 JrnlID 11412_ArtID 9106_Proof# 1 - 24/12/2010

Computer-Supported Collaborative Learning
DOI s11251-9144/s11412-010-9106-4

Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment

Manu Kapur

Received: 10 February 2010 / Accepted: 20 December 2010

Abstract This paper argues for a need to develop methods for examining temporal patterns in computer-supported collaborative learning (CSCL) groups. It advances one such quantitative method—Lag-sequential Analysis (LsA)—and instantiates it in a study of problem-solving interactions of collaborative groups in an online, synchronous environment. LsA revealed significant temporal patterns in CSCL group discussions that the commonly used "coding and counting" method could not reveal. More importantly, analysis demonstrated how variation in temporal patterns was significantly related to variation in group performance, thereby bolstering the case for developing and testing temporal methods and measures in CSCL research. Findings are discussed, including issues of reliability, validity, and limitations of the proposed method.

Keywords Temporality · Lag-sequential analysis · Collaborative learning · Temporal methods · Event-based process analysis Q1

Introduction

I advance a quantitative method for characterizing and analyzing the temporal patterns in computer-supported, collaborative learning (CSCL) and problem solving. This paper comes in response to recent calls for a greater focus on temporality; underpinning this work is the belief that learning in general, and problem solving in particular, is a continuous, dynamic process that evolves over time. In a seminal paper on temporality published in *ijCSCL*, Reimann (2009) argued that, "Temporality does not only come into play in quantitative terms (e.g., durations, rates of change), but order matters: Because human learning is inherently cumulative, the sequence in which experiences are encountered affects how one learns and what one learns" (p. 1). Understanding how such processes evolve in time and how variation in this evolution explains learning outcomes ranks among the most important

M. Kapur (✉)
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk,
Singapore 637616, Singapore
e-mail: manu.kapur@nie.edu.sg

challenges facing educational research—and temporal methods that expand the methodological toolkit are needed (Akhras and Self 2000; Kapur et al. 2007; McGrath and Tschan 2004; Suthers et al. 2007).

It is not surprising therefore that there has been a push in CSCL research towards unpacking temporal patterns in group interactions and understanding how these patterns relate to group and individual performance and learning (Stahl 2005; Suthers 2006). Reimann (2009) argued "although CSCL researchers are privileged in the sense that they have direct access to processes as they unfold over time (via recordings), there is comparatively little research that makes use of the information contained in the order and duration of events" (p. 1). This presents a unique challenge to traditional analytical measures and methods for analyzing group processes because, for the most part, existing methods continue to take cumulative accounts of member interactions (e.g., categorization of interactional content, rating of discussion quality, member perceptions, and so on) and relate them to group performance. While these accounts are certainly useful, they fail to fully utilize the temporal information embedded in the data. A failure to utilize temporal information naturally reduces the power of the analysis, which may, in turn, limit the validity of the conclusions drawn. Therefore, measures and methods for characterizing and analyzing the evolution of problem-solving processes are needed (Arrow et al. 2000; Collazos et al. 2002; Kapur et al. 2007; Kapur et al. 2005; McGrath and Tschan 2004; Reimann 2009). Therein lies the need for the research reported in this paper.

This manuscript is organized into four sections. (i) I start with a brief review of how interaction has been studied in CSCL research and argue for a need for temporal measures. A substantial amount of literature attempts to understand group processes using qualitative analytical methods, which provide insightful and meaningful microgenetic accounts. Indeed, these methods are highly useful for revealing temporal patterns over multiple timescales. For the present purposes, however, my analysis is delimited to quantitative approaches,[1] typically involving quantitative content analysis (QCA) of interactional data (Chi 1997). The use of QCA, also commonly known as "coding and counting," is pervasive in examining the nature of interaction in CSCL research (Rourke and Anderson 2004; Suthers 2006). (ii) Next, I derive a measure—*transition patterns*—for characterizing the problem-solving process. (iii) I then situate the discussion and illustration of the proposed measure in a study of problem-solving interactions of collaborative groups in a CSCL environment. (iv) Finally, I discuss issues of reliability, validity, and limitations of the proposed temporal method.

The need for temporal measures

Because learners interact with and influence each other in the process of problem solving, these interactions form important units of analyses for research; problem-solving interactions have been used extensively in investigating productivity conditions of collaborative groups (e.g., Barron 2003; Cohen et al. 2002; Jonassen and Kwon 2001; Kapur and Kinzer 2007; Lee et al. 2006; Scardamalia 1989, 1992; Schellens and Valcke 2006; Spada et al. 2005; Zumbach et al. 2005). Because these studies seek to understand group processes and outcomes, they attempt to explain the variation between groups, e.g.,

[1] My focus on quantitative methods should not be mistaken for a singular commitment to or reliance on such methods, nor is it something that I suggest others should do. Indeed, I have advocated and used qualitative methods in my earlier work as part of a larger mixed-method analytical commitment.

why do some groups collaborate better than others? Why are some groups more productive than others? Naturally, part of the variation in the productivity of group is a function of the group process, an analysis of which typically informs a process theory[2] that mediates between input variables (e.g., conditions, context, manipulations) and outcome variables (e.g., performance, learning) (Reimann 2009).

As an example of how experimental manipulation influences group processes, and how these processes in turn influence outcomes, consider my earlier work on *productive failure* (Kapur 2008; Kapur and Kinzer 2009). In a randomized controlled experiment, I assigned student triads to solve either well- or ill-structured physics problems. Analysis suggested that the groups in the two conditions differed significantly in both their processes and outcomes. Compared to groups who solved well-structured problems, groups who solved ill-structured problems expectedly struggled with defining and analyzing the problems. Their discussions were significantly more complex, chaotic, and divergent, resulting in poor group quality of solutions produced in the shorter term. However, despite failing in their collaborative, problem-solving efforts, these students outperformed their counterparts from the well-structured condition on both well- and ill-structured problems subsequently, suggesting a latent productivity in what seemed to be failure initially. In other words, group processes that led to failure initially were more productive for individual learning in the longer term than group process that led to performance success.

