Time is precious: Why process analysis is essential for CSCL (and can also help to bridge between experimental and descriptive methods)

Peter Reimann, University of Sydney, Research Centre for Computer-supported Learning and Cognition (CoCo), Education Building A35, Sydney, NSW 2006, Australia
Email: p.reimann@edfac.usyd.edu.au

Abstract: Although *process* 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 increasingly the case. The quantitative method dominant in the social and learning sciences—variable-centered variance theory—is of limited value, so we argue, for studying change on longer time scales. We introduce event-centered process analysis as a more generally applicable approach, not only for quantitative analysis, but also for providing closer links between qualitative and quantitative research methods. We conclude with suggestions on how nomothetic, idiographic, and design-oriented research interests can become better integrated in CSCL.

Goals

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 Koschman (2001) suggested, it might be a defining element of CSCL that it is about "...studying learning in settings in which learning is observably and accountable embedded in collaborative activity" and that learning within 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 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 socio-cultural as well as the individual-cognitive, the nature of the process remains: learning is a process that unfolds over time; hence order matters.

The analysis of processes 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 can be seen from a short analysis of all empirical studies reported in the last CSCL conference (Koschmann, Suthers, & Chan, 2005). As Table 1 shows, the majority of studies 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').

Table 1: Duration of group lifetime in studies from the CSCL 2005 conference

"Lifetime" of groups studied	No. of studies	Percentage
Single session (20-180 minutes)	25	35%
2-6 days	5	7%
1-4 weeks	7	10%
Longer (1.5 months – 1 year)	34	48%
Total	71	

In studies where interaction and learning is distributed over multiple sessions, the research process does not only become more challenging for logistical reasons, but also because core assumptions of the experimental, treatment-oriented methods no longer hold. For instance, it becomes implausible that a treatment factor (be that a technical feature or a pedagogical measure) is acting *continuously* over time, an assumption that is fundamental to any experimental design and statistical method related to analysis of variance. Furthermore, as time increases, non-controlled 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'. Order effects as well as non-linear changes will become more pronounced because of the self-sustaining feedback processes at work in groups over time (Arrow, McGrath, & Berdahl, 2000). All of these problems constitute serious challenges for any theory and method that either ignores time completely or is based on the variance analysis model.

These challenges might partially account for the fact that although CSCL researchers are privileged in the sense that they have access to processes as they unfold over time, there is comparatively little research that makes use of the information contained in the order and duration of events. As a case in point, by my count only one study (Kapur, Voiklis, & Kinzer, 2005) out of 71 from the 2005 CSCL conference made use of statistical analysis methods that take time into account. Not only is the information contained in the order of events unused, there is also the risk that the results found using data subject to order effects are of limited value when order is ignored.

Since this is certainly not an ideal state of affairs, this paper sets out to accomplish two goals. The first goal is method-oriented: to provide the reader with some information on how sequential analysis can be conducted and appropriate methods that may apply, in particular for cases where the duration of group processes is long. This will be kept short, though, because good introductions into sequential data analysis exist (e.g., Sanderson & Fisher, 1994).

The second goal is a methodological one: This paper can be seen as continuing the discussion started by Dan Suthers (2005) on What To Study in CSCL research and How to Study It. With respect to the What, he suggests that research on "…processes of intersubjective learning, and how technological affordances mediate or support such processes" be privileged (p. 669). With respect to the How, he proposes hybrid methodologies that combine the strengths of experimental, descriptive, and interactive design approaches. However, integrating methods from such diverse paradigms is challenging due to the tensions arising from the differences in research interests. Experimental methods (along with analysis of variance as the predominant statistical method) have been developed in the tradition of the nomothetic, 'law-searching', quantitative paradigm, while Descriptive and Design approaches can be seen as variants of the idiographic, qualitative paradigm. While idiographic methods can make important contributions to improving computer tools and pedagogical designs, their contribution to theory building and testing, i.e. the nomothetic research program, has often been challenged (Goldthorpe, 2000).

I attempt in this paper to identify ways in which, for the field of CSCL, descriptive and experimental methods can be best aligned, starting from a discussion of the obstacles a purely variable-centered approach (of which the Experimental Method is an instance) faces for theorizing and analyzing change and learning processes in groups. Building on a reconsideration of what should count as process and process analysis, *event analysis* is suggested as most appropriate for law-searching research in CSCL because it can deal with change processes of various forms, provides a research logic that integrates qualitative-descriptive with quantitative-nomothetic accounts, and is at least somewhat informative in the design of software tools and pedagogical strategies.

