In P. Dillenbourg, A. Eurelings, K. Hakkarainen (Eds.) *European Perspectives on Computer-Supported Collaborative Learning, Procedings of the First European Conference on Computer-Supported Collaborative Learning*, Universiteit Maastricht, Maastrict, the Netherlands, March 22-24 2001, pp. 577-584.

Learning by Constructing Collaborative Representations: An Empirical Comparison of Three Alternatives

DANIEL D. SUTHERS, CHRISTOPHER D. HUNDHAUSEN
Laboratory for Interactive Learning Technologies
Department of Information and Computer Sciences
University of Hawai`i at Manoa
Honolulu, HI 96822 USA
{suthers,hundhaus}@hawaii.edu

Abstract: Given the explosive growth in the use of computer media for learning and the wide range of choices available to designers of online learning tools, it is crucial to understand how these design choices may influence learning. This study evaluated the influence of tools for constructing representations of evidential models on collaborative learning processes and outcomes. Pairs of participants worked with one of three representations while investigating complex science and public health problems. Dependent variables included quantity of discourse about evidential relations ("for" and "against") and two learning outcome measures. Significant effects of tools on learning processes were found, although there appears to have been insufficient time for these process differences to influence learning outcomes.

Keywords: collaborative representations, representational guidance

1. Introduction

The importance of social processes to learning, including the potential utility of collaborative learning, is well established (Brown & Campione, 1994; Lave & Wenger, 1991; Slavin, 1990; Scardamalia & Bereiter, 1991; Webb & Palincsar, 1996). Likewise, prior work has shown the importance of representational aids to individual understanding and problem solving (Kotovsky & Simon, 1990; Larkin & Simon, 1987; Novak, 1990; Novick & Hmelo, 1994; Zhang, 1997). Yet few studies have addressed the combination of these factors, namely the role of representational aids in supporting group learning processes. The problem can be approached from one of two directions. Working *from collaborative learning to representations,* we might ask how to design representations that will enable learners to easily record their deliberations (discourse and problem solving processes) and conclusions for subsequent reflection and assessment. Working *from representations to collaborative learning,* we might ask how to design representations that guide and support collaborative learning processes (such as discourse) in a positive way. The study that is reported in this paper takes this second approach.

Our study focused primarily on representations of evidential relationships between hypotheses and empirical information, particularly visual representations accessible to middle school to undergraduate students. Our work has several converging motivations. Prior work by the first author on Belvedere, a networked environment for collaborative construction of "evidence maps" (Suthers, et al. 1997) suggested that the representational bias of tools such as Belvedere may influence students' discussion. At that time, other researchers were using different representations for similar objectives (supporting epistemic reasoning in science). For example, SenseMaker (Bell, 1997) used a container representation (in which data is sorted into theory containers), WebCamile (Guzdial, et al. 1997) and SpeakEasy (Hoadley, et al., 1995) used threaded discussions, and Puntambeker et al. (1997) experimented with matrix representations. However, few systematic comparisons of the effects of representations on collaborative learning had been undertaken. Exceptions include Baker & Lund (1997) and Guzdial (1997). Theoretical inspirations for such a comparison came from Roschelle's (1994) observation that shared representations (animations and simulations in his case) serve to mediate collaborative inquiry; and from Collins & Fergusons' (1993) discussion of representations as "epistemic forms" with associated "epistemic games." Other literature suggests that representational guidance has it origins in *constraints:* limits on expressiveness, and on the sequence in which information can be expressed (Stenning & Oberlander, 1995) and *salience:* how the representation facilitates processing of certain information (Larkin & Simon, 1987).

Our work is based on the premise that representational tools mediate collaborative learning interactions by providing learners with the means to express their emerging knowledge in a persistent medium, inspectable by all participants, where the knowledge then becomes part of the shared context. We hypothesize that representational guidance constrains which knowledge can be expressed in the shared context, and makes some of that knowledge more salient and hence a likely topic of discussion. This representational guidance influences discourse and learning outcomes in ways that can be predicted from the constraints and salience of the notation. A review of related work and theoretical background may be found in Suthers (1999).