Invariant across the abovementioned studies, including my own, is a quantitative method for unpacking group processes, commonly called the “coding and counting” method (Suthers 2006). This method involves applying one or more coding schemes (as informed by some process theory) to the interaction data resulting in a cumulative or relative frequency distribution of interactions across the categories of the coding scheme (hereinafter referred to as process categories, e.g., depth of explanations, functional content of interactions, misconceptions) (Strijbos et al. 2006). These distributions essentially tally the amount, proportion, and type of interactions vis-à-vis the interactional coding scheme. Significant links are then sought between quantitatively-coded interactional data and outcomes, such as quality of group performance and group-to-individual transfer (see Rourke and Anderson (2004) for a discussion on the validity of QCA). For example, consider a prototypical case wherein such an analysis may reveal that successful groups had greater proportions of explanation and critique than the less successful groups. If so, this would bolster a process theory that emphasizes the role of explanations and critique in learning (Chi et al. 1989, 1994). Clearly, such “coding and counting” analyses are useful because they help explain how the distribution of process categories relates to outcome variation (Reimann 2009).

Notwithstanding the empirically supported significant links between the distribution of process categories and group performance and learning, interpreting findings from interactional coding schemes is limited by the very nature of the information tapped by these measures. These measures tell us *that* a certain proportion of interactional content was coded in a particular process category but nothing about the sequence or order of these categories. By aggregating counts over time, information about temporal variation is lost. Such an analysis, therefore, does not take the temporality of interactions into account. For example, in the prototypical case earlier, there could be two groups with similar proportions of explanation and critique in their discussions. However, these groups could be very different when the temporal information is factored in. For one group, it could be that

[2] Reimann (2009) provides an excellent description of the how temporal events in group processes mediate between input factors and outcome variables.

explanations were followed by more explanations, and likewise for critique. For the other group, it could be the case that explanations followed critiques that in turn led to more explanations and critique. In other words, for the first group, the learning mechanisms invoked by explanations and critique could be independent of each other whereas for the second group, they could be co-evolving and dependent. By simply coding and counting, an explanation that follows an explanation is accorded the same weight as one that follows critique—an assumption of *temporal homogeneity* (Kapur et al. 2008) that is rarely valid in light of the complexities of group dynamics.

The above example illustrates that groups with similar frequency distribution of process categories may well have contrasting temporal dynamics of those process categories. More importantly, this temporal contrast may be germane to a process theory of group learning and performance. After all, evidence suggests explanations that follow critiques or impasses are more likely to invoke processes that are germane for learning (Van Lehn et al. 2003). Therefore, methods that take temporality into account stand to not only add to the methodological toolkit of the researcher but also help in building a better process theory of group learning and problem solving (Reimann 2009).

CSCL research on temporality using quantitative approaches

There is a small but growing body of CSCL research that is beginning to develop temporal measures to better understand CSCL processes using quantitative methods. I describe a few illustrative (but not exhaustive) examples, including my own initial forays.

Soller and colleagues (2002) used Hidden Markov Modeling (HMM; Rabiner 1989) to analyze and assess temporal patterns in on-line knowledge-sharing conversations over time. Their HMM model could determine the effectiveness of knowledge-sharing episodes with 93% accuracy, that is, 43% above what one would expect by chance. They argued that understanding the temporal dynamics of how groups share, assimilate, and build knowledge together is important in building a process theory of facilitation to increase the effectiveness of the group interactions.

Employing a different analytical method—Time Series Analysis—Muukkonen and colleagues (2007) modeled changes in students' emotions as they engaged in their ongoing projects and collaboration. Students responded to survey queries through their mobile phones five times a day for a period of two weeks. Student interviews and query data were used to form a picture of the variation of daily routines, challenges, and reflections of one's own activities, and more importantly, the extent to which this variation related to their engagement in learning.

Jeong (2005) illustrated how exploratory sequential analysis can be used to measure the likelihood of a message receiving a response in computer-mediated discussion boards, the types of such responses (e.g., challenging, giving evidence, explaining), and whether sequences of responses (e.g., claim → challenge → explain) evidence theoretically-conjectured sequences that are germane for problem-solving and learning. Olson et al. (1994) described a similar approach wherein they examined how sequential interactional patterns differed between supported (with electronic representational tools) and unsupported collaborative groups. In both Jeong's and Olsen et al.'s work, one finds a careful consideration and development of process categories, sequences of which are then examined to detect patterns that occur significantly above chance level. Much as the detection of sequential patterns is important in and of itself, neither study described analytical procedures for relating these sequential patterns to group performance.

Q2

My earlier attempts at examining temporal patterns of CSCL problem-solving groups entailed using a *Random Walk* model (Ross 1996) to model convergence in group discussions (Kapur et al. 2008). Analysis revealed that high (low) quality member contributions made earlier in a discussion did more good (harm) than those made later on. A differential temporal impact of member contributions suggested a high sensitivity to early exchange. More importantly, by relating this process sensitivity to eventual group performance, analysis showed that group performance could be predicted based on what happened in the first 30–40% of a discussion. From the standpoint of a process theory of facilitating and scaffolding group problem solving, findings suggested a greater emphasis on the earlier phases of a group discussion.

While each of the abovementioned examples examined temporality in CSCL groups, they did so using different analytical methods and in different contexts. However, common across these examples is an emphasis on: a) understanding temporal variation to uncover patterns that may otherwise not be possible using coding and counting methods, and b) informing a process theory of collaborative learning and performance. I believe this dual emphasis is important and necessary, and one that can be set as broad criteria which temporal methods, including the one that is advanced in this paper, must minimally meet.

An illustrative study

The purpose of this paper is to advance a method for analyzing the temporal patterns in CSCL discussion. The focus is squarely methodological. I situate the discussion and illustration of the proposed method in a study of CSCL problem-solving interactions, which was part of a larger program of research on productive failure described earlier (Kapur 2008; Kapur and Kinzer 2009). I briefly describe the context of the study in which the methodology was instantiated before illustrating the methodology.

Research context and data collection

Participants were 177, 11th-grade science students (120 male, 57 female) from two co-educational, English-speaking high schools in the National Capital Region of India. Students in the science stream typically study Mathematics, Physics, Chemistry, and English as their main academic subjects. The school's curriculum was prescribed by the Central Board of Secondary Education (CBSE) of India. These schools were of similar academic standing. Using data from the 10th-grade CBSE national standardized test scores in science, an ANOVA did not find any significant difference between the two schools in terms of student ability in science, p=.227. As is typically the case, students came from upper-middle class families and were considered technologically savvy. The study was designed to reflect the schools' science curricula. Prior to the study, all students had completed the curricular unit on Newtonian kinematics—the targeted conceptual domain of the study.

