Metadata of the article that will be visualized in Online

1	Article Title		**Time is precious: Variable- and event-centred approaches to process analysis in CSCL research**
2	Article Sub- Title		
3	Article Copyright - Year		**International Society of the Learning Sciences, Inc.; Springer Science + Business Media, LLC 2009 (This will be the copyright line in the final PDF)**
4	Journal Name		International Journal of Computer-Supported Collaborative Learning
5	Corresponding Author	Family Name	**Reimann**
6		Particle	
7		Given Name	**Peter**
8		Suffix	
9		Organization	University of Sydney
10		Division	Research Centre for Computer-supported Learning and Cognition (CoCo)
11		Address	Education Building A35, Room 244, Manning Road (Darlington Campus), Sydney 2006, NSW, Australia
12		e-mail	p.reimann@usyd.edu.au
13	Schedule	Received	9 October 2008
14		Revised	
15		Accepted	29 May 2009
16	Abstract		Although temporality is a key characteristic of the core concepts of CSCL—interaction, communication, learning, knowledge building, technology use—and although CSCL researchers have privileged access to process data, the theoretical constructs and methods employed in research practice frequently neglect to make full use of information relating to time and order. This is particularly problematic when collaboration and learning processes are studied in groups that work together over weeks, and months, as is often the case. The quantitative method dominant in the social and learning sciences—variable-centred variance theory—is of limited value for studying change on longer time scales. We introduce the event-centred view of process as a more generally applicable approach, not only for quantitative analysis, but also for providing closer links between qualitative and quantitative research methods. A number of methods for variable- and event-centred analysis of process data are described and compared, using examples from CSCL research. I conclude with suggestions on how experimental, descriptive, and design-oriented research orientations can become better integrated.
17	Keywords separated by ' - '		Process analysis - Qualitative methods - Quantitative methods - Research methods
18	Foot note information		

Computer-Supported Collaborative Learning
DOI 10.1007/s11412-009-9070-z

Time is precious: Variable- and event-centred approaches to process analysis in CSCL research

Peter Reimann

Received: 9 October 2008 / Accepted: 29 May 2009

Abstract Although temporality is a key characteristic of the core concepts of CSCL—interaction, communication, learning, knowledge building, technology use—and although CSCL researchers have privileged access to process data, the theoretical constructs and methods employed in research practice frequently neglect to make full use of information relating to time and order. This is particularly problematic when collaboration and learning processes are studied in groups that work together over weeks, and months, as is often the case. The quantitative method dominant in the social and learning sciences—variable-centred variance theory—is of limited value for studying change on longer time scales. We introduce the event-centred view of process as a more generally applicable approach, not only for quantitative analysis, but also for providing closer links between qualitative and quantitative research methods. A number of methods for variable- and event-centred analysis of process data are described and compared, using examples from CSCL research. I conclude with suggestions on how experimental, descriptive, and design-oriented research orientations can become better integrated.

Keywords Process analysis · Qualitative methods · Quantitative methods · Research methods

Time (and order) matters

CSCL is concerned with technology-mediated learning as it takes place in groups. Independently of the context of the learning—on the level of the individual, the group, the situation, or in the interaction of these—the main object of analysis in CSCL is a process, something that unfolds over time. As Koschmann (2001) suggested, it might be a defining element of CSCL that it is about "...studying learning in settings in which learning is observably and accountably embedded in collaborative activity" and that learning within

P. Reimann (✉)
Research Centre for Computer-supported Learning and Cognition (CoCo), University of Sydney, Education Building A35, Room 244, Manning Road (Darlington Campus), Sydney, NSW 2006, Australia
e-mail: p.reimann@usyd.edu.au

these settings is to be conceptualized as an "unfolding process of meaning making" (p. 19). More recently, Stahl argues that one can meaningfully speak about group cognition as different from the sum of individual cognitions (Stahl 2006). This is substantially different from the psychological notion of learning as a basically unobservable process, taking place in the mind/brain, a process we can observe only indirectly by measuring learning outcomes. However, for both views of learning, the sociocultural as well as the individual-cognitive, the nature of the process remains temporal: Learning unfolds over time.

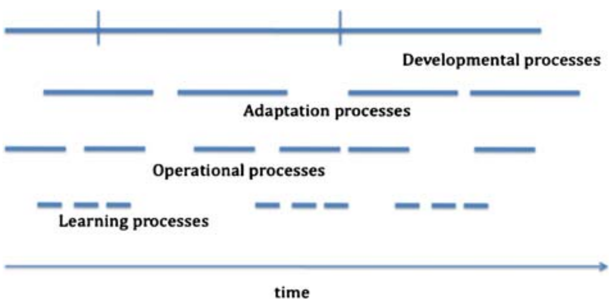
Temporality does not only come into play in quantitative terms (e.g., durations, rates of change), but *order* matters: Because human learning is inherently cumulative, the sequence in which experiences are encountered affects how one learns and what one learns (Ritter et al. 2007). This can certainly be generalized to learning in groups, and to the communication and interaction processes that take place in groups in addition to learning.

Groups are subject to, and subject themselves to, change processes of various kinds. In a book that is dedicated to discern these types of processes, McGrath and Tschan (2004) distinguish four categories: (a) developmental processes, which are inherent to the system; (b) adaptational processes "generated by the system's response to (actual or anticipated) changes in the embedding context" (p. 6); (c) learning processes, which are based on a system's experience and reflection thereof; and (d) the system's operational processes, actions, and activities, which are hierarchically and sequentially related. Learning is seen as different from adaptation in as much as it requires intentional reflection. We can speak, with McGrath and Tschan, of different types of "forces" that are responsible for these types of processes, but need to keep in mind that these forces refer to different types of causality. The developmental force would be akin to Aristotelian *formal* causality; adaptational forces are at least partially of the "push" causality type; the "operational" forces are mainly *teleological* in nature because they involve a strong element of goal orientation, of purpose.

All four sets of forces are intrinsically temporal, and can operate simultaneously, as illustrated with Fig. 1. While a group is performing a certain task, it is also in a certain developmental stage, reacting to environmental changes, and learning from aspects of the task performance. McGrath and Tschan see all forces as acting *continuously*, but I suggest reserving this assumption for the developmental forces only. While they see process in terms of variables, and are, hence, "forced" to assume continuity of causation, the event perspective on process introduced below allows us to relax this assumption.

Taking time and order into account becomes particularly relevant, but also more challenging, as the time frame considered for analysis grows. That CSCL is as much concerned with long-term collaboration as with short-term collaboration (e.g., talk) can be seen from an analysis of all empirical studies reported in the last two CSCL conferences (Chinn et al. 2007; Koschmann et al. 2005). As Table 1 shows, the majority of studies

Fig. 1 Types of processes effective in groups on different time scales



Q3

Table 1 Duration of group lifetime in studies from the CSCL 2005 and 2007 conferences

	CSCL 2005		CSCL 2007	
"Lifetime" of groups studied	No. of studies	Percentage	No. of studies	Percentage
Single session (20–180 min)	25	35%	32	45%
2–6 days	5	7%	6	8%
1–4 weeks	7	10%	10	15%
Longer (1.5 months–1 year)	34	48%	23	32%

analyze group interactions that extend beyond a couple of hours and almost 50% of the studies concern groups that learned together for more than a month (of course, the duration assessed is not commensurate to "time on task").