We tested two specific hypotheses regarding the effects of three alternative representational environments (text, graph, and matrix) on participants' collaborative discourse and learning outcomes. Our first hypothesis predicted that participants who construct matrices would talk more about evidential relations than participants who construct graphs, and that both of these groups would talk more about evidential relations than participants who construct plain text documents. This prediction was made because the representation of evidential relations is no more salient than anything else in a textual representation; while graphs represent relations with an explicit object (a link) and carry with them the expectation that one construct such links; and matrices prompt for all possible relationships with empty fields. Our second hypothesis predicted that these process differences would lead to significant differences in learning outcomes, with those who construct matrices remembering more data, hypotheses, and evidential relations than those who construct graphs, and those who construct graphs remembering more data, hypotheses, and evidential relations than those who construct plain text documents. This prediction was made because those representations that prompt for increased consideration of evidential relations are in effect prompting students to elaborate on the information being considered. This elaboration in turn should lead to increased memory for the information.

2. Design

We employed a single-factor, between subjects design with three participant groups defined by the software they used: Matrix, Graph, and Text. All three groups were given the identical task of exploring an unsolved science challenge problem—presented as a series of textual web pages—by recording data, hypotheses, and evidential relations as they encountered them. Dependent measures included (a) the percentage of utterances and participant actions in the software focused on evidential relations; (b) ability to recall the data, hypotheses, and evidential relations explored, as measured by a multiple-choice test; and (c) ability to list, in a written essay, the data, hypotheses, and evidential relations that a scientist familiar with the problem would deem important. Pairs of participants were randomly assigned to the three treatment groups. There were no significant differences between the groups' mean grade point average.

2.1 Participants

We recruited 60 students (32 women, 28 men) in self-selected, same-gender pairs, out of introductory biology, chemistry, physics, and computer science courses at the University of Hawai`i. Participants were all under 25 years of age, and had a mean grade point average of 2.99 (on a 4-point scale). All but three participants were native English speakers. (The three non-native speakers were fluent.) Participants were paid a $25 honorarium for their participation.

2.2 Materials

Figures 1, 2, and 3 present software used by pairs of participants in the Matrix, Graph, and Text groups. The left window contains a tool for representing data, hypotheses, and evidential relations. In the Matrix version (Figure 1), the left window contains a spreadsheet-like tool that enables one to type in data items along the left-hand column, and hypotheses along the top row. Clicking on an internal cell of the matrix brings up a pop-up menu with three choices of evidential relations: "+" (i.e., supports), "–" (conflicts), and "?" (unsure or unrelated). Choosing one of these items causes the corresponding symbol to appear in the cell, thus relating the data item and hypothesis in the corresponding row and column.

In the Graph version (Figure 2), the left window contains a tool that enables one to build a graph of nodes (data items and hypotheses) and links (evidential relations). To create a data item node (a pink rectangle), one types the data text into the text field centered above the graph drawing area, clicks on the "Add Data" button, and finally drags and drops the node in the graph drawing area. One creates a hypothesis node (a green oval) in the same way, except that one clicks on the "Add Hypothesis" button. Finally, to create a link, one clicks on the "Add + link," "Add – link," or "Add ? Link" button, and then clicks, in sequence, on the two nodes between which the link is to be positioned.

The left window of the Text version (Figure 3) contains a simple word processor. Text may be formatted in the usual way—by highlighting the text and then clicking on a formatting button (boldface, italics, underline) or choosing one of several fonts from a list menu.