Variable-centered Process Analysis

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 on-line 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 the three dimensions. The coding scheme is developed and applied following best practice (e.g., Strijbos, Martens, Prins, & Jochems (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 on-line 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 post-tests are conducted to assess individual learning gains and during the pre-test phase a number of other individual factors are assessed, including metacognitive capabilities. As outcome measures, individual learning is assessed with a pre- and post-test, and knowledge building is assessed by analyzing the discussion board entries.

How these data are analyzed will depend largely on what the researcher considers a process to be. Two conceptualizations of process will be distinguished here. The first one, *variable-centered*, relates to analysis of variance (or, as we prefer to call it, the *variance method*, because it includes the design of experiments, not only the analysis of data). The second one with roots in historical and organisational research, is called *event-centered analysis* or *event analysis* for short. I use the terminology suggested by Abell (1987) and in particular by Poole, van der Ven, Dooley, & Holme's (2000) excellent treatment of process analysis in the social sciences informed many parts of this paper.

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 income 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 & Jenkins, 1976). Instead, a word on the assumptions behind the variable-centered 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 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 on-line, 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-centered 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*). Since we will find, in any group, processes unfolding on different time scales (McGrath & Tschan, 2003), relating them in one model is a challenge indeed. And as was mentioned before, the variable-centered 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 than before.

The main argument I want to put forward is that, for situations similar to our scenario, which is typical for CSCL research, the variable-centered approach is of limited value, and needs to be extended by an *event-centered approach* that can more comprehensively account for the change processes under study in CSCL. Not only do I argue for event analysis because it adds important information to our understanding of learning and change, I also argue that it offers a bridge between qualitative and quantitative research methods, a bridge that seems particularly valuable for CSCL where research is conducted in both traditions.

The Event-centered Approach to Process Analysis

Our short sketch of the event-centered approach builds on Abbott (1990) and Abell (1987; Abell, 2004), who, among others, noted the differences between scientific explanations cast in terms of independent variables causing changes in a dependent variable, and explanations that provide a narrative in order to explain how a sequence of events unfold to produce an observed outcome.

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 (i.e., 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 on-line 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,* i.e. 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 towards an end in terms of only efficient causality would fail to identify the goal orientation.

The event analysis approach to be introduced now encompasses all three kinds of causality: efficient, formal and final. (As we don't go 'down' to the neurological level, we leave out Aristotle's fourth type, *material* cause, for explaining individual and group behavior.) A pivotal difference to the variable-centered method is that event analysis does not start by framing 'the world' in terms of variables, i.e. 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).

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*. 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 disciplinary background.

In our 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 etc. 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 narrative for each case, a case being a single person or a group, dependent on the level of analysis chosen. We note further that in *narrative* *explanations* the three types of causality are usually combined (Abell, 1987). The format of a narrative explanation is not only used by people when explaining other peoples' behavior, but also frequently employed by social scientists, for instance historians and political scientists.

We 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, Martens, Prins, & Jochems, 2006; Wever, Schellens, Valcke, & Keer, 2006). However, it is important to keep in mind that events are *not* treated as variables in event analysis, i.e. they are not aggregated (by coding category) into counts. Correspondingly, the process researcher does not look for co-variance between the values of independent and dependent variables, but "explains outcomes as the result of the order in which the events unfold and of particular conjunctions of events and contextual conditions" (Poole *et al.*, 2000, p. 36). The explanation takes essentially a *narrative* form and works with a *historical* logic: In order to explain any event in the scope of the study, that event will need to be related to events that took place (potentially a long time) before, not only to contextual factors (such as tool variation in our scenario). The order in which events occur and the conjunctions between different lines of events are essential to narrative explanations. Dan Suther's recent analysis of 'uptake' actions (2006) can be seen as an instance of such a type of analysis applied to a collaborative learning situation.