The sequential organization of behavior has received extended treatment in Conversation Analysis, with one of its founders recently developing a whole book to the subject (Schegloff 2007). Sequential organization refers "...to any kind of organization which concerns the relative positioning of utterances or actions" (p. 2). Turn taking, that is, the relative order of speakers, is a kind of sequential organization (of talk), and the most extensively studied one. However, methods developed in Conversation Analysis and ethnomethodology, with their focus on talk and embodied physical (inter-)action (Goodwin 2000), do not carry over to interactions that are fragmented over time, stretch over longer durations, or are mediated by artifacts rather than talk. In situations where the people meet repeatedly (e.g., teamwork instead of helpline conversations), it is not only the context that matters, but each group has a history, and this history affects their activities and their learning (McGrath and Tschan 2004). Mercer (2008) picks up on this in a recent publication for the case of communication and learning in the classroom: "Analytical methods that do not recognize or deal with the temporal development of talk, its reflexivity, and its cohesive nature over longer timescales than one episode or lesson will inevitably fail to capture the essence of the educational process" (p. 56).

In studies where interaction and learning is distributed over multiple sessions, establishing internal validity becomes difficult. For instance, as time increases, noncontrolled factors will come into play with a higher probability than is the case for short-term collaboration, and changes in group membership become more frequent, thus qualitatively changing the experimental "unit." Nonlinear changes will become more pronounced because of the self-sustaining feedback processes at work in groups over time (Arrow et al. 2000); that is to say, small differences can have large effects. Development in groups progresses in general in a nonlinear fashion, so that both the nature of the data as well as the nature of the underlying processes make it necessary to employ advanced statistical methods (Sloane and Kelly 2008). In general, order effects will become more pronounced as groups construct their histories and make use of them, through communication, as resources for interpreting events and planning future actions.

Challenges such as these might partially account for the fact that, 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. For instance, Kapur et al. (2005) made use of statistical analysis methods that take time into account, and Schümmer et al. (2005) employ a similarity based metric to identify the similarity of change processes in log files. Muukkonen et al. (2007) use time series analysis, and Suthers et al. (2007)

apply graphical techniques for analysis. There is perhaps a trend, as exemplified by the Muukonen et al. study as well as by Mochizuki et al. (2007), that CSCL researchers who employ mobile devices also use these devices to systematically gather data over time with more regular measurement intervals than is the case in studies with stationary technology such as PCs.

The goal of this paper is, in particular, to identify methods appropriate for the analysis of long-time changes (on a scale of days, weeks, months). The kind of data and the manner in which these data are analyzed and theorized, need to change substantially when moving from an analysis of short-duration sequences to long-term processes. One of the reasons being that group development processes come into play, another that multiple levels of analysis have to be considered now (Cress 2008; Sloane 2008). This requires extending the range of methods considered beyond those covered in reviews such as Sanderson and Fisher's (1994) or Olson et al.'s (1994). In Organizational Science as well as in Sociology, and in particular in History, the challenges of analyzing processes that capture longer stretches of time and develop on multiple levels have been intensively discussed, covering a wider range of methods than is typically done in psychology and education (see also Langley 1999).

Building on this literature, I argue in this paper that an event-based view of process and change is an important addition to the variable-centric approach. Variables are attributes of fixed entities defined by measurement (e.g., with a scale) or by a coding and counting procedure. It is important to realize that the decision to phrase research questions in terms of variables and relations between them is a very decisive one, because many other decisions depend on this one, both metaphysical (e.g., regarding type of causality) as well as methodological (e.g., methods of analysis) ones. Because variables are not the only means to formulate and test (quantitative) hypotheses about time-dependent processes and data, this paper develops the case for making more use of methods in CSCL research that take *events* as the basic unit of analysis. This not only allows us to include qualitative methods, but also to add additional quantitative and computational methods to the repertoire of CSCL.

The paper continues by further elaborating the difference between variable- and event-based approaches for the analysis of temporal data, at the same time illustrating some of the typical methods. I then look into the question of how one can establish causality in the event-oriented approach, and touch on the relation between explanation and generalization. How to generalize over event sequences is a topic I cover because this is more challenging than in the variable approach and because a comparison of the respective generalization strategies further helps to come to terms with the differences between the two approaches. Methods for sequence mining, pattern identification, and process mining are introduced, all helpful for the business of generalizing from individual sequences. I close by identifying opportunities for combining variable- and event-based methods and pointing to possible next steps to advance process research in CSCL.

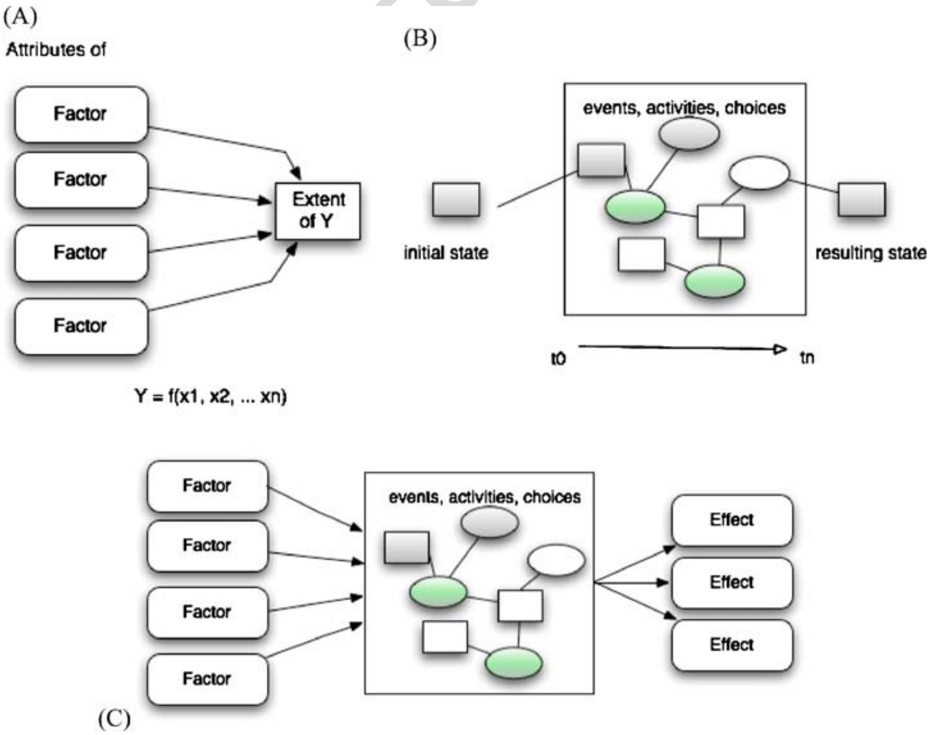
The unit of analysis: Variables versus events

Q1

In order to illustrate our discussion, let us sketch a hypothetical, but prototypical scenario. The situation that we want to address is one where the researcher is interested in interaction and learning processes as they take place in online groups over time. The researchers want to test a process theory, one that says that groups need to go through a cycle of definition, conflict, and synthesis repeatedly in order to successfully engage in and learn from discussion activities. Therefore, they have developed a coding scheme that can be applied to the content of the discussion board entries and categorize them in respect to these three dimensions. The coding scheme is developed and applied following best practice (e.g., Strijbos et al. (2006)). Let us further assume that the researchers are interested in design issues pertaining to the visualization of argument threads. For this purpose, they have developed a new version of the discussion board, one that includes a graphical display of the argument structure.

