Assessing Skills in Scientific Inquiry, Argumentation, and Causal Modeling

Paper presented at AERA 2008 in New York, NY

By Allan Jeong (Florida State University)

Abstract

To jointly assess students' performance in scientific inquiry/argumentation and students' ability to construct causal diagrams/theories/beliefs that explain complex phenomenon (and their impact on one another), new software tools are needed to help us: a) code and identify similarities between the diagrams of multiple learners; b) measure the extent to which changes in diagrams of the individual or the collective group progress toward group consensus or target models; and c) determine which and to what extent particular dialog moves sequences observed in group discourse (e.g., present consensus data → propose causal link with supporting evidence → cross-examine presented evidence → counter-argument → amend proposal) trigger targeted changes in students' causal diagrams. This paper presents preliminary findings from two case studies that used a newly developed software tool, jMAP, that enabled: a) students to individually produce and electronically submit their diagrams, download and aggregate the diagrams of multiple learners, and generate aggregated diagrams and matrices to reveal similarities between learners' diagrams, the percentage of diagrams sharing particular causal links, average causal strength assigned to each link, and degree of match between the diagrams of the collective group and the target/expert diagram; and b) the author to identify the dialogic processes of scientific inquiry and argumentation (focused around the jMAP data and recorded in online threaded discussions) that trigger changes in students' causal diagrams over one time period. This paper presents the preliminary findings, evaluations of jMAP, and areas for future research.

Introduction

Each one of us holds many different beliefs and theories about the world. Theories can be conceived and examined in the form of causal diagrams - a network of events (nodes) and their causal relationships (links). Some causal diagrams may be more accurate than others—depending on the presence and/or absence of supporting evidence; and some diagrams and the causal links within the diagrams may be more or less firmly held—depending on both the strength of the supporting evidence and the strength of specific causal relationships. Furthermore, causal diagrams are not fixed and unchanging. Instead, they are incomplete and constantly evolving; may contain errors, misconceptions, and contradictions; may provide simplified explanations of complex phenomena; and may often contain implicit measures of uncertainty about their validity (Ifenthaler & Seel, 2005; Seel, 2003). As a result, causal diagrams can change, but usually not randomly. That is, there are typically events that trigger and provide the impetus for change. Although causal diagrams are being increasingly used to help learners articulate and assess learners' understanding of complex domains and their progress towards increased understanding (Spector & Koszalka, 2004), few studies have examined changes and the pedagogical discourse that trigger changes in the strength of causal links in learners' causal diagrams (Shute, Jeong, & Zapata-Rivera, in press).

As a result, *one goal* of this paper is to present a computer-based learning environment, jMAP, which enables learners to articulate and apply scientific inquiry and argumentation to construct and refine theories to explain complex phenomena (see Figure 1). jMAP enables learners to individually produce and electronically submit causal diagrams, download and aggregate diagrams of all or selected learners to capture the group's collective understanding, generate matrices to compute and report the percentage of students' diagrams that share each causal link (including the average strength of each link observed across all learners' diagrams), and superimpose his/her own causal diagram over the aggregate map to visually identify similarities and differences between the causal diagrams of all learners (Jeong, 2008).

The *second goal* is to demonstrate how jMAP can be used by researchers and possibly teachers to: (a) graphically superimpose selected diagrams (individual, aggregate, expert/target) to highlight changes occurring over time in the causal diagrams of an individual or groups of learners; (b) determine the extent to which the observed changes progress toward a target or collective model; (c) determine precisely where and when changes occur in the causal diagrams; and most importantly, (d) identify which, how, and to what extent specific events (e.g., viewing consensus data, discussing evidence, engaging in specific and argumentation patterns) trigger changes in the causal links and strength of links in learners' causal diagrams (see Figure 2).