In addition to formulating such narratives for the change processes observed in the cases under study, the process researcher can test *general* theories, i.e. add a nomothetic dimension. This, as well as the use of quantitative methods, distinguishes event analysis from purely descriptive methods, such as ethnomethodology (Garfinkel, 2002) and conversation analysis (Schegloff, 1996). Generalizations are performed in two ways. Firstly, by identifying general, prototypical event sequences; looking across the event sequences from a number of cases (*all* groups in our scenario study), a process researcher would look for sequences or cycles that occur within and across groups with some regularity. Secondly, process research of this kind entails the need to account for the observed (sequence) regularities in terms of *generative mechanisms*, in terms of "motors" that "drive" change. To the extent that these generative models can themselves be related to a typology of classes of change models, generalizations can be performed not only on the level of sequence descriptions, but also on the level of generative theories/models. To give an example for such a typology: van de Ven & Poole (1995) have developed a typology of process theories that identifies four (ideal) types of theories of social change: (1) life cycle (e.g., Piaget's stage model of ontogenetic development); (2) evolution (e.g., Darwinian evolution in biology); (3) dialectic (e.g., Dialectical Materialism in economy/history), (4) teleology (e.g, Mead's Symbolic Interaction theory in sociology). To the extent that this typology is complete (for social sciences), any specific generative account for a change process can be expressed as a variant of one of these theory types, or as hybrid model: a combination of two or more of the theory types.

To relate this to CSCL: most of the change processes observable in on-line learning groups will incorporate elements of a life-cycle motor because groups will comply to some extent with the pedagogical or experimental design imposed on them. In addition, they might incorporate elements typical for a dialectical motor, for instance in settings where argumentation is important (Wegerif, 2005), or elements of an teleological motor, for instance for groups where problem solving is the main task (Zumbach, Hillers, & Reimann, 2003). An evolutionary motor may be found in groups that deal with design challenges (Kolodner *et al.*, 2003), for instance.

Independently of how appropriate one considers a specific combination of change motors to be for specific observations (an empirical issue), the point we want to make here is that the framing of a specific model in terms of more fundamental (generative) theories— for instance in terms of the four families of change theories—constitutes a powerful explanatory strategy, well aligned with—if not prototypical for—the scientific method in general. Unlike variance theory event analysis can deal with change where there is no consistent 'push' force and where the entities under study change qualitatively over time (are not uniform). While for the variable-centered approach generality depends on *uniformity* of the identified relation between variables across contexts and cases, a event approach theory aims for *versatility*, "...the degree to which it can encompass a broad domain of developmental patterns without modification of its essential character" (Poole et al., 2000, p. 43)

The reason process theories can be considered to be closer to the causally effective processes has to do with the definition of events as those incidents that are enacted by and happening to the central subject. This is a central feature of narrative explanations (Abbott, 1988). Narrative explanations apply also to situations where not only attributes of entities (central subjects) change, but the entity itself changes— for instance, through transformation into a different entity, through division, mergers, or dissolution. For CSCL research, where a group will more often than not be the central subject under study, this flexibility is a great advantage because it allows us to deal with all those change processes that affect a group qualitatively, such as changes in membership or major changes in groups' mission. For variable-centered theories, changes in the qualitative nature of variables are a non-issue: we would no longer measure the same (latent) concept. Of course, any method allowing for qualitative change of the central subject needs to find a way to distinguish between what constitutes a 'legitimate' qualitative change (that needs to be accounted for by theories dealing with that central subject) and the case where a theory no longer applies. Historians, where narrative explanations are ubiquitous, have found ways to deal with this challenge by explicating the idea of a *coherent central subject*, making not similarity, but spatio-temporal continuity the criterion: "...for any historical entity to remain the same entity, no degree of similarity between earlier and later stages in its development is required, as long as this development is spatio-temporally continuous" (Hull, 1975, p. 256).

Quantitative Methods for Event-centered Theory Testing

Although generalizing across cases and testing generalizations against cases does not require statistical methods (see for instance Abell (1987) and Heise (1990) for alternatives), we will only discuss the statistical methods in order to continue the comparison with the variable-centred method. An element of probability needs to be introduced when we move to testing general models. The reason for this is that predicting singular events based on a deterministic model requires the assumption that all factors other than those included in the deterministic model are constant. This is not realistic in most cases in the social sciences, certainly not in the situation considered here with a minimum of experimental control and a long duration.

Event analysis does not reject quantitative methods. Quite to the contrary, they form an important element for the purposes of generalizing across cases and testing process theories. Event analysis makes use of statistical methods that are appropriate for event data, i.e. do not require the data to be represented as variables. An example for such *stochastic* methods is Markov Chain modelling. Stochastic modelling methods have a fairly long tradition in the social sciences and psychology (e.g., Coleman (1964); Suppes & 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 modelling in any detail, but in order to provide a flavour, 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 events 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)) could be used, or Hidden Markov modeling (Rabiner, 1989).