All students took a twenty-item, multiple-choice pretest on the targeted concepts (*Cronbach alpha* = .81). The 177 students were first randomly grouped into triads, resulting in 59 groups. These groups were then randomly assigned to one of two conditions: an ill-structured (IS) problem-solving condition (28 groups) or a well-structured (WS) problem-solving condition (31 groups). A post-randomization check revealed that there was no significant difference between students on the pretest, p=.317. They were instructed to collaborate with their group members to solve either a well-structured (WS) or an ill-structured (IS) problem scenario as appropriate to their assigned condition. The study was carried out in the school's computer laboratory, where group members communicated with one another only through synchronous, text-only chat. The chat application allowed groups to privately and simultaneously engage in synchronous discussions and automatically archived the transcript of their discussion as a text file. These 59 transcripts, one for each group, contained the problem-solving interactions of group members as well as the final solutions produced by the groups and formed the data source. All the materials—the pretest and the problem scenarios—can be found in Kapur (2008), and are therefore not reproduced here.

Procedure

A WS and an IS problem scenario were developed consistent with Jonassen's design theory typology for problems (2000). Both problem scenarios dealt with a car accident scenario and targeted the same concepts from Newtonian Kinematics and Laws of Friction to solve them. Content validation of the two problem scenarios was achieved with the help of two physics teachers from the school with experience in teaching those subjects at the senior secondary levels. The teachers also assessed the time students needed to solve the problems. Pilot tests with groups of students from the previous cohort further informed the time allocation for the group work, which was set at 1.5 h. Ultimately, all groups completed the problem in the allotted time.

The study was carried out in the school's computer laboratory. The online synchronous collaborative environment was a Java-based, text-only chat application running on the Internet. Despite these participants being technologically savvy in using online chat, they were familiarized with the use of the synchronous text-only chat application prior to the study. Group members could only interact within their group. Each group's discussion and solution were automatically archived as a text file to be used for analysis. A seating arrangement ensured that participants of a given group were not proximally located so that the only means of communication between group members was synchronous, text-only chat.

To mitigate status effects, I ensured that participants were not cognizant of their group members' identities; the chat environment was configured so that each participant was identifiable only by an alphanumeric code. Cross-checking the transcripts of their interactions revealed that participants followed the instruction not to use their names and instead used the codes when referring to each other. No help regarding the problem-solving task was given to any participant or group during the study. Furthermore, no external member roles or division of labor were suggested to any of the groups. The procedures described above were identical for both WS and IS groups. The time stamp in the chat environment indicated that all groups made full use of the allotted time of 1.5 h and solved their respective problems.

Data coding of problem-solving interactions into process categories

Quantitative Content Analysis (QCA) (Chi 1997) was used to segment and code utterances. The unit of analysis was semantically defined as the function(s) that an intentional utterance served in the problem-solving process. Bransford and Nitsch (1978) support the case for semantically-defined units by viewing meaning-making and understanding as functions of the interdependence between interaction and context. Thus, every utterance was segmented into one or more interaction unit(s), and coded into process categories adapted from the

Functional Category System (FCS)—an interaction coding scheme developed by Poole and Holmes (1995). Accordingly, each interaction unit was coded into one of seven problem-solving process categories:

1. Problem Analysis (PA): Statements that define or state the causes behind a problem (e.g., "*I think the man was driving too fast*"),
2. Problem Critique (PC): Statements that evaluate problem analysis statements (e.g., "*how can you be sure that the man was driving fast?*"),
3. Orientation (OO): Statements that attempt to orient or guide the group's process, including simple repetitions of others' statements or clarifications; statements that reflect on or evaluate the group's process or progress (e.g., "*lets take turns giving our opinions*"),
4. Criteria Development (CD): Statements that concern criteria for decision making or general parameters for solutions (e.g., "*we need to find the initial speed of the car*"),
5. Solution Development (SD): Suggestions of alternatives, ideas, proposals for solving the problem; statements that provide details or elaborate on a previously stated alternative. They are neutral in character and provide ideas or further information about alternatives (e.g., "*use the* 2nd *equation of motion*"),
6. Solution Evaluation (SE): Statements that evaluate alternatives and give reasons, explicit or implicit, for the evaluations. This also included statements involving simple agreement or disagreement with criteria development or solution suggestion statements especially since these statements were frequently coupled with evaluative responses. Statements that state the decision in its final form or ask for final group confirmation of the decision. (e.g., "*yes, but how do we get acceleration?*"), or
7. Non-Task (NT): Statements that do not have anything to do with the decision task. They include off-topic jokes and tangents (e.g., "*lets take a break!*").

After an initial training phase, two trained doctoral students independently coded the interactions with an inter-rater reliability (*Krippendorff's alpha*) of .88. The result of coding the problem-solving interactions was a representation of each problem-solving discussion as an ordered sequence of FCS process categories.

Results

I present the analyses and corresponding results in three major sections.

First, I present a typical coding and counting multivariate analysis. Recall that the coding of discussion data reduced each discussion to a string of FCS process categories. By calculating the relative frequency of each process category, a multivariate analysis shed light on the functional content of WS and IS group discussions. This coding and counting analysis served as a baseline against which the *added value* of the temporal method could be evaluated.

Second, given the stated limitations of a coding and counting analysis, I examine the transition patterns between process categories. Using an analytical technique called Lag-sequential Analysis (LsA; Bakeman and Gottman 1997; Wampold 1992), I examine the likelihoods of how some process categories follow or are followed by other process categories. LsA revealed significant insights into the temporal patterns in the transitions between process categories, including how these patterns differed between WS and IS groups.

Finally, I relate the significant temporal patterns to group performance, as measured by the quality of group solutions. I first establish that there is in fact a significant difference in group performance between WS and IS groups. I then demonstrate that the strongest interactional predictor of group performance is the transition pattern between process categories.

Taken together, I describe one way in which examining temporal patterns can reveal significant insights over and above the coding and counting method, and contribute to a larger process theory of CSCL. It is worth reiterating that I am not proposing to replace the coding and counting method with LsA. I advance LsA as a value-added, which can provide additional insights over and above those provided by the coding-and-counting methods.