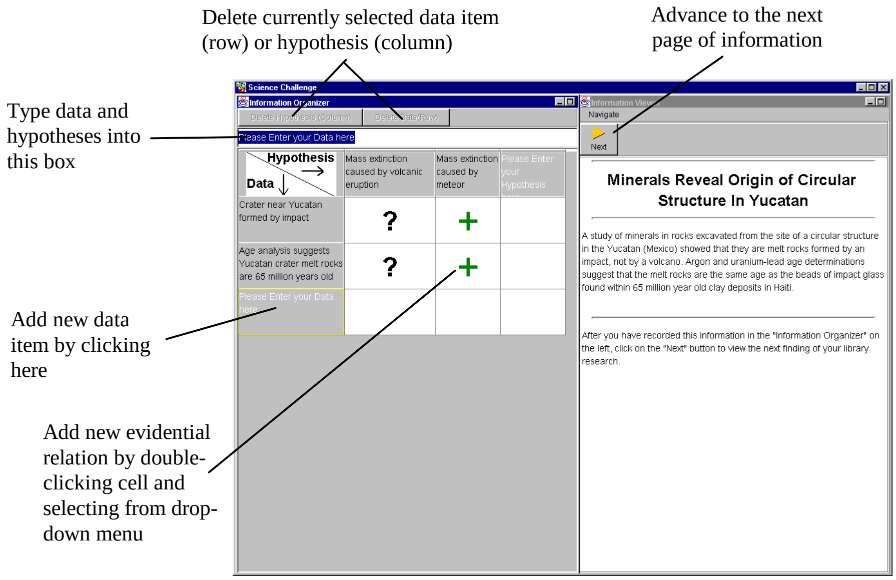


Figure 1, The Matrix version of the software

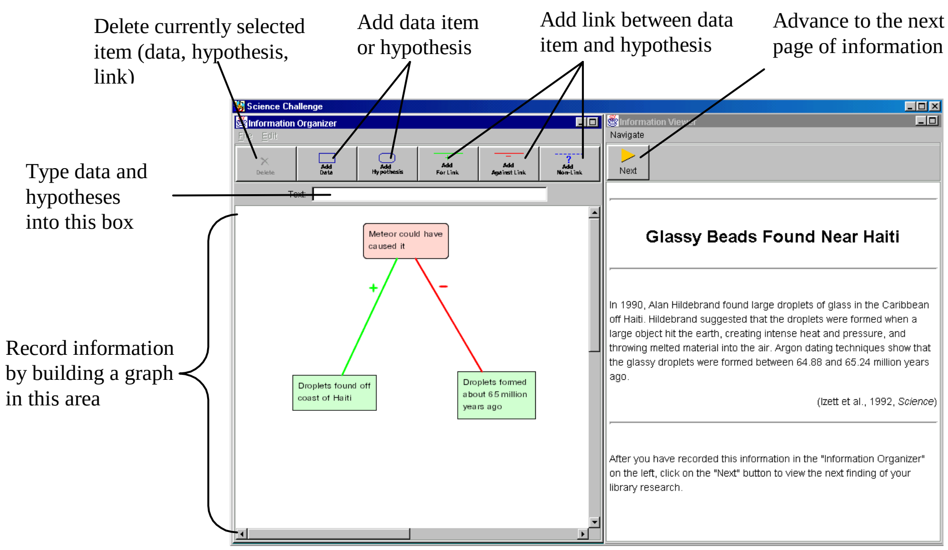
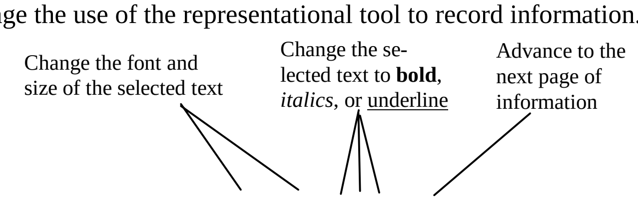


Figure 2. The Graph version of the software

The right-hand window of all three software versions is identical. This window enables one to advance through a series of 15 textual pages. Each of these pages presents a piece of information pertaining to one of two problems: the cause of mass extinctions at the end of the Cretaceous, or the unsolved mystery of ALS-PD, a neurological disease combining symptoms of Parkinsonism and dementia which has an unusually high occurrence on the island of Guam. One clicks on the "next" button to advance to the next page. The software prohibits the user from revisiting previously encountered pages, an experimental design choice intended to encourage the use of the representational tool to record information.