These matters can not be discussed further here (see Soller, Wiebe, & Lesgold (2002) for an example of Hidden Markov modelling 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 time, meaning that events do not have to form a continuous stream), information we sometimes have available in log files. Finally, Markov modelling has been generalized to deal with continuous time.

A question we have not tackled yet is: Where do the process models come from? No surprises here, at least for those researchers who work nomothetically: from theory. One should have theory-based expectations as to the changes one would expect in groups before one engages in an empirical study. Of course, after testing the theory-driven hypotheses, few researchers would resist exploring if there are other interesting change processes hidden in the data. Identifying interesting new narratives that apply to event sequences not (adequately) covered by the theoretical expectations will often need to be done by researchers 'manually'. To some extent, data mining methods can help in this inductive phase. Kay, Maisonneuve, Yacef & Zaiane (2006), for instance, present a nice example of how applying a data mining algorithm (the Frequent Sequential Pattern algorithm, first introduced by Agrawal & Srikant (1995)) to more or less unprocessed (at least not human-coded) log file data covering students' long-term interactions in a realistically complex socio-technical setting can result in interesting discoveries. In this case, systematic differences between successful and less successful teams in (asynchronous) interaction sequences where found in a corpus of about 10.000 incidents.

Combining Variable-centered and Event-centered Methods

For the purpose of clarity, I have juxtaposed the variable- and event-centered methods, focusing on their differences and ignoring their commonalities. The main commonality they share is their nomothetic character; like with variable-centered methods, event analysis can be used to test law-like explanations. (For a deeper analysis of the fact that variance explanations are preceded by generalizing, whereas narrative explanations logically come before (optional) generalizing, see Abell, (1987) The criterion for generality is different, though (versatility instead of uniformity). Like the variable-centered approach, event analysis incorporates quantitative methods and embraces probabilistic concepts. Indeed, event analysis can be said to be *more* quantitative than the variable-centered approach because it aims to apply mathematical methods to phenomena where not only effective causation is at work, but formal and final causation as well. This suggests, despite the many differences, that the two methods can also fruitfully be combined, forming a general process analysis method.

The variable-centered, variance-oriented approach works perfectly well for research questions that involve relationships among variables. An event analysist has nothing against variables, as long as they are not seen as the *only* way to describe and explain change. We already mentioned that stochastic event sequence analysis can incorporate information that takes the form of values of variables by employing non-homogeneous Markov models. But the potential for method integration is not exhausted here. 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 (see above) can be used. Finally, 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-centered 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. What event analysis reminds us, though, is that we should not harbor overly simplistic assumptions as to the causal relations between such treatments and groups' behavior, in particular when groups interact with technology over longer stretches of time. Table 2 summarizes the research steps that are shared and unique, respectively, between variable- and event-centered approaches.

Conclusions: What is gained?

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 (quantitative) 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—if we take the 2005 conference as representative—deal with change processes that have a duration of weeks and months. If individual and group processes are analysed on such a scale without taking into account history, sequence, dynamics, in short: time, then many of the resulting findings are of limited value. We argued further that for studies that aim to analyse change unfolding over days, weeks and months, the quantitative method dominant in the social and learning sciences—variable-centered 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. We introduced a general process approach that builds on the notion of narrative explanations. I identified the main differences between variance and event analysis, provided arguments why the event analysis suits the need of CSCL research better, and concluded with an illustration of the type of quantitative analysis the event analysis allows, in addition to the many features it shares with qualitative methods.

Table 2: Method integration

Nomothetic		**Idiographic**
Variable-Centered	**Event-Centered**	
Research design		
Operationalisation of theoretical constructs into variables	Identification of central subject(s); definition of event types; Optional: Definition of variables	-/-
Realisation of Treatment Conditions	optional	-/-
Randomisation	optional	-/-
Control of external factors	Optional; Recording/Documentation of changes in the environment of the central subject	-/-
Data analysis		
Coding: Classification of events		Optional
-/-	Establishing of narrative explanations for sequences and conjunctions per case	Qualitative, often narrative, "thick" accounts
-/-	Identification of patterns in sequences across cases	Optional
-/-	Identification of change motor	Optional
-/-	Stochastic Modelling	-/-
Aggregation of codes into counts		-/-
Analysis of Variance		-/-
Time Series Analysis		-/-
Reporting		
Variable-related	Case-and variable related	Case-related