Coding and counting

The proportion of interactional activity in the six functional categories PA, PC, OO, CD, SD, and SE formed the six dependent variables in the coding and counting analysis. Proportion of NT was very small, and was excluded from the analysis. Controlling for the effects of school and group prior knowledge (mean score of group members on the pretest), a MANCOVA revealed a significant multivariate effect of WS vs. IS groups on the functional content of their discussions, $F(6, 50)=3.46$, $p=.006$, partial $\eta^2=.29$. As a rule of thumb, partial $\eta^2=.01$ is considered a small, .06 medium, and .14 a large effect size (Cohen 1977). Table 1 presents the descriptive statistics.

The six univariate Levene's tests for equality of error variances were statistically not significant. Univariate analyses showed that IS groups had significantly greater proportion of activity centered on:

- PA: problem analysis, $F(1, 55)=16.81$, $p<.001$, partial $\eta^2=.23$,
- PC: problem critique, $F(1, 55)=12.27$, $p=.001$, partial $\eta^2=.18$, and
- CD: criteria development, $F(1, 55)=3.79$, $p=.047$, partial $\eta^2=.06$.

In contrast, WS groups had significantly greater proportion of activity centered on:

- SD: solution development, $F(1, 55)=4.37$, $p=.041$, partial $\eta^2=.07$.

There was no significant difference in the OO and SE activity between WS and IS groups. IS groups had greater proportion of interactional activity centered on PA, PC, and CD whereas WS groups had greater proportion of interactional activity centered on OO, SD, and SE, although OO and SE did not reach significance.

Table 1 Descriptive statistics for functional content of WS and IS group discussions

Functional Category	WS Groups		IS Groups	
	M	SD	M	SD
PA: Problem Analysis	.046	.022	.081*	.031
PC: Problem Critique	.032	.016	.053*	.020
OO: Orientation	.355	.128	.382	.079
CD: Criteria Development	.045	.019	.053*	.018
SD: Solution Development	.354*	.126	.272	.087
SE: Solution Evaluation	.151	.052	.143	.046

*$p<.05$

Transitions between process categories

The coding and counting analysis does not provide any indication or measure of the temporal patterns in the process categories. For example, what are the transitions between process categories? Are some process categories more likely to follow or be followed by other process categories? Are these likelihoods different for WS and IS groups? These questions seek to understand the transition probabilities between process categories, and in doing so, uncover temporal patterns in the input sequence of process categories (recall that the data coding reduced each discussion to a temporal string of process categories).

One method of quantitatively examining these transition probabilities is Lag-sequential analysis (LsA)—a technique increasingly being used to detect such patterns. LsA treats each interactional unit (defined earlier) as an observation; a coded sequence of these observations forming the problem-solving sequence of a group discussion (Erkens et al. 2003). It detects the various non-random aspects of interactional sequences to reveal how certain types of interactions follow others more often than what one would expect by chance (Wampold 1992). It accomplishes this by comparing the expected and actual transition probabilities between process categories to identify *statistically significant transitions* from one type of interactional activity to another; statistical significance corresponding to an alpha level of .05 (for a fuller treatment of LsA and related methods that are beyond the scope of this paper, see Bakeman and Gottman 1997; Sanderson and Fisher 1994; Wampold 1992). These transition probabilities can then be converted into odds ratios or likelihoods for comparison.

The software program Multiple Episode Protocol Analysis (MEPA) was used for carrying out LsA.[3] LsA revealed significant differences between the discussions of WS and IS groups (see Fig. 1). In Fig. 1, a category with an arrow directed to itself means that groups in that condition were *at least twice* as likely to sustain that type of activity, i.e., the activity was at least twice as likely to appear in coherent clusters rather than throughout the discussion. For example, PA was at least twice as likely to be followed by more PA in WS groups than in IS groups; attempts at problem analysis were followed by more problem analysis. An arrow from one category to another represents a directed transition. For example, PA activity was at least twice as likely to be followed by PC activity in IS groups; attempts at problem analysis were followed by problem critique, which in turn were followed by even more critique.

Figure 1 suggests that with regard to how groups sustained different types of activities, IS groups were at least twice as likely to sustain PC and SE activities. For example, sequences where PC was followed by PC, and inductively, more PC, were twice as likely to be found in IS group discussions than in WS group discussion. In contrast, WS groups were at least twice as likely to sustain PA, CD, and SD activities. With regard to transitions, there were no significant transitions that WS groups were more likely to exhibit. In contrast, the discussions of IS groups were more likely to exhibit many significant transitions (PA-PC, PA-CD, and CD-SD) as well as feedback loops (SE-PA and SE-PC).

Inducting on the transition likelihoods, it follows that the discussions of WS groups were more likely marked by interactional sequences: PA-PA-PA, CD-CD-CD, SD-SD-SD (three instances of a process category are chosen just as an illustration of the pattern; sequences can be shorter or longer depending upon the transition probabilities). Discussions of IS groups, by contrast, were more likely marked by sequences: PA-PC-PC-PC, PC-PC-PC,

[3] MEPA was developed by Dr Gijsbert Erkens. For more information, see http://edugate.fss.uu.nl/mepa/index.htm.

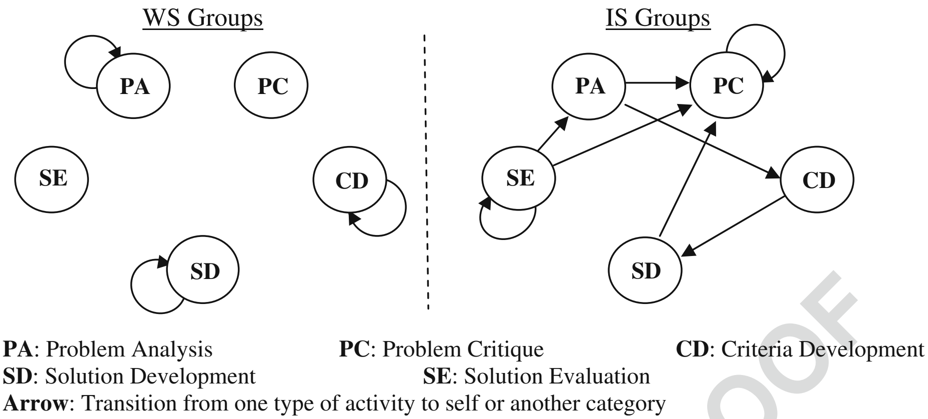

Fig. 1 Likely sequential patterns in the discussions of well- vs. ill-structured groups

PA-CD, PA-CD-SD, CD-SD, CD-SD-PC, CD-SD-PC-PC-PC, SD-PC-PC-PC, SE-SE-SE, SE-PA, SE-PA-PC, SE-PA-PC-PC-PC, SE-PA-CD, SE-PA-CD-SD, SE-PA-CD-SD-PC-PC-PC-PC, SE-PC, and SE-PC-PC-PC.