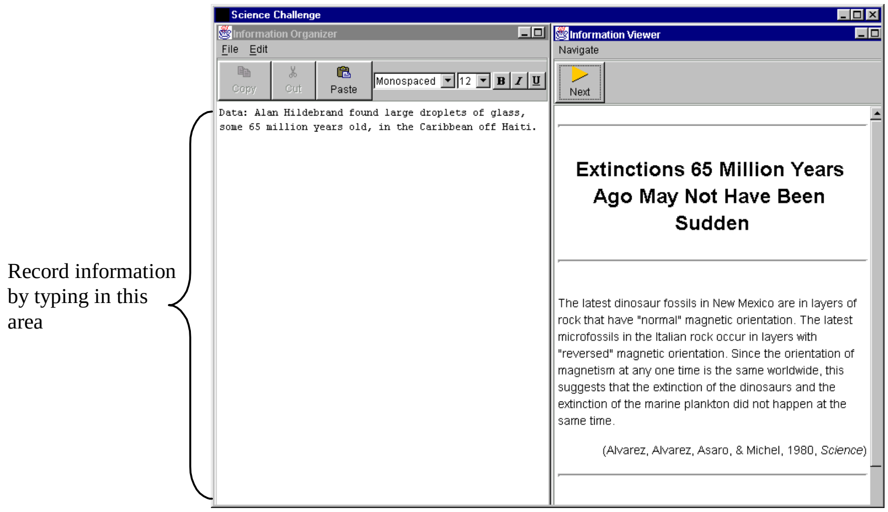


Figure 3. The Text version of the software

2.3 Tasks

Participants were given a "mission statement" explaining that they were to prepare for an imaginary field trip to Guam by studying some background research on the ALS-PD disease, with the ultimate goal of discovering the cause of the disease. They would reach this goal by formulating a set of hypotheses regarding the cause of ALS-PD, and evaluating data for and against those hypotheses. Participants were instructed that the right-hand window would present background research, one page at a time. They were instructed to record the information on each page using the software tool in the left-hand window and that they would not be able to revisit a page once they clicked on the "next" button. Finally, they were advised that, upon recording all of the pages of background research, they would individually take a short, multiple-choice test designed to evaluate their recollection of the information they explored, and then work with their partner to write an essay summarizing the results of their research. The essay instructions asked participants to write (a) a brief paragraph describing each hypothesis they formulated, and summarizing the evidence for and against the hypothesis, and (b) a concluding paragraph that identifies the hypothesis or hypotheses that they believe were best supported by the evidence, and justifies this decision.

2.4 Procedure

After the introduction to the study, we provided participants with a brief (10-minute) introduction to the software they would be using; the experimenter read aloud and performed a demonstration while participants followed along. So that they could become acquainted with the software and the information-recording process, participants then worked on a warm-up science challenge problem (on mass extinction), which was completely unrelated to the main problem. After 15 minutes, participants were instructed to stop work on the warm-up problem, and to move on to the main problem (ALS-PD). Participants were given as much time as they needed to explore all 15 informational pages on the ALS-PD problem. When they reached the page that informed them that there were no more pages left, the experimenter asked them whether they felt they were done. Some participant pairs decided that they wanted to work further; they were given as much additional time as they needed. Once a participant pair declared themselves done, the experimenter instructed them to turn off their computer screen, at which point they were given 20 minutes to complete a multiple-choice post-test, and 30 minutes to complete a collaborative essay.

3. Transforming the raw data into analyzable data

We performed some non-trivial transformations of the data to obtain the dependent measures for process and learning outcomes.

3.1 Transforming process data

We recorded participants' talk in stereo, with the participants' voices recorded in opposite channels. In addition, we recorded two video streams of participants' interaction. A camera positioned behind participants captured their gestures on the screen. A video card output the participants' computer monitor to a second video track. Using a picture-in-picture device, we merged the audio stream with the two video streams such that the behind-the-participants image was inset within the monitor image.

In addition to the video record, we collected automated software logs containing time-stamped records of participants' high-level actions within the software. Using these logs as a starting point, we transcribed the 30 participant sessions. Our transcripts included not only all participant interaction and gestures, but also their actions in the software. Participant utterances were broken up into *segments*, based on the principle that a single proposition or idea should occupy a single segment. Likewise, each high-level action in the software (for example, creating a new data item or hypothesis) was a single segment.