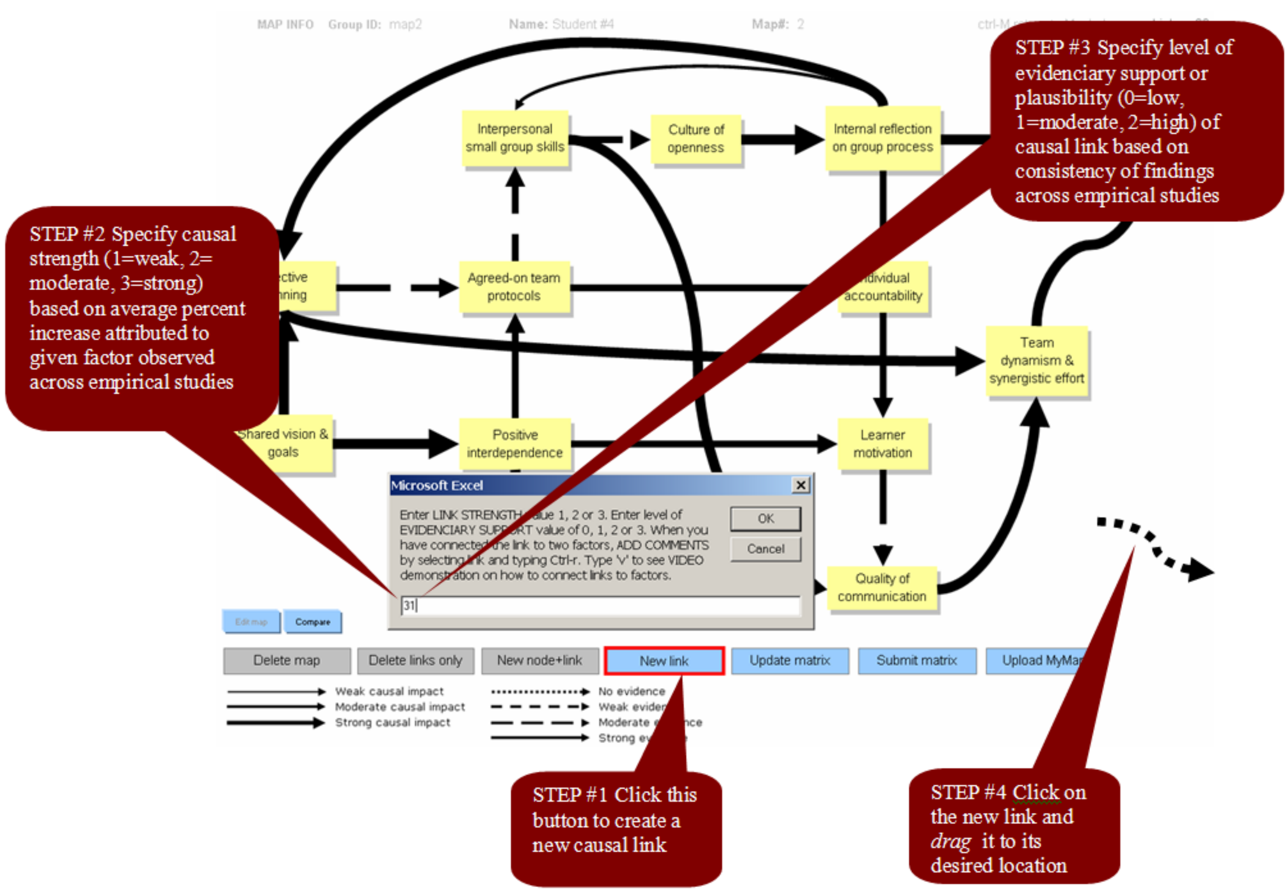


Figure 1. Causal diagram produced with jMAP

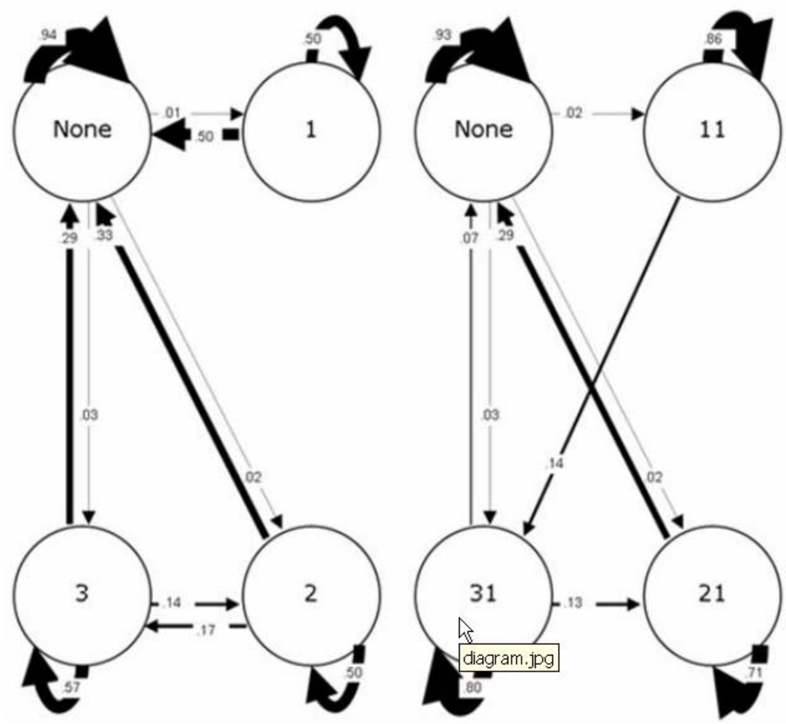


Figure 2. Direction and likelihood of changes in causal strengths when links were presented without vs. with supporting evidence

Figure 1 reveals one of the findings from a recent case study (Shute, Jeong, & Zapata-Rivera, in press) that examined the processes of collaborative theory construction among a group of learner's in an online course on instructional technology. In this study, each student used jMAP to individually construct and articulate a theory (at the beginning, middle, *and* end of the semester) that explains the complex events and conditions (including intermediate events and their causal relationships) that determine when media technology is or is not effective in increasing learning and achievement. In the causal maps, students were required to assign a strength value to each causal link (1 = weak, 2 = moderate, 3 = strong impact) based on personal experiences with collaborative learning and based on empirical findings and theories examined in the course. In addition, students were instructed to specify with each causal link added to the causal diagram the quality of evidence they possessed and/or compiled to justify the plausibility of each given link (0 = no evidence, 1 = weak evidence, 2 = moderate evidence, 3 = strong evidence). In this preliminary case study, the experimenter coded all the maps by hand and causal links into adjacency matrices. Once coded, jMAP was used to tabulate the sequential changes in causal links observed in each student's causal diagrams produced *prior* to and *subsequent* to collaborative work (identifying factors; collecting, annotating, and sharing supporting evidence; cross-examining the evidence; interpreting the evidence; consensus making).