Our hypothetical research team has access to students in an online university course who are working together in several small groups. About half of the groups work with the old, run-of-the-mill discussion board, whereas the other half of the groups uses the new version. Data are recorded electronically in the form of the discussion board log file, so that we know who contributed what and when. Pre- and posttests are conducted to assess individual learning gains and during the pretest phase a number of other individual factors are assessed, including metacognitive capabilities. Knowledge building is assessed by analyzing the discussion board entries.

How these data are analyzed will depend largely on what the researcher considers to be the main unit of analysis. Two conceptualizations can be distinguished here. The first one, *variable-centred*, relates to analysis of variance. The second one we call *event-centred analysis* or *event analysis* for short. I use the terminology suggested by Abell (1987) and in particular by Poole et al. (2000), whose excellent treatment of process analysis in the social sciences informed many parts of this paper. Figure 2a, b depict graphically the difference between these two views, and Fig. 2c their combination.



Q3

Fig. 2 Illustration of the variable (**a**) and event approach (**b**) and their combination (**c**)

Variable-based account Q1

For the experimentalist, being trained in the variance method, a process takes the form of a *category of concepts* that mediate between independent and dependent variables. In CSCL, variables such as communication frequencies, learning techniques, and group decision-making techniques can play this role. Such "process concepts" are distinguished from other concepts considered to be static, such as individual learning capabilities, group makeup, or learning outcomes. A process theory for the experimentalist takes the form of a causal relationship between input and outcome variables mediated by process variables. The process concepts, like the static concepts, are operationalized as constructs and measured as variables, as fixed entities, the *attributes* of which can vary from low to high along numerical scales. A typical question that could be analyzed with this framework is the extent to which individual learning skills (exogenous independent variable) can predict learning outcomes (dependent variable), dependent on more or less successful group communication (endogenous independent variable).

For our scenario, the initial analysis would be fairly straightforward. The experimentalist would "code and count": code the data stored in the discussion board log, and count, yielding frequencies for the process categories (definition, conflict, synthesis). Then these measures can be set in relation to the treatment (tool variation) as well as in relation to other variables assessed, in particular to the dependent variables: individual learning and group knowledge building. A typical analysis of variance would yield results that show if the difference in the dependent variables can be related statistically to the variation in the tool, if this relation is mediated by the process variables, and if there are (statistical) interactions with the other variables assessed (for instance, metacognitive competence).

In order to test the process theory in more detail—which says that we should see, in successful groups, cycles of issue definition followed by conflict among positions followed by synthesis/integration of positions—the researcher could treat each of these categories as a variable, using the categories' frequencies assessed at regular intervals (daily, say) as the quantitative attribute, and treat them as three time series. For each individual time series, curve fitting can be performed to test if they form a sine wave—as they should if the assumption of "repeated cycles" is correct. Having established this (and, before that, having established that the time series variables follow approximately a Gaussian distribution), the researcher could go ahead and use multivariate time series (ARIMA) models to test the dependencies between the three time series (they should follow each other and "peak" with a certain time lag, but in the order definition-conflict-synthesis) and to test if and to what extent extraneous factors, in particular, the type of discussion board, affect the time series. Based on the same logic, one could also look for the effects of differences between groups (using a criterion for "successful" and "less successful" groups, for instance) and for differences between individuals (using metacognitive competence as a criterion, for instance).

There is neither need nor space for statistical details here (see e.g., Box and Jenkins 1976). Instead, a word on the assumptions behind the variable-centred research method may be in order. A basic assumption that underlies any research logic based on the analysis of variance is that independent variables are acting *continuously* on the dependent variables. I would argue that this basic assumption is, for CSCL scenarios, often not met. Obviously, students in our scenario will, over the duration of the semester, do many things other than the type of activities captured by the measurements. Even when they are actively engaged online, only a small set of the factors represented as independent variables might be effective at any point in time; for instance, the students using the enriched discussion board might not attend to the information offered on the visualizations. This *fragmented* nature of the underlying causal processes is not easily captured in variable-centred models.

Another thorny problem in process studies arises from the fact that all variables must be measurable at the same time point, and the temporal unit or measurement must be equal for all variables (*minimal unit of time*). Because we will find, in any group, processes unfolding on different timescales (McGrath and Tschan 2004), relating them in one model is a challenge indeed. As was mentioned before, the variable-centred method cannot accommodate qualitative changes in the variables. For instance, when a group loses a member or gets a new member, it is not clear if variables that build on group activities can be considered to be qualitatively the same as before.

Event-based account

Processes can be analyzed with statistical methods that do not require the data to be represented as variables. An example for such *stochastic* methods is Markov Chain modeling. Stochastic modeling methods have a fairly long tradition in the social sciences and psychology, for example, Coleman (1964), and Suppes and Atkinson (1960), yet are not as widely taught and used in learning research as are variance analysis methods and other members of the General Linear Model family.

This is not the place to introduce stochastic modeling in any detail, but in order to provide a flavor, a simple example might be appropriate. Let us again assume that we want to test if the life cycle model that presupposes that (successful) groups will go through a cycle of Definition-Conflict-Synthesis is supported by the data. One can also see this as a dialectical model if the cycle is not imposed on groups by the pedagogical design or strongly afforded by tool design but emerges out of the interactions. We could have coded incidents directly in these terms, yielding a event sequence in each group of a form like DDDCCDCCSCCSSSS… , with D for Definition, C for Conflict, S for Synthesis. To test if this mini-theory describes the behavior in the groups adequately, one could use a Markov Chain model. Markov chains belong to the class of homogenous Markov models, which are appropriate for cases where time can be considered as consisting of discrete intervals and where the only aspect we need to know about an event is when it was present in time. Being stochastic, Markov models do not predict the occurrence of a specific event, but predict the probability distribution of a set of possible events at a given point in time. The Markov chain predicts the probability of occurrence of an event at time t as a function of the event occurring immediately before. No other information is taken into account.

A more complex, but also more realistic case is one where we do not define events in terms of the comprehensive descriptors (Definition, Conflict, Synthesis) directly, but code on a finer level of analysis. For instance, we could code the interactions in the groups with a taxonomy that is inspired by speech act or dialogue act theory (adapted to the asynchronous case). We would use, say, a coding scheme with 12 different categories, c1 to c12 (omitting any further details here). We would then look at sequences in the groups of the form like … c3c1c1c5c3c12c3c6c6c6c1c2c6…. To test our mini-theory of the three phases in this case, phasic analysis (e.g., Holmes 1997) or Hidden Markov modeling (Rabiner 1989) could be used.