Note that the greater the number of significant transitions and feedback loops, the greater the number of possibilities in which the discussion could unfold from any given point/event in the discussion. This, in turn, suggests not only greater interactional complexity but also more divergent temporal event trajectories. Of course, an intuitive way of understanding this is to realize that the greater the number of interactions between the components (process categories) of a given system (group discussion), the greater is its complexity (Holland 1995; Kauffman 1995).

Thus conceived, LsA can be analogously related with Kauffman's (1995) measure of complexity for the evolution of Boolean networks. The relation becomes clearer when one conceives the FCS process categories as component states of an evolving Boolean network; Boolean in the sense that, at any point in time, a component state (PA, PC, etc.) may be present or absent in the group discussion, and interactions between the component states may be represented in terms of probabilistic logical functions. This, in many ways, is what LsA attempts to do; it looks at the probability of how certain process categories (or component states) follow others at a rate that is significantly above chance level: an *if-then* probabilistic logical function. As a result, the collaborative process can be examined as an evolving, multi-state Boolean network, and the greater the number of significant transitions between the component states, the greater the complexity of and divergence in the temporal trajectories of its evolution (for a fuller treatment, see Bar-Yam 2003; Kauffman 1995).

Figure 1 suggests that the IS group discussions seemed to exhibit greater divergence and complexity relative to those of WS groups. Based on the argument above, this complexity was a direct function of the number of statistically significant transitions between process categories (Bar-Yam 2003; Kauffman 1995). Therefore, I compared the number of such significant transitions between IS and WS groups.

IS groups, M=8.27, SD=3.00, had a greater number of significant transitions between process categories than WS groups, M=3.27, SD=2.52. Controlling for the effect of group prior knowledge (as measured by the pretest), an ANCOVA suggested that this difference was statistically significant, $F(1, 56)=53.05$, $p<.001$, partial $\eta^2=.49$. Levene's test for violation of homogeneity of variance was not significant, $p=.217$.

Note how LsA revealed differences in interactional patterns that the coding-and-counting analysis of functional content did not. First, coding and counting revealed that IS groups spent a greater proportion of their interactional activity on problem analysis and criteria development than WS groups. However, what coding and counting did not reveal was that WS groups' problem analysis and criteria development were more clustered together rather than spread throughout the discussion—a temporal pattern revealed by LsA. Second, coding and counting revealed that IS groups had a greater proportion of problem critique than WS groups. What it did not reveal was that such critique in IS groups tended to be more clustered as well, rather than spread throughout the discussion—another temporal pattern revealed by LsA. Third, coding and counting revealed that WS groups spent greater proportion on solution development than IS groups. What it did not reveal was that such solution development in WS groups tended to be clustered rather than spread throughout the discussion—another temporal pattern revealed by LsA. Fourth, coding and counting did not reveal any difference in solution evaluation between WS and IS groups. However, LsA revealed that solution evaluation in IS groups tended to be more clustered than in WS groups. Finally, LsA was also able to pick several significant transitions between the process categories, which coding and counting simply could not.

In sum, where coding and counting suggested differences in the relative proportions of the process categories, LsA suggested further differences in the temporal patterns of these process categories. Where coding and counting did *not* suggest a difference in the process categories, LsA was able to pick differences in temporal pattern in these categories. In other words, where things appeared to be different, LsA differentiated them even further, and where there appeared to be no difference, LsA revealed important differences. Therefore, it is reasonable to suggest that LsA revealed temporal patterns in WS and IS groups that coding and counting could not reveal.

However, revealing temporal patterns alone is not sufficient. These patterns must be of some value, both theoretically and empirically. Theoretically, it is not unreasonable to argue that the above temporal patterns are in fact important for learning. Research strongly suggests that processes of defining the problem, critique, questioning, elaboration, and explanation are germane for learning (Anderson 2000; Chi 1989). Consequently, the sequences in which these processes unfold are naturally germane for learning (Barron 2003; Kapur 2008; Van Lehn et al. 2003). This is because these sequences represent how learners (in groups) explored the problem and solution spaces for representations and methods to solve the complex problem.

Therefore, on the premise that the above temporal patterns are theoretically important, they must at least be able to significantly explain some variance in group performance. If not, empirical evidence for their argued theoretical importance would be weak in the present case. Furthermore, an even more stringent empirical test would compare the predictive power of coding and counting patterns with temporal patterns in explaining variation in group performance. This is precisely the purpose of the analyses described in the following section.

Analyzing and explaining group performance

The purpose of analyzing group performance was to relate it to the coding and counting as well as temporal patterns found in the preceding sections. I first examine differences in the group performance of WS and IS groups, and then demonstrate that the strongest interactional predictor of group performance is the temporal pattern of transitions between process categories.

The measure of group performance was operationalized as the quality of solution produced by the group. In consultation with the teachers, the strategy adopted was to focus on the extent to which groups were able to support their decisions through a synthesis of both qualitative and quantitative arguments, and supporting them with justifiable assumptions. The extent to which groups were able to accomplish this was rated on a scale from 0 to 4 points in units of 0.5 using a holistic rubric shown in Table 2. Two trained doctoral students scored the solutions with an inter-rater reliability (*Krippendorff's alpha*) of .93.