The Discussion Analysis Tool (Jeong, 2005) was then used to sequentially analyze the data to produce the transitional state diagram presented in Figure 2. The state diagram on the left, for example, shows that 50% of all causal links that were assigned a strength value of one remained the same between the first and second, and between the second and third causal diagrams, when *no* evidence was presented nor discussed in the online group discussions to establish the plausibility and the strength of the link. In contrast, the state diagram on the right shows that when evidence was presented with the causal links, these same links with strength value of one were more likely to remain the same (86% instead of 50%). Overall, this preliminary study illustrates how the jMAP environment – when combined with sequential analysis – can produce a potentially powerful method to studying the *processes* of theory construction and the factors and conditions that both support and inhibit the process.

To further explore the potential instructional and research applications of jMAP and to identify areas for future study on theory construction (and the processes of scientific inquiry and argumentation that support theory construction), a second case study was conducted with a more advanced version of jMAP (e.g. automatic coding to adjacency matrices, networked for more frequent map sharing and comparisons) to address the following research questions:

1. Do students tend to add links to their causal diagrams to conform to the majority when students use jMAP to determine the most common and least commonly accepted causal links?

2. How do the reported percent agreements trigger argumentation over the merits of each causal link? Do lower or higher levels of agreement trigger more or less argumentation, respectively?