In the next step of our analysis, we performed a content analysis of participants' learning processes by coding all segments in the 30 transcripts into 8 mutually exclusive "topic" categories:

- *Evidential relation*. These segments consider whether data and hypotheses are consistent, that is, whether a data item supports or conflicts with a hypothesis.
- *Epistemic classification*. These segments classify information as either empirical or theoretical—that is, as either data or hypothesis.
- *Metacognitive*. In these segments, participants assess what they know so far—for example, how much they believe an explanation, or what information is needed but lacking.
- *Warrant*. These segments provide justification for an evidential relation previously cited.
- *Tool talk*. These segments discuss some aspect of the software. For example, participants may ask how to complete some specific task with the software, such as adding a data or hypothesis, or they may share their emerging understandings of how the software works.
- *Domain talk*. These segments discuss the domain of the science problem that participants are exploring. Since this is the loosest of the *Topic* categories, it had the lowest precedence: We coded segments into this category only if they could not be coded into one of the five categories above.
- *On-task*—Segments that did not fall into any of the first six categories, but that could still be considered on-task, were placed into this category.
- *Off-task*—Segments that were deemed not to be focused on participants' learning task were coded into this category.

Our current analyses focus on the first category above. In addition, we coded topic segments into four "modifier" categories, according to whether they were

- *verbal* or *representational*—spoken or represented using the software;
- *recited* or *non-recited*—quoted verbatim from the information pages, or not quoted;
- *introduced* or *repeated*—the first occurrence of an idea within a given conversation, or a reintroduction of an idea already brought up within a given conversation;
- *conceptual* or *tool-based*— described in conceptual terms, or described with reference to the software.

Our current analyses utilize the first two modifiers. In order to verify the reliability of our coding system, we had two independent analysts code 20% of the transcripts. With respect to the eight mutually exclusive "topic" categories, our analysts attained 89% overall agreement, and 0.86 kappa. With respect to the four modifier categories, agreement levels ranged from 88% (0.77 kappa) for *introduced* vs. *repeated*, to 100% agreement (0.99 kappa) for *verbal* vs. *representational*. Given these high levels of agreement, we decided that our coding system was sufficiently reliable, and we had a single analyst code the remaining 80% of the transcripts.

3.2 Transforming outcome data

Each of 13 multiple-choice test questions had four possible answers plus "none of the above." We instructed participants that any given question could have more than one answer. To score the tests, we gave one point for each correctly circled answer, and we subtracted one-half point for each incorrect answer, subject to the constraint that it was not possible to score below 0 on any given question.

We began essay scoring by performing an expert analysis of the information by doing essentially the same task participants completed in the study. Each time we visited a page of information, we recorded any new data items and hypotheses in an "expert matrix," and we noted any relationships between the newly created data items and hypotheses and previously created items. Our expert matrix identified 22 different evidential relations in the information trail. For each such evidential relationship, we provided estimates of the following three measures:

- *Evidential strength*—the strength of the evidential relationship, on a scale of 0 to 4, with + indicating a positive (supporting) relationship, and – indicating a negative (conflicting) relationship: 0 = neutral, 1 = apparently relevant because it was mentioned in materials; 2 = weak correlation or expert opinion; 3 = strong correlation or expert opinion, and 4 = demonstration of causality.
- *Inferential difficulty*—the number of information pages that must be accessed in order to infer the relationship, with 0 indicating that the relationship is explicitly stated in the material, and 1 indicating that the relationship can be inferred from a single information page.
- *Inferential spread*—the difference (in pages) between the first and last page needed in order to infer the relationship. This is a measure of how well participants integrate information given at different pages, which should be sensitive to the utility of the representation.

To complete the scoring, we scanned participant essays for evidential relations cited in the text. For each evidential relationship that sufficiently matched one in the expert matrix, we added the corresponding evidential strength, inferential difficulty, and inferential spread values to a running total for the participant pair. These totals are analyzed in the next section.

4. Results

Table 1 presents the mean time that participants in the three conditions took to complete the learning task. On average, the Text group finished over six minutes faster than the Graph group, and over eight minutes faster than the matrix group; however, an analysis of variance found no significant differences between the groups ($df = 2$, $F = 1.38$, $p < 0.27$).

Table 1. Mean time to complete learning task, in minutes and seconds. Standard deviations are in parentheses.

