feedback merely suggests the possibility of a problem, but asks the student to decide whether there is a problem that needs to be addressed (see examples in Table 2).

We conjecture that reflective feedback is more effective than directive feedback in prompting students to reconsider their ideas and revise what they have written (processes consistent with a knowledge-transforming mode of writing). More specifically, we predicted that reflective feedback would prompt more deletion of words. Reflective feedback would not necessarily prompt more addition of words, because adding words (expanding a text) can be performed without much reconsideration of text already produced.

Study design and participants

The participants were (N=45) undergraduate and postgraduate engineering students from The University of Sydney. All participants were enrolled in a project-based course where the main activity was to develop a web application. Students had to individually write a proposal that would become the basis of their software development project. The semester was 13 weeks long; the writing activity occurred in the first half of the semester and spanned 32 days.

Table 1. Timeline of the writing assignment

<table>
<thead>
<tr>
<th>Group</th>
<th>Day 0</th>
<th>Day 10</th>
<th>Day 12</th>
<th>Day 17</th>
<th>Day 19</th>
<th>Day 27</th>
<th>Day 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (n=22)</td>
<td>Assignment opened</td>
<td>1st draft submitted</td>
<td>Directive feedback sent</td>
<td>2nd draft submitted</td>
<td>Reflective feedback sent</td>
<td>Peer feedback sent</td>
<td>Final deadline</td>
</tr>
<tr>
<td>B (n=23)</td>
<td>Reflective feedback sent</td>
<td>Directive feedback sent</td>
<td>Reflective feedback sent</td>
<td>Directive feedback sent</td>
<td>Reflective feedback sent</td>
<td>Reflective feedback sent</td>
<td>Reflective feedback sent</td>
</tr>
</tbody>
</table>

In this context, we conducted a crossover field experiment where students were randomly assigned to receive either reflective (n=22) or directive (n=23) tutor feedback in response to the first drafts of their assignment. The first draft was worth 3% of the course mark. Students then had five days to revise before submitting their second draft, which was not assessed. Following this, students who initially received reflective feedback were provided with the directive feedback for their second draft (and vice versa). The crossover setup
helped satisfying ethical requirements by ensuring that students were not systematically disadvantaged by the type of tutor feedback. In addition, after the second tutor feedback, the students had to review their peers’ second drafts. The final submission was worth 5% of the course mark.

**Tools and measurement**

**The writing environment**
Writing activities were managed through iWrite (Calvo, O&Rourke, Jones, Yacef, & Reimann, 2011). Students write on Google Docs, a cloud-based application, and revisions of the documents are retrieved using Google’s API. Tools and heuristics developed to automatically recognize collaborative writing activities (Southavilay, Yacef, & Rafael A. Calvo, 2010) was used here to explore what students did after reading the tutor’s feedback.

**Tutor feedback**
The tutor feedback was prepared by the three instructors of the course (who each generated feedback for roughly 1/3 of the students). Feedback was written within the reviewers’ page in iWrite. In order to better align the feedback provided by the three instructors, feedback samples were kept in a shared document. An email announcing that feedback was available at a certain webpage was sent to all students simultaneously. Students were then able to access the feedback on the same interface. Between 3 and 6 feedback items were provided to each student in each phase (M=4.4, SD=1.12). Reflective feedback items had M=96.52 (SD=26.10) number of words at the first release and M=86.86 (SD=23.44) number of words at the second release. Directive feedback items had M=63.77 (SD=16.78) at the first release and M=58.47 (SD=20.66) at the second one.

**Data analysis**

**Log preparation**
While users write, Google Docs (GD) saves the document frequently and stores all its revisions. This feature provides a history of the document, where each version of the document can be retrieved with its timestamp. All major revisions available for the 45 documents (an average of N=7.76, SD=4.33 per document) were used. The content of the revisions was used for detecting the type of text change operations. In process mining terms, the history each document is a process instance (case),

In addition, based on the log file of iWrite we used a log file containing all the students’ interactions with the website, with a timestamp each time they accessed the feedback page. This log revealed when and how often individual students access and read their feedback. The feedback-reading log also has one process for each document (student), in total of 45 processes. The revision and feedback-reading logs were merged for each student. A process mining tool, ProM (ProM., 2010) was used to analyze how students accessed feedback and revised their documents. A Dot Chart Analysis (Song & van der Aalst, 2007) was implemented as a plug-in in ProM to extract a snapshot of student activities of revising their documents and accessing the feedback. The result is shown in Figure 1.

**Detecting text change operations**
The model developed by Boiarsky (Boiarsky, 1984) was used to analyze semantic changes in the writing process. In particular, the 12 types of text change operations (Southavilay, Yacef, & Rafael A. Calvo, 2010) shown in Table 3 were used. Text change operations at both paragraph and document levels were considered. At the paragraph level, beyond surface changes, the other types of text changes are listed in Table 3.

A text comparison utility was used to identify text change operations (Southavilay, Yacef, & Rafael A. Calvo, 2010). The utility is based on the Unix Diff utility, which takes two revisions of text and produces a difference statement with insertions, deletion, and replacements between the two. The text comparison utility uses both paragraph and word-differencing algorithms to detect text changes operations at paragraph and word levels. For each document each two consecutive revisions were compared and the differences computed. First,
the utility uses the paragraph differencing algorithms to discover the addition of new paragraphs, the deletion of existing paragraphs, and the alteration of existing paragraphs. Based on the paragraph-differencing algorithm, we detect text change operations of adding, deleting, moving/reordering, merging, and splitting of existing paragraphs. For each altered paragraph in the later revision, the utility then uses the word differencing algorithm to compare it to the corresponding paragraph in the former revision in order to detect text change operations of moving/reordering, replacing, inserting, deleting, and appending words in the altered paragraphs.

Table 3. Types of text changes automatically identified

<table>
<thead>
<tr>
<th>Text structure</th>
<th>Content change within individual paragraphs (i.e. word level changes)</th>
<th>Content change at the paragraph level</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Moving/reordering paragraphs</td>
<td>• Moving or reordering words in a paragraph</td>
<td>• Adding new paragraphs</td>
</tr>
<tr>
<td>• Merging paragraphs</td>
<td>• Replacing words in a paragraph</td>
<td>• Deleting paragraphs</td>
</tr>
<tr>
<td>• Splitting paragraphs</td>
<td>• Inserting words in the middle of paragraphs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Deleting words in a paragraph</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Adding words at the end of a paragraph</td>
<td></td>
</tr>
</tbody>
</table>

Findings and discussion

The dotted chart (Figure 1) shows the timing of each student’s access to the tutor feedback, and also the timing of revisions performed in Google Docs. Each row represents a process of one student revising his or her document and accessing the tutor feedback. One “column” (in between white vertical lines) represents a period of one week. Directive feedback is represented by triangles, reflective feedback by squares, and revisions by dots. Group A (who received directive feedback first) occupies the upper half of the chart.

Figure 1. Dot chart showing when each student accessed directive feedback (red triangles) and reflective feedback (blue squares), and also instances of revisions (green dots).

We refer to the period between first and second tutor feedback as Phase 1, and the period between the second feedback and final submission as Phase 2. The black line represents a date when the first tutor feedback sent and the brown line represents a date when the second feedback sent. Several observations can be made from Figure 1. While the feedbacks were sent at the same time, students accessed them at different times (most accessed the feedback on the same day they received it, but some took one or more days). Also, almost all students accessed the feedback multiple times (except for one student who accessed the feedback only once). On average students in Group A accessed the directive feedback 5.8 times and the reflective feedback 4.6 times. In Group B, on average students accessed the reflective feedback 6.3 times, and the directive feedback 4.9 times. Students in both groups revisited the first tutor feedback during Phase 2 (i.e. after receiving their second feedback), as can be seen from the mix of red triangle and blue squares on the dot chart. Hence, students’ behaviors in Phase 2 need to be seen as potentially influenced by both types of tutor feedback. An examination of the timing the changes shows that students took longer to respond to reflective feedback than to directive feedback (1 day longer in Phase 1, and 2 days longer in Phase 2). While not predicted theoretically, this difference is understandable, as reflective feedback requires students to think for themselves, as opposed to being a direct instruction.

In order to see how students revised their documents according to feedback types, we analyzed four types of revision types (adding paragraphs, deleting paragraphs, adding words and deleting words in existing paragraphs) and the corresponding number of words added and deleted. For each phase, we compared the
number of words added and deleted in the four types of revision types in the two experimental conditions. In order to distinguish between headings (not considered in the analysis) and content paragraphs, we defined a paragraph as containing at least 7 words. This resulted in several findings.

First, not many students performed major revisions, especially in Phase 1. For instance, in Phase 1, only 13 students from Group A and 12 from Group B added new paragraphs, and only 10 from Group A and 9 from Group B added words to existing paragraphs. Even fewer students revised their documents by deleting words: in Phase 1, only 4 students from Group A and 5 from Group B deleted paragraphs, and only 8 from Group A and 3 from Group B deleted words in existing paragraphs. In other words, most of the students in both groups performed only minor revisions in Phase 1. In Phase 2, more students did revise their documents. They mainly did so by adding paragraphs and by adding words to existing paragraphs, rather than deleting paragraphs or deleting words within paragraphs. Over the two phases, only about half of the students (in both groups) performed revisions by deleting words. This indicates that only about half of the students reconsidered and revised what they have already written (which are indicators of knowledge transforming writing processes).

A second set of observations are related to the differences, or lack thereof, between Groups A and B. Although in terms of average number words (see Table 4) Group B seemed to have performed more additions and deletions (in both phases), the box plots above show that there is more variation within rather than between the groups. That is, the difference in the average seemed to be inflated by several students who performed much more extensive revisions than most other students. This difference of amount of revision between the groups largely disappeared if we consider the median (as shown in the box plots). Groups A and B are also similar in terms of the number of students performing the three of the four types of revisions shown in the box plots. There seems to be a difference between the groups in Phase 1, in terms of the deletion of words in existing paragraphs: 8 students in Group A, compared to 3 in Group B (although these 3 students made more extensive deletions compared to the 8 students from Group A).

Together, these observations suggest that the feedback (both reflective and directive) failed to prompt major revisions for most of the students. This is an important pedagogical point: that students have the liberty to take into account, or not, their tutors' feedback. Most students did not revise extensively, despite the feedback given, the time students had to address the feedback, and also the fact that students accessed the feedback several times. Furthermore, the observations indicate that our theoretical predictions about the impact of different feedback types were not supported by the data. This could have been due to several reasons which are more to do with the methods, rather than the hypotheses themselves. One possible explanation is that not all of the reflective feedback was of the same quality: an initial inspection of the reflective feedback items indicated that some items were more directive (instructing students to do specific things). Another possible reason for the lack of difference between the groups’ revision behavior is that too few students engaged in extensive revisions during Phase 1, mainly because the students had little incentive to perform those revisions (as indicated above, students’ behavior in Phase 2 cannot be taken as indicators of the influence of different feedback types).

References

Acknowledgments
This project has been funded by Australian Research Council DP0986873 and a Google Research Award.
Variation in Fifth Grade Students’ Propensities for Managing Uncertainty during Collaborative Engineering Projects

Michelle E. Jordan, Arizona State University, Mary Lou Fulton Teachers College, Farmer Education Building 1050 S Forest Mall, Tempe, AZ 85287–1811, michelle.e.jordan@asu.edu

Abstract: Uncertainty is ubiquitous to learning, particularly in collaborative learning contexts, which are likely to induce uncertainty related to both social and task issues. The purpose of this qualitative study was to explore how students vary in their individual propensities for managing uncertainty they experience while working with peers. Using observations and interviews I examined how students in one fifth grade class managed uncertainty while collaborating on robotics engineering projects. Relying on techniques of grounded theory and microanalysis of discourse, I addressed the question: How do students vary in their individual propensities for managing uncertainty they experience during collaborative academic tasks? In particular, I examined variation in the size and composition of the set of strategies from which students drew, changes in students’ use of strategies across projects, and their willingness to entertain or take up uncertainty.

Introduction
Learning in academic contexts involves experiencing psychological uncertainty, and the ways students manage their uncertainty influences their behavior and their learning in collaborative contexts (Huber, Sorrentino, Davidson, Epplier, & Roth, 1992). Although learning is often portrayed as a process of reducing uncertainty, it is long recognized that learning may also require the cultivation of uncertainty (Bereiter & Scardamalia, 1993; Piaget, 1972; Sieber, 1974). Furthermore, when academic tasks entail collaborating with peers, uncertainty may be particularly prevalent as it is likely to stem from diverse social and task issues. If uncertainty is a pervasive experience in collaborative learning contexts, it is important for educators and educational researchers to understand learners’ responses to this experience in order to develop strategies for helping students increase their skills at effectively managing the uncertainty they experience while working with peers. Educational philosophers the likes of Dewey (1929) and Bruner (1996) have extolled the place of psychological uncertainty in learning; however, the experience of uncertainty has most often been discussed tangentially to concepts of more focal interest to educational researchers. Wishing to focus directly on how students manage uncertainty during collaborative academic tasks, I conducted a qualitative study relying on observations and interviews with fifth grade students collaborating in small groups to design and build robots. Specifically, I addressed the question, how do students vary in their propensities for managing uncertainty during collaborative tasks?