3. How does argumentation affect changes in percent agreement in subsequent diagrams?

Method

Participants. Nineteen graduate students (8 male, 11 female) enrolled in a Masters level online course on computer-supported collaborative learning at a large southeastern university participated in this study. The participants ranged from 22 to 55 years in age, and the majority of the participants were enrolled in a Master's level program in instructional systems/design.

Procedures. The course examined factors that influence collaborative learning and instructional strategies associated with each factor. In week 2, students used a Wiki webpage to share and construct a running list of factors believed to influence the level of learning or performance achieved in group assignments. Students classified and merged the proposed factors, discussed the merits of each factor, and submitted votes on the factors believed to exert the largest influence on the outcomes of a group assignment. The votes were used to select a final list of 14 factors that students individually organized into causal diagrams.

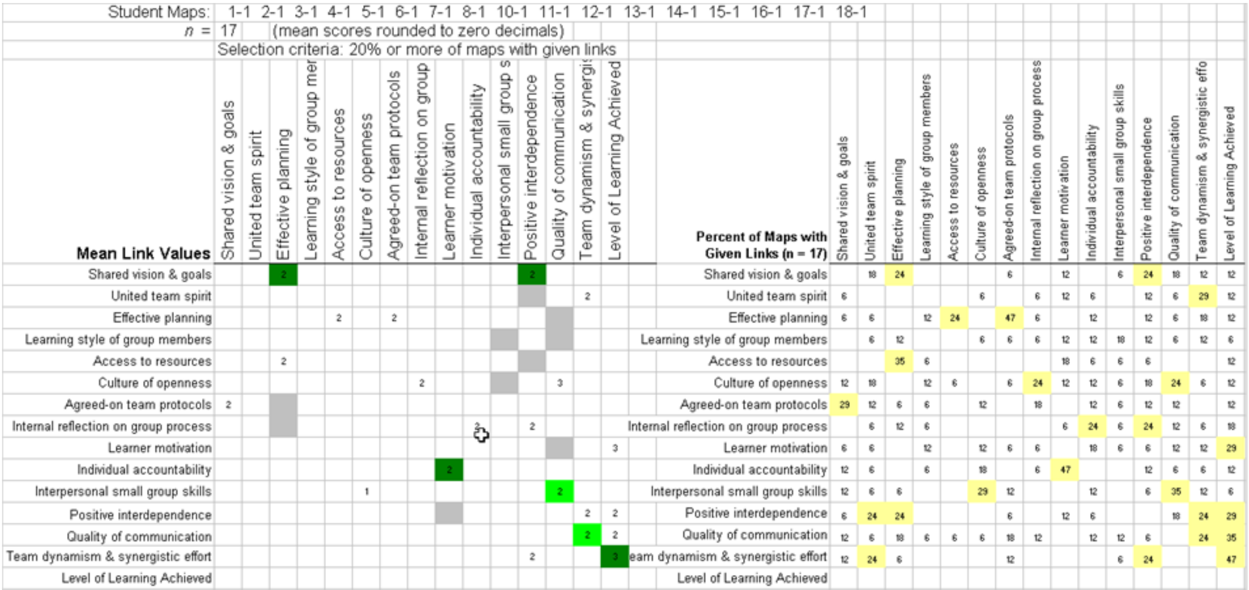
In week 3, students were presented six example diagrams to illustrate the characteristics and functions of causal diagrams. Students were provided a MS Excel-based software program called jMAP (pre-loaded by the instructor with nodes for each of the 14 selected factors) to construct their first causal diagram (map 1). The purpose of map 1 was to graphically explain their understanding of how the selected factors influence learning in collaborative settings. Using the tools in jMAP, students connected the factors with causal links by: (a) creating each link with varying densities to reflect the perceived *strength* of the link (1 = weak, 2 = moderate, 3 = strong); and (b) selecting different types of links to reveal the level of evidentiary support (from past personal experiences) for the link. The map in Figure 2 was constructed by a student in the course who omitted three of the 14 factors from the map because he/she did not believe that the omitted factors affected the learning outcome. Personal diagrams were completed and electronically uploaded within a one-week period to receive class participation points (class participation accounted for 25% of the course grade). The diagrams were also used to complete a written assignment describing one's personal theory of collaborative learning (due week 4, and accounting for 10% of course grade).