An ANCOVA, $F(1, 56)=4.61$, $p=.036$, partial $\eta^2=.11$, revealed that the quality of solution produced by WS groups, $M=2.84$, $SD=1.26$, was on average significantly better than that of IS groups, $M=1.29$, $SD=1.08$, controlling for group prior knowledge. Levene's test for violation of homogeneity of variance was not significant, $p=.426$. This difference in solution quality was not particularly surprising given that IS groups had to solve a problem that was more complex and ill-structured. Analysis of functional content and transition patterns supports this contention. After all, IS groups spent more effort analyzing and critiquing the problem, setting appropriate criteria for a solution than actually developing a solution, resulting in poor group performance. WS groups, on the other hand, solved a problem that afforded a more defined problem and solution space. Thus, WS group discussions were, on average, more coherent, less complex, and less likely to exhibit transitions or feedback loops. WS groups found it relatively easier to analyze the problem, set appropriate criteria, and develop a solution, which, in turn, resulted in relatively higher group performance.

From a methodological standpoint, it is important to explain this variation in group performance. Recall that if the methodological arguments thus far hold, then temporal patterns should not only explain this variation significantly but also perform better than the functional content in explaining this variation. To test this, I employed a regression analysis with group solution score as the dependent variable. The predictors entered (in the order they are mentioned) were: problem type, proportion of FCS process categories in a group discussion, and number of significant transitions between process categories in a group discussion.

The regression model was significant, $F(7, 51)=7.56$, $p<.001$, with $R^2=.509$ and adjusted $R^2=.442$. Table 3 presents the model summary. As can be seen, the temporal

t2.1 **Table 2** Rubric for coding quality of group solution

	Quality	Description
t2.2	Quality	Description
t2.3	0	Solution weakly supported, if at all
t2.4	1	Solution supported in a limited way relying either on a purely quantitative or a qualitative argument with little, if any, discussion and justification of the assumptions made
t2.5	2	Solution is only partially supported by a mix of both qualitative and quantitative arguments; assumptions made are not mentioned, adequately discussed, or justified to support the decision
t2.6	3	Solution synthesizes both qualitative and quantitative arguments; assumptions made are not adequately discussed and justified to support the decision
t2.7	4	Solution synthesizes both qualitative and quantitative arguments; assumptions made are adequately discussed and justified to support the decision

Mid-point scores of .5, 1.5, 2.5, and 3.5 were assigned when the quality of solution was assessed to be between the major units 0, 1, 2, 3, and 4

Table 3 Model summary for predicting group performance

Source	*B*	SE	Beta	*t*	*p*
Constant	3.472	.969		3.582	.001
Problem Type	.707	.432	.288	1.636	.108
PA	−1.008	5.096	−.031	−.198	.844
PC	−12.258	8.607	−.218	−1.424	.161
CD	−.065	6.670	−.001	−.010	.992
SD	.879	1.304	.081	.674	.503
SE	−2.124	2.534	−.100	−.838	.406
No. of Sig. Transitions	−.252	.053	−.721	−4.716	<.001

pattern predictor (number of significant transitions) is the only significant predictor of group performance, thereby withstanding the empirical test in the present case.

Discussion

The purpose of this paper was specific and modest. I argued for the need to develop methods for uncovering temporal patterns in CSCL groups, and advanced one such measure to examine transition patterns between process categories. Using LsA, analysis revealed significant temporal patterns that the typical coding and counting method could not reveal. More importantly, analysis demonstrated how variation in temporal patterns was significantly related to variation in group performance. In fact, the temporal pattern predictor was the only interactional predictor of group performance. This analysis, therefore, bolsters the case for more work on developing and testing temporal methods and measures in CSCL research.

It was not surprising that the measure of temporal patterns emerged as a more powerful predictor of group performance. One only needs to compare the nature of information tapped by coding and counting with that tapped by temporal patterns to explain why this was the case. As argued earlier, coding and counting only captures every instance of occurrence of a particular process category in the discussion. By aggregating this data, it gives no indication of when or where in the discussion the process category occurred. LsA, however, provides that temporal information. To calculate significant transitions, LsA necessarily has to take into account the number of instances of a particular process category, or else the transition probabilities could not be calculated. Furthermore, it also examines process categories before and after a given process category. Therefore, LsA not only takes into account the information that coding and counting captures, but it goes further and captures information about the order and sequencing of the process categories. Inductively, it includes information from all preceding process categories because the likelihood of a given process category is an inductive function of all the preceding likelihoods. Given this, the greater predictive power of temporal patterns over the coding and counting patterns makes sense conceptually.

What also needs explanation is the predictive relation between variation in temporal patterns and group performance. The negative coefficient for the predictor—number of significant transitions—in the regression model (see Table 3) suggested that the greater the number of significant transitions between process categories in a group discussion (as was the case for IS groups), the lower the group performance. This may seem to contradict my theoretical arguments. After all, part of my argument emphasized the theoretical importance of the very kinds of sequences of process categories that were more likely in IS groups. If such sequences are theoretically germane for learning, then how is it that IS groups with a greater likelihood of these sequences performed worse than WS groups who had a lower likelihood of these very theoretically-important sequences?

The answer to this question lies at the incommensurability between performance and learning (Clifford 1984; Schmidt and Bjork 1992). I have discussed this incommensurability in greater detail elsewhere as part of my research program on productive failure (see Kapur 2008, 2009, 2010; Kapur and Kinzer 2009). For the present purposes, I provide only a brief explanation: Increasingly there is a realization that conditions that maximize performance (e.g., solving well-structured problems) may not necessarily be the ones that maximize learning (e.g., as useful for solving complex, ill-structured problems). Even though IS groups had significantly lower group performance, process analysis suggested that they engaged in processes (of analyzing, critiquing, explaining, evaluating, etc.) in ways that were germane for learning. Therefore, processes that seemed to lead to failure (in group performance) initially constituted the locus of powerful learning in the longer term. In other words, failure in the shorter term can be productive in the longer term insofar as learners engage in processes that are germane for learning; processes that may not necessarily lead to successful performance at first (Clifford 1984; Schwartz and Bransford 1998).

Reliability and validity

Inferences drawn from new measures are strong in so far as the coding scheme is reliable and valid. Because LsA runs statistical operations on the sequence of FCS process categories, the reliability and validity of interpretations derived from LsA are, in part, a function of the reliability and validity of the FCS process categories. In this study, I opted to use an existing coding scheme, namely the functional category system (FCS) developed by Poole and Holmes (1995). The reasons for choosing the FCS as the interaction coding protocol for this study are:

i. The FCS was developed specifically for the purpose of studying small-group collaborative interactions in problem-solving contexts.
ii. The FCS categories are theoretically well grounded in the cognitive and educational theories of problem solving, thereby increasing their content validity.
iii. The FCS has been tried and tested in several research studies (for example, Poole and Holmes 1995; Jonassen and Kwon 2001; Kapur 2008; Kapur and Kinzer 2009), making it more reliable and stable than developing an entirely new coding scheme (Gall et al. 1996).
iv. From a broader perspective of research design and measurement, using a pre-existing interaction coding scheme adds to the validity of the inferences drawn from the results (Rourke and Anderson 2004).