	MATRIX	GRAPH	TEXT
Time to complete task	44:15 (15:19)	46:51 (12:13)	38:04 (7:36)

Since our hypotheses address participant talk and activities dedicated to evidential relations, we present the mean number of segments concerned with evidential relations in Table 2. This measure is given for each of the three conditions, both as raw counts and as percentages of the total on-task, non-recited segments. These counts and percentages are further broken down according to whether they are *representational* (i.e., actions leading to representation changes in the software tool used) or *verbal* (i.e., spoken).

Table 2. Mean evidential relations segments as counts and as percentages of the total, verbal, and representational on-task segments. Standard deviations are in parentheses.

		MATRIX		GRAPH		TEXT	
Total On-Task Segments		510.0 (163.1)		454.5 (232.3)		392.5 (145.8)	
	Verbal	396.2 (132.6)		365.8 (195.0)		305.0 (126.1)	
	Representational	113.8 (61.2)		88.7 (38.8)		87.5 (20.3)	
		Count	**%**	**Count**	**%**	**Count**	**%**
Evidential Segments		139.2 (97.0)	27.3 (13.4)	60.3 (19.1)	15.1 (6.3)	37.7 (27.2)	9.6 (6.3)
	Representational	67.4 (60.7)	59.2 (19.9)	25.6 (10.2)	29.9 (7.3)	15.5 (7.9)	17.7 (11.1)
	Verbal	71.8 (44.9)	18.1 (9.3)	34.7 (12.3)	11.3 (6.1)	22.2 (21.7)	7.3 (4.8)

To test for differences between group percentages, we used a non-parametric Kruskall-Wallis test. Beginning with the evidential relations, we found significant differences with respect to overall percentages of evidential relation segments ($df = 2$, $H = 8.712$, $p < 0.013$), and with respect to the percentages of verbal evidential relations ($df = 2$, $H = 12.56$, $p < 0.0019$). (Statistical comparisons of verbal events were performed on percentages of total verbal segments, not percentages of total segments.) A post-hoc Fischer PLSD test determined that, in both cases, the significant differences were between Matrix and Graph ($p < 0.05$), and between Matrix and Text ($p < 0.05$).

Moving to our learning outcomes results, we present in Table 3 the average scores and percent correct on the post-test, and the average essay scores with respect to the three measures described earlier: evidential strength, inferential difficulty, and inferential spread. An analysis of variance revealed no significant differences between the three groups on the post test ($df = 2$, $F = 0.046$, $p < 0.96$). With respect to the essay scores, analyses of variance revealed no significant differences between the groups' evidential strength scores (*df* = 2, *F* = 0.63, $p < 0.54$), inferential difficulty scores (*df* = 2, *F* = 0.084, $p < 0.92$), inferential spread scores (*df* = 2, *F* = 0.83, $p < 0.45$), and the sum of these three scores (*df* = 2, *F* = 0.74, $p < 0.49$), although trends for all three were in the predicted direction.

Table 3. Mean post-test and essay scores, with standard deviations in parentheses

	Matrix	Graph	Text
Post-test			
Score (out of 22)	15.2 (3.48)	15.18 (3.41)	14.90 (3.53)
% Correct	69.09%	68.98%	67.73%
Essay			
Evidential Strength (out of 51)	25.0 (11.20)	22.5 (11.51)	19.4 (10.80)
Inferential Difficulty (out of 41)	16.0 (7.63)	13.7 (7.29)	12.3 (7.04)
Inferential Spread (out of 81)	38.4 (17.01)	31.3 (17.84)	28.7 (17.42)
Total (out of 173)	79.4 (35.2)	67.5 (36.2)	60.4 (34.5)

5. Discussion

The results confirmed our prediction that the extent to which collaborating learners discuss evidential relationships can be influenced by the extent to which the representational toolkit used by the learners prompts for consideration of evidential relations. This effect was obtained both when considering all propositional acts (whether verbally expressed or by a modification to the representation), and when considering only verbally expressed propositions. These results indicate that further investigation of the ways in which representational features guide learning interactions is merited.