These matters cannot be discussed further here (see Soller et al. 2002 for an example of Hidden Markov modeling in CSCL). Suffice it to say that further generalizations of Markov models have been developed. For instance, nonhomogeneous Markov processes add variables other than the events to the model. With them, we could test if the two tool conditions (conventional vs. enhanced discussion board) make a difference, or if individual differences add predictive power. So called semi-Markov process models allow information about the *duration* of events to be included (still assuming discrete event).

Like variable-based modeling, Markov models entail the assumption that history does not matter: "The entire influence of the past occurs through its determination of the immediate present, which in turn serves (via the process) as the complete determinant of the immediate future." (Abbott 1990, p. 378). Histories are a kind of "surface reality" (Abbott) that are generated by deeper, underlying probabilistic processes that find expression in the value of variables or the conditional probabilities of event transitions. In the variable-based case, this "deep structure" is expressed in terms of linear transformations; in the event-based case, as transition probabilities. For situations where history (and/or anticipated future) does matter, we need to find different forms of modeling a process.

There are other views of process as event sequence that do not depend on this limited view of history entailed in the Markov assumption. These will be discussed further below. But before that, I will further elaborate the difference between the variable- and event-based views of process.

By using stochastic modeling, an important decision has been made: The phenomenon under study is not phrased in terms of variables and their relations. We are not primarily looking at how quantitative attributes change their value over time, but deal with a (constructed) event sequence directly. The limitations of the variable-centered approach (in the social sciences) to describe change processes are mainly due to a restricted view of causation. Independent variables are seen as "acting on" dependent variables; the underlying process is supposed to operate continuously over time; the nature of the variables does not change over time—all that can change are the values of the quantitative attributes used to operationalize the variable—and no qualitatively different kinds of forces are deemed necessary to explain changes in the dependent variables. If too much variance remains unexplained, one has to look for additional independent variables and/or include specifications of relationships (statistically: interactions) between the variables. The underlying notion of causality is *efficient causality*, the "push" type causality that has been so instrumental for theories in physics.

To account for group (and in general, for social) phenomena, a process method should, in addition to *efficient cause*, be able to deal with at least two other kinds of causes (of the four Aristotle identified overall [Aristotle 1941]), namely: *formal* cause, referring to the patterns of which things are made, and *final* cause, the end for which things are made, or a teleological "pull." In groups, formal causality is at work whenever constraints—as imposed on them in terms of workflow, scripts, or roles—are effective. For instance, many events taking place in online learning groups are a consequence of the manner in which groups have been set up (scripts, roles, workflow, deadlines). In organizations, the way team members interact with each other and with other teams is to some extent affected by the organizations' design and their business processes, all best captured as *formal cause*, and not requiring reduction to efficient causes (where the invariants and the explanatory power would be lost because many efficient cause processes can instantiate a single formal cause relation). Similarly, explaining human behavior (in various levels of aggregation: individuals, pairs, groups, and larger structures) in terms of *goals,* that is, driven by an *end,* adds considerable explanatory power, in particular for the (rather typical) cases where a goal can be reached in many different ways. Any account of these different paths toward an end in terms of only efficient causality would fail to identify the goal orientation.

Viewing a process in terms of sequences of events provides space to consider all four kinds of causality: efficient, formal, final, and material. Efficient cause can be modelled in terms of variables defined over events, formal cause in terms of event configurations (such as routines, procedures, practices, scripts) and final cause in terms of narrative structures that capture the goal orientation (purpose, ends) of a sequence. Material causes can be expressed in terms of constrains that the physical environment (e.g., architectural features) impose on actions and sequences of actions.

A pivotal difference to the variable-centred method is that event analysis does not start by framing the world in terms of variables, that is, fixed entities with varying attributes. Instead, event analysis "...conceptualizes development and change processes as sequences of events which have unity and coherence over time" (Poole et al. 2000, p. 36). While variable- and event-centred analysis can be combined (see below), conceptually they are quite different and these differences are important to keep in mind (Mohr 1982).

What counts as an event is basically up to the researcher, constrained by theory and informed by research goals; events are not "raw data," or incidents. In particular, events need to be defined dependent on the identification of the *central subject* under study because *entities participate in events* (for a more systematic treatment of the process of defining events, or colligation, see Abbott 1984). The central entity in event analysis is some kind of "actor," but the "actor" does not have to be a person; it can also be a group, an organization, a nation, an idea, a technology—dependent on research question and level of analysis. I will not go into more details with respect to event coding here, because this kind of content analysis is well understood and has recently been the subject of methodological reflection in CSCL (Strijbos et al. 2006; Wever et al. 2006). Although I gloss over these issues, it needs to be kept in mind that the conceptualization of what counts as an event and what event types to distinguish, as well as the measurement of the occurrences of an event—which is a theoretical structure, a concept, hence not "identical" with its occurrences—play a critical role in the research process because they determine to a large extent the quality of the analyses and reflections that build on observations of events.

In our hypothetical CSCL study scenario, the main entities are individuals and groups. That implies then that events are constrained to those incidents in which either individuals or groups can participate. For our scenario, a process researcher would focus on the sequences of activities, incidents, crises, or stages that unfold in the groups over the duration of the semester. An explanation for an observed chain of events would take the form of a narrative that explains how event e(t) is related to events e(1) ... e(t-1) in terms of the actors' goals, motives, moves, and so forth, and would keep track of how events happening outside the groups might affect them. The process is conceptualized here as a *developmental event sequence*, not a change in values of process variables. The research process yields a kind of narrative for each case, a case being a single person or a group, dependent on the level of analysis chosen.

Causation and generalization in the event-centred approach Q1

While there can be little doubt that the event-based ontology of process and change has merits, it is less than straightforward to employ it as an explanatory device: In what sense can we say that a sequence of antecedent events can explain a certain state of the world? In other words, how can we establish the claim that a chain of antecedent events causes a target event? Questions of causation are tightly linked to generalizations.

Causation: Covering laws versus narrative structures Q1

A major advantage of the quantitative-experimental, variable-based approach is that generalizing is straightforward and testing for the validity of the generalization is well understood. In a sense, generalizations are built in right from the start, as soon as variables are used and as soon as explanations take the form of quantitative relations between antecedent and dependent variables. Furthermore, the General Linear Model sees the value of a particular variable y as a function of a set of antecedent variables x_1 to x_n, plus an error term: $\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{e}$ (with **y**, **b** and **e** being vectors, **X** a matrix with dimensions number of cases (m) * number of antecedent variables (n)). While the value of y for each case may vary, dependent on the values of the antecedent variables x_1 to x_n (captured in the vector **b**), the relation between y and the antecedent variables is the same for all cases. This is, of course, nothing else than a particular mathematical form of the *covering law* principle that sees a particular observation as explained if (and only if) this observation can be derived from a general law (in the quantitative realm: if the value of a particular variable for a particular case can be expressed as a function of a set of antecedent variables).