Theoretical Framework
Uncertainty is an individual’s psychological experience of doubting, being unsure, or wondering about how the future will unfold, what the present means, or how to interpret the past. Uncertainty can pertain to one’s self, other individuals, or other aspects of the environment. Rather than a coldly cognitive phenomenon, uncertainty is a cognitive feeling (Clore, 1992) that can be more or less conscious and more or less tied to emotions. Uncertainty management is behavior students engage in to facilitate action in the face of uncertainty. Although uncertainty management is often conceptualized as entailing tactics to reduce uncertainty, it also pertains to efforts aimed at ignoring, maintaining, or intentionally increasing uncertainty (Jordan, 2010; Babrow & Matthias, 2009). Which strategies might be appropriate in a given situation depends on multiple factors, including the social context and social norms for managing uncertainty that have been established in a particular community of practice (Goldsmith, 2001; Lingard, Garwood, Schryer, & Spafford, 2003).

Self-report scales and experimental methods to measure individual orientations to various forms of uncertainty can be traced from the mid-20th century (e.g., Budner, 1962; Debacker & Crowson, 2006). Such approaches may not reflect the complicated nature of uncertainty management as observed by teachers and as experienced by students working with peers. Therefore, I used a methodological approach that allowed me to observe the rich nature of uncertainty as it is experienced and managed by students in collaborative contexts.

Even young students can recognize their uncertainty (Metz, 2004), but they often fail to experience uncertainty when uncertainty is warranted (Sieber, 1974). The amount of uncertainty students experience during academic tasks can be influenced by task characteristics such as evaluation systems (Doyle & Carter, 1984), novelty (Herbst, 2003), task framing (Schauble, Klopfer, & Raghavan, 1991) and discourse (Hierbert et al., 1996). Therefore, I chose to study uncertainty in a collaborative design setting, open and ambiguous, ill-structured and generative (Jonassen, 2000; Kolodner et al., 2003).
Method
The study was initiated at the beginning of instruction in robotics engineering and preceded through the completion of three collaborative projects (14 sessions each); two were more-structured and close-ended and one was an ill-structured task (Spiro, Feltovich, Jacobson, & Coulson, 1991). All projects utilized LEGO Mindstorms materials. Students worked in groups of three or four, changing group membership for each project. Participants were 24 fifth graders in a public suburban classroom. The study concentrated on 15 focal students purposefully selected for diversity of gender (eight female), ethnicity (seven Black, four White, two Hispanic, two Asian), and academic achievement (five received special education services; two gifted-and-talented).

I conducted naturalistic observations (1½-2 hours each) and interviews. Data sources include expanded field notes, audio and video recordings, transcripts for 12 groups across three projects, and interview transcripts. Data analysis was inductive and interpretive, relying on techniques of grounded theory (Corbin & Strauss, 2008) and influenced by Erickson’s (1992) ethnographic microanalysis of discourse and interactional sociolinguistics (Mercer, 2000). Member checking with the teacher and peer debriefing occurred throughout the study.

A two-pronged strategy was used to identify individual propensities. First, drawing on knowledge of linguistic, paralinguistic, and gestural markers of uncertainty (e.g., Feldman & Wertsch, 1976; McNeill, 1992), I conducted open coding of uncertainty experiences and management using transcripts of small group sessions and student interviews randomly selected from projects 1 and 3. Using axial coding to refine categories, I identified 20 management strategies and categorized them as tactics for reducing, ignoring, maintaining, or increasing uncertainty (see Table 1). Re-reading the data and focusing on individuals, I determined which strategies dominated each focal student’s approach to uncertainty. I then organized these data in a strategy-by-student matrix to identify students whose overall patterns were similar.

<table>
<thead>
<tr>
<th>Table 1: Strategies and Categories for Managing Uncertainty.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reduce</strong></td>
</tr>
<tr>
<td>Analyze issues</td>
</tr>
<tr>
<td>Trial-and-error</td>
</tr>
<tr>
<td>Seek consensus</td>
</tr>
<tr>
<td>Seek expert other</td>
</tr>
<tr>
<td>Request info.</td>
</tr>
<tr>
<td>Refer to authority figure</td>
</tr>
<tr>
<td>Seek information from materials or texts</td>
</tr>
</tbody>
</table>

In a separate analysis, I created “mini-portraits” of fifteen focal students (Do & Schallert, 2004), for whom I had data from at least two robotics projects (the exception was “Nathan,” for whom I had data from only one project). Creating case studies in NVIVO, I first selected five to ten episodes from small-group sessions in which the focal student played a major part in group interaction. I conducted microanalysis of transcripts of each episode, examining uncertainty (and certainty) and uncertainty management and comparing with field notes and recordings. I read through all interview transcripts with each focal student multiple times and integrated information across data sources, making extensive memos and tuning the picture of students’ propensities. I then compared these portraits with the matrix to determine consistency between the two analyses.

Findings
Of the 20 management strategies identified, 16 were used by all 15 focal students at least once. Despite the fact that all students used a majority of the strategies, I was able to identify for each student his/her propensity for responding to uncertainty. There were five propensities in all (see Table 2).

<table>
<thead>
<tr>
<th>Table 2: Characteristic Patterns of Uncertainty Management.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Pause for Reflection</td>
</tr>
<tr>
<td>Demetre Berta</td>
</tr>
</tbody>
</table>

Pause to reflect. Students in this category accepted uncertainty as a normal and inevitable aspect of academic tasks. They had little reticence about expressing uncertainty, recognizing and acknowledging it openly. They exhibited a desire to collect more information and reflect prior to action, often using strategies...
categorized as maintaining uncertainty. They asked for explanations of members’ ideas. Rather than address and resolve an uncertainty only in relation to a present situation, they reflected on its greater meaning.

Seek a plausible explanation. These students expected uncertainty and did not seem threatened by it. Their uncertainty management was instrumental, focused on achieving the task. They were proactive in using group members, seeking confirmation for potential actions and comparing perspectives to create plausible explanations. They tended to interpret outcomes of action in a positive light, thus sustaining task focus. They planned for and took tentative action and observed and analyzed the outcome of each step.

Take action. Students with this set of characteristics were gung-ho, forward-oriented, and impatient with uncertainty. They oscillated between managing uncertainty by risk-taking/trial-and-error experimentation and dismissing uncertainty by making positive predictions. They rarely consulted about their own or their group members’ decisions. When pressed for explanations, they often replied, “Ms. Katell said...” Eager to experiment and see what happened, these students rarely let analysis stand in the way of the next test, either prior to or subsequent to action. Rare reflective actions were usually based on immediate past experience.

Can somebody help me? These students preferred to fly under the radar. They were willing to acknowledge uncertainty but tried to avoid it. They frequently turned tasks over to group members, relied on outside expertise, and requested scaffolding from their group members for their project participation. These students requested help with immediate uncertainty, rarely considering the long-range use of that help.

What, me uncertain? Unique in their dogged denial of uncertainty, these students minimized or avoided acknowledgment of uncertainty. They were quick to blame their peer collaborators rather than puzzle over a problem. They made confident claims, (e.g., “I know what to do”) rather than mitigate their claims with uncertainty markers. They prepared for failure by expressing certainty, for instance, “It’s not going to work.”

One way to better understand variation is to examine the interaction of students with different propensities. There was always tension between action and reflection as students worked toward joint academic goals, a tension that impacted learning and project success. The ways students experienced and managed uncertainty played a large role in how this tension played out. This was the case with Nathan and Demetre, group members for the first robotics project. Demetre’s propensity for managing uncertainty was to pause for reflection. He often used analogies in his attempts to reduce uncertainty as represented in his response to my inquiry about his group’s problems programming the obstacle course.

Demetre’s response to his uncertainty about programming his group’s robot was to collect more information and reflect on similar problems before taking action. Contrast with Demetre’s need for reflection is Nathan’s propensity to take action. Nathan operated from a model of risk-taking in the face of uncertainty.

We did this new program that actually worked, but at first I was a little bit unsure that it would work. And it ended up coming through for us when it hit the wall... Well since I wasn't very sure, testing stuff, well, that program was kind of strange. But a scientist needs to take his risks... Sometimes when people do experiments and they're not sure, they take a risk and most of the time it comes through.

Nathan managed uncertainty by mimicking one way he saw other scientists manage uncertainty. He attributed his group’s success not to knowing what to do and doing the right thing, but to their risk-taking. These contrasting strategies for managing uncertainty caused friction throughout the project, with Nathan wanting to take action and Demetre pushing for reflection. There was little explanation or discussion to create sense together that might have helped these students resolve uncertainty, learn, and progress with the project.

I also organized the data in terms of whether students used many/few strategies and whether they did/did not easily acknowledge uncertainty (see Table 3).

Table 3: Range of Strategies and Recognition of Uncertainty

<table>
<thead>
<tr>
<th></th>
<th>Many Strategies</th>
<th>Few Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easily Acknowledged Uncertainty</td>
<td>Satya, Berta, Ray, Isabel</td>
<td>Alexis, Sierra, Shamitra</td>
</tr>
<tr>
<td>Did Not Easily Acknowledge Uncertainty</td>
<td>Derrick, Trevor, Becky, Kisha, Bobby, Luis, Nathan</td>
<td></td>
</tr>
</tbody>
</table>
The reader will note that cell 3 is empty. Note also that students with propensities and used many strategies for managing uncertainty. Students with propensities, and used few strategies. Those in cell 4 not only made frequent use of many strategies, but also tended to draw from a diverse range of strategies that included ignoring, reducing, and maintaining uncertainty.

Students with propensities, , and tended not to acknowledge uncertainty, whereas students with propensities and did. For example, Satya’s willingness to take up uncertainty was exhibited when she responded to an announcement that Kisha and Becky had dismantled the trailer the group had labored on for days, intending to make a robo-claw instead. Satya was visibly upset by the decision her group members’ made without her. Yet, even while being thrown off-kilter by the change of design, Satya rather quickly surrendered her commitment to her past work, distracted by the new set of problems.

Satya: But is it going to be able to move, go like this and then go like this.
Kisha: I don’t know.
Satya: But if it doesn’t do that we just/
Kisha: //Oh, yeah, it has a motor. //Yeah, it should, yeah.
Satya: No, it needs two motors then, right. It needs to be able to move go like this and then move this way like this.
Kisha: It should be.
Satya: You’re sure, sure.
Kisha: I’m sure.
Satya: Well then.
Satya: I’m going to cry. Wait, how many motors does it need.

Satya’s response could have been to be quite certain that the new idea would not work. Instead, she was willing and able to entertain uncertainty about the situation in which she found herself.

Significance
Students brought with them to the collaborative engineering setting tendencies for managing uncertainty that had developed through experience, their habits and histories of participation in prior groups. The students all used a range of strategies, but at the same time they had individually identifiable patterns of uncertainty management. Some students had many ways to manage uncertainty; some students had only a few ways. Some students easily acknowledged uncertainty; others seemed to resist taking up uncertainty when uncertainty seemed warranted to one or more of their group members, to their teacher, or to me as an observer.

Although the number of participants from this study is too small to make generalizations, it is interesting that the three students in the category were female and that all three student in the category were male. Since previous research has noted that the expression of uncertainty can vary by gender and socioeconomic status (see, ), future studies need to explore the association between these characteristics and uncertainty propensities. Future research should also examine how propensities for managing uncertainty in collaborative environments are related to learning and achievement.

Such of the research on uncertainty management has conceived of it as an individual level of analysis; however, some see uncertainty management as a social issue (see, ). A majority of activities are likely to face as adults are collaborative tasks, and most of the learning they do will be of a social nature (Bruner, ). For most individuals, most of the time, the primary resource they have is each other, whether in face-to-face or virtual interaction. Thus, how one manages uncertainty in a collaborative context is important. Given the central position of relationships in determining how social systems emerge and unfold (Arrow, c Berdahl, ), it is likely that the nature of interdependencies and interactions within a collaborative peer group will influence the ways individuals in that group manage uncertainty (ordan, cDaniel, under review) and that the propensities of individual group members will likewise influence the quality of interactions in the group. The study described here contributes to understanding of these issues.

If uncertainty is required for learning and is a ubiquitous experience in collaborative learning contexts, then it behooves educational researchers to understand how students manage uncertainty and ways in which the interaction among students with different propensities influences the quality of peer collaboration. Doing so could allow us to identify instructional strategies for helping students increase their range of strategies, willingness to acknowledge uncertainty, and abilities to recognize when different strategies are appropriate.

References

© ISLS 2012  Volume 2: Short Papers  424


Breeding Birds to Learn about Artificial Selection:
Two Birds with One Stone?

Aditi Wagh & Uri Wilensky, Northwestern University, 2120 N Campus Drive, Evanston, IL 60208, USA, aditiwagh@northwestern.edu, uri@northwestern.edu

Abstract: Recent research in Learning Sciences has drawn attention to the affordances of enabling students to learn about scientific phenomena through a complex systems lens. In this study, we adopt a complex systems perspective in helping students to learn about artificial selection by using an agent-based participatory simulation - Bird Breeder. Our goal is to identify the strategies students used when using this simulation, and investigate the kinds of abstractions they made about the underlying agent-level mechanisms and emerging population-level outcomes of the model. We answer these questions using data collected from three high school biology classes that used this simulation. Our findings indicate that Bird Breeder was effective in fostering learning about the agent-level mechanisms that drive artificial selection. In contrast, the effectiveness of Bird Breeder at helping students learn about population-level outcomes of artificial selection is less clear. The findings have provided insight into possible design revisions to this simulation.