The lack of significance of learning outcomes was disappointing but not surprising. The total amount of time spent working with the tool is less than an hour (15 minutes of warm-up added to the figures in Table 1). This is not enough time for learning outcomes to develop fully. However, the trends in Table 3 are encouraging.

This research resulted in a rich data set, which we continue to analyze. We will examine whether users of a given tool are more likely to refer back to earlier information items during their discourse (similar to the concept of inferential spread in our essay analysis). We will also conduct a qualitative analysis to understand the participants' own semantics for the representations and how they appropriated the representations for their own communications.

We have demonstrated effects of representations on collaborative learning interactions in proximal (co-present) learning situations. Future research in our laboratory will focus on the design of representational guidance for distance and asynchronous discourse; particularly discourse focused on knowledge artifacts being constructed by learners. This work will address the need for empirically validated theories of collaborative representation design, mandated by the explosive growth in the use of computer media for learning.

6. Acknowledgments

We are grateful to Laura Girardeau for running experimental sessions and transcribing and coding the videotapes, Michelene Chi for her guidance and advice on the design and analysis of the experiment, and to Martha Crosby for tomatoes, eggplants, and help with statistical analyses. This work was supported by the National Science Foundation under Grant No. 9873516. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

7. References

Baker, M. & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13:* 175-193.

Bell, P. (1997, December). Using argument representations to make thinking visible for individuals and groups. In *Proc. Computer Supported Collaborative Learning Conference '97,* 10-19. University of Toronto.

Brown, A. L. & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Practice* (pp. 229-270). Cambridge: MIT Press.

Collins, A. & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28*(1): 25-42.

Guzdial, M. (1997, December). Information ecology of collaborations in educational settings: Influence of tool. In *Proc. 2nd International Conference on Computer Supported Collaborative Learning (CSCL'97) (pp. 91-100).* Toronto.

Guzdial, M., Hmelo, C., Hubscher, R., Nagel, K., Newstetter, W., Puntambekar, S., Shabo, A., Turns, J., & Kolodner, J. L. (1997, December). Integrating and guiding collaboration: Lessons learned in computer-supported collaborative learning research at Georgia Tech. In *Proc. 2nd International Conference on Computer Supported Collaborative Learning (CSCL'97)* (pp. 91-100). Toronto.

Hoadley, C. M., Hsi, S., & Berman, B. P. (1995). The multimedia forum kiosk and SpeakEasy. In P. Zellweger (Ed.), *ACM Multimedia '95* (pp. 363-364). San Francisco, CA: ACM Press.

Kotovsky, K. and H. A. Simon (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*: 143-183.

Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science 11*(1)*:* 65-99. 1987.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Novak, J. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*(10): 937-49.

Novick, L. R. & Hmelo, C. E. (1994). Transferring symbolic representations across nonisomorphic problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6): 1296-1321.

Puntambekar, S., Nagel, K., Hübscher, R., Guzdial, M., & Kolodner, J. (1997, December). Intra-group and intergroup: An exploration of learning with complementary collaboration tools. In *Proc. Computer Supported Collaborative Learning Conference '97* (pp. 207-214). University of Toronto.

Roschelle, J. (1994, May). Designing for cognitive communication: Epistemic fidelity or mediating collaborative inquiry? *The Arachnet Electronic Journal of Virtual Culture*.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences, 1*(1): 37-68.

Slavin, R. E. (1990). *Cooperative learning: Theory, research, and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Stenning, K. & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science 19*(1): 97-140. 1995.

Suthers, D. D. (1999, January). Representational support for collaborative inquiry. In *Proceedings of the 32nd Hawai`i International Conference on the System Sciences (HICSS-32*). (CD-ROM). Maui, Hawai`i: Institute of Electrical and Electronics Engineers, Inc. (IEEE). Available: http://lilt.ics.hawaii.edu/lilt/papers/hicss99.pdf.

Suthers, D., Toth, E., and Weiner, A. (1997, December). An integrated approach to implementing collaborative inquiry in the classroom. In *Proceedings of the 2nd International Conference on Computer Supported Collaborative Learning (CSCL'97)* (pp. 272-279). Toronto.

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Education Research,*13:21-40, 1989.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*(2): 179-217.