How do we establish causality in the event-based case? In this case, we can explain the event y as being *brought about* by a sequence of events, but this kind of explanation is obviously very different from employing a covering law. In particular, the causality at work here is not of a hypothetico-deductive or inductive-probabilistic covering law model type, but one of action causality (Abell 2004). This type of causality can be invoked when changes in the world are linked together by (human) actions. To the extent that one has evidence that a state of the world is transformed through the direct or indirect evidential action(s) of individual or collective agents, the causality in the particular case has been observed (Abell 2004, p. 293). Instead of a covering law, a *narrative structure* is invoked in order to establish causality. This kind of explanation is typically not used in a predictive manner, but the narrative formulation takes place after the transformation of world states is observed.

As Abell (2004) observes, the pivotal difference between the covering law model and the narrative structure one (or between nomothetic and idiographic causality) is that in the covering law case, proper explanations are only possible *after* generalizations and comparisons have been performed, whereas in the narrative account the explanation (for a specific case) comes *first*, followed by attempts to generalize to other cases (if and when the researcher is inclined to do so).

The issue of single-case causality has received extensive analysis (for an overview see Danto 1985) that has led to a certain consensus amongst philosophers of science that singular causes might exist. More problematic than this ontological claim is the epistemological aspect (Abell 2004, p. 294): How can we ground claims that we *know* they exist? In other words, how can we distinguish, for a single case, between a *consequence* and a mere sequence? Before that: Where does a consequential chain of events begin? It turns out that answering such questions typically requires referring to generalizations.

Hence, even in cases where the covering law model is (for good reasons) not accepted on ontological grounds, it is difficult to avoid generalizations and case comparisons altogether when one wants to establish claims regarding causal connections between events. Different from the variable-based approach, for the event-based approach generalizing does not come "automatically" and it is not as "straightforward." In particular, a dimension or metric, a distance measure, needs to be established along which to generalize from single sequences to patterns.

Generalizing by pattern extraction Q1

Abbott (1990) provides an overview of methods useful for finding patterns in sequence data, distinguishing between methods employing or not employing *inter-event distance* *measures*. An example for the later is calculating a simple Spearman rank correlation coefficient as a measure of the resemblance of one sequence to another. Repeating this for all pairs of observed sequences yields an *inter-sequence* distance matrix that can then be subjected to any standard classification technique, such as cluster analysis, to identify groups of sequences. Calculating the correlation coefficient as a direct measure of similarity between two sequences is, however, only meaningful for *non-recurrent* sequences in which every event is observed once and only once. Furthermore, as permutation statistics, Spearman rank correlation coefficients and similar measures have problems with extensive ties and missing events.

The already mentioned Markov modeling method is an approach to identify patterns in sequences that can be applied to *recurrent* sequences without employing any notion of inter-event distance. We have described its main assumption already—sequences are explained as random realizations of an underlying stochastic process—and need to mention here that in order to estimate the parameters of a Markov model, large data sets are needed (Kemeny and Snell 1976).

However, most pattern searching methods for sequence data build on an inter-event distance matrix. There are three general ways to measure inter-event distance (Abbott 1990). Firstly, one can use temporal distance between events across cases. This is often done for non-recurrent sequences. Secondly, one can use categorical resemblance and measure inter-event similarity analogous to kinship in a family tree. This can be done for recurrent sequences, where temporal distances are problematic to use as a distance measure, but requires, of course, that a (hierarchical) category system for event coding exists and can be reliably applied. The third type of distance measure builds on sequence transformation costs, using so-called *optimal matching* or *alignment* techniques. These can be applied to recurrent sequence data, and have seen widespread use in (molecular) biology (Miura 1986). The main idea is straightforward: For any two sequences, the distance between the two is determined by calculating the "cost" of transforming one (by insertion, deletion, and substitution) into the other. Different costs can be associated with the three types of transformations, and/or with the event types subject to the transformation. Also, the total costs of a sequence transformation can be combined algebraically in different forms (e.g., total, mean,…). In any case, the resulting distance matrix can be used for classification (e.g., clustering) as well as scaling (e.g., Multi-Dimensional Scaling) to identify families of sequences and dimensions of differences, respectively (Abbott and Hrycak 1990).

Pattern extraction is one way of generalizing from particular event sequences while sticking to an event ontology: The generalization is accomplished without using variables, that is, attributes of an event. I want to introduce another approach—Process Modeling—that can be used for the same purpose, but is in interesting ways different from pattern analysis: It can deal with information about concurrent events (parallelism), and it employs two levels of description, a *model* of a process, and *instances* of the model.

Generalizing by process modeling

Process Models are interesting conceptually because they describe processes holistically, incorporating a priori assumptions about the form a process and all its instantiations can take. This makes Process Models suitable to describe *designed* processes, with the design effecting process enactment through prescriptions (e.g., collaboration scripts) and/or through constraints built into the collaboration software (e.g. an argumentation ontology, or specific features in the user interface). Process Models are interesting, furthermore for practical reasons, as they can under certain circumstances be identified automatically from log data.

A Process Model in the meaning intended here is a formal model, a parsimonious description of all possible activity sequences that are compatible with a model. (Note that I use capitals to distinguish Process Modeling/Models from other forms or modeling process, such as mathematical ones). Processes can be modeled in many forms, for example, using a system dynamics formalism for continuous process models. The class of Processes Models that I want to concentrate on here pertain to the large class of discrete event systems (Cassandras 1993). Finite state machines are one type of modeling language that can be used to describe and analyze discrete, sequential-event systems (Gill 1962). Another one is the language and theory of Petri nets (Reisig 1985) which present the advantage of modeling concurrency in addition to sequentiality.

Petri nets can be mathematically described as bipartite directed graph with a finite set of places P, a finite set of transitions T, both represented as nodes (round and rectangular, respectively), two sets of directed arcs, from places to transitions and from transitions to places, respectively, and an initial markup of the nodes with tokens (usually representing resources). The Petri net shown in Fig. 3 for instance, expresses the fact that all process instances start with A and end in D. It also expresses the fact that the only predecessor to B is A, the B can only be followed by D, and that possible predecessors for D are B, C, and E. Furthermore, it shows that B, C, and E can be executed in parallel, or in any order. The black token in the initial node represent a token, which enables the transition A to be fired. Petri nets are nondeterministic but a transition can only be fired if all the predecessor nodes have at least one token. (Two "technical" transitions are included in the net, and *And Split* (AS) and an *And Join* (AJ) in order to express formally the parallelism between activities B and C.)

Process Model representations that take the form of Petri nets and similar formalisms have several interesting features. For instance, since they have formal semantics, they can be used to determine computationally if a specific activity sequence is commensurate with a model or not; like a grammar, a model can "parse" an activity sequence. For the same reason, one can use them to simulate potential (non-observed) model behavior computationally, and to compare different models with respect to certain formal parameters. The fact that they come with a graphical notation can be exploited for learning purposes: The graphical representations could be made an object for comparison and reflection for the group members, that is, serve as a mirroring or feedback device (Kay et al. 2007).