Using jMAP, the instructor downloaded and *aggregated* all diagrams (*n* = 17) to produce and share with students a matrix conveying the percentage of diagrams that possessed each causal link. To illustrate (see Figure 3), the causal link between 'Individual Accountability' and 'Learner Motivation' was observed in 47% of students' diagrams.

The links highlighted in *yellow* in the matrix above (on the right) identifies the *common links* observed in 20% or more of the students' diagrams (note: this criterion was specified by the instructor when aggregating diagrams in jMAP). Presented in the left matrix are the mean strength values of only those links observed in 20% or more of the diagrams. The highlighted values reveal links that are present or absent in the expert's map (i.e., ***dark green*** = links and strength values match, ***light green*** = links match, but strength values do not, ***gray*** = missing target links).

In week 9, students were presented the matrix revealing the percentage of diagrams (map 1) that possessed each link. Students posted messages in online threaded discussions to explain the rationale behind each proposed link (Figure 3). Each posted explanation was labeled by students with the tag 'EXPL' in message subject headings. Postings that questioned or challenged explanations were tagged with 'BUT.' Postings that provided additional support were tagged with 'SUPPORT.' In weeks 9 and 10, students searched and reported quantitative findings from empirical research in a Wiki to determine the instructional impact of each factor. Students were received instructions on how to use jMAP to *superimpose* their own map over the group map (figure 4) to visually identify similarities and differences between their own vs. the collective conception of the causal relationships among factors and outcomes.

Figure 3. Scaffolded discussion of each proposed causal link in a threaded discussion forum