Within the constraints and limitations of a singly study, the above reasons positively influence the reliability and validity of the quantitative content analysis and LsA.

Limitations

As with any new method, its repeated application and modification over multiple data sets is needed before strong and valid inferences about the underlying cognitive processes can be made (Rourke and Anderson 2004). Furthermore, I delimited the scope of this paper to quantitative analysis, mainly to make the argument as succinct as possible. This delimitation is not to be mistaken as a belief in a singular reliance on quantitative methods alone, for I strongly believe in triangulating findings with microgenetic qualitative analysis. Elsewhere, I have carried out LsA as part of a mixed-method analysis (Kapur 2008). Indeed, this only bolsters the reliability of the LsA analysis discussed earlier.

Another limitation includes the requirement of capturing data in which there is sufficient manifestation of process categories; the greater the number of process categories, the greater the requirement of manifestation. While capturing the data was made easy due to the technology itself, analyzing the data was time consuming. But the marginal effort over and above what one would have done for coding and counting was minimal because the output of coding and counting—a representation of group discussion as a temporal sequence of process categories—formed the input data for LsA. As such, LsA can be a fairly useful addition to the methodological tool-kit of CSCL researchers who already plan to carry out coding and counting types of analysis. More importantly, inferences drawn by researchers from a combined analysis will have meaningful implications for the design of CSCL environments, especially with regard to the design and scaffolding of instruction and learning environments for problem-solving tasks (Dillenbourg 2002; Reimann 2009).

Finally, a commonly-held but (I believe) misconceived limitation of LsA and event-based analyses in general must also be addressed. Although Reimann (2009) advanced the argument for event-based analyses, he did not directly address this misconception, and so it needs to be addressed here. It is commonly argued that because the transition probabilities for a particular event are calculated based only on the previous event in the sequence, that event-based analysis such as LsA throw away information about preceding temporal events or sequences, which makes such methods overly simplistic and impoverished. In the words of Suthers et al. (2007), these methods "use a state-based representation that reduces the sequential history of interaction to the most recently occurring event category." In other words, it is a huge error to model an event that depends upon a cumulative series of events leading up to it as though it depends *only* upon the preceding event.

I believe this commonly-held limitation arises from a confusion of the difference between a *transition probability* and the *probability of an event*. The transition probabilities are calculated pair-wise, that is, the probability of an event occurring immediately after a given event; it is a conditional probability. The probability of an event however is an inductive function of the various transition probabilities preceding it. For example, an event has a certain probability of occurring after another event, which in turn has a certain probability of occurring after yet another event, and so on. Mathematically inducting on these transition probabilities suggests that the probability of an event occurring is not simply the transition probability but instead a function of the preceding transition probabilities. Therefore, the occurrence of an event is a function of the very information from its history that it is mistakenly criticized for having been thrown away. Even so, one has to acknowledge that quantitative event-based analyses do represent reductions of the richness and complexity of group processes, which forms an inherent limitation. This is why I advocate their use as part of a more comprehensive mixed-method analytical regime, so as to achieve greater reliability and validity of interpretation derived from these methods.

Future directions

Moving forward, there is a need to apply the proposed method in other contexts and settings. Better indications of the validity and reliability will emerge from a repeated application and modification of the measures in triangulation with other quantitative and qualitative analyses over multiple data sets. In turn, this will lead to fine-tuning of the measures in an iterative fashion.

Concomitantly, there is also a need for developing new measures, especially at a macroscopic level of analysis, in particular, the stable interaction phases that a discussion goes through. In other words, a problem-solving discussion can be conceptualized as a temporal sequence of phases. One can use several methods to isolate temporal phases, including measures of genetic entropy (Adami et al. 2000), intensity of mutation rates (Burtsev 2003) or, in the case of problem interactions, the classification of coherent phases of interaction. Whether these phases involve genetic mutations or stable interactions, sequences of fluctuations often alternate between stable phases, with chaotic phases interspersed throughout. These often correspond to low vs. high mutation rates, clustered vs. unclustered interactions. With the phases identified, one can calculate and predict the probabilities of moving from one phase to another using, for example, Hidden Markov Models (HMM). Unlike LsA and Markov Models that work on transitions between states visible to the observer or categories coded by the observer, HMMs allow for a detection of latent or hidden patterns that are not directly visible to, or coded by the observer. As a result, one may begin to understand when and why phase transitions, cascades, and catastrophes (sudden mass change), as well as stable phases emerge. More importantly, one may begin to understand how the configuration (not just the presence) of one phase may influence the likelihood of moving to any other phase. Whether one can control or temper these phases, or whether such control or temperance would prove a wise practice remains an open question, which, even if only partially answered, will be a major breakthrough in characterizing and modeling the problem solving process (Kapur et al. 2006; Voiklis et al. 2006).

Conclusion

Temporality clearly matters. By emphasizing the need for temporal measures as well as developing them, CSCL researchers who wish to study problem-solving processes will find choices among several lenses at varying resolutions. Used in addition to coding and counting methods, temporal measures can reveal information about sequences and transitions that are important for learning. When carried out as part of a comprehensive mixed-method analytical program, one can zoom from micro- to macroscopic properties and behaviors of the problem-solving process, which would be critical to building a more powerful process theory of collaborative problem solving.

References

Adami, C., Ofria, C., & Collier, T. C. (2000). Evolution of biological complexity. *Proceedings of the National Academy of Sciences, 97*, 4463–4468.

Akhras, F. N., & Self, J. A. (2000). Modeling the process, not the product, of learning. In S. P. Lajoie (Ed.), *Computers as cognitive tools, volume two: No more walls* (pp. 3–28). Mahwah: Lawrence Erlbaum Associates.

Anderson, J. R. (2000). *Cognitive psychology and its implications*. New York: Worth.