In terms of the terminology introduced in this paper, Process Models (e.g., expressed as a Petri net) constitute a holistic view of a process: A process has a beginning and an end, it comprises events (activities), and the possible event/activity sequences are subject to more or less numerous constraints. Even a simple Petri net is a basic, but powerful language to represent, for instance, the logic of a group script. While Petri nets are one out of many possible formalisms to express a process succinctly, they have another advantage: They can be automatically discovered from performance data.

A specific class of data mining methods can be applied in situations where we can expect that a group realizes a multistep process over time. This would be the case, for instance, when the group behavior is controlled by a script (Dillenbourg and Hong 2008;

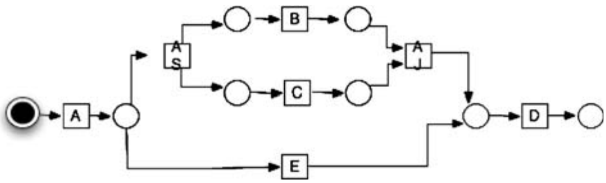


Fig. 3 Example for a Petri net description of a process

Kollar et al. 2006), or when the nature of the task suggests a specific sequence of activities, such as phases of a decision-making process (Poole and Roth 1989). Process Model mining (or process mining, for short) assumes that (a subset of) observed activities can be related to one or more processes, or in other words that (a subset of) observed activities constitutes an instance of a process. Collaboration scripts frequently used in CSCL, for instance, can be seen as processes, and the activities performed by students enacting the scripts being the process instance. Another example: normative models of group decision making can be seen as constituting processes, the enacted decision processes being instances thereof.

Process mining can serve a number of purposes, among them: (a) Discovery—No a priori model exists. Based on an event log, a model is constructed; (b) Conformance—An a priori model exists. Event logs are used to determine the extent to which the enacted collaboration corresponds to the model; (c) Extension—An a priori model exists. The goal is not to test but to extend the model, for instance with performance data (e.g., durations of activities). Extended models can then be used, for example, to optimize the process (Van der Aalst and Günther 2007).

We look here only at the discovery task because it is conceptually and computationally the most demanding one, although conformance checking is of obvious relevance, for instance in the context of our hypothetical study. The input to process mining is an event log, as shown in hypothetical form in Fig. 4. The result of process mining is a process model, for instance, a Petri Net as shown in Fig. 3.

The cases in the event log refer to different instances of the same process. Different cases can result from different groups enacting the same process, or the same group enacting the process at different times. The example event log in Fig. 4 illustrates the later case: A team formed by six members enacts a process composed of activities A–E five times. For example, if we look at the first enactment, it takes the form ABCD. The fifth enactment takes the form AED. The Petri net in Fig. 3 is a formal representation of the process logic underlying the activities depicted in the event log. For instance, it expresses the fact that all process instances start with A and end in D. It also expresses the fact that the only predecessor to B is A, the B can only be followed by D, and that possible predecessors for D are B, C, and E. Furthermore, it shows that B, C, and E can be executed in parallel, or in

Q5

Fig. 4 An event log example (from Van der Aalst and Günther 2007)

case id	activity id	originator	time stamp
case 1	activity A	John	9-3-2004:15.01
case 2	activity A	John	9-3-2004:15.12
case 3	activity A	Sue	9-3-2004:16.03
case 3	activity B	Carol	9-3-2004:16.07
case 1	activity B	Mike	9-3-2004:18.25
case 1	activity C	John	10-3-2004:9.23
case 2	activity C	Mike	10-3-2004:10.34
case 4	activity A	Sue	10-3-2004:10.35
case 2	activity B	John	10-3-2004:12.34
case 2	activity D	Pete	10-3-2004:12.50
case 5	activity A	Sue	10-3-2004:13.05
case 4	activity C	Carol	11-3-2004:10.12
case 1	activity D	Pete	11-3-2004:10.14
case 3	activity C	Sue	11-3-2004:10.44
case 3	activity D	Pete	11-3-2004:11.03
case 4	activity B	Sue	11-3-2004:11.18
case 5	activity E	Clare	11-3-2004:12.22
case 5	activity D	Clare	11-3-2004:14.34
case 4	activity D	Pete	11-3-2004:15.56

Q3

any order. It is assumed that two activities are parallel, or concurrent, if they appear in any order in the log.

An important difference between visualizations resulting from process mining (as in Fig. 3) and visualizations such as an uptake graph (Suthers 2006) is that the former are constructed automatically, whereas most visualizations used in qualitative research, such as uptake graphs, are constructed manually. Generating graphical process representations automatically has the obvious advantage of saving researchers' time, and in addition opens up the possibility to use them as a resource in the hand of teachers and teams. However, the transformation of input data from a log file into a meaningful process representation can, at this stage, not be fully automatized unless the event data come from a highly structured workflow environment. For the kind of data typical for CSCL research, in most cases various steps of data cleaning, event identification performed by human raters, and tuning of parameters of process mining algorithms are required. Furthermore, for "real" CSCL data process model types such as Petri nets with a formal semantics will regrettably not be suitable, among other reasons, because they are overly deterministic. One will have to employ heuristic methods, which are more complex both algorithmically and with respect to interpretation of results (for an example involving the analysis of chat data, see Reimann et al. 2009).

These practical obstacles not withstanding, process modeling in the sense introduced here can be an interesting component in the methodological arsenal of the CSCL researcher because it combines a holistic view of a process with graphical representations on an algorithmic basis (with at least in some cases clear semantics). When used in an inductive, process mining mode, it adds to the repertoire of sequence mining methods applicable to CSCL data (Kay et al. 2006), but the approach can also be used to formulate process models and test them in a more deductive fashion, typical for experimental studies. For those CSCL researchers particularly interested in collaboration scripts, process models offer ways to formulate scripts as well as to computationally assess, based on log data, to what extent scripts have been enacted.

Combining variable-centred and event-centred methods Q1

The variable-centred approach works well for research questions that involve relationships among variables. An event analyst has nothing against variables, as long as they are not seen as the *only* way to describe and explain change. I already mentioned that stochastic event sequence analysis can incorporate information that takes the form of values of variables by employing non-homogeneous Markov models. However, the potential for method integration is not exhausted there. While process analysis makes use of stochastic modeling methods because they use event type directly and thus preserve the nominal character of events and the integrity of event sequences unfolding over time, it can also employ *event variables*. Event variables are quantitative aspects of events, such as duration and intensity, or any other quantitative dimension that can be associated with an event. For such variables, variants of time series analysis can be used. Also, variables can be used in process research that describe the *characteristics of event sequences*, such as their periodicity, and these variables can figure as independent or dependent variables in theories of how such characteristics affect outcomes or are affected by other factors, respectively.

Since event analysis is more of a generalization of, rather than an antagonist to, the variable-centred method, experimental design with its meticulous control of external variables can be integrated. This is important for CSCL when we are interested in experimental trials of pedagogies and technical tools. There is no reason why such treatments should not be realized and included in process analysis, both in its narrative part as well as in the statistical analysis. Variables can represent contextual factors, and/or experimental conditions. What event analysis reminds us, though, is that we should not harbor overly simplistic assumptions as to the causal relations between "treatments" and groups' behavior, in particular when groups interact with technology over longer stretches of time.