Literature review

The study of complex systems is increasingly becoming a new strand of literacy (Jacobson & Wilensky, 2006). We consider a complex system to be an emergent system in which population-level trends emerge from individual-level mechanisms. In the educational context, the study of complex systems as a literacy entails empowering students with the opportunity to understand the complexities of a system at varying levels of the system (Wilensky, 2003; Wilensky & Resnick, 1999).

Biology is replete with instances of complex systems in which phenomena can best be understood by grappling with relationships between levels in the system (Wilensky & Reisman, 2006). Evolution, which is central to the study of biological sciences, is a complex emergent phenomenon, often grossly misunderstood by the public at large (Alters & Nelson, 2002). Broadly viewed from a complex systems perspective, evolution is a process of change at the level of populations, such as species, resulting from mechanisms at work at the individual level such as organisms, alleles and so on.

Popular representations of evolutionary change such as cladograms and graphic simulations have been widely studied for their impact on student learning (Ainsworth, 2009; Evans et al, 2010; Soderberg & Price, 2003). While these representations depict change over time, they fail to provide insight into the mechanistic underpinnings of evolutionary change. They do not represent how populations change over time. Agent-based representations are more effective at helping students learn about how population-level changes emerge from individual-level interactions (Centola, McKenzie & Wilensky, 2000; Wilensky & Novak, 2010; Yang & Passmore, 2010). Most of this work has been done using agent-based models in which students explored models of evolutionary processes as an observer of the system.

In this paper, we examine the use of agent-based models in which students are participants in the system. This is done using agent-based participatory simulations, which enable students to step into an agent-based model by becoming an agent in the system (Wilensky & Stroup, 1999a). By doing this, students discover strategies that characterize the agent-level mechanistic underpinnings in an emerging system. In this way, learners increasingly appropriate the relationship between emergent aggregate-level change in the system and the agent-level mechanisms that drive the change (Wilensky & Stroup, 1999a).

A participatory simulation is likely to be a felicitous environment for a notoriously challenging topic such as evolution because it can leverage students’ agent-level cognitive resources related to evolution (Wagh & Wilensky, 2010). The goal of this paper is to study this potential in the context of a participatory simulation, the Bird Breeder (Wilensky & Novak, in preparation) which forms one activity in the agent-based modeling curriculum for evolution, BEAGLE (Biological Experiments in Adaptation, Genetics, Learning & Evolution) (Wilensky & Novak, 2010).

Research Questions

In this paper, we investigate two questions in the context of students participating in the Bird Breeder simulation. First, we examine the strategies students spontaneously employed when participating in the environment to breed certain variations of birds. Second, we investigate what students learned about the agent-level mechanisms and population-level outcomes of artificial selection from participating in this environment.

Methods
Data Collection
This study was conducted in the context of a two-week long implementation of BEAGLE in three high-school biology classes in a mid-western town in the United States. Students used a participatory simulation of artificial selection, Bird Breeder developed in the HubNet module (Wilensky & Stroup, 1999b) of NetLogo (Wilensky, 1999). The students spent one class period working on Bird Breeder. On the second day, the teacher led a whole class discussion to synthesize what students had learned from the model. Then she asked students to answer one of two questions: to note down either three mechanisms or three outcomes of artificial selection that they learned from the model. The students worked alone and were given about 7-10 minutes to work on this task. Responses of 38 students were collected and transcribed. Students were also video-recorded as they worked in their groups on this model. Ideas of three groups were closely analyzed.

Bird-Breeder differs from other participatory simulations in BEAGLE in that the class is divided up into groups of four, with each group engaging in their own enactment of the participatory simulation rather than the entire class participating in a single enactment. Within their groups, each student assumed the role of a bird breeder and was randomly assigned three or four birds at the start of the simulation. These birds differed from each other with respect to four traits: the color of their crest, tail, breast and wings. Through the course of the simulation, students could breed birds in the communal breeding site. Students could also release birds they did not want into the wild. Each group’s goal was to breed three pairs of birds with a red tail and wings, a purple breast and a blue crest.

Data Analysis
Two coding schemes were developed to analyze the data. The first coding scheme comprises codes that identify strategies students used when breeding birds. These strategies were identified through student utterances were coded as a strategy when students verbalized a move by giving directions to another student or justifying their own move.

The following codes emerged from open-coding of student talk in their groups. 1) Elimination: A student move was so coded when a certain bird was released from the population. 2) Selective focusing: A student move was so coded when students focused on one or two traits for breeding at a time. 3) Suitable mate: A student move was coded such when students purposefully picked a certain bird as a good mate for another bird. 4) Purebred: A student move was so coded when students bred 2 goal birds assuming that the offspring would be a goal bird. 5) Survey: A move was so coded when students reviewed the population of birds they collectively had in their group.

A second coding scheme was developed to analyze student responses to the question posed by the teacher. The codes are: 1) Model-centric: A response was so coded if it included details about the model without identifying a mechanism or outcome of artificial selection. 2) Agent-level mechanism: A response was so coded if it described one or more of the underlying agent-level rules for artificial selection. 3) Population-level outcome: A response was coded such if it described what was happening to the population over time. 4) Other: A response that could not be categorized in any of the previous categories.

Each student response was coded and responses in each category were counted up for analysis.

Results
Question 1: What strategies emerged as students participated in this simulation?
As students engaged in this simulation, they spontaneously enacted mechanisms of artificial selection in the form of strategies used to breed the goal birds. Four mechanisms were most commonly adopted by the groups first, students eliminated birds that did not have the variations needed in the goal birds second, students actively strategized about which birds would be suitable mates to breed third, students selectively focused on one or two traits to breed at a time and last, once a male and female goal bird was bred, students used those birds for further breeding.

Across the three groups, students tended to release birds when they ran out of space in the breeding site. In the beginning of the simulation, students retained birds that had even one of the traits they needed. As the simulation progressed, students tended to retain birds with two or three of the desired variations.

The following is an excerpt from Group 3. Group 3 had little space left in their breeding grounds to breed more birds. However, they had not yet bred the goal birds and needed to continue breeding birds.

S1: Oh. Um, the problem is that you don’t have any space.
S2: Oh, I don’t have any space.
S1: Uhhh yeah. So get rid of your two, your two, that has no color, and spots, and make them fly away.

In this excerpt, S1 asks S2 to release the two birds that do not have any of the variations they need. All groups commonly pursued this strategy of releasing birds. Moreover, once a group managed to breed several of the goal birds, they were willing to release even a goal bird. This excerpt, taken from Group 3, illustrates this strategy.
S3: We just keep- get rid of some of your birds, S1.
S1: Yeah, I know.
S3: Be careful you don’t get rid of the good female. (The group laughs)
S3: I’d rather get rid of a good male because we have extra.

In this excerpt, S3 is willing to forego a good male (male goal bird) rather than a good female (female goal bird) because they have extra males. The development of this strategy and the willingness to release even goal birds indicates that students were learning something about how the pool of the breeding population determines the kinds of offspring that would result from it.

A second strategy students commonly pursued was to purposefully select which birds to mate rather than randomly mating birds. Male and female birds with the maximum number of desired variations were selected for mating. The following excerpt, taken from Group 1, illustrates the use of this strategy.

S3: Well, I've got a red tail and a red wing. So, what we need to do-
S2: Alright, S3, you’ve got purple
S1: Oh, hold on S3, right off the top, the one you have selected-
S4: I have red, I have ..
S1: The one you have selected is going to be a good one to start with because three of the traits are already there, homozygous recessive so we just need to breed that with, does anybody have ..
S4: I’ve got a red wing.
S2: Yeah, yeah, let’s try to focus on the red wing.

In this excerpt, S1 encourages S3 to begin mating with the selected bird because it already had 3 of the 4 variations they wanted in the goal bird. Moreover, this group uses the third common strategy used by groups, selectively focusing on one or more traits at a time. This group decides to focus on a single trait, color of the wing in further breeding. It is important to note that Group 1 was using a feature of the model they had been asked to deactivate, the genotype display. Students had been asked to turn it off. This group accidentally left it on, and was using that information as a scaffold to guide their activity.

These two strategies, purposeful mating and selective focusing were seen across all three groups. In this excerpt, one sees students using both these strategies: they select a suitable mate for the bird they want to breed, and make that decision based on the traits they plan to focus on, purple breast and red tail.

S1: And then after they make love, and have kids, then we’re going to have to breed your other one.
S2: Dang it.
S1: Yessss
S3: Closer.
S1: Now we gotta breed that bird with your, uh, other one.
S2: Alright take, now take that one, and breed it with-
S1: Your right one-
S2: No, the same one you bred it with earlier.
S1: No, it needs a red tail.
S2: Oh. Yeah-
S1: And that one has a purple breast too, so just breed that one and you can possibly knock out two birds with one stone. (laughs). Pun.
S3: Oho Two birds with one stone.
S1: Now we got the purple breast, now we need a red wing and tail-

Finally, the fourth strategy that emerged was that once a group bred the goal bird, they exclusively used it for further breeding. The following excerpt is from Group 3, taken from just before they bred their first goal bird.

S2: Grrr. This needs a blue cap. God this stupid c'mon, c'mon, c'mon □YES□
S1: YES□
S2: Now we just breed that with your good one- with your good male- I mean, female- and we just keep doing it over and over again until we have □
S3: We just keep- get rid of some of your birds, S1.
S1: Yeah, I know.
S3: Be careful you don’t get rid of the good female.
(The group laughs.)
S3: I’d rather get rid of a good male because we have extra.
S1: Yes. We got our two. We got two. (Students give each other high fives)
S1: We have extra females.
S3: And then, at the end, you could just move, uh-
S2: C'mon, c'mon, c'mon- dangit. That’s a good female.
S3: Alright, now we just do this again.□

At the start of this excerpt, Group 3 needed only a blue cap to breed the goal bird. Their last breeding
gave them a goal bird that they called the "good" male. S2 suggests breeding the newly bred "good" male with the good female they already have, over and over again. Their plan is to breed a "good female" and then breed the two over and over again to get more goal birds.

An interesting point to note here is that though the group correctly strategized that breeding two purebreds with the four required variations of traits will result in another goal bird, they are inaccurately assuming that the two goal birds they have are indeed purebred. That is, they are assuming that the two birds are homozygous recessive without testing their assumption. Though this group has discovered a mechanism that is key to artificial selection, it is manifesting levels confusion (Wilensky & Resnick, 1999) by taking the phenotype of a bird as an indicator of being a purebred rather than its genotype. This observation has led us to re-think the design of the activity to enable students to view the already existing feature in the model, viewing the genotype of a bird.

In conclusion, the findings suggest that engaging in the Bird Breeders model facilitated discovering agent-level mechanisms that contribute to change in populations through artificial selection.

**Question 2: What did students learn about the mechanisms and outcomes of artificial selection?**

Of the 38 student responses analyzed, 22 students responded to the question about mechanisms of artificial selection while 16 students responded to the question about outcomes of artificial selection.

Of the 22 students who responded to the mechanisms question, 16 students described one or more mechanisms of artificial selection. For instance, one of the students responded, "Breeding lots of purebreds. Removing specimen with undesirable traits. Mating specimen with some desired traits with specimen with other desired traits." Another student described the mechanisms as: "breeding several purebreds, removing individuals with unwanted traits from the breeding population, selectively breeding individuals with required traits, and conducting several matings for each pair."

Of the 16 students who responded to the question about outcomes of artificial selection, 7 students described the outcomes in model-specific terms, by describing the goal birds in the simulation. For example, one of these students responded, "Blue head, red tail with red wing, purple chests. Blue head red tail wing purple chest. Blue head with red wing tail. Purple chest same outcome. A normal bird. In other words, these 7 students described the outcome in terms of the goal birds they tried to breed in the simulation. Only 5 students described the outcome in terms of population-level trends. For instance, one of the students responded, "Lowered genetic diversity in the species." Another student responded, "Less durability in the species." These findings indicate that while a majority of students reported agent-level mechanisms of artificial selection, fewer students described a population-level trend as an outcome of intentional selection.

**Discussion & Implications**

The findings of this paper indicate that Bird Breeder was a felicitous learning environment for students to learn about mechanisms underlying artificial selection. Participation in the environment served to foreground the mechanisms of artificial selection for students. It is important to note that students had not been taught these strategies before using the simulation, or been explicitly asked to use them. By spontaneously generating strategies to breed the goal birds, students learned about mechanisms of artificial selection.

Fewer students explicitly expressed their understanding of how populations change through artificial selection. We think there are three possible explanations for this. It is possible that the phrasing of the question did not trigger student thinking about what was happening at the population level and led students to respond in terms of model-specific outcomes (description of individual goal birds). Another possible explanation relates to the nature of representations included in the model. Unlike other HubNet models in BEAGLE, this model did not have a population-level representation that depicted the shifting trend in the population over time. This lack of an explicit aggregate-level representation might have made it difficult for students to synthesize how the population had changed over the course of the simulation run. Another explanation might be that Bird Breeder was a small-scale participatory simulation consisting of 4 students rather than the entire class. This small-scale HubNet model might have made it difficult for students to notice the aggregate-level trends that were emerging from their interactions.

This study has opened a few exciting avenues for possible design revisions to the activity and future work. With respect to the design, we would like to include an aggregate-level representation in the model to depict how the population changes over time in the model so students can track shifting trends in the population. We would also like to give students access to the genotype indicator of each bird for a part of the activity. We
think that information will give students the opportunity to reason at a third level of complexity, individuals’
genomes, to reason about how change at that level influences change at the level of individual phenotypes, and the
population. Finally, we would like to compare the affordances and constraints of exploring an agent-based
model as an observer vs. a participant.