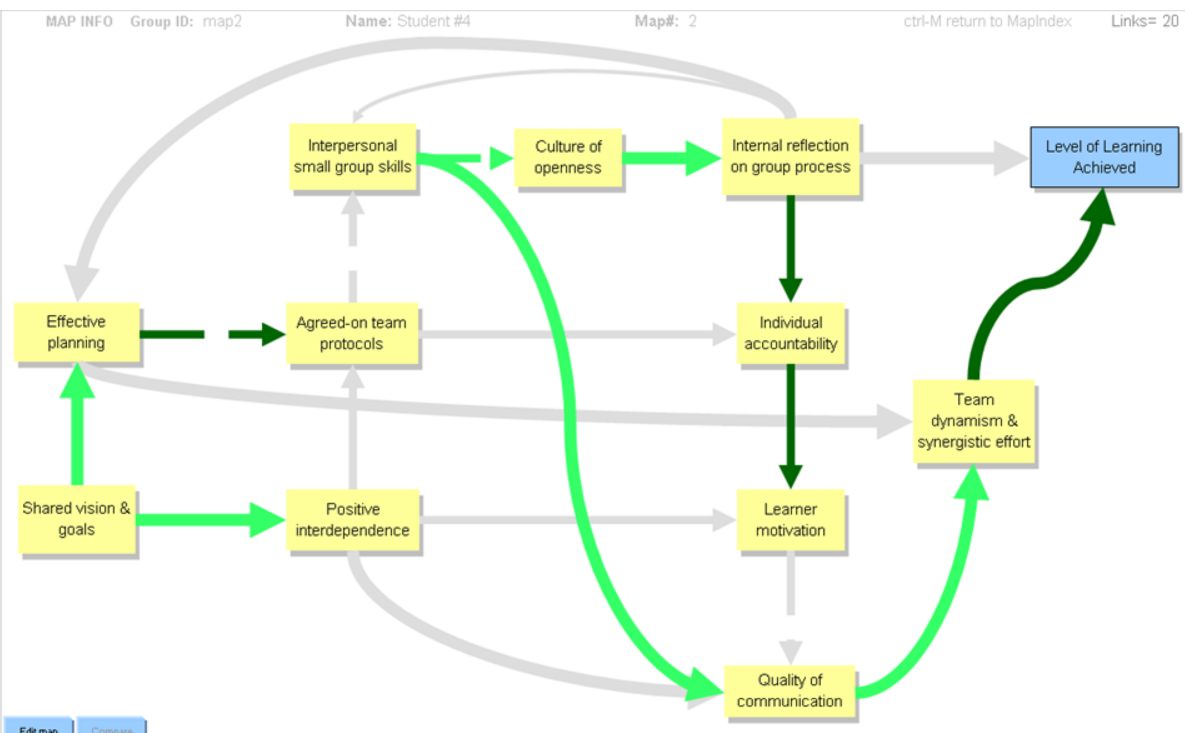


Student Maps: 1-1 2-1 3-1 4-1 5-1 6-1 7-1 8-1 10-1 11-1 12-1 13-1 14-1 15-1 16-1 17-1 18-1
n = 17 (mean scores rounded to zero decimals)
Selection criteria: 20% or more of maps with given links

Mean Link Values	Shared vision & goals	United team spirit	Effective planning	Learning style of group mer	Access to resources	Culture of openness	Agreed-on team protocols	Internal reflection on group	Learner motivation	Individual accountability	Interpersonal small group s	Positive interdependence	Quality of communication	Team dynamism & synergis	Level of Learning Achieved
Shared vision & goals			2									2			
United team spirit														2	
Effective planning					2		2								
Learning style of group members															
Access to resources			2												
Culture of openness								2					3		
Agreed-on team protocols	2														
Internal reflection on group process										2		2			
Learner motivation															3
Individual accountability									2						
Interpersonal small group skills						1							2		
Positive interdependence														2	2
Quality of communication														2	2
Team dynamism & synergistic effort												2			3
Level of Learning Achieved															

Percent of Maps with Given Links (n = 17)	Shared vision & goals	United team spirit	Effective planning	Learning style of group members	Access to resources	Culture of openness	Agreed-on team protocols	Internal reflection on group process	Learner motivation	Individual accountability	Interpersonal small group skills	Positive interdependence	Quality of communication	Team dynamism & synergistic effo	Level of Learning Achieved
Shared vision & goals		18	24				6		12		6	24	18	12	12
United team spirit	6					6		6	12	6		12	6	29	12
Effective planning	6	6		12	24		47	6		12		12	6	18	12
Learning style of group members		6	12			6	6	6	12	12	18	12	6	12	6
Access to resources			35	6					18	6	6	6			12
Culture of openness	12	18		12	6		6	24	12	12	6	18	24	6	12
Agreed-on team protocols	29	12	6	6		12		18		12	6	12	12		12
Internal reflection on group process		6	12	6					6	24	6	24	12	6	18
Learner motivation	6	6		12		12	6	6		18	6	6	12	12	29
Individual accountability	12	6		6		18		6	47			12	6	6	12
Interpersonal small group skills	12	6	6			29	12			12		6	35	12	6
Positive interdependence	6	24	24				6		12	6			18	24	29
Quality of communication	12	6	18	6	6	6	18	12		12	12	6		24	35
Team dynamism & synergistic effort	12	24	6				12				6	24			47
Level of Learning Achieved															

Figure 4. Mean causal link strengths across all maps and percent of maps with given links



Navigational Tools for Visually Comparing Maps			
View my matrix	*ctrl-t*		
View selected matrix	*ctrl-y*	*ctrl-u*	*ctrl-i*
Maps to view & compare with my map	g1	1-1	2-1
ctrl-g & ctrl-l scrolls thru MapIndex	*ctrl-h*	*ctrl-j*	*ctrl-k*
Number of maps listed in MapIndex = 38			

Quantitative Measures	
4	Links with strength values matching my map
6	Links strength values different from my map
10	Total number of shared links with my map
20	Number of links in my map (15 nodes)
50.00%	Percent of shared links with my map

Figure 5. Comparison of a student's revised map superimposed over other individual or aggregated maps

In week 10, students reviewed the discussions from week 9. Within each discussion thread for each examined link, students posted messages to report whether they rejected or accepted the link (along with explanations). At the end of week 10, students posted a revised causal diagram based on their analysis of the arguments presented in class discussions (see Figure 5).