Conclusions

Starting from the observation that the analysis of change processes—in individuals in the form of learning, in groups in the form of participation and knowledge building—is a central concern for CSCL and that CSCL researchers have privileged access to detailed change data, we have noticed a lack in the use of (formal) methods that take the core dimension of change—time—into account. This is a particular concern in light of the fact that the majority of studies conducted in CSCL deal with change processes that have a duration of weeks and months. If individual and group processes are analyzed on such a scale without taking into account history, sequence, dynamics, in short: time, then many of the resulting findings are of limited value. I argued further that for studies that aim to analyze change unfolding over days, weeks, and months, the quantitative method dominant in the social and learning sciences—variable-centred variance theory—is of limited value, not only because of the problems arising from "controlling" extraneous variables over longer stretches of time, but more importantly because of problems with the fundamental notion of variable and process. Methodologies that focus solely on order in short-term interactions, such as Conversation Analysis, are also not applicable to the analysis of processes that unfold over long-term, and fragmented, forms of interaction.

Therefore, I introduced a more general process approach that builds on the notion of event and narrative explanation. CSCL research can gain from an adoption of a wider range of process methods in a number of ways. By the adoption, group process research gets a sound methodological foundation, descriptive and experimental approaches can be better integrated, and insights informative for design can be derived. As has been the main argument on these pages, the variable-centred method, dominant in most experimental learning research, is not the only method for conducting (formal, quantitative) process research in CSCL. It makes many restrictive assumptions on the kind of data useful for analysis (namely variables only) and on the kinds of causation allowed to explain change. Adapting the more general stance to process analysis described above, we gain a more widely applicable yet by no means less rigorous method to analyze group processes.

Event analysis holds the potential to provide a methodological link between those researchers in CSCL who are producing descriptive, "thick," interpretive accounts of observations on learners' computer-mediated interactions, and those in the research community who work experimentally and quantitatively. The link results mainly from the fact that the event-centred approach makes extensive use of event descriptions: They enter into narrative accounts and, optionally, into statistical analysis without losing their distinctiveness. Hence, independent of the research orientation (descriptive, experimental, design-oriented), activities such as defining, identifying, distinguishing events and event sequences as well as providing qualitative, narrative accounts of events and sequences are part of a common set of research activities and become shareable. The fact that there are many common elements to the research "work" across different epistemological orientations is better exploited in the event-view of process than with the variable-centred perspective alone.

By the same token, the event-centered method can contribute substantially to design-oriented research. A comprehensive, detailed descriptive account of how individuals and groups interact with technology over time is an important component to inform instructional and software designers in the early stages of the development process, and it provides opportunities in the trial phase to gauge for (positive as well as negative) side effects of introducing methods and technologies. An example for the value of employing (qualitative) process studies for information technology design is the research on structuration and appropriation processes (Poole and DeSanctis 2004). But it needs to be said that this line of research has less implications for interface design than for organizational design and change management.

Analyzing the effects of specific tool and design decisions over longer stretches of time is also important for a realistic assessment of costs and benefits; for instance, Zumbach and Reimann (2003) observed that providing feedback to group members on interactional aspects was much more effective in the early stages of groups' lifetime than later and that, hence, this information should be phased out over time in order to reduce the cognitive load (the "costs"). Still, the contribution to design, in particular to "interface" design, is the least satisfying aspect of the strategy for method combinations suggested here. While researchers both in the nomothetic and idiographic tradition might appreciate some of the suggestions, the Great Unified Methodology for CSCL that pays due respect to all three epistemic orientations—nomothetic, idiographic, and design-perspective—is not identified here.

While there is ample concern for sequence data analysis in psychology (and to some extent in CSCL research), the analysis of long-term change processes is mainly taking place in disciplines such as organizational research and history, as well as developmental psychology. However, understanding organizational change processes and how they affect and are affected by collaborative technologies will become very important when (and if) CSCL follows the proposal that CSCL needs to concern itself more with processes that take place on a *meso level*, a level "...intermediate between small scale, local interaction and large-scale policy and institutional processes" (Jones et al. 2006, p. 37). In general, when collaboration tools are used over extended periods of time, as they increasingly are, due to the ubiquity of technology for collaboration and learning, then knowledge about how our technologies and tools affect individuals and groups over time becomes essential. As we move out of the laboratory and provide people with tools for their daily use, some of the most interesting processes are those that unfold over time (such as appropriation moves, Poole and DeSanctis 2004). They are not observable in the usability lab or the short-term study looking into second-by-second interactions and immediate (learning) effects.

Researchers in the learning sciences, education, and psychology know of the pivotal role of time and process in their areas of research. Logistical hurdles have been reduced to a significant extent, certainly in CSCL, where recording collaboration automatically is the rule rather than the exception. This does by far not solve all questions of data acquisition over time, but it provides a good basis for progress (Markauskaite and Reimann 2008). Problems remain in research training: The almost exclusive focus on variable-centred methods in quantitative training, as well as the almost total lack of concern for formal analysis of qualitative data are both not productive. Problems remain in the area of research dissemination and publication. Editors and reviewers of the leading journals need to be aware of methodological alternatives to canonical quantitative and qualitative research, and be perhaps more critical when evaluating claims regarding the analysis of "process." Longitudinal research designs, including design-based research are hard to fit into the

 JrnlID 11412_ArtID 9070_Proof# 1 - 02/06/2009

conventional journal paper format, in particular when involving case studies, data in multimedia format, and narrative accounts. We cannot let it come to the point where the last step in the research process, the formal publication, blocks innovation in research methods and strategies. Publishers need to extend their services, and need to extend the very notion of what it means to "publish," or they will be increasingly side-stepped. Problems remain regarding the adequacy of modeling methods for dealing with the complexities of human communication, cognition, and group behavior. However, different from the first two problems areas, these are "productive problems": They drive the research process forward.

Time is indeed precious. Too precious to be ignored or not treated adequately when formulating and testing theories of working and learning collaboratively. But the time of CSCL researchers is also precious; process studies are very work intensive, thus any method that can help us to share the workload and to conduct research cooperatively across epistemic interests and paradigms, without forcing us to gloss over fundamental differences, should be welcomed by the field. As a next step, shared online collections of (annotated) sequence data could be created that can be analyzed from multiple perspectives and with various methods or tools. The time gained might be most profitably spent on developing and testing process models and theories, of which there is a genuine lack in CSCL. While this paper has little to say on substantive theories of change in (learning) groups, it is obvious that existing process models in CSCL, which are predominantly describing short-term interactions, will need considerable theoretical extensions to connect with theories of long-term change.

Acknowledgements I would like to thank Anindito Aditomo and Fides-Ronja Voss for conducting the duration analysis on the CSCL 2005 and 2007 conference papers, and a number of colleagues for feedback on earlier versions of this paper.