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation*. Thousand Oaks: Sage Publications.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Barab, S. A., Hay, K. E., & Yamagata-Lynch, L. C. (2001). Constructing networks of action-relevant episodes: An in-situ research methodology. *The Journal of the Learning Sciences, 10*(1&2), 63–112. Q4

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.

Bar-Yam, Y. (2003). *Dynamics of complex systems*. Reading: Addison Wesley.

Bransford, J. D., & Nitsch, K. E. (1978). Coming to understand things we could not previously understand. In J. F. Kavanaugh & W. Strange (Eds.), *Speech and language in the laboratory, school, and clinic*. Harvard: MIT.

Burtsev, M. S. (2003). Measuring the dynamics of artificial evolution. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in artificial life*. *Proceedings of the 7th European Conference on Artificial Life*, Dortmund, Germany, September 14–17.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 5*, 145–182.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.

Clifford, M. M. (1984). Thoughts on a theory of constructive failure. *Educational Psychologist, 19*(2), 108–120.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic.

Cohen, E. G., Lotan, R. A., Abram, P. L., Scarloss, B. A., & Schultz, S. E. (2002). Can groups learn? *Teachers College Record, 104*(6), 1045–1068.

Collazos, C., Guerrero, L., Pino, J., & Ochoa, S. (2002). *Evaluating collaborative learning processes. Proceedings of the 8th International Workshop on Groupware (CRIWG'2002)*. Heidelberg: Springer.

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

Erkens, G., Kanselaar, G., Prangsma, M., & Jaspers, J. (2003). Computer support for collaborative and argumentative writing. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merrienboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 157–176). Amsterdam: Elsevier Science.

Gall, M., Borg, W., & Gall, J. (1996). *Education research: An introduction*. White Plains: Longman.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. New York: Addison-Wesley.

Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367–383.

Jonassen, D. H. (2000). Towards a design theory of problem solving. *Educational Technology, Research and Development, 48*(4), 63–85.

Jonassen, D. H., & Kwon, H. I. (2001). Communication patterns in computer-mediated vs. face-to-face group problem solving. *Educational Technology, Research and Development, 49*(1), 35–52.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424.

Kapur, M. (2009). Productive failure in mathematical problem solving. *Instructional Science*. doi:10.1007/s11251-009-9093-x.

Kapur, M. (2010). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*. doi:10.1007/s11251-010-9144-3.

Kapur, M., & Kinzer, C. (2007). The effect of problem type on interactional activity, inequity, and group performance in a synchronous computer-supported collaborative environment. *Educational Technology, Research and Development, 55*(5), 439–459.

Kapur, M., & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 21–46.

Kapur, M., Voiklis, J., & Kinzer, C. (2005). Problem solving as a complex, evolutionary activity: A methodological framework for analyzing problem-solving processes in a computer-supported collaborative environment. *Proceedings the Computer Supported Collaborative Learning (CSCL) Conference*. Mahwah, NJ: Erlbaum.

Kapur, M., Voiklis, J., Kinzer, C., & Black, J. (2006). Insights into the emergence of convergence in group discussions. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the International Conference on the Learning Sciences* (pp. 300–306). Mahwah: Erlbaum.

Kapur, M., Hung, D., Jacobson, M., Voiklis, J., Kinzer, C., & Chen, D.-T. (2007). Emergence of learning in computer-supported, large-scale collective dynamics: A research agenda. In C. A. Clark, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the International Conference of Computer-Supported Collaborative Learning* (pp. 323–332). Mahwah: Erlbaum.

Kapur, M., Voiklis, J., & Kinzer, C. (2008). Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups. *Computers and Education, 51*, 54–66.

Kauffman, S. (1995). *At home in the universe*. New York: Oxford University Press.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Student assessment of collaborative learning in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 57–87.

McGrath, J. E., & Tschan, F. (2004). *Temporal matters in social psychology: Examining the role of time in the lives of groups and individuals*. Washington: American Psychological Association.

Muukkonen, H., Hakkarainen, K., Konsonen, K., Jalonen, S., Heikkil, A., Lonka, K., et al. (2007). Process- and context-sensitive research on academic knowledge practices: Developing CASS-tools and methods. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Minds, mind, and society. Proceedings of the 6th International Conference on Computer-supported Collaborative Learning (CSCL 2007)* (pp. 541–543). New Brunswick: International Society of the Learning Sciences.

Olson, G. M., Herbsleb, J. D., & Rueter, H. H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. *Human-Computer Interaction, 9*, 427–472.

Poole, M. S., & Holmes, M. E. (1995). Decision development in computer-assisted group decision making. *Human Communications Research, 22*(1), 90–127.

Rabiner, L. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286.

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239–257.

Ross, S. M. (1996). *Stochastic processes*. New York: John Wiley.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology, Research and Development, 52*(1), 5–18.

Sanderson, P., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction, 9*, 251–317.

Scardamalia, M. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research, 5*(1), 51–68.

Scardamalia, M. (1992). Educational applications of a networked communal database. *Interactive Learning Environments, 2*(1), 45–71.

Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education, 46*, 349–370.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.

Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach to assessing knowledge sharing during collaborative learning activities. In G. Stahl (Ed.), *Proceedings of Computer Support for Collaborative Learning* (pp. 128–137). Hillsdale: Erlbaum.

Spada, H., Meier, A., Rummel, N., & Hauser, S. (2005). A new method to assess the quality of collaborative process in CSCL. *Proceedings of the Computer-supported Collaborative Learning conference 2005*, Taipei, Taiwan.

Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning, 21*, 79–90.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29–48.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 315–337.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). A framework for eclectic analysis of collaborative interaction. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Minds, mice, and society. Proceedings of the 6th International Conference on Computer-Supported Collaborative Learning (CSCL 2007)* (pp. 694–703). New Brunswick: International Society of the Learning Sciences.

Van Lehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249.

Voiklis, J., Kapur, M., Kinzer, C., & Black, J. (2006). An emergentist account of collective cognition in collaborative problem solving. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 858–863). Mahwah: Erlbaum.

Wampold, B. E. (1992). The intensive examination of social interaction. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 93–131). Hillsdale: Erlbaum.

Zumbach, J., Schonemann, J., & Reimann, P. (2005). Analyzing and supporting collaboration in cooperative computer-mediated communication. *Proceedings of the Computer-supported Collaborative Learning Conference 2005*, Taipei, Taiwan.