Data Analysis. To measure the level of change in learners' diagrams, link frequencies from map #2 (n = 15) were aggregated into a matrix to determine the percentage of diagrams that shared each link. Differences in the reported percentages between diagrams 1 and 2 were computed (see Figure 6 below). Overall, the percentages in 19 of the 24 commonly shared links (in yellow) increased by an average of 26%. Five of these links (black boxes) decreased by an average of -8.4%.

The level of argumentation produced within each discussion on each link was determined by the counting the number of EXPL-BUT, BUT-BUT, BUT-EXPL or SUPPORT, and BUT-SUPPORT exchanges observed within each discussion. Challenges to explanations, and explanatory responses to challenges were used as a measure of level of argumentation because explanations, when generated in direct response to conflicting viewpoints, have been shown to improve learning (Pressley et al., 1992). Pearson correlations between level of agreement in shared causal links and indicators of argumentation are presented below in Table 1.

Change in percentage of maps sharing links in map 2 from map 1	Shared vision & goals	United team spirit	Effective planning	Learning style of group members	Access to resources	Culture of openness	Agreed-on team protocols	Internal reflection on group process	Learner motivation	Individual accountability	Interpersonal small group skills	Positive interdependence	Quality of communication	Team dynamism & synergistic effo	Level of Learning Achieved
Shared vision & goals		-11	63				-6	7	-5		1	16	-18	-12	-12
United team spirit	-6					-6		-6	-5	-6		-5	1	11	2
Effective planning	1	-6		-12	10		40	-6		8		8	-6	-18	-12
Learning style of group members		-6	-5			-6	-6	-6	-12	-12	-4	-12	-6	-12	-6
Access to resources	7	7	11	-6				7	-11	-6	-6	-6			-12
Culture of openness	-5	-18		-12	-6		-6	-10	-12	-12	1	-18	43	1	-12
Agreed-on team protocols	-16	2	-6	-6		2		2		-5	1	-5	15	7	-12
Internal reflection on group process	7	-6	-5	-6					-6	16	-6	16	-5	-6	9
Learner motivation	-6	1	7	-5		-5	-6	-6		-18	1	-6	-12	2	17
Individual accountability	-12	-6		-6		-11		-6	40			-12	-6	1	-5
Interpersonal small group skills	-5	-6	-6			11	-12			-5		1	51	2	-6
Positive interdependence	-6	30	16				7		-5	1			-18	-4	4
Quality of communication	-12	-6	-18	-6	-6	1	-18	8		-12	-12	-6		70	-2
Team dynamism & synergistic effort	-12	16	-6				-12	7			-6	-10	7		26
Level of Learning Achieved									7					7	

Figure 6. Change in percent of maps sharing selected links

		LevelAgree	CritDisc	%Change	Expl	But	Support	Expl-But	But-Ex/Sup	But-But	Expl-Sup
LevelAgree	*r*	1	.385	-.089	.233	.328	.291	.330	.365	.177	.153
	signif		.063	.679	.272	.118	.168	.115	.079	.409	.476
	n	24	24	24	24	24	24	24	24	24	24
CritDiscourse	*r*	.385	1	-.152	.339	.921	.120	.867	.921	.494	-.135
	signif	.063		.478	.105	.000	.575	.000	.000	.014	.530
	n	24	24	24	24	24	24	24	24	24	24
PercentChange	*r*	-.089	-.152	1	-.058	-.173	.313	-.051	-.167	-.219	.386
	signif	.679	.478		.788	.420	.136	.814	.435	.304	.063
	n	24	24	24	24	24	24	24	24	24	24

Table 1. Correlations between level of agreement in shared causal links and indicators of argumentation

Results

Adding links and conforming to early majority vote. The 24 commonly shared links observed in all initial maps (map #1) were divided into low vs. high percent agreement based on the median percent agreement. No significant differences were found relating to the increases in percent agreement in subsequent diagrams between links with initially high vs. low levels of agreement, t_{22} = .42, p = .68. The average change in link strength among links with low agreement vs. high agreement was 21.33 (n = 12, SD = 26.10) and 17.42 (n = 12, SD = 19.11). This finding suggests that the opinions of the majority did not unduly influence students' decisions to add new links to their diagrams.

Relationship between initial agreement and level of argumentation. Differences approaching statistical significance were found in the level of argumentation produced in the discussion of *map 1* links with low vs. high level of agreement, t_{22} = -1.96, p = .06. The mean number of critical exchanges produced in links with lower vs. higher agreement was .58 (n = 12, SD = 1.44) and 2.17 (n = 12, SD = 2.40), respectively. This finding suggests that providing students with information on the links that are observed most frequently in other students' diagrams can influence the extent to which students critically examine the causal links – particularly the links found in the majority of students' diagrams.

Effects of argumentation on changes in agreement. No significant differences were found in the change in agreement between links discussed at low (no critical exchanges) vs. high (number of critical exchanges > 0) levels of argumentation, t_{22} = .36, p =.73. The average change in percent agreement with low vs. high argumentation was 20.67 (n = 15, SD = 22.93) and 17.22 (n = 9, SD = 22.86), respectively. However, post-hoc analysis revealed that higher frequencies of EXPL-SUPP exchanges were correlated (r = .386, p = .063) with larger increases in subsequent levels of agreement. Discussions with 0 to 1 vs. 2 to 4 EXPL-SUPP exchanges produced the average increase in levels of agreement by 14.29 (n = 17, SD = 18.85) vs. 31.71 (n = 7, SD = 27.14) percentage points, t_{22} = -1.81, p =.08. These findings suggest that responses that support (not question/challenge) other students' explanations are the events that persuade students to add new links.

Not presented here due to space limitations are (a) findings on what particular forms of argumentation trigger *decreases* in agreement, and (b) transitional state diagrams that convey how likely *causal links change in strength* depending on events observed in the online discourse.

Discussion

Although the findings reported in this paper are inconclusive due to insufficient sample sizes, the methods and preliminary findings presented here illustrates how tools like jMAP can support the use of causal diagrams to facilitate collaborative learning, and support research on the interplay between tools and dialogue. The preliminary findings show that initial levels of agreement did not affect subsequent levels of agreement in subsequent diagrams, but *did* affect the level of argumentation prior to student revisions of their causal diagrams. These findings, taken together, suggest that student access to initial levels of agreement (and the pressures of social conformity) do not bias students' decisions when revising their causal diagrams. Instead, the findings suggest that access to initial levels of agreement help direct students' attention to the causal links perceived to be most important and links that deserve more serious examination and discussion.

Given that this study is still ongoing and due to space limitations, the complete version of this paper will provide more information about the tools and functions of jMAP (e.g., sequentially analyzing changes in the strengths of each link over time), additional data and findings (e.g., excerpts from discussion transcripts, inter-rater reliabilities, discourse patterns that explain or contributed to the observed *decreases* in percent agreement, etc.), full discussion of the methods, its limitations, and ideas for future research.

References

Ifenthaler, D., & Seel, N. M. (2005). The measurement of change. Learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning, 2*(4), 317-336.

Jeong, A. C. (2008). jMAP. Retrieved July 24, 2008 http://jmap.wikispaces.com

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging Mindful Use of Prior Knowledge: Attempting to Construct Explanatory Answers Facilitates Learning. *Educational Psychologist*.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning, 1*(1), 59-85.

Shute, V. J., & Jeong, A. C., & Zapata-Rivera, D. (in press). Using flexible belief networks to assess mental models. In B. B. Lockee, L. Yamagata-Lynch, and J. M. Spector (Eds.), *Instructional Design for Complex Learning*. New York, NY: Springer.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Final report the National Science Foundation Evaluative Research and Evaluation Capacity Building).* Syracuse, NY: Syracuse University.