References
Museum of Natural History, Cambridge, MA.
the evolution of cooperation and altruism through multi-agent modeling The EACH Project. Proceedings
of the Fourth Annual International Conference of the Learning Sciences, Ann Arbor, MI, June 14-17.
Do pictorial representations of evolution help or hinder museum visitors’ understanding. Understanding
the Tree of Life Conference, Carnegie Museum of Natural History, Pittsburgh.
http://ccl.northwestern.edu/netlogo/models/BirdBreeder. Center for Connected Learning and
Computer-Based Modeling, Northwestern University, Evanston, IL.
and evolution using EOLIE, a computer simulation. *International Journal of Science Education*,
25(1), 35-55.
Computer-Based Modeling, Northwestern University. Evanston, IL.
*International Journal of Computers for Mathematical Learning*, 8(1), 141 (special issue on agent
based modeling).
the World. *Journal of Science Education and Technology*, 8(1).
Modeling, Northwestern University.
Wilensky, U., & Stroup, W. (2002). *Participatory Simulations: Envisioning the networked classroom as a way to
support systems learning for all*. Paper presented at the annual meeting of the American Educational
Constructing and Testing Computational Theories -- an Embodied Modeling Approach. *Cognition &
Instruction*, 24(2), pp. 171-209.
models of evolutionary dynamics. In R.S. Taylor & M. Ferrari (Eds.), *Epistemology and Science
Learning and Computer-based Modeling.
pieces of knowledge about natural selection. Paper presented at Jean Piaget Society Conference,
Berkeley, June 2-4.
Xiang, L. & Passmore, C. (2010). The Use of Agent-based Programmable Modeling Tool in 8th Grade Students’
Tracing Ideas and Participation in an Asynchronous Online Discussion across Individual and Group Levels over Time

Alyssa Friend Wise, Ying-Ting Hsiao, Farshid Marbouti, Yuting Hao
Simon Fraser University, 250-13450 102 Avenue, Surrey BC, Canada, V3T 0A3
Email: afw3@sfu.ca, yha73@sfu.ca, fmarbout@sfu.ca, yza174@sfu.ca

Abstract: This paper advances a methodology to support a coordinated multi-level analysis of individual actions and group dynamics in asynchronous online discussions. The approach uses log-file data to examine group and individual participation patterns, and argumentation coding of post contents to probe developmental trajectories of individual and collective understandings. Importantly, these traces of ideas and behaviors are coordinated within and across levels. To illustrate the method, the paper presents an analysis of five undergraduate students taking part in a five-day online discussion to address a business challenge. Findings provided insight into the ways in which phenomena at the group and individual levels interrelated and drove each other, specifically the complex interdependency between group and individual willingness to engage in debate, and how one dominant individual’s unfounded presumption of consensus led to early abandonment (rather than conscious rebuttal) of ideas.

Introduction
Asynchronous text-based discussion forums are often employed to facilitate online collaboration among students. Temporal flexibility gives students time to reflect on their ideas and those of others, thus theoretically affording opportunities for the group to collectively construct knowledge and for individuals to develop their personal understanding (Stahl, 2005). However, in practice, studies have found that learners often exhibit low cognitive engagement (Hew et al., 2010) with shallow and unfocused argumentation (Nussbaum, 2005). The problems of superficial discussion and weak argumentation can be considered on both the individual level (students don’t engage with or challenge each other’s ideas) and that of the group (teams engage in consensus building without considering multiple ideas first). These levels are inherently interconnected as the group’s process emerges from the accumulation of individual actions while also situating and constraining them (Suthers & Teplov, 2011). Similarly, efforts to support productive dialogue in online discussions may be targeted at one of these levels, but will necessarily have interdependent effects on both. For example, requiring a discussion group to come to a consensus (group-level intervention) will affect the ways individuals contribute to the conversation – conversely giving students specific responsibilities in a discussion (individual-level intervention) will impact their group dynamics. For this reason, understanding how a pedagogical intervention affects participation and argumentation processes in online discussions requires a coordinated examination of individual- and group-level phenomena as they evolve over time.

Attention to the multiple levels at which collaborative learning can be studied is growing (Stahl & Hesse, 2008) as is the use of statistical techniques that account for both individual and group influences on collaboration (Cress, 2008). However little work has attempted to examine how phenomena at these levels mutually depend on and affect each other. In addition, recent research in collaborative learning has emphasized the importance of studying the flow of student interactions over time (Reimann, 2009). This paper advances a methodology to support a coordinated multi-level temporal analysis of individual actions and group dynamics in online discussions. To illustrate the method, we present an analysis of a group of five students taking part in a five-day online discussion to reach consensus on how to address an organizational business challenge.

A Coordinated Multi-level Microanalytic Approach
To understand how conversation flows in an asynchronous online discussion we track trajectories of participation and ideas across group and individual levels and examine how they interrelate over time. The detailed traces of student activity created in online discussions present a rich source of data for such examination. Specifically, log-file data of actions (e.g. how students make posts and access those of others) facilitate analysis of group and individual participation patterns, while the content of posts supports examination of collective and individual understandings. While previous approaches have studied these data sources in aggregate (e.g. Palmer et al., 2008), we have found that unpacking data to provide an on-the-ground view of particular students and groups provides insight not available from collectivized data. This work draws on our previous research developing different analytic techniques to study how groups and individuals participate and develop understanding over time (Wise & Chiu, 2011; Wise et al., 2012). The important addition here is the attempt to coordinate analyses across levels. Such coordination is challenging for several reasons. First, while a group’s participation is spread over hours and days, individuals experience participation in discrete episodes lasting minutes thus differing timescales must be considered (Lemke, 2000). Second, the artifact of a group’s
complete discussion is not how the living conversation appeared to individual participants at any given point in time. Finally, there is the difficulty of inferring relationships between phenomena that exist at different ontological levels. We address these challenges through three sequential analytical steps.

Log-file data analysis
Log-file data is extracted by participant. Three types of actions are coded: “views” (opening others’ posts), “posts” (creating a post), and “reviews” (revisiting one’s own posts). Durations are calculated by subtracting time between subsequent actions. Views are further categorized as scans or reads using a cut-off maximum reading speed of 6.5 words per second (Hewitt et al., 2007). Adjustments are made for unlogged system exits and actions are grouped into sessions (a series of consecutive actions). To trace the group’s timeline of participation, we visualize the log-file data to show patterns of group participation over threads and time (Wise & Padmanabhan, 2009; see Fig. 1) an interpretive narrative is then created. Separately, we create a narrative reconstructing the experience for each individual action-by-action using a dynamic discussion map to provide the context of how a discussion appeared at the time a behavior was taken. The combination of these data sources supports a meaningful reconstruction of individuals’ behaviors in the forum (Wise et al., 2012).

Post content analysis
Our examination of post contents draws on work from the field of argumentation. A coding scheme adapted from Clark and Sampson (2008) is used to evaluate when students propose, support and rebut ideas using the post as the unit of analysis (multiple codes can be applied to each post). The scheme also captures if students analyze the problem and if they attempt to synthesize the group’s position. With this approach we can look at how the group negotiates ideas to move to a common conclusion, how individuals’ ideas change over time, and how these relate to the ways in which students position themselves with respect to the group’s dialogue. Other schemes and models of knowledge construction could certainly be used within our overall methodology as well. Once posts are coded, we create a collective visual map of positions on each idea over time to trace the group’s idea trajectory (see Fig. 2a). To examine the evolution of individuals’ ideas, a second map for each student is drawn (see Fig. 2b). Both diagrams are interpreted to create initial idea development narratives.

Coordination of data sources within and across levels
While log-file and post content data are initially analyzed separately, the data streams are highly interrelated. For example, an individual’s new idea might be driven by their actions (e.g. reading a certain post) and highly contested ideas might trigger a collective burst of activity. Thus our first coordination step is to integrate the analyses of actions and ideas at each level (individual and group) into coherent narratives. We then use the narratives at each level to inform each other: contextualizing the participation and ideas of each individual within the group’s trajectory, and tracing how patterns of group activity emerged from actions of particular individuals. Thus for a group of N students, the result is N + 1 accounts of a discussion, one from the perspective of each student and one for the group as a whole. While the narratives at each level necessarily refer to each other, they are distinct in the primacy given to the perspective of the group or a specific individual.

Sample Analysis
We present a sample analysis using our multi-level microanalytic methodology with data from a group of five students (Arlene, Ben, Celia, David and Evan) solving a collaborative task that was seeded with divergent opinions but required consensus. Each student’s trajectory of participation and ideas was traced along an individual timeline, as was the collective process of participation and argumentation for the group as a whole. These parallel narratives were then compared and coordinated to examine the ways in which processes and phenomena at the individual and group levels interrelated and drove each other.

Learning Context and Participants
Data was taken from a larger study comparing the impact of different task-types on online discussion in a blended undergraduate business course. Students were randomly assigned to small groups, each responsible for discussing one business challenge online (Friday to Tuesday) and leading a subsequent tutorial discussion. The challenge in the week studied was communication. A junior executive was told that her response to criticism of her presentation by senior executives was too forceful. Two possible solutions were given for discussion: either email the senior executives to explain why she spoke up or accept the feedback and adjust her future behavior. Additional solutions could also be proposed. Students were asked to come to consensus on a solution with a rationale using theories from class. Posting guidelines and criteria (worth 5% of course grade) were provided.

Process and Findings
Log-file data from the chosen discussion was extracted and processed as described above. Contents of all posts were coded by two researchers; agreement was 89% or higher for all categories except problem analysis (73%).
Group and individual timelines of ideas and participation were visualized in diagrams. Due to space limitations in a short paper, in this presentation we focus on the participation of three students (Arlene, Ben, Celia) over two key days as an illustration of the insight that the multi-level microanalytic approach can provide.

Figure 1. Time-Event Diagram. Left side shows all group posts in threaded structure distributed over time on the vertical axis. Letters are student initials, P professor. Right side shows each individual’s reading actions by day. Column height indicates the total time spent reading. Labels are of sessions, of posts read.

The Group

The group’s discussion focused primarily on three solutions, two provided in the task (accept the feedback and email the executives) and a variation on the latter of these (talk to the executives). The participation peak for the discussion (greatest number of posting and reading actions) occurred on Sunday (see Fig. 1) however argumentation in the discussion was greatest one day prior in a back-and-forth sequence between Ben and Arlene debating the merits of the three solutions. All three other students also logged-in on this day and read the debate posts, but none joined in the debate. This appears to have been a pivotal moment for the group’s discussion; the choice of other students to not join the debate while it was “hot” rendered what could have been a central locus of argumentation peripheral in the discussion. On Sunday many posts were made in support of the three solutions, but none rebutted the opposing alternatives. That evening, Arlene made a post presuming the group’s consensus was Talk. This was a second pivotal moment for the group as all subsequent comments unquestioningly accepted this as the group’s solution and focused on how to implement it or how to present it in the class tutorial. None questioned its status as the solution. Thus the group’s argumentation was effectively completed on Sunday evening without clear rebuttal of all unchosen alternatives.

Figure 2. Timeline of argumentation (a) for group by idea (b) for individuals Arlene, Ben, and Celia (A, B, C).

Arlene

On Friday, Arlene initiated the first thread analyzing the problem and listing three possible solutions (Accept, Email, Talk) without indicating a preference. On Saturday, she logged in the forum and spent just one minute to read and respond to a post by Ben supporting Email. Her position in this post seems to have developed in reaction to Ben’s stance; she rebutted his argument by pointing out that the executives could easily misinterpret an email, thus she advocated Talk. That evening, Arlene read and replied back to a new counterargument Ben had made by restating the possibility of miscommunication via Email and her preference for Talk. However, she also suggested that maybe no action was necessary (Accept). On Sunday afternoon, Arlene returned to the
discussion, quickly skimmed a few posts and then posted in a thread analyzing the problem in light of upward-downward communication theory. An hour later, she logged-in again and immediately started a new thread that attempted to connect her preferred solution (Talk) to her problem analysis. She posted “earlier we agreed there are 3 main problems... [but] I just realized that in our solution, we actually only address problem #2...The face-to-face meeting with the senior exec might clear up this misunderstanding.” This post assumed that the rest of the group agreed with her about the problems identified (“we agreed”) though two of these had only been mentioned by her, and about the solution of Talk (“our solution”) though a group consensus had not been explicitly discussed and both support and dissent for Talk had been expressed. Her reasons for presuming consensus are unclear. It may be that she took Ben’s abandonment of the debate as tacit agreement for Talk, or that since she spent very little time reading others’ ideas she was unaware that other solutions still had support. In this post she also shifted away from her earlier support of Accept as a possible solution without explanation.

Ben
Ben first contributed to the discussion on Saturday, reading all the previously made posts and taking a stand for Email for the employee to “clear-up any misunderstandings...and show that she is capable of what she is doing.” That evening he spent almost a half hour reading a rebuttal responded by Arlene and composing his own counterclaim in support of Email stating that Talk was unpractical based on the theory of power distance in organizations. He also rebutted another peer’s post suggesting to Accept the feedback since this would leave a bad impression of the employee. Ben logged in again later that night and read Arlene’s subsequent rebuttal of his defense; however, he did not continue the debate with her. On Sunday, Ben was active in reading others’ posts, but not in making his own. He made only one post, and in it did not advocate for or against any solutions. This might be a result of his minority position in the discussion, a reaction to the general lack of dissent, or his realization (revealed in a later post accepting Talk) that the group should agree on a solution. This final possibility suggests Ben interpreted the need for consensus as a reason not to debate different options.