CSCL research can gain from an adoption 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 information informative for design can be derived. As has been the main argument on these pages, the variable-centered method, dominant in most experimental learning research, is not the best (nomothetic) method for conducting process research in CSCL. It makes too 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 analyse 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-centered approach makes extensive use of event descriptions: they enter into narrative accounts and, optionally, into statistical analysis without loosing their distinctiveness. Hence, independent of the research orientation (nomothetic, idiographic, design-oriented), activities such 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 than is the case for variable-centered methods (see also Table 2).

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 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 tools 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 & DeSanctis, 2004). But it needs to be said that this line of research has less implications for interface design than for organisational design and change management.

However, understanding organisational 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, Dirckinck-Holmfeld, & Linstroem, 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). They are not observable in the usability lab or the short-term study looking into immediate (learning) effects. Analysing 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 & 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.

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 side effect, shared on-line collections of (annotated) sequence data could be created that can be analysed from multiple perspectives and with various methods or tools. The time gained might be most profitably be spent on developing generative process models and theories, of which there is a genuine lack in CSCL.

References

Abbott, A. (1988). Transcending general linear theory. *Sociological Theory, 6*, 169-186.

Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives*. Oxford: Clarendon Press.

Abell, P. (2004). Narrative explanations: An alternative to variable-centered explanation? *Annual Review of Sociology, 30*, 287-310.

Agrawal, R., & Srikant, R. (1995). *Mining sequential patterns.* Paper presented at the Proceedings of the International Conference on Data Engineering (ICDE95).

Aristotle. (1941). *The basic works of Aristotle (ed. R. McKeon)*. New York: Random House.

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems*. Thousand Oaks: Sage.

Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Coleman, J. S. (1964). *Introduction to mathematical sociology*. New York: Free Press.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism*. Lanham, MD: Rowman & Littlefield.

Goldthorpe, J. H. (2000). *On sociology: Numbers, narratives, and the integration or research and theory*. Oxford: Oxford University Press.

Heise, D. R. (1990). Modeling event structures. *Journal of Mathematical Sociology, 16*, 142-159.

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

Hull, D. L. (1975). Central subjects and historical narratives. *History and Theory, 14*, 253-274.

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, NJ: Lawrence 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.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-based learning meets case-based reasoning in the midlle-school science classroom: Putting learning by design into practice. *The Journal of the learning sciences, 12*(4), 495-547.

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

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, NJ.: Lawrence Erlbaum.

McGrath, J. E., & Tschan, F. (2003). *Temporal matters in social psychology: Examining the role of time in the lives of groups and individuals*. Washington, DC: American Psychological Association.

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

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.

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

Schegloff, E. A. (1996). Confirming allusions: Towards an empirical account of action. *American Journal of Sociology, 104*, 161-216.

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, NJ.: Erlbaum.

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

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, CA: Stanford University Press.

Suthers, D. D. (2005). Technology affordances for intersubjective learning: a thematic agenda for CSCL. In T. Koschmann, D. Suthers & T.-W. Chan (Eds.), *CSCL 2005 - the next 10 years. Proceedings of the International Conference on Computer Supported Collaborative Learning 2005* (pp. 562-571). Mahwah, NL: Lawrence Erlbaum.

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.

van de Ven, A., & Poole, M. S. (1995). Explaining development and change in organizations. *Academy of Management Review, 20*, 510-540.

Wegerif, R. (2005). Towards a dialogic understanding of the relationship between CSCL and teaching thinking skills. In T. Koschmann, D. Suthers & T.-W. Chan (Eds.), *Computer-supported collaborative learning 2005: The next 10 years!* (pp. 707-716). Mahway, NJ: Lawrence Erlbaum.

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., Hillers, A., & Reimann, P. (2003). Supporting Distributed Problem-Based Learning: The Use of Feedback in Online Learning. In T. Roberts (Ed.), *Online Collaborative Learning: Theory and Practice* (pp. 86-103). Hershey, PA: Information Science Publishing.

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.

Acknowledgements

I would like to thank Anindito Aditomo for conducting the duration analysis on the CSCL 2005 conference papers, and him as well as Rob Kildare for providing me with feedback on earlier versions.