References

Abbott, A. (1984). Event sequence and event duration: Colligation and measurement. *Historical Methods, 17*, 192–204.
Abbott, A. (1990). A primer on sequence methods. *Organization Science, 1*(4), 375–392.
Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology, 96*, 144–185.
Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives*. Oxford: Clarendon.
Abell, P. (2004). Narrative explanations: An alternative to variable-centered explanation? *Annual Review of Sociology, 30*, 287–310.
Aristotle. (1941). In R. McKeon (Ed.), *The basic works of Aristotle*. New York: Random House.
Arrow, H., McGrath, J. E., & Behrdal, J. (2000). *Small groups as complex systems: Formation, co-ordination, development and adaptation*. Thousand Oaks: Sage.
Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
Cassandras, C. G. (1993). *Discrete event systems*. Homewood: Richard D. Irwin, Inc.
Chinn, C., Erkens, G., & Puntambekar, S. (Eds.). (2007). *CSCL 2007–The Computer Supported Collaborative (CSCL) Conference 2007 (Rutgers, NJ, July 16th–21st, 2007)*: International Society of the Learning Sciences.
Coleman, J. S. (1964). *Introduction to mathematical sociology*. New York: Free.
Cress, U. (2008). The need for considering multilevel analysis in CSCL research–An appeal for the use of more advanced statistical methods. *International Journal of Computer-supported Collaborative Learning, 3*, 69–84.
Danto, A. C. (1985). *Narration and knowledge*. New York: Columbia University Press.
Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-supported Collaborative Learning, 3*(1), 5–24.
Gill, A. (1962). *Introduction to the theory of finite-state machines*. New York: McGraw-Hill.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522.

Holmes, M. E. (1997). Optimal matching analysis of negotiation phase sequences in simulated and authentic hostage negotiations. *Communication Reports, 10*, 1–9.

Jones, C., Dirckinck-Holmfeld, L., & Linstroem, B. (2006). A relational, indirect, meso-level approach to CSCL design in the next decade. *Computer-supported Collaborative Learning, 1*(1), 35–56.

Kapur, M., Voiklis, J., & Kinzer, C. K. (2005). Problem solving as a complex, evolutionary activity: a methodological framework for analyzing problem-solving processes in a computer-supported collaborative environment. In T. Koschmann, D. Suthers & T.-W. Chan (Eds.), *Computer-supported collaborative learning 2005: The next 10 years!* (pp. 252–261). Mahway: Erlbaum.

Kay, J., Maisonneuve, N., Yacef, K., & Zaiane, O. R. (2006). *Mining patterns of events in students' teamwork data.* Paper presented at the Proceedings of Educational Data Mining Workhop held in conjunction with ITS'06, Jhongli, Taiwan.

Kay, J., Yacef, K., & Reimann, P. (2007). Visualisations for team learning: small teams working on long-term projects. 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. 351–353). New Brunswick: International Society of the Learning Sciences.

Kemeny, J. G., & Snell, J. L. (1976). *Finite Markov chains*. New York: Springer.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychological Review, 18*, 159–185.

Koschmann, T. (2001). *Revising the paradigms of instructional technology.* Paper presented at the Meeting at the Crossroads. 8th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).

Koschmann, T., Suthers, D., & Chan, T.-W. (Eds.). (2005). *CSCL 2005–the next 10 years. Proceedings of the International Conference on Computer Supported Collaborative Learning 2005*. Mahwah: Erlbaum.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review, 24*(4), 691–710.

Markauskaite, L., & Reimann, P. (2008). *Enhancing and scaling-up design-based research: The potential of e-research*. Paper presented at the International Learning Science Conference–ICLS 2008, Utrecht, The Netherlands (24–28 July 2008) from http://www.fi.uu.nl/en/icls2008/343/paper343.pdf

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.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences, 17*, 33–59.

Miura, R. M. (1986). *Some mathematical questions in biology: DNA sequence analysis (Vol. 17).* Providence: American Mathematical Society.

Mochizuki, T., Kato, H., Yaegashi, K., Nishimori, T., Nagamori, Y., & Fujita, S. (2007). ProBoPortable: Does the cellular phone software promote emergent division of labor in project-based learning? 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. 512–514). New Brunswick: International Society of the Learning Sciences.

Mohr, L. (1982). *Explaining organizational behavior.* San Francisco: Jossey-Bass.

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*(3/4), 427–472.

Poole, M. S., & DeSanctis, G. (2004). Structuration theory in information systems research: methods and Q6 controversies. In M. E. Whitman & Woszcynski (Eds.), *The handbook of information systems research* (pp. 206–249). Idea Group Publishing.

Poole, M. S., & Roth, J. (1989). Decision making in small groups IV: A typology of group decision paths. *Human Communication Research, 15*(3), 323–356.

Poole, M. S., van de Ven, A., Dooley, K., & Holmes, M. E. (2000). *Organizational change and innovation processes. Theories and methods for research*. New Oxford: Oxford University Press.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286.

Reimann, P., Frerejean, J., & Thompson, K. (2009). *Using process mining to identify decision making processes in virtual teams (accepted)*. Paper presented at the International Conference on Computer-supported collaborative learning (CSCL2009). Rhodes/Greece.

Reisig, W. (1985). *Petri Nets. An introduction*. Berlin: Springer.

Ritter, F. E., Nerb, J., Lehtinen, E., & O'Shea, T. (eds). (2007). *In order to learn: How the sequences of topics influences learning*. Oxford: Oxford.

Sanderson, P. M., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction, 9*(3/4), 251–317.

Schegloff, E. A. (2007). *Sequence organization in interaction*. New York: Cambridge University Press.

Schümmer, T., Strijbos, J.-W., & Berkel, T. (2005). A new direction for log file analysis in CSCL: Experiences with a spatio-temporal metric. In T. Koschmann, D. Suthers & T. W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The next 10 years!* (pp. 567–576). Mahwah: Erlbaum.

Sloane, F. C. (2008). Multilevel models in design research. A case from mathematics education. In A. E. Kelly, R. A. Lesh & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 459–476). New York: Routledge.

Sloane, F. C., & Kelly, A. E. (2008). Design research and the study of change. In A. E. Kelly, R. A. Lesh & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 441–448). New York: Routledge.

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.

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

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.

Suppes, P., & Atkinson, R. C. (1960). *Markov learning models for multiperson interactions*. Stanford: Stanford University Press.

Suthers, D. D. (2006). A qualitative analysis of collaborative knowledge construction through shared representations. *Research and Practice in Technology Enhanced Learning, 1*(2), 115–142.

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, mind, 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 der Aalst, W. M. P., & Günther, C. W. (2007). Finding structure in unstructured processes: the case of process mining. In T. Basten, G. Juhas & S. Shukla (Eds.), *Proceedings the 7th International Conference on Applications of Concurrency to System Design (ACSD 2007; Bratislava, Slovak Republic)* (pp. 3–12). Los Alamitos: IEEE Computer Society.

Wever, B. d., Schellens, T., Valcke, M., & Keer, H. v. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education, 46*, 6–28.

Zumbach, J., & Reimann, P. (2003). Influence of feedback on distributed problem based learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments* (pp. 219–228). Dordrecht: Kluwer.