Celia
Celia made her first post on Saturday evening, after spending considerable time that day and the one before reading the posts of others. Though the debate between Ben and Arlene was in progress, she did not join in, instead starting a new thread analyzing the problem. On Sunday morning, Celia again spent time reading others’ posts and then made another post emphasizing that the group should identify the problem before trying to solve it. After, she responded to the arguments in the Ben-Arlene debate thread by explaining her preference for Accept, but also indicating support for Arlene’s idea of Talk. While she did not support Email, she also did not refute it; instead she expressed understanding about why Ben might take that position. Thus Celia’s response to the debate dynamic was diplomatic, acknowledging the value of each of the ideas involved. Celia did not log in again till Monday night, when she spent the better part of an hour catching up the discussion before writing a short reply accepting the group’s consensus of Talk.

Discussion and Conclusions
The coordination of multiple levels of analysis helps shed light on specific ways in which individual actions influenced the group’s process and how the group collectively set the stage for individual actions. Close to the start of discussion one student took a stand for a solution, prompting another to rebut his idea and initiate a debate dynamic at the group level. However the reluctance of others to join-in, perhaps because they valued reaching consensus over persuading others of their opinions (Nussbaum, 2005), created a collective lack of dissent. Left in the minority, the first student abandoned his argument. The following day the group shifted to initiating many new threads. This created a disjointed discussion, making it possible for one dominant student’s unjustified presumption of consensus to go unchallenged. As a result, the group stopped considering other solutions, a deviation from dialogical argumentation (Clark & Sampson, 2008). While one student seemed to equally influence the group and be influenced by it, another was insistent in driving the group, and the third seemed content to follow the group’s will. This highlights differences in the reciprocity of influence between the group and individuals. Several effects of timing also seemed to mediate: first, comments of individuals who were active early set the stage for individual actions. Close to the end of discussion one student took a stand for a solution, prompting another to rebut his idea and initiate a new thread analyzing the problem. On Sunday morning, Celia

This paper presented a methodology for coordinated multi-level analysis to examine interrelationships between individual actions and group dynamics in online discussions. The methodology is still in development testing and revision with additional data is needed. While the approach appears fruitful for generating insight into relationships between individuals’ trajectories of behavior and understanding, and groups’ patterns of dialogue and idea-development, its fine-grained, interpretive and time-intensive nature raises important questions about intersubjectivity, generalizability and scalability. Here, we present our current thinking around these issues. First, with respect to intersubjectivity, our initial participation narratives and diagrams are based on log-file data and constructed objectively as neutral, straight-forward reports of the temporal sequence of actions.

© ISLS
in which an individual engaged or in which posts were contributed by the group. Content analysis of posts involves researcher subjectivity. We address this through the well-established practices of using a clear and theoretically-grounded coding scheme, choosing an unambiguous unit of analysis, and reporting inter-rater reliability (De Wever et al., 2006). The major interpretive steps occur when the initial idea and participation analyses for an individual or the group are integrated into coherent narratives and when these interpretive narratives are used to contextualize and inform each other. Making inferences about how different kinds of entities (ideas and behaviors) affect each other and relationships between phenomena that exist at different ontological levels inherently engages the researcher as meaning-maker and thus necessarily involves attention to intersubjectivity. For this reason, the interpretive narratives are repeatedly negotiated by multiple researchers throughout their creation, review and revision.

Nonetheless, as with any fine-grained analysis that focuses on a small number of participants, the goal is not to make statistical generalizations or suggest that what was found in one particular case will likely be found in another. Instead, the objective is theoretical generalization (Seale, 1999): that is to use the deep examination of a particular occurrence to develop theoretical understandings that can present a useful lens for interpreting other situations and suggest implications for educational practice. For example, this sample analysis highlighted the importance of how and why ideas stop being considered, the notion of reciprocity between group and individual levels and the importance of temporal sequence and pace. Finally there is the question of scalability. The multi-level microanalytic approach involves multiple steps of technical data processing, content analysis, and interpretation. The time to conduct the complete analysis of this one discussion by five students over five days is estimated at 150 hours. Thus, use on a large scale is prohibitively expensive in terms of man-hours. Instead, this methodology might be most effectively used in combination with existing temporally-sensitive large-scale methods (e.g. Chiu & Khoo, 2005) which can identify important moments or phases in collaboration worthy of the investment in in-depth multi-level microanalysis.

References


Acknowledgments
This work was supported by the Social Sciences and Humanities Council of Canada.
Scaffolding a Knowledge Community for High School Physics

Mike Tissenbaum, James D. Slotta, University of Toronto, 252 Bloor St. W., Toronto, Canada
(miketissenbaum@gmail.com, jslotta@gmail.com)

Abstract: This paper presents a design study of a collective inquiry model for high school Physics, where student-contributions are captured, aggregated, tagged and represented in a coherent visualization. We have developed a flexible technology layer that supports the aggregation of peer responses, including the collection of student explanations and semantic tags. We investigate collaborative inquiry scripts and discuss how they must comprise both the (macro) scripts that are enacted over a long-term curriculum, and the (micro) scripts that are enacted in class. We outline our rationale for inquiry design in Physics, the role scripting and orchestration play in the successful implementation of this curriculum, the role of the “smart classroom” in their enactment, and three successive iterations of our curriculum.

Knowledge Communities for 21st Century Learning

As we move further into the “knowledge age” today’s workplace is shaped by new technologies, where activities are increasingly data-driven, collaborative, and predicated on a set of fundamental skills commonly referred to as information literacies (Livingstone, 2008). This shift is particularly pronounced across STEM disciplines, where workplace practices are shifting towards large, data-intensive, multidisciplinary collaborations across ever-widening spatial and temporal scales (e.g., the Human Genome project).

A theoretical perspective from the learning sciences that is well suited to learning and instruction in the knowledge age is that of knowledge communities, as exemplified by the Fostering Communities of Learners (FCL) project (Brown & Campione, 1996), and Knowledge Building (Scardamalia & Bereiter, 1996), amongst others. These researchers have advanced an epistemological perspective where students consider learning as a social process, and value the collective knowledge of their peers. Although difficult to enact (Sherin et al, 2004), the knowledge community approach has garnered renewed attention, partly as a result of Web 2.0 capabilities, which can support complex pedagogical constructs (Slotta & Najafi, 2010).

Using “Web 2.0” technologies, students are active participants in a knowledge community that is engaged in the production, aggregation, and assessment of science topics, with an emphasis on inquiry and collaboration (Peters & Slotta, 2010). The socially-oriented process of adding of meta-data (tags) to user-contributed content can provide opportunities to reveal meaningful connections and flexible real-time representations (Mathes, 2004). However, questions still remain about how such collections of content can best serve student learning and foster knowledge communities. Reflection and scripting, are two important aspects of such research and will be discussed in the present paper

Reflection, Discourse, Scripting and Orchestration in Learning Activities

An important dynamic within most inquiry or knowledge community research is that of reflection, which is typically embedded within student learning activities (Slotta & Linn, 2009). While generally accepted as an essential part of the learning process, reflection takes on particular significance in digitally-mediated learning environments (Johnson & Aragon, 2003), where many interactions take place asynchronously. Students are provided with increased opportunities to reflect on their own understanding, think critically about their peers’ ideas, construct coherent arguments, and reconcile misconceptions, before adding to the public discourse (Garrison, 2003; Chi, 2000).

Another topic of interest to learning scientists is the notion of scripting and orchestration (Dillenbourg, Jarvela & Fischer, 2009), where specified learning and interaction designs (i.e., “the script”) are enacted (“the orchestration”) by teachers and students. The script can be seen as a formalism that captures the pedagogical structure of a learning design, including a wide range of interaction patterns among students, their peers, and the teacher. When user-contributed materials are introduced, the script becomes more open-ended (Peters & Slotta, 2010), and any inquiry design must be left somewhat “unbounded” to allow for emergent themes, directions or content. While teachers are the primary “orchestrators” of the script, this role is also shared amongst students. In technology-enhanced learning environments teachers receive real-time feedback about student ideas, resulting in opportunities for evidence-based decisions that can influence the script itself (i.e., real-time “course corrections”), and provide opportunities for teacher professional development (Dillenbourg & Jerman, 2007; Lui, Tissenbaum & Slotta, 2011; Slotta & Linn, 2009).

Research Context

In this paper we present a design study of a collective inquiry model, where student-contributions are captured, aggregated, tagged and represented in a coherent visualization in the context of a high school Physics course. We have developed a flexible technology layer that allows the investigation of collaborative inquiry scripts to
support the aggregation of peer responses, including the collection of student explanations and semantic tags. In the sections below, we outline our rationale for inquiry design in Physics, we describe the role scripting and orchestration play in the successful implementation of this curriculum, and we discuss the role of the “smart classroom” in supporting complex interactions in three successive iterations of our curriculum.

The research progressed through three design-based iterative advancements. The first two were formative, providing important information about student collaboration using real-time digital features. In iteration three, we dramatically expanded our designs, moving from single session smart classroom scripts, to a persistent digital layer that supported periodic inquiry and collaboration for the duration of the Physics class, both at home and in the field. We worked closely with the teacher to develop designs, including a repository of user-contributed materials, and social and semantic tags, which facilitated the development of new scripts for teachers and students alike. Our specific research questions are as follows: How can the aggregated products of student inquiry cultivate a knowledge community in high school Physics? What kinds of scripts can best aid students in leveraging user-contributed materials towards creating deeper understandings of Physics? What technology supports can aid teachers in the scripting of curriculum within an emergent knowledge community?

Method
This research employs a design-based method, involving successive cycles of design, enactment, analysis, and redesign within authentic classroom settings (DBRC, 2003). Using a co-design approach (Roschelle, Penuel, & Schectman, 2006) our team of researchers, technologists, and teachers worked together developing technologies, curricular materials, activities, and interaction patterns. The study was set within an urban high school, with all activities occurring as part of students’ regular homework and school activities. All materials and interactions reported in this paper were delivered using SAIL Smart Space (S3)– a technology infrastructure for smart classrooms and knowledge communities (Slotta, Tissebaum & Lui, 2011).

Iteration One – Developing a Cross-Context Physics Problem Solving Activity
The aim of the first design was to investigate how the aggregation and representation of peers’ work generated outside the classroom (i.e., at home) could be leveraged for in-class knowledge building, and how different technologies could aid the teacher in gaining insight into the state of class knowledge in support of different scripting and orchestration moves. Students first individually solved a set of multiple-choice Physics problems as a homework activity. In a follow-up synchronous activity, small groups re-solved these problems using the aggregated responses of their peers from the homework stage. To support the process, we developed a portal allowing the teacher to customize the activity (i.e., the number of questions to be served, and the type of questions presented), a visualization displaying the student-negotiated answers and relationships (tags) between problems, and an aggregated report of students’ homework responses. By viewing the report and visualization before class, the teacher could adjust the upcoming class script based on his perception of the students’ understanding. The teacher could also use both tools during live classroom activities to gain insight into student group work in real-time, or as a post-activity discussion tool.

Design:
Our study consisted of two Physics classes, with twenty students (n=20) in the first trial and sixteen (n=16) in the second trial. During the enactment, the teacher logged into the portal and uploaded five homework questions. Upon receiving an email alert, each student Tagged, Answered, and provided a Rationale (TAR) for his or her answer before the start of the next class (two days later). In advance of the in-class session, the teacher logged into the portal to view the aggregated student work to develop a sense of the class’ understanding of the ideas present in the homework. During the in-class activity, student groups viewed aggregated answers of the whole class, reached consensus, and decided whether or not to re-TAR the question.

Data Sources:
Data were drawn from three sources: 1) All student and group tags, answers, and rationales were captured by the system; 2) Researchers collected in-class activity field notes; 3) A follow-up debriefing with the teacher. The captured student data was examined to reveal changes in the accuracy of responses between students answering individually versus groups, and in their rationales. The field notes provided us with an understanding of how the students were engaging with the curriculum and their peers. Finally, the follow-up with the teacher provided insight into his perceived effectiveness of the added technology scaffolds in meeting curricular goals.

Findings:
Overall, groups faired significantly better at solving problems (97% correct) than individuals at home (80% correct), with t=2.02, df=41, and p<0.05. One problem, for example, had marked improvement with 45% of students answering incorrectly at home, while 100% answered correctly in groups. A potential confound is that the groups were solving problems they had seen individually as homework. However the addition of the whole class’ rationales made it worthwhile to re-engage the students with the same set of problems.
Throughout the in-class activity, groups read their peers’ tags and rationales, and attempted to make sense of differences through discussion. Overall, student groups recorded forty-eight rationales with their answers. A comparison of individual answers versus group answers showed that in 24 cases (55%) the groups’ rationales were unique – not identical or nearly identical to any individual answer (with an intercoder agreement of 83%) – suggesting that students did not simply reiterate individual ideas during the group activity (although, it is possible they ignored those ideas, which is equally problematic). For rationales that were identical or nearly identical to an individual response, it was unclear if this was due simply to re-stating the original idea, or if they really believed the original answer was best. These outcomes suggest the potential benefits of a script that engages students groups around the aggregated individual contributions of their peers.

In the de-briefing, the teacher stated that he found the real-time reports useful in understanding where students were having problems with the content prior to conducting the class. During the first class, the teacher decided to allow the lesson to run without changing the script, preferring to see how students fared in their small groups. However, seeing students struggle on a particular question prompted him to intervene (i.e., adapt his script). Drawing from this insight, he then adapted the class orchestration more readily on the second day to address the issue as it arose, reducing the potential for student frustration. The teacher revealed that although the visualization was useful, the need to refresh the screen made it difficult for him to know what was happening in true “real-time”. Instead, he relied on the large format displays to inform his in-class orchestration.

The results of this preliminary study point to the efficacy of certain tools for gaining insight into the state of student knowledge at different points within the script, and in aiding the teacher in making necessary adjustments. This study also highlights the limitations of certain tools within a live activity – particularly the visualization – given its inability to provide information at key moments within the script. Future designs need to better coordinate the flow of information based on teacher needs.

**Iteration Two – Adding student expertise groups and real-time teacher feedback**

The second iteration aimed to improve students’ use of the aggregated work of their peers and to further aid the teacher in orchestrating the script. Building upon the first iteration we wanted to further understand what scripts guided students to in depth reflections, and also the role aggregated information played in forming these reflections. In addition, we designed a teacher report application for a tablet computer that used a colour-coded matrix to display group performance on problem solving in real-time. The teacher could touch the icon of any problem to bring up an individual groups’ TAR response, helping inform his understanding of the group’s progress. We were interested in how this tablet application could provide new opportunities for understanding the state of the class’ knowledge and how this affected the orchestration of class activities.

**Design:**

The teacher uploaded thirty-five homework questions, representing five distinct topic areas. Each student was assigned to one topic area, and received five (of seven in that area) problems for homework. During the smart classroom activity, students were placed in groups of five (one student from each topic area), and given five questions (one from each topic area) that no member had seen before as homework. This complex tracking of prior exposure to problems, and selection of suitable items, made possible by the S3 framework, would have been challenging using the traditional approach. On the first day, students were not provided with aggregated homework content; rather, they relied on group negotiation to solve the problems. During the second day, groups were supplied with their peers’ aggregated answers from both classes. The teacher was also given slightly different conditions: on day one, he only had access to the large-format displays in the smart classroom for information about class activities; on day two, he was provided with the tablet for real-time updates.

**Data Sources:**

Data collection for this iteration was identical to iteration one. Further, student TAR data was examined for changes in the correct responses between students’ answering individually compared to in groups without the aggregated work of their peers (Day 1) and in groups with the aggregated work of their peers (Day 2). Individual student and group rationales were also evaluated using a four-point scale, developed conjunction with the teacher, to evaluate the depth of student understanding.

**Findings:**

Two researchers evaluated all student and group responses using the co-developed scale (intercoder agreement of 83%). Overall, the group on Day 2 (score of 2.0) significantly outscored both the individual students during the homework phase (score of 1.32) (t=4.13, p<0.01, df=51), and the groups from Day 1 (score 1.21) (t=4.19, p<0.01, df=50). In groups, students got more questions correct both days (Day 1=83%, Day 2=84%) versus individually (71%), however both cases were only marginally significant. Taken together, these findings suggest that students in groups perform better than individuals in terms of correct answers, without or without access to the broader class’ ideas, but that access to this information helps in the depth and quality of their reflections.
Similar to the previous iteration, the teacher actively moved throughout the class, interacting with students where necessary. At several points during the activity, the teacher read the rationales being written by the groups (projected on the large format displays) and prompted them to refine their thinking towards focusing on the deeper principles relevant to solving and understanding the problems. As the intervention progressed, the teacher adopted a catch phrase “words more than numbers”, in response to what he observed in the class.

The teacher’s interactions with the tablet elicited surprising results. Initially, he was very engaged with the tablet, clicking on group responses, reading rationales and watching for wrong answers. Eventually however, he abandoned the tablet, stating that it divided his attention and hampered his ability to monitor the class. He noted that although useful for seeing group errors, the information came too late to intervene at critical moments, and he could more effectively monitor the class by watching the large displays. This underscored the importance of thoughtful design, and cultivating a deep understanding of the interaction patterns that are most relevant or helpful within an inquiry script.

**Iteration Three: A persistent, multi-context collaborative inquiry environment**

In our most recent design we sought to add a dimension of student-contributed content, while still allowing the teacher to insert new materials, based on emergent patterns within the script. In PLACE.Web (Physics Learning Across Contexts and Environments) students capture examples of Physics in the world around them (through pictures, videos, or open narratives), which they explain, tag, and upload to a shared social space. Within this knowledge community, peers respond, debate, and vote on ideas, toward gaining consensus about the phenomena being shown. Student work was visualized as a complex interconnected web of social and semantic relations, allowing students to filter the information to match their own interests and learning needs. The visualization also became the focal point of real-time smart classroom activities in which students leveraged the products of the class in the creation of challenge problems. S3 pedagogical agents coordinated these activities and informed the teacher of student knowledge at-a-glance. The tablet application was redesigned as a student application. Tools were also developed to allow the teacher the ability to assess student work, review individual student progress, and to record their own reflections on class understanding to aid scripting decisions.

Building upon the findings of the earlier studies, we designed six activities for the teacher to enact within the curricular script: (1) student-developed phenomena for peer debate, voting, and discourse; (2) teacher-created homework problems; (3) student-created challenge problems for their peers; (4) smart classroom activities drawn from the collective knowledge base; (5) in-class discussion around artifacts within the knowledge base; and (6) group development of shared narratives within the smart classroom. Although initial scripts were co-designed with the teacher, the resulting activities that emerged were flexible, to allow the teacher to adjust their implementation as he saw fit within the context of the class’ emerging knowledge.

**Design:**

This iteration involved two grade 11 physics classes (n=20, n=25), spanning three separate curricular units over a six-moth period: Kinematics; Force and Motion; and Energy, Work, and Power. The units were thematically connected, allowing content to carry over between units. To start, each student was given one concept (tag) from a list of 13. The teacher considered these concepts to be “fundamental” to the understanding of the grade 11 curriculum. Over the six-month period, students were given more concepts to work with. Students focused on their assigned concept when capturing examples for inclusion in the community knowledge base. To assess how the depth of the negotiated discourse of the knowledge community approached expert descriptions, a selection of student-submitted examples were given to graduate physics students (as experts), who were asked to tag (from the list of ‘fundamental concepts’) and describe in words how those concepts were being exhibited.

For each of the units, we altered the script design towards formalizing a set of interactions that (1) fostered knowledge community growth, (2) supported the teacher in altering the scripts to address student needs, and (3) helped students to use the knowledge base for their own individual constructivist learning. Across the three units two scripts were enacted that engaged students in the activities mentioned above. The “collective inquiry cycle,” (CIC) where (1) Students submitted inquiry items to the knowledge base; (2) Collectively (at home or in class) examine and tag peers’ work, adding comments to explanations; (3) Teacher reviews the community knowledge, (4) In-class activity engages students with collective knowledge artifacts chosen by teacher; (5) Students reflect individually. The second, “Revisit, Reexamine, Reflect” (RRR) where: (1) Student revisits previously uploaded examples; (2) Reexamines the example for new insights based on his/her evolving understanding of the curriculum; (3) And submits a reflection on his/her new understandings. In the third unit a culminating smart classroom activity was added where students leveraged the collective knowledge base to collaboratively highlight and correct violations of Physics in Hollywood films. During this activity the S3 agent framework responded students work in real-time, altering group configurations, sending specific and timely content to students and the teacher on their tablets, and to the interactive large-format displays around the room.

**Findings:**
While the third iteration is still in progress, findings are encouraging, as students are submitting content to the community, debating and voting on the work of their peers. Examinations of the discourse taking place suggests an evolving understanding of science content and inquiry processes. Several times the teacher used information from PLACE.Web to inform class activities, including using student-submitted examples as a starting point for discussion. Student responses to questions suggest a deeper level of connection to physics principles.

Conclusion
These studies have begun the process of formalizing a set of scripts that successfully engage students in a knowledge community while providing teachers with tools to adjust the scripts in response to emergent ideas within that community. These scripts must take into account both the longer (macro) scripts that are enacted over a curriculum, and consider how they can support (and be supported by) in-class (micro) scripts. Further we have begun to formalize an understanding of the informational needs of the teacher executing these scripts, and the role technology can play in both helping, and hindering, this execution. We are developing an understanding of the important role that a smart classroom infrastructure plays in supporting the orchestration and coordination of real-time knowledge building activities in ways that were simply not possible with traditional pen and paper approaches. These affordances allow students to coordinate their information seeking practices with a group or class towards a common goal. The results from the three studies also promote the role of community voting, debate, and individual reflection around user-created artifacts, as an effective means for developing deeper understandings of the curriculum. Visualizations of this work can also provide powerful means for filtering, sense-making and the re-application of aggregated student work in structured knowledge building activities.

References
Examining system dynamics models together: Using variation theory to identify learning opportunities in online collaboration

Anindito Aditomo, CoCo Research Centre, The University of Sydney, Australia, and The University of Surabaya, Indonesia
Email: aadi4954@uni.sydney.edu.au

Kate Thompson and Peter Reimann, CoCo Research Centre, The University of Sydney, Australia
Email: kate.thompson@sydney.edu.au, peter.reimann@sydney.edu.au

Abstract: This study applies variation theory to examine the discourse of three online groups learning about the structure and behavior of simple population models. Students’ discussions were examined to reveal whether the discussions provided opportunities to become aware of possible variations corresponding to those critical dimensions. The findings indicated that only in Group 1 did the discussion open variations corresponding to critical dimensions of the model structure and growth patterns. We propose that variation theory is useful as an analytic lens for researchers interested in collaborative learning particularly in online environments, and also as an instructional design tool for teachers in designing collaborative tasks.

The goal of this paper is to examine how to identify what the students may have learned (about a given topic) from their online collaboration. The process of online collaborative learning can be analyzed from many perspectives. With regards to making claims about learning from group interaction data, one common approach is to see the collaborative discourse as reflecting cognitive processes, some of which are more germane for learning than others (Hara, Bonk, & Angeli, 2000; Schellens & Valcke, 2006). Another approach to making claims about learning from collaboration discourse is informed by a more socio-cultural view of learning as participating in social practices. This approach is more concerned with the qualities of the discourse at the group level, rather than with learning at the individual level (Kennedy-Clark & Thompson, 2011; Stahl, 2006).

These two approaches to analyzing online collaboration discourse have their own limitations. The cognitive approach allows inferences about individual students’ relative amounts of learning (or potentially germane cognitive processes). However, it provides little insight into what or how students might have learned about the topic. The second approach is useful to highlight group-level processes and properties, but it shies away from making inferences about individual students’ possible learning outcomes. We propose that variation theory (Marton & Pang, 2006; Marton, Runesson, & Tsui, 2004) offers an approach to analyzing collaboration discourse that complements the two approaches above.

Variation theory has been successfully applied to analyze classroom discourse (Ling, Chik, & Pang, 2006; Pang & Marton, 2003; Runesson, 2005), but has not been widely used to examine collaborative learning, particularly in an online learning environment (for exceptions, see Aditomo & Reimann, 2009; Booth & Hulten, 2003). The theory is rooted in phenomenography (Marton, 1981, 1992), which asserts that awareness always has an object, and hence knowing or learning is always knowing and learning about something. Phenomenographic studies have shown that objects (both material and conceptual) can be understood, experienced, or conceived in several qualitatively different ways (Marton & Booth, 1997). Learning, in this view, is becoming able to see, understand, or experience an object in a new and more powerful way. The central tenet of variation theory is that to discern an aspect of something, a person needs to experience variation corresponding to that aspect (Marton & Pang, 2006). It is the awareness of possible variation in certain aspects of an object that enables one to discern those aspects. In this study, we apply variation theory to examine the discourse of three groups learning about system dynamics modeling.

Method

Participants
Data were obtained using a tool called Snooker (Ullman, Peters, & Reimann, 2005) from a postgraduate class. The students worked in groups: Group 1 consisted of two female students and one male student, Group 2 two males and one female, and Group 3 two males and one female.

Topic and task
Students were learning about the structure and behavior of complex systems. Features that distinguish complex systems include emergence, feedback, time-delays, and non-linear cause and effect relationships (for further explanation of these features, see (Sterman, 2000)). The topic addressed in this session was a population model that included the carrying capacity and the model’s behavior (which was an S-shaped growth pattern). Upon
completing the task, students were expected to understand the relationship between the structure of a system and its behavior: that a model with a carrying capacity produces an S-shaped growth pattern. Implicit, but crucial, in this outcome is the understanding that carrying capacity is an emergent property, one which arises out of the interaction of several components of the model: the birth rate, habitat size, density, and death rate. In the example in Figure 1, the carrying capacity is 200; that is where the population stabilizes. This specific carrying capacity is a result of the specific birth rate, habitat size, density, and death rate values in the model; altering any of these would change the carrying capacity in ways that are difficult to predict.

Rather than a fixed value, the death rate is formulated as a function of density (i.e. death rate increases as the habitat becomes more dense or populated). Hence, in the first phase of the system’s growth, density was low and the death rate was lower than the birth rate. This produced exponential population growth. However, as the population and density rose, the death rate also rose, which slowed the population growth (the system still grew, but not exponentially). When the density reached a certain point (the carrying capacity), the death rate was equal to the birth rate and that is why the population stabilizes. Phrased differently, this type of growth is the result of a non-linear relationship between the system’s positive loop (in this case, the birth cycle) and negative loop (in this case, the death cycle).

Figure 1: The deer population model (left) and the S-shaped population growth (right).

Analysis and findings
From a variation theory perspective, analysis of learning opportunities starts with the identification of critical aspects of the intended object of learning (i.e. the aspects of the topic which students need to discern, to be able to see it in a way intended by the instructor). The intended object of learning was to understand that a system dynamics model that includes a carrying capacity produces S-shaped growth. To achieve this, students needed to discern two broad aspects simultaneously: structural components of the model, and the corresponding behavior of the model. The structural components were: (1) the death rate formula, which depends on density and area (this is the part of the model which embody the system’s carrying capacity), and (2) the relative strengths of the birth and death rates. Aspects of model behavior relevant to this task were: (3) the shape of the population growth (S-shaped vs. exponential), and (4) the phases in an S-shaped growth (increasing in the first phase, and decreasing in the second).

Did the students enact the task in a way which made possible the discernment of critical aspects of the topic? (There are issues related to the establishment of a common ground necessary for online communication to occur, but we have analyzed this elsewhere (Reimann, Aditomo, & Thompson, 2009)).

Enactment of Problem 1
“This model includes a carrying capacity. What are the implications of this for the behaviour of the model?”
In discussing this problem, Group 1 discussed the meaning of carrying capacity in a qualitative sense, rather than the formal sense represented in the model. For example, Christine started the discussion by saying “the
point of carrying capacity is that the land can only support so much life” (Line 46)1, and “so if there are too many deer, there won’t be enough food for them all and some will die” (Line 47). Judy, also offered her general sense of carrying capacity by saying that “it have a limit of carrying how much dear in the same area” (Line 50), and “if the Habitat Area is wilder, then the dear can live longer” (Line 60). In this group, only Christine related carrying capacity to the model, by mentioning “death rate” and “density” in Line 51, and again mentioning “death rate” in Line 62. The remainder of Group 1’s discussion of Problem 1 (Lines 63 to 76) centered on whether it was better to have a higher or lower carrying capacity.

In Group 2, Problem 1 was discussed in the presence of the lecturer (Lines 47-61). As in Group 1, Group 2 focused on the concept of carrying capacity in a qualitative sense. Geoff (Line 48) asked the lecturer what carrying capacity meant, and the lecturer answered “carrying capacity is the amount that the system can cope with” (Line 53), and provided an analogy with the computer room they often used that has a carrying capacity of X number of students. Geoff indicated he understood, and said that it was “straightforward” (Line 57). Another member, Teresa, also indicated that she understood (Lines 58). After the lecturer left, the students began Problem 2, indicating that they felt Problem 1 to have been adequately addressed.

In Group 3, discussion of Problem 1 began with each member offering their interpretation of the problem. Luke asked, “Are we supposed to explain the model?” (Line 31). Anne, responded “the first question means how does the carrying capacity of the area affect the model, i think” (Line 32). To this, Paul responded with a qualitative definition of carrying capacity: “In farming terms we have a carry capacity of so many sheep per acre” (Line 34). Anne also offered her definition: “yes, so there must be a limit to the number of deers” (Line 35). Subsequently, only Luke referred to the model itself, or what it meant for a model to have a carrying capacity. Luke described the relation between density and death: “if density is over the carrying capacity, then death rate will rise” (Line 36), “which in turn lowers density” (Line 47). Anne added that “in addition to increasing death rate, carrying capacity also decreases birth rate” (Line 45), which indicates a potential misunderstanding because in the model, the concept of carrying capacity was embodied in the variable death rate (it is a function of density and habitat size).

Enactment of Problem 2
“Change the birth rate and death rate in order to find a combination that will result in a decline in the deer population despite unlimited habitat.”

Group 1’s discussion started when Frank said that it fixes the death rate, and hence they can only manipulate the birth rate (Lines 79 to 81). Frank’s statement indicated he was examining the Stella model, unlike Judy, who was using an online simulation. Judy responded by saying she had already obtained a correct combination: “just put 3% for birth rate and 50% for death rate” (Line 82). The online simulation, which was introduced in a previous task by the lecturer, allowed users to manipulate the death and birth rate in percentage forms (not decimals, as in the model). Furthermore, the simulation did not represent a model with a carrying capacity. This caused obvious confusion, as Christine and Frank tried to understand what Judy meant by “3%”. The group began Problem 3 without resolving this communication problem. Hence, for Christine and Frank, but not Judy, addressing Problem 2 enabled the discernment of the difference between constant converters (the birth rate) and graph function converters (the death rate).

Group 2’s discussion started with Geoff’s comment “wouldnt we just have to increase the death rate above the birth rate to have a declining population?” (Line 93), which was approved by the other two members (Lines 94 and 95). Geoff continued: “because it doesnt really matter how big the area is...if births are less than deaths...or the ratio is leaning that way...its pretty logical” (Line 97). Here, Geoff made sense of Problem 2 in a qualitative way; population size is simply a function of the relative sizes of the birth and death rates. This is qualitatively correct, but it trivialized the problem and diverted attention away from the components of the model that formally represented the system’s carrying capacity (i.e. the area, density, and death rate converters). Without attending to these components, students would not have the opportunity to discern that the death rate could be represented in different ways (as a fixed value or as a function of density). Furthermore, without examining the density and death rate components, it would be difficult to guess the birth rate that would decrease the population, as Nathan discovered. Approximately 30 seconds later, Nathan said “but my numbers dont seem to change the graph” (Line 98). This indicates that Nathan was examining the model. This prompted Geoff to think about how density affects the system, as reflected in his next utterance, 40 seconds later: “a higher deer density can mean both less food and more protection in numbers from hunters etc” (Line 99). This move could have directed the group’s attention to how carrying capacity is represented in the model, however the group decided to end their discussion due to time considerations.

In Group 3, discussion of Problem 2 started with Anne expressing her puzzlement about the problem’s suspicious simplicity: “i dont understand the second question? because whenever the birth rate is smaller than death rate, it will decline. there must be a catch that i dont get ... despite unlimited habitat” (Lines 52 and 53).

1 Due to space limitations, the data to which this line number refers are not included in the paper, however the line numbers were kept in the paper to allow the reader to have a sense of the order in which statements were made.
The discussion then focused on what model should be examined. When they returned to Problem 2, Luke reported that he didn’t succeed in lowering the population, despite having specified a very small birth rate (Line 70). Surprised, Anne reported “oh, i just put 25% birth and 30 % death - for example-“ (Line 71). This statement indicated that Anne was examining the online simulation, instead of the Stella model, without a carrying capacity. After Luke elaborated what he was seeing in his model (Line 75; that death rate is determined by density), Anne realized that she needed to examine the model (Line 79), at which point the discussion of Problem 2 ended due to time constraints. Overall, enacting this task enabled Group 3 (or at least the students who examined the model) to discern the model components that represented carrying capacity.

**Enactment of Problem 3**

*Specify the habitat size (choose a specific value for the “Square miles” converter), “What kind of growth does this illustrate? What is the carrying capacity of your habitat?”*

Only Group 1 had enough time to discuss this problem. The group agreed to try 600 as the habitat size. Christine then said that the system produced an S-shaped growth (Line 114). Judy, on the other hand, said that the population would *increase* if the birth rate were more than 10% (Lines 112-113). Christine expressed her confusion at Judy’s prediction (Line 116). This exchange indicates that Judy was still looking at the online simulation, whereas Christine was examining the Stella model. The group ended their discussion without clearly resolving this issue, and without examining the carrying capacity of their system.

**Summary of learning opportunities**

Table 1 summarizes the learning opportunities that were opened in the groups’ enactment of the task. If a student stated a variation of an aspect of the topic, or referred to an outcome or event (e.g. model behavior) that requires the discernment of that aspect, these were taken as evidence of a learning opportunity for that student (but not necessarily for other students in the group). Hence, Table 1 lists the names of each student whose utterance provided evidence for the learning opportunity.

<table>
<thead>
<tr>
<th>Critical aspects of the intended object of learning</th>
<th>Learning opportunities (variation which enabled discernment of each critical aspect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Death rate formula (fixed vs. varying as a function of density and habitat size)</td>
<td>Christine (in Problems 1 &amp; 2), Frank (in Problem 2) - Luke (in Problems 1 and 2)</td>
</tr>
<tr>
<td>b. Relative strength of birth and death cycles</td>
<td>Christine, Frank, and Judy (in Problem 2) Nathan (in Problem 2) - Luke and Anne (in Problem 2)</td>
</tr>
<tr>
<td>c. Population growth (exponential vs. S-shaped pattern)</td>
<td>Christine and Frank (in Problems 2 &amp; 3) - -</td>
</tr>
<tr>
<td>d. Phases in an S-shaped growth (exponential increase vs. decrease)</td>
<td>- - -</td>
</tr>
</tbody>
</table>

**Discussion and Conclusion**

The groups’ enactment of the task did not open the complete set of variations necessary to understand the intended object of learning. In discussing the model, Groups 2 and 3 focused on the model’s structure (aspects (a) and (b) in Table 1), but did not mention any aspect of the model’s behavior (aspects (c) and (d) in Table 1). Only members of Group 1 discussed an aspect of the model’s behavior (exponential and S-shaped growth), but not even this group examined the properties of the growth pattern in more detail.

The activity did enable students to discern the meaning of carrying capacity in a general sense. For students who examined the model, the activity presented an opportunity to discern the model component that represented carrying capacity: the type of death rate (constant vs. function of density). For students who examined the online simulation, the activity presented an opportunity to simultaneously discern the relative magnitude of death and birth rates, as well as the resulting exponential growth pattern. Students did not have the opportunity to simultaneously discern the critical or defining structural feature of population models with and without carrying capacity, along with their associated growth patterns (exponential vs. S-shaped).

The most obvious challenge in this task was that of communication. Establishing adequate common ground is a well-known problem for online chat groups (Reimann, et al., 2009). In this case, the difficulty was
due to one member in both Groups 1 and 3 looking at an online simulation, while other members examined a Stella model. Only the Stella model included carrying capacity. Communication problems do not, however, explain Group 2’s enactment of the task, in which the evidence suggests that members were examining the same model. We propose that another important challenge for students was concerned with the dual meaning of “carrying capacity”. This concept can be understood at a general, commonsense level, as members from all groups expressed in this task. However, the objective of the task was to understand carrying capacity in a formal sense, as it was represented in the model; as a property of the model’s behavior that emerges from the interaction of the model’s structural components. Group 2 almost exclusively focused on their discussion on the general meaning of carrying capacity, and did not inspect the Stella model. Members of Groups 1 and 3 also expressed this confusion, although in these groups, some of the members did inspect the model.

To conclude, using variation theory allowed us to examine online group interactions to make inferences about individual learning opportunities. This gave us useful insights into the design of the task that were implemented in subsequent years. This paper illustrates the utility of variation theory as both an analytic lens and as a conceptual tool for teachers, who often feel that learning theories are too abstract, too complex, while offering few practical guidelines (Yanchar, South, Williams, Allen, & Wilson, 2010). One feature that makes variation theory a potentially practical tool is that it focuses on the subject matter or topic itself, which is something that teachers are familiar with.

References
Effects of Computer-Supported Collaboration Scripts on Domain-Specific and Domain-General Learning Outcomes: A Meta-Analysis

Freydis Vogel, Ingo Kollar, Frank Fischer, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
freydis.vogel@psy.lmu.de, ingo.kollar@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: CSCL scripts are an effective approach to structure collaborative learning processes in a beneficial way and herewith to foster learning outcomes in various domains. As the amount of research on CSCL scripts grows, it is the right time for a meta-analytical integration of the results. Using a random effects model, this meta-analysis integrated effect sizes of 19 comparisons derived from 13 studies about CSCL scripts. The average effects of CSCL scripts on domain-specific as well as domain-general learning outcomes were estimated. Additionally the meta-analysis targeted the role of additional content support as a potential moderator. Results indicate that CSCL scripts have moderate effects on learning outcomes while additional content support positively influences the effectivity of CSCL scripts on domain-specific learning outcomes.

Introduction

Collaboration scripts for computer-supported collaborative learning (CSCL scripts) are used to support learners’ acquisition of domain-specific knowledge and domain-general skills (Kollar, Fischer, & Hesse, 2006). As learners are often not spontaneously able to engage in a deep elaborative dialogue while learning collaboratively (Cohen, 1994), collaboration scripts may be used to guide students through a meaningful and beneficial collaborative learning process which may result in deeper individual learning compared to unstructured collaboration (King, 2007). In CSCL scripts, the guidance usually is implemented by distributing roles and activities among the students as well as by sequencing of activities and role switches (Kobbe, et al. 2007).

Over the last years a variety of CSCL scripts have been developed and also to some extent been analyzed in empirical studies. In these studies the distribution of roles and activities is mostly implemented in an explicit way by means of assigning and introducing roles directly to the learners (e.g. Ertl, Reiserer, & Mandl, 2005; Schellens, Van Keer, De Wever, & Valcke, 2007) and requesting the learners to perform specific activities that are attached to these roles (e.g. Weinberger, Stegmann, & Fischer, 2010). Sometimes, roles are also induced in an implicit way by creating resource interdependence through the distribution of complementary material among the collaborators (e.g. Molinari, Sangin, Dillenbourg, & Nussli, 2009). Further, CSCL scripts may group learning partners in a way that takes advantage of differences in their prerequisites, e.g. their previous knowledge or their individual attitudes, as proposed by Dillenbourg and Hong (2008). When considering the variety of empirical studies, it also becomes obvious that CSCL scripts have been used in a broad range of domains which provide the content of the domain-specific knowledge the learners are supposed to acquire. The domains range from biology (e.g. Hron, Hesse, Reinhard, & Picard, 1997) over philosophy (e.g. Pfister, Mühlpfordt, & Müller, 2003) to psychology (e.g., Ertl, Kopp, & Mandl, 2006).

Beyond domain-specific knowledge, some studies also aim at learners’ acquisition of domain-general skills. In this meta-analysis, domain-general skills are understood as the learners’ capability to handle script-immanent mechanisms, such as the construction of arguments when a CSCL script guides learners through the sequence of argumentation (e.g. Weinberger et al., 2010). Since CSCL scripts typically have learners repeatedly carry script-immanent mechanisms out, we assume that these mechanisms will be automatized and internalized by the learners so that internal scripts (Kollar, Fischer & Slotta, 2007) are gradually developed (Schank, 1999).

In addition, some studies use multi-factorial designs that compare collaboration scripts with other forms of instructional support, more precisely additional content support like content schemes (e.g. Kopp, Ertl, & Mandl, 2006). By definition, CSCL scripts structure interactions on a content-free level and do not give any support regarding the learning content. Yet, learners might also require support to structure the learning content in a meaningful way in order to benefit from the collaborative learning process. Therefore, it is worthwhile to analyze if additional content support can advance the effectiveness of the use of CSCL scripts.

Since research on CSCL scripts has been flourishing over the past few years, the time seems right to provide a statistical meta-analysis that integrates the results of the different studies and analyzes the effects of learning with collaboration scripts in CSCL settings on learners’ domain-specific knowledge and domain-general skills, summarizes the single effects by estimating average effect sizes, examines the impact of additional content support as moderator, and critically discusses the effectiveness of the approach. To provide such a statistical meta-analysis is the main purpose of this paper.
This meta-analysis is concerned with three research questions regarding the effects of collaboration scripts in CSCL settings on learning outcomes that have been reported in previous studies.

R1: What is the mean effect of CSCL scripts on the acquisition of domain-specific knowledge, compared to unstructured CSCL, and to what extent are the integrated effects homogeneous?

CSCL scripts usually aim to induce deeper elaboration of domain-specific learning material and herewith foster the acquisition of domain-specific knowledge. Thus, we hypothesize a positive effect of CSCL scripts on domain-specific knowledge compared to unstructured CSCL. As the realizations of the CSCL scripts differ between the included studies in a broad range, a rather high heterogeneity of the effects of CSCL scripts on domain-specific knowledge is expected.

R2: What is the effect of combining CSCL scripts with additional content support on domain-specific knowledge acquisition?

Collaboration scripts only provide support for collaborative interactions, i.e. they lack content-specific support. The additional use of content support could pre-structure the content of the learning material, so that learners are better capable to examine the material in the way suggested by the script and herewith benefit more from the use of a script regarding their domain-specific knowledge. Therefore, we hypothesize that the use of additional content support maximizes the effect of CSCL scripts on learners’ domain-specific knowledge.

R3: What is the mean effect of the use of CSCL scripts on the acquisition of domain-general skills compared to unstructured CSCL, and to what extent are the integrated effects homogeneous?

As collaboration scripts provide learners with the opportunity to repeatedly practice script-immanent domain-general activities, a positive effect of CSCL scripts on domain-general skills is expected as compared to unstructured CSCL. Also for the effects of CSCL scripts on domain-general skills we expect a rather high degree of heterogeneity.

Method

Criteria for Inclusion

All studies that were included in this meta-analysis met the following requirements:

1. Method: Only empirical studies investigating the effects of at least one experimental condition that involved the use of a collaboration script in a between-subjects design were included in this meta-analysis. In order to distinguish collaboration scripts from other kinds of instructional support, the term “script” had to be used by the author(s).

2. Dependent variable: To be included, at least one dependent variable needed to be a measure of a learning outcome (domain-specific and/or domain-general) with corresponding parameters reported.

3. Domain: The study needed to be conducted in a CSCL context.

4. Source: To reach a high quality standard, only studies that were published in peer-reviewed journals were included.

5. Language: Only studies written in English or German were included.

6. Sample: For each sample of data only the effects of one published study could be included. If more than one study reported findings based on identical data from an identical sample, only the study reporting the most precise data was included.

Sample of Included Studies

To attain a substantial sample of studies, two steps of data collection were conducted. In the first step data were collected by searching for relevant studies in three different bibliographical databases (OvidSP, ISI and ERIC, all with the same search terms, Boolean operators and search properties ((collaborat* OR cooperat*) AND (computer* OR CSCL) AND (script* OR scaffold* OR structur*) AND learn*) within the search fields Title, Abstract, or Keywords). The database search provided an amount of 242 articles. After scanning title and abstract of each article, 122 articles could be identified as being unrelated to CSCL (e.g. articles about scripts in Theater Arts). By reading the remaining articles we filtered 45 articles which were classified as relevant as they were concerned with collaboration scripts in a computer-supported learning setting. In a second step, additional 14 relevant articles were found by scanning the reference lists of the 45 relevant articles.

After inspection of the 59 relevant articles, 46 of them had to be excluded due to one or more criteria for inclusion being violated. Finally, the sample for this meta-analysis includes 13 articles that met all criteria for inclusion. In summary, 1563 learners participated in the studies that are reported in these 13 articles.
Variables

This meta-analysis includes three different variables that were coded from the primary studies. As the focus lies on the effects of collaboration scripts on learning outcomes, each primary study provides at least one measure of domain-specific knowledge or domain-general skills. If more than one measure for the assessment of a variable was given, the most general measure was selected to be included into the meta-analysis in order to achieve the most comparable measures (e.g. if the measure for domain-specific knowledge consisted of one test asking broadly for concepts within the whole domain and a second test asking for knowledge about a very specific topic within the domain, the first test would have been selected). Finally, for each sample that was derived from the primary studies the presence of additional content support was coded. In the following, the single variables are described in more detail.

Domain-specific knowledge: As domain-specific knowledge, all outcomes were used which assessed learners’ domain-specific knowledge in a post-test administered after the intervention. The assessment could have been done in an open or closed format and should aim at the knowledge that students were expected to acquire within the domain the CSCL setting was designed for.

Domain-general skills: As indicators for domain-general skills, all variables were selected that assessed learners’ domain-general skills in a post-test accomplished after the intervention. To count as domain-general skills, the used tests were required to aim at knowledge about mechanisms suggested by the script. For instance, this could be knowledge about aspects of good collaboration or knowledge about the construction of argumentation sequences if the script aimed to structure related activities within the collaborative learning process.

Additional content support: Some studies reported a multi-factorial design which analyzed the effect of collaboration scripts in combination with a content support (e.g. content schemes, content-related scaffolds). In this case, the studies delivered two different sub-samples, which could be integrated into the meta-analysis. For each sub-sample included into the meta-analysis, the presence of additional content support was coded.

Statistical Analysis

As estimation for the single effect sizes of the primary studies, the unbiased estimate as proposed by Hedges (1981), also “called Hedge’s g” (Borenstein, Hedges, Higgins, & Rothstein, 2009, p. 27) was used. To estimate the average effect size \( d \) by integration of the single effects, the random-effects model (Borenstein, et al., 2009, pp. 69ff) was used because the features of the studies and the kind of measures for the variables varied between the studies and a common true effect size for all studies could not be assumed.

Results

Research Question 1: Effects of CSCL Scripts on Domain-Specific Knowledge

To estimate the average effect of CSCL scripts on the acquisition of domain-specific knowledge (compared to unstructured CSCL), 19 single effect sizes derived from the 13 selected primary studies were integrated. The integration yielded a moderate estimated average effect \( d = 0.36; SE = 0.09; CI_{95\%} = [0.21; 0.51]; p < .01, \text{ one-tailed} \). As hypothesized, a positive effect of CSCL scripts on domain-specific knowledge compared to unstructured CSCL occurred. Regarding the analysis of heterogeneity of the true effect sizes of CSCL scripts on domain-specific knowledge, as hypothesized, there was considerable variation between the integrated studies with a moderate proportion of real variation within the observed variance \( Q(df = 18) = 31.05, p < .01, T = 0.24, F^2 = 42\% \). As there is a substantial amount of unexplained variance between the integrated studies, it is worthwhile to investigate which moderators are appropriate to explain this heterogeneity.

Research Question 2: Effects of Combining Content Support with CSCL Scripts

To analyze the impact of additional content support on the effects of CSCL scripts on the acquisition of domain-specific knowledge, two subgroups were formed out of the single effect sizes already reported in RQ 1 representing samples with vs. without additional use of content support. As hypothesized, the estimated average effect of CSCL scripts in the subgroup with additional content support \( d = 0.59; SE = 0.16; CI_{95\%} = [0.32; 0.86]; p < .01, \text{ one-tailed} \) outperformed the estimated effect of CSCL scripts in the subgroup without additional content support \( d = 0.12; SE = 0.16; CI_{95\%} = [-0.15; 0.39]; \text{ n.s.}, \text{ one-tailed} \) significantly with a difference of \( d_{\text{diff}} = 0.47; SE = 0.23; Z_{\text{diff}} = 2.01; p < .05, \text{ one-tailed} \).

Research Question 3: Effects of CSCL Scripts on Domain-General Skills

Effects of CSCL scripts on domain-general skills were reported in four out of the 13 selected primary studies. To estimate the average effect on domain-general skills, a total of seven single effect sizes derived from the four studies were integrated. As hypothesized, the effect of CSCL scripts compared to unstructured CSCL is positive with a large estimated average effect \( d = 1.07; SE = 0.25; CI_{95\%} = [0.66; 1.47]; p < .01, \text{ one-tailed} \). Regarding the analysis of heterogeneity of the true effect sizes of CSCL scripts on domain-general skills, the statistics also
point to a considerable amount of variation between the integrated studies with a moderate to high proportion of real variation within the observed variance ($Q(df = 1) = 15.26, p < .01, T = 0.50, I^2 = 61\%$), as hypothesized.

**Discussion**

Generally, the results of this meta-analysis support our hypotheses that CSCL scripts exert positive effects on specific learning outcomes when compared to unstructured CSCL. Regarding the first research question the results indicate that CSCL scripts on average provide more substantial support for learners’ domain-specific knowledge acquisition than unstructured CSCL. Thus, it can be presumed that CSCL scripts stimulate collaboration processes which are beneficial for domain-specific learning (Illenbourg & Hong, 2008; Ling, 2007). Additionally we found quite a high variance of the true effect sizes between studies. This finding shows that there is still a considerable amount of variance left between the integrated studies that cannot be explained by the CSCL scripts. To investigate which variables might explain the remaining variance and moderate the effects, further analyses like the comparison of subgroups that was addressed by RQ 2 are required. The analysis of the impact of additional content support on the effect of CSCL scripts on domain-specific knowledge revealed a positive effect compared to CSCL scripts that are not combined with content-related support. Therefore, the use of additional content support strengthens the effects of CSCL scripts on learners’ domain-specific knowledge. It may thus be suggested that learning with CSCL scripts can be enhanced by the supplement of additional content support appropriate to the targeted domain.

As hypothesized, the meta-analysis also revealed a large positive average effect of learning with CSCL scripts on learners’ domain-general skills, compared to unstructured CSCL. The repeated practice of script-immanent mechanisms should have led to internalizing and automatizing the script (Schank, 1999). This finding supports the approach of using CSCL scripts to foster the acquisition of skills specified within the script itself. This approach is primarily reasonable when the script triggers skills meant to be internalized (e.g. argumentation). Most studies that reported findings on domain general skills, used CSCL scripts for argumentation (e.g. Collar, et al. 2007; Stegmann, et al. 2007; Einberger, et al. 2010). This limits the external validity of the finding to some extent. It would be desirable if further studies investigating effects of CSCL scripts on domain-general skills would focus also on other learning-relevant processes beyond argumentation.

High variances of the true effect sizes between the integrated studies were found for the effects on domain-specific knowledge and domain-general skills. Therefore, it can be assumed that the different CSCL scripts used in the integrated studies vary in their impact on learning outcomes due to their distinct features. Consequently, further meta-analytical investigation should focus on the identification of features within the CSCL script studies that might moderate the strength of their impact on learning outcomes.

One limitation of this meta-analysis lies in its rather strict criteria for inclusion. Certainly, a less strict set of criteria would lead to a larger amount of studies that could be integrated (e.g., studies investigating instructional interventions similar or equal to CSCL scripts that are not labelled as such; e.g. Repman, 1993; Stri Bos, Hartens, ochens, & Broers, 2007). Also, studies that are concerned with CSCL scripts without analyzing domain-specific learning outcomes (e.g. ever, can ever, Schellens, & Catke, 2010; Rummel, Spada, & Hauser, 2009; Schoonenboom, 2007) could be integrated into an enhanced meta-analysis. In addition, several qualitative studies and case studies could not be integrated into this meta-analysis, although they might help to reach a deeper understanding of how CSCL scripts work (e.g. H. & Arva, 2009). Further research should try to find ways to further integrate research on CSCL scripts. From our perspective, this meta-analysis can serve as a starting point for such efforts.

**References**

References marked with an asterisk (*) indicate that studies were integrated in the meta-